

Fred Brauer
Carlos Castillo-Chavez
Zhilan Feng

Mathematical Models in Epidemiology

Texts in Applied Mathematics

Volume 69

Series Editors

Anthony Bloch, University of Michigan, Ann Arbor, MI, USA
Charles L. Epstein, University of Pennsylvania, Philadelphia, PA, USA
Alain Goriely, University of Oxford, Oxford, UK
Leslie Greengard, New York University, New York, NY, USA

Advisory Board

J. Bell, Lawrence Berkeley National Lab, Berkeley, USA
R. Kohn, New York University, New York, USA
P. Newton, University of Southern California, Los Angeles, USA
C. Peskin, New York University, New York, USA
R. Pego, Carnegie Mellon University, Pittsburgh, USA
L. Ryzhik, Stanford University, Stanford, USA
A. Singer, Princeton University, Princeton, USA
A. Stevens, Max-Planck-Institute for Mathematics, Leipzig, Germany
A. Stuart, University of Warwick, Coventry, UK
T. Witelski, Duke University, Durham, USA
S. Wright, University of Wisconsin, Madison, USA

The mathematization of all sciences, the fading of traditional scientific boundaries, the impact of computer technology, the growing importance of computer modeling and the necessity of scientific planning all create the need both in education and research for books that are introductory to and abreast of these developments. The aim of this series is to provide such textbooks in applied mathematics for the student scientist. Books should be well illustrated and have clear exposition and sound pedagogy. Large number of examples and exercises at varying levels are recommended. TAM publishes textbooks suitable for advanced undergraduate and beginning graduate courses, and complements the Applied Mathematical Sciences (AMS) series, which focuses on advanced textbooks and research-level monographs.

More information about this series at <http://www.springer.com/series/1214>

Fred Brauer • Carlos Castillo-Chavez • Zhilan Feng

Mathematical Models in Epidemiology



Springer

Fred Brauer
University of British Columbia
Department of Mathematics
Vancouver, BC, Canada

Zhilan Feng
Purdue University
Department of Mathematics
West Lafayette, IN, USA

Carlos Castillo-Chavez
Mathematical and Computational Modeling
Center (MCMSC)
Department of Mathematics and Statistics
Arizona State University
Tempe, AZ, USA

ISSN 0939-2475
Texts in Applied Mathematics
ISBN 978-1-4939-9826-5
<https://doi.org/10.1007/978-1-4939-9828-9>

ISSN 2196-9949 (electronic)
ISBN 978-1-4939-9828-9 (eBook)

Mathematics Subject Classification: 92D30

© Springer Science+Business Media, LLC, part of Springer Nature 2019
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Science+Business Media, LLC, part of Springer Nature.

The registered company address is: 233 Spring Street, New York, NY 10013, U.S.A.

Foreword

The subject of infectious diseases has been one of the richest areas of application of mathematics in biology, dating back to Sir Ronald Ross' classic contributions to the management of malaria at the beginning of the last century. Infectious diseases obviously impose great burdens of morbidity and mortality on humanity, as well as on non-human populations of importance to us; and hence, the ability to use formal models to reduce those burdens has attracted the attention of exceptional mathematicians, including all three of the authors of this volume. All have made major contributions to the subject, independently and in collaboration with each other, and the current volume profits from their complementary talents and expertise. The authors combine broad expertise built on excellent scholarship, their own fundamental research and masterful exposition, and I am delighted to have been asked to write this Foreword.

The subject of mathematical epidemiology has seen great progress in the century since Ross, as new diseases like SARS and HIV–AIDS have appeared and as older ones, like measles and TB, have waned and waxed and re-emerged in places where they were thought to be defeated. But our battle with infectious diseases can never be won completely; like Lewis Carroll's Red Queen said, we must keep running just "to keep in the same place"; and we will need to run even faster if we are to make progress in this perpetual struggle. Highly adaptive agents like influenza A continue to evolve in a continual struggle with immune defenses, fostering increased efforts to develop a universal vaccine; similarly, pathogenic bacteria have evolved resistance to the drugs that had at first tamed them. At the same time, a growing and increasingly mobile world population has accelerated the spread of disease, including novel diseases that previously would have remained endemic or died out. Climate change has added yet another challenge, as the vectors of diseases, from insects to migrant birds, shift their ranges to carry diseases into areas whose populations are unprepared to deal with them. As discussed in the next paragraph, social influences raise new challenges for management and for mathematical modeling. Even our own microbiomes have been dramatically affected by environmental influences, antibiotic uses, and medical practices such as C-sections, changing the background against which management must operate.

Medical science has responded brilliantly to these challenges, through the development of new antibiotics, vaccines, and treatment protocols, guided along the way by improved mathematical modeling. Such models have needed to address the multi-scale dynamics of an interacting system of within-host immune dynamics, the population-level modeling of disease spread, the dynamics of vector populations, and the socioeconomic context within which diseases spread. The socioeconomic issues include not only the economics of control measures but also the behavioral responses before and during outbreaks, from vaccine hesitancy to the avoidance of social contacts involving infectious individuals. The overuse of antibiotics and underuse of vaccines have created new dilemmas for practitioners and new avenues for mathematical modeling to make contributions, guiding individual decisions, medical practice, hospital decision-making, and public policy at all levels.

This volume is a comprehensive treatment of the classical theory and modern extensions of that theory. It is constructed as a textbook for students with solid mathematical foundations, including problem sets that illuminate the issues. As such, it should find wide use as the basis for beginning and advanced courses in this important subject and serve as a complement and update to the excellent 2001 text by Brauer and Castillo-Chavez; clearly, a lot has happened in the field in the nearly two decades since that book appeared, and those advances are well documented in this book. But this volume, as the earlier one, also will be a treasure trove for mathematicians and other researchers looking for an authoritative introduction to the literature and to open problems to the solutions of which they might contribute their skills. It should find its way onto the shelves of every researcher interested in the topic.

Princeton University
Princeton, NJ, USA

Simon Levin

Preface

The goal of this book is to interest students of mathematics and public health professionals in the modeling of infectious diseases transmission. We believe that some knowledge of disease transmission models can give useful insights to epidemiologists and that there are interesting mathematical problems in such models. The mathematical background necessary for studying this book is a knowledge of calculus, some matrix theory, and some ordinary differential equations, specifically approximate and qualitative methods. Our emphasis is on describing the mathematical results being used and showing how to apply them rather than on detailed proofs. We hope that eventually the education of biologists and public health professionals will include these mathematical topics in the first 2 years of university so that this book can be accessible to upper-level undergraduate students.

A course based on this book should cover the first section on basic concepts and some of the chapters on specific diseases in the second section. In addition, some of the more advanced topics from the third section, depending on the interests of the students and the time available, can be included. Throughout the book, there are suggested projects giving a sense of research topics that could be attacked by students in groups. In the first section of the book, there are also numerous exercises, and answers to selected exercises are available. We believe that a book should go beyond a course based on it, to lead to further reading, and the last three chapters of the book and the epilogue are intended to suggest some topics for further thought.

In order to make the book more accessible to the readers with minimal mathematical background, we have put some mathematical review material on matrix algebra, differential equations, and systems of differential equations into some appendices to this book. We have also established a website

<https://mcmsc.asu.edu/content/mathematical-models-epidemiology>

containing review notes covering some topics in calculus, matrix algebra, and dynamical systems (ordinary differential equations and difference equations). We have also starred difficult exercises and difficult sections which are intended for the readers with strong mathematical backgrounds.

Stochastic disease transmission models form an important area that we have largely omitted in order to keep the size of this book within bounds. For readers who may wish to explore this subject, we have included some references near the end of Chap. 1. Discrete disease transmission models are not studied in the main text, but there are several projects introducing the subject in addition to some references near the end of Chap. 1.

Our emphasis in this book is on relatively simple models whose goal is to describe general qualitative behavior and establish broad principles. Public health professionals are more concerned with detailed models to make short-term quantitative predictions. Such detailed models are generally difficult or impossible to solve analytically, but the relatively recent development of the availability of high-speed computing has made detailed models available for quantitative predictions and comparison of different possible management strategies. The emphasis in this book is on relatively simple models, but readers should remember that there is another important aspect of the subject.

Vancouver, BC, Canada
Tempe, AZ, USA
West Lafayette, IN, USA

Fred Brauer
Carlos Castillo-Chavez
Zhilan Feng

Acknowledgments

It is always hard to write acknowledgments because so many have contributed in a multitude of ways to the content of his book, and so we hope that our public recognition in this section is accepted by those who are not mentioned explicitly. Thanks for your understanding. This book, like our textbook *Mathematical Models in Population Biology and Epidemiology* by FB and CCC, has finally been completed after many detours and delays. This volume has been enriched by the inclusion of a third author, Zhilan Feng (ZF). The three of us have been involved in mathematical epidemiology and its applications for roughly three decades, an effort that was intensified by the establishment of the Mathematical and Theoretical Biology Institute (MTBI) by Carlos Castillo-Chavez (CCC) in 1996. This is a summer institute that has mentored nearly 500 undergraduates (152 have completed a PhD), nearly 180 graduate students, several postdocs, and a dozen young faculty. The energy of MTBI and its regular faculty has been a source of inspiration, energy, and ideas for this project. The staff of the Simon A. Levin Mathematical, Computational, and Modeling Sciences Center (SAL-MCMSC) has been involved in the logistics of the MTBI and of this project. We are particularly thankful to Kamal Barley for his invaluable help in drawing most of the figures; to Baltazar Espinoza and Victor Moreno for their assistance in indexing, in various chapters, and some of the computations and graphics; and to MTBI alumni for their useful comments. In addition, Baltazar wrote much of the material on age distributions in Chap. 13. All books have misprints. This one has many fewer misprints than it might otherwise have had because of the proofreading by Baltazar Espinosa, Ceasar Montalvo, Anarina Murillo, and Marisabel Rodriguez.

CCC wants to thank his three children, Carlos W, Gabi, and Melissa, and his wife, Nohora, for their support, ideas, and inspiration. ZF wants to thank her daughter, Haiyun, and son, Henry, for all their love and support. FB wants to thank his wife, Esther; his children, David, Deborah, and Michael; and his grandchildren, Noah, Sophie, Benjamin, and Evan, for their support and tolerance. One of the lessons learned from the MTBI is that a cooperating group can accomplish much more than any of the individuals in the group, and this applies to families even without direct involvement.

We would like to restate our appreciation for the deliberate or unsuspecting mentoring that we have received over the years from our undergraduate, graduate, and postdoctoral students, colleagues, collaborators, visitors, and researchers in the field across the world. This book would have not completed without the support that we have received over the years by the National Science Foundation, DOD, and occasionally by the NIH, as well as the Canadian group MITACS. The MITACS (Mathematics of Information Technology and Complex Systems) program in response to the SARS epidemic of 2002–2003 has instigated much cooperation of mathematical modelers and public health professionals to learn to communicate with one another. The support that the SAL-MCMSC has received from the Offices of the President and Provost, the PhD Program in Applied Mathematics in the Life and Social Sciences, and the School of Human Evolution and Social Change at Arizona State University have been critical for this effort. We thank them all.

Contents

Foreword	v
Preface	vii
Acknowledgments	ix

Part I Basic Concepts of Mathematical Epidemiology

1 Introduction: A Prelude to Mathematical Epidemiology	3
1.1 Introduction	3
1.2 Some History	4
1.2.1 The Beginnings of Compartmental Models	5
1.2.2 Stochastic Models	7
1.2.3 Developments in Compartmental Models	9
1.2.4 Endemic Disease Models	11
1.2.5 Diseases Transmitted by Vectors	12
1.2.6 Heterogeneity of Mixing	14
1.3 Strategic Models and This Volume	15
References	16
2 Simple Compartmental Models for Disease Transmission	21
2.1 Introduction to Compartmental Models	23
2.2 The <i>SIS</i> Model	26
2.3 The <i>SIR</i> Model with Births and Deaths	30
2.4 The Simple Kermack–McKendrick Epidemic Model	35
2.5 Epidemic Models with Deaths due to Disease	43
2.6 *Project: Discrete Epidemic Models	46
2.7 *Project: Pulse Vaccination	47
2.8 *Project: A Model with Competing Disease Strains	49
2.9 Project: An Epidemic Model in Two Patches	52
2.10 Project: Fitting Data for an Influenza Model	53
2.11 Project: Social Interactions	54
2.12 Exercises	55
References	60

xi

3	Endemic Disease Models	63
3.1	More Complicated Endemic Disease Models	64
3.1.1	Exposed Periods	64
3.1.2	A Treatment Model	65
3.1.3	Vertical Transmission	67
3.2	Some Applications of the <i>SIR</i> Model	68
3.2.1	Herd Immunity	68
3.2.2	Age at Infection	69
3.2.3	The Inter-Epidemic Period	71
3.2.4	“Epidemic” Approach to Endemic Equilibrium	72
3.3	Temporary Immunity	73
3.3.1	*Delay in an <i>SIRS</i> Model	75
3.4	A Simple Model with Multiple Endemic Equilibria	78
3.5	A Vaccination Model: Backward Bifurcations	81
3.5.1	The Bifurcation Curve	87
3.6	*An <i>SEIR</i> Model with General Disease Stage Distributions	88
3.6.1	*Incorporation of Quarantine and Isolation	93
3.6.2	*The Reduced Model of (3.42) Under GDA	95
3.6.3	*Comparison of EDM and GDM	96
3.7	Diseases in Exponentially Growing Populations	100
3.8	Project: Population Growth and Epidemics	102
3.9	*Project: An Environmentally Driven Infectious Disease	107
3.10	*Project: A Two-Strain Model with Cross Immunity	111
3.11	Exercises	112
	References	114
4	Epidemic Models	117
4.1	A Branching Process Disease Outbreak Model	118
4.1.1	Transmissibility	123
4.2	Network and Compartmental Epidemic Models	125
4.3	More Complicated Epidemic Models	129
4.3.1	Exposed Periods	129
4.3.2	A Treatment Model	131
4.3.3	An Influenza Model	133
4.3.4	A Quarantine–Isolation Model	134
4.4	An <i>SIR</i> Model with a General Infectious Period Distribution	137
4.5	The Age of Infection Epidemic Model	139
4.5.1	A General <i>SEIR</i> Model	141
4.5.2	A General Treatment Model	144
4.5.3	A General Quarantine/Isolation Epidemic Model	146
4.6	The Gamma Distribution	148
4.7	Interpretation of Data and Parametrization	151
4.7.1	Models of <i>SIR</i> Type	152
4.7.2	Models of <i>SEIR</i> Type	154
4.7.3	Mean Generation Time	156

4.8	*Effect of Timing of Control Programs on Epidemic Final Size	159
4.9	Directions for Generalization.....	161
4.10	Some Warnings	161
4.11	*Project: A Discrete Model with Quarantine and Isolation.....	162
4.12	Project: Epidemic Models with Direct and Indirect Transmission	167
4.13	Exercises	171
	References.....	176
5	Models with Heterogeneous Mixing	179
5.1	A Vaccination Model	179
5.2	The Next Generation Matrix and the Basic Reproduction Number	182
5.2.1	Some More Complicated Examples	188
5.3	Heterogeneous Mixing	188
5.3.1	*Optimal Vaccine Allocation in Heterogeneous Populations	202
5.4	A Heterogeneous Mixing Age of Infection Model	209
5.4.1	The Final Size of a Heterogeneous Mixing Epidemic ..	212
5.5	Some Warnings	216
5.6	*Projects: Reproduction Numbers for Discrete Models	217
5.7	*Project: Modeling the Synergy Between HIV and HSV-2.....	221
5.8	Project: Effect of Heterogeneities on Reproduction Numbers	223
	References.....	225
6	Models for Diseases Transmitted by Vectors	229
6.1	Introduction	229
6.2	A Basic Vector Transmission Models	230
6.2.1	The Basic Reproduction Number.....	232
6.2.2	The Initial Exponential Growth Rate.....	233
6.3	Fast and Slow Dynamics	234
6.3.1	Singular Perturbations	237
6.4	A Vector-Borne Epidemic Model	239
6.4.1	A Final Size Relation	240
6.5	*Project: An <i>SEIR/SEI</i> Model	241
6.6	*Project: Models for Onchocerciasis	241
6.7	Exercises	244
	References.....	244

Part II Models for Specific Diseases

7	Models for Tuberculosis.....	249
7.1	A One-Strain Model with Treatment.....	251
7.2	A Two-Strain TB Model	252
7.3	Optimal Treatment Strategies	256

7.4	Modeling of the Long and Variable Latency of TB	260
7.5	Backward Bifurcation in a TB Model with Reinfection	263
7.6	Other TB Models with More Complexities.....	265
7.7	Project: Some Calculations for the Two-Strain Model	267
7.8	Project: Refinements of the One-Strain Model	268
7.9	Project: Refinements of the Two-Strain Model.....	269
	References.....	271
8	Models for HIV/AIDS	273
8.1	Introduction	273
8.2	A Model with Exponential Waiting Times	275
8.3	An HIV Model with Arbitrary Incubation Period Distributions ..	278
8.4	An Age of Infection Model.....	281
8.5	*HIV and Tuberculosis: Dynamics of Coinfections.....	284
8.6	*Modeling the Synergy Between HIV and HSV-2	292
8.6.1	Reproduction Numbers for Individual Diseases	295
8.6.2	Invasion Reproduction Numbers	297
8.6.3	Influence of HSV-2 on the Dynamics of HIV	299
8.7	An HIV Model with Vaccination.....	299
8.8	A Model with Antiretroviral Therapy (ART).....	301
8.9	Project: What If Not All Infectives Progress to AIDS?	302
	References.....	305
9	Models for Influenza	311
9.1	Introduction to Influenza Models	311
9.2	A Basic Influenza Model	312
9.2.1	Vaccination	316
9.3	Antiviral Treatment	319
9.4	Seasonal Influenza Epidemics	323
9.4.1	Season to Season Transition	326
9.5	Pandemic Influenza	326
9.5.1	A Pandemic Outbreak	327
9.5.2	Seasonal Outbreaks Following a Pandemic	328
9.6	The Influenza Pandemic of 2009	330
9.6.1	A Tactical Influenza Model	330
9.6.2	Multiple Epidemic Waves	331
9.6.3	Parameter Estimation and Forecast of the Fall Wave	334
9.7	*An SIQR Model with Multiple Strains with Cross-Immunity ..	337
9.7.1	*An SIQR Model with a Single Infectious Class	338
9.7.2	*The Case of Two Strains with Cross-Immunity	342
9.8	Exercises	347
	References.....	348
10	Models for Ebola	351
10.1	Estimation of Initial Growth and Reproduction Numbers	351
10.1.1	Early Detection	358

10.2	Evaluations of Control Measures	360
10.3	The Legrand Model and Underlying Assumptions	362
10.3.1	The Legrand Model	362
10.3.2	A Simpler System Equivalent to the Legrand Model	364
10.4	*Models with Various Assumptions on Stage Transition Times	365
10.5	Slower than Exponential Growth	382
10.5.1	The Generalized Richards Model	383
10.5.2	The Generalized Growth Model	383
10.5.3	The IDEA Model	384
10.5.4	Models with Decreasing Contact Rates	384
10.6	Project: Slower than Exponential Growth	385
10.7	Project: Movement Restrictions as a Control Strategy	386
10.8	Project: Effect of Early Detection	387
	References	388
11	Models for Malaria	391
11.1	A Malaria Model	391
11.2	Some Model Refinements	395
11.2.1	Mosquito Incubation Periods	395
11.2.2	Boosting of Immunity	396
11.2.3	Alternative Forms for the Force of Infection	397
11.3	*Coupling Malaria Epidemiology and Sickle-Cell Genetics	398
	References	407
12	Dengue Fever and the Zika Virus	409
12.1	Dengue Fever	409
12.1.1	Calculation of the Basic Reproduction Number	411
12.2	A Model with Asymptomatic Infectives	412
12.2.1	Calculation of the Basic Reproduction Number	412
12.3	The Zika Virus	414
12.4	A Model with Vector and Direct Transmission	414
12.4.1	The Initial Exponential Growth Rate	417
12.5	A Second Zika Virus Model	419
12.6	Project: A Dengue Model with Two Patches	421
12.7	Exercises	422
	References	422
	Part III More Advanced Concepts	
13	Disease Transmission Models with Age Structure	429
13.1	Introduction	429
13.2	Linear Age Structured Models	430
13.3	The Method of Characteristics	432
13.4	Equivalent Formulation as an Integral Equation Model	433
13.5	Equilibria and the Characteristic Equation	435
13.6	The Demographic Model with Discrete Age Groups	436

13.7	Nonlinear Age Structured Models	437
13.8	Age-Structured Epidemic Models.....	439
13.8.1	*Age-Dependent Vaccination in Epidemic Models	440
13.8.2	Pair-Formation in Age-Structured Epidemic Models....	445
13.9	A Multi-Age-Group Malaria Model	447
13.10	*Project: Another Malaria Model	448
13.11	Project: A Model Without Vaccination	451
13.12	Project: An <i>SIS</i> Model with Age of Infection Structure	452
13.13	Exercises	453
	References.....	453
14	Spatial Structure in Disease Transmission Models	457
14.1	Spatial Structure I: Patch Models	457
14.1.1	Spatial Heterogeneity	457
14.1.2	Patch Models with Travel	459
14.1.3	Patch Models with Residence Times	461
14.2	Spatial Structure II: Continuously Distributed Models	464
14.2.1	The Diffusion Equation	465
14.2.2	Nonlinear Reaction–Diffusion Equations	467
14.2.3	Disease Spread Models with Diffusion	468
14.3	Project: A Model with Three Patches	471
14.4	Project: A Patch Model with Residence Time	474
	References.....	475
15	Epidemiological Models Incorporating Mobility, Behavior, and Time Scales	477
15.1	Introduction	477
15.2	General Lagrangian Epidemic Model in an <i>SIS</i> Setting	479
15.3	Assessing the Efficiency of Cordon Sanitaire as a Control Strategy of Ebola	481
15.3.1	Formulation of the Model	481
15.3.2	Simulations	482
15.4	*Mobility and Health Disparities on the Transmission Dynamics of Tuberculosis	485
15.4.1	*A Two-Patch TB Model with Heterogeneity in Population Through Residence Times in the Patches	486
15.4.2	*Results: The Role of Risk and Mobility on TB Prevalence	486
15.4.3	The Role of Risk as Defined by Direct First Time Transmission Rates	487
15.4.4	The Impact of Risk as Defined by Exogenous Reinfection Rates.....	488
15.5	*ZIKA	490
15.5.1	*Single Patch Model	491

15.5.2 *Residence Times and Two-Patch Models	492
15.5.3 What Did We Learn from These Single Outbreak Simulations?	500
References	501

Part IV The Future

16 Challenges, Opportunities and Theoretical Epidemiology	507
16.1 Disease and the Global Commons	508
16.1.1 Contagion and Tipping Points	508
16.1.2 Geographic and Spatial Disease Spread	509
16.2 Heterogeneity of Mixing, Cross-immunity, and Coinfection	510
16.3 Antibiotic Resistance	510
16.4 Mobility	512
16.4.1 A Lagrangian Approach to Modeling Mobility and Infectious Disease Dynamics	513
16.5 Behavior, Economic Epidemiology, and Mobility	521
16.5.1 Economic Epidemiology	521
16.5.2 Lagrangian and Economic Epidemiology Models	523
References	524
Epilogue	533
References	537

A Some Properties of Vectors and Matrices	539
A.1 Introduction	539
A.2 Vectors and Matrices	540
A.3 Systems of Linear Equations	544
A.4 The Inverse Matrix	545
A.5 Determinants	546
A.6 Eigenvalues and Eigenvectors	548
B First Order Ordinary Differential Equations	553
B.1 Exponential Growth and Decay	553
B.2 Radioactive Decay	555
B.3 Solutions and Direction Fields	557
B.4 Equations with Variables Separable	562
B.5 Qualitative Properties of Differential Equations	569
C Systems of Differential Equations	579
C.1 The Phase Plane	579
C.2 Linearization of a System at an Equilibrium	581
C.3 Solution of Linear Systems with Constant Coefficients	584
C.4 Stability of Equilibria	593
C.5 Qualitative Behavior of Solutions of Linear Systems	598
Index	605

Part I

Basic Concepts of Mathematical

Epidemiology

Chapter 1

Introduction: A Prelude to Mathematical Epidemiology



1.1 Introduction

Recorded history continuously documents the invasion of populations by infectious agents, some causing many deaths before disappearing, others reappearing in invasions some years later in populations that have acquired some degree of immunity, due to prior exposure to related infectious pathogens. The “Spanish” flu epidemic of 1918–1919 exemplifies the devastating impact of relatively rare pandemics; this one was responsible for about 50,000,000 deaths worldwide, while on the mild side of the spectrum we experience annual influenza seasonal epidemics that cause roughly 35,000 deaths in the USA each year.

Communicable diseases have played a significant role in shaping human history. The Black Deaths (probably bubonic plague) spread starting in 1346, first through Asia and moving across Europe repeatedly during the fourteenth century. The Black Death has been estimated to have caused the death of as much as one-third of the population of Europe between 1346 and 1350. The disease reappeared regularly in various regions of Europe for more than 300 years, a notable outbreak being that of the Great Plague of London of 1665–1666. It gradually withdrew from Europe afterwards.

Some diseases have become endemic (“permanently” established) in various populations causing a variable number of deaths particularly in countries with inefficient or resource-limited health care systems. Even within the twenty-first century, we see that millions of people die of measles, respiratory infections, diarrhea, and more. Individuals still die in significant numbers from diseases that are no longer considered dangerous, diseases that are easily treated or that have been well managed by resource-rich societies or by nations that invest in public health and prevention systematically. Some highly prevalent old foes include malaria, typhus, cholera, schistosomiasis, and sleeping sickness, endemic diseases in many parts of the world; diseases that have a significant negative impact on the mean life span of a population as well as on the economy of afflicted countries due to their impact

on the health of the population. The World Health Organization has estimated that in 2011 there were 1,400,000 deaths due to tuberculosis, 1,200,000 deaths due to HIV/AIDS, and 627,000 deaths due to malaria (but other sources have estimated the number of malaria deaths to be more than 1,000,000). In short, HIV, malaria, and TB account for at least 9,000 deaths each day. The impact of vaccines can be dramatic, for example, there were 2,600,000 deaths due to measles in 1980 but only 160,000 by 2011. The development and availability of the measles vaccine led to a reduction in the number of deaths due to this childhood disease of nearly 94%.

Epidemiologists, in response to a health emergency or as a result of systematic surveillance, first obtain and analyze observed data. They use data, observations, science, and theory as they work at identifying a pathogen (when unknown) behind an observed disease outbreak or as they proceed to plan or implement policies that ameliorates its impact. Naturally, understanding the causes and modes of transmission of each disease is central to forecasting or mitigating its impact within and across populations at risk. Mathematical models have played a substantial role in both short and long term planning for controlling the dynamics of a disease.

This volume provides a guided tour of the role that mathematical models have played in epidemiology and public health policy. This tour introduces a wide range of models and tools that have proven useful in the study of disease dynamics and control. This book provides a framework that will position those interested in the use of modeling and computational tools in epidemiology, public health, and related fields, in a position to contribute to the study of the transmission dynamics and control of contagion.

1.2 Some History

The study of infectious disease data began with the work of John Graunt (1620–1674) in his 1662 book “Natural and Political Observations made upon the Bills of Mortality.” The Bills of Mortality were weekly records of numbers and causes of death in London parishes. The records, beginning in 1592 and kept continuously from 1603 on, provided the data that Graunt used to begin to understand or identify possible causes of observed mortality patterns. He analyzed the various causes of death and gave a method of estimating the comparative risks of dying from various diseases, giving the first approach to a theory of competing risks.

In the eighteenth century smallpox was endemic and, perhaps not surprisingly, the first model in mathematical epidemiology was tied in to the work that Daniel Bernoulli (1700–1782) carried out on estimating the impact of inoculation against smallpox. Variolation, essentially inoculation with a mild strain, was introduced as a way to produce lifelong immunity against smallpox, but with a small risk of infection and death. There was heated debate about variolation, and Bernoulli was led to study the question of whether variolation was beneficial. His approach was to calculate the increase in life expectancy if smallpox were to be eliminated as a cause of death. His approach to the question of competing risks led to the publication of

a brief outline in 1760 [7] followed in 1766 by a more complete exposition [8]. His work received a mainly favorable reception; research that has become known in the actuarial literature rather than in the epidemiological literature. More recently his approach has been generalized [31].

Another valuable contribution to the understanding of infectious diseases prior to our understanding of disease transmission processes was gained from the study of the temporal and spatial pattern of cholera cases during the 1855 epidemic in London carried out by John Snow. He was able to pinpoint the Broad Street water pump as the source of the infection [54, 71]. In 1873, William Budd was able to gain a similar understanding of the spread of typhoid [17]. Statistical theory also moved forward with William Farr's study of statistical returns in 1840, a study that had as its goal the discovery of the laws that underlie the rise and fall of epidemics [36].

Many of the early developments in the mathematical modeling of communicable diseases are due to public health physicians. The first known result in mathematical epidemiology, as noted before, is a defense of the practice of inoculation against smallpox in 1760 by Daniel Bernoulli, a member of a famous family of mathematicians (eight spread over three generations) who had been trained as a physician. The first contributions to modern mathematical epidemiology are due to P.D. En'ko between 1873 and 1894 [30], and the foundations of the entire approach to epidemiology based on compartmental models were laid by public health physicians such as Sir R.A. Ross, W.H. Hamer, A.G. McKendrick, and W.O. Kermack between 1900 and 1935, along with important contributions from a statistical perspective by J. Brownlee.

1.2.1 *The Beginnings of Compartmental Models*

In order to describe a mathematical model for the spread of a communicable disease, it is necessary to make some assumptions about the means of spreading infection. The idea of invisible living creatures as agents of disease goes back at least to the writings of Aristotle (384–322 BC). The existence of microorganisms was demonstrated by van Leeuwenhoek (1632–1723) with the aid of the first microscopes. The first expression of the germ theory of disease by Jacob Henle (1809–1885) came in 1840 and was developed by Robert Koch (1843–1910), Joseph Lister (1827–1912), and Louis Pasteur (1822–1875) in the late nineteenth and early twentieth centuries. The modern view is that many diseases are spread by contact through a virus or bacterium. We focus in this book on the problem of understanding the spread of disease at a population level. Similar modeling approaches can be used to study the dynamics of infection within a host for diseases including HIV. This area is the backbone of the field of mathematical and computational immunology and viral dynamics. An introduction to immunology may be found in the book by Nowak and May [67].

In 1906, W.H. Hamer argued that the spread of infection should depend on the number of susceptible individuals and the number of infective individuals [44]. He

suggested a mass action law for the rate of new infections, and this idea has been basic in the formulation of compartmental models since that time. It is worth noting that the foundations of the entire approach to epidemiology based on compartmental models were laid, not by mathematicians, but primarily by public health physicians such as Sir R.A. Ross, W.H. Hamer, A.G. McKendrick, and W.O. Kermack between 1900 and 1935.

A particularly instructive example is the work of Ross on malaria. Sir Ronald Ross was awarded the second Nobel Prize in Medicine in 1902 for his demonstration of the dynamics of the transmission of malaria between mosquitoes and humans. He discovered the malarial parasite in the gastrointestinal tract of the *Anopheles* mosquito from which he was able to characterize the life cycle of malaria. He concluded that this vector-borne disease was transmitted by the *Anopheles* mosquito and in the process he developed a program for controlling or eliminating it at the population level.

It was generally believed that, so long as mosquitoes were present in a population, malaria could not be eliminated. Ross introduced a simple compartmental model [69] that included mosquitoes and humans. He showed that reducing the mosquito population below a critical level would be sufficient to eliminate malaria. This was the first introduction of the concept of the basic reproduction number, a central idea in mathematical epidemiology since that time. Field trials supported Ross' conclusion leading sometimes to brilliant successes in malaria control.

The basic compartmental models to describe the transmission of communicable diseases are contained in a sequence of three papers by W.O. Kermack and A.G. McKendrick in 1927, 1932, and 1933 [55–57]. The first of these papers described epidemic models.

The Kermack–McKendrick epidemic model, introduced in Chap. 2 and studied in more detail in Chap. 4, included dependence on age of infection, that is, the time since becoming infected, and can be used to provide a unified approach to compartmental epidemic models.

Various disease outbreaks including the *SARS* epidemic of 2002–2003, the concern about a possible *H5N1* influenza epidemic in 2005, the *H1N1* influenza pandemic of 2009, and the Ebola outbreak of 2014 have reignited interest in epidemic models, with the reformulation of the Kermack–McKendrick model by Diekmann, Heesterbeek, and Metz [27] highlighting the importance of looking at the foundational work. Chapter 4 contains a study of epidemic models.

In the work of Ross and Kermack and McKendrick there is a threshold quantity, the basic reproduction number, which is now almost universally denoted by \mathcal{R}_0 . Neither Ross nor Kermack and McKendrick identified this threshold quantity or gave it a name. It appears that the first person to name the threshold quantity explicitly was MacDonald [60] in his work on malaria.

The basic reproduction number, \mathcal{R}_0 (referred to as the basic reproductive number by some authors), is defined as the expected number of disease cases (secondary infections) produced by a “typical” infected individual in a wholly susceptible population over the full course of the disease outbreak. In an epidemic situation, in which the time period is short enough to neglect demographic effects and all

infected individuals recover with full immunity against reinfection, the threshold $\mathcal{R}_0 = 1$ is the dividing line between the infection dying out and the onset of an epidemic. In a situation that includes a flow of new susceptible individuals, either through demographic effects or recovery without full immunity against reinfection, the threshold $\mathcal{R}_0 = 1$ is the dividing line between an approach to a disease-free equilibrium and an approach to an endemic equilibrium, in which the disease is always present. This situation is studied in detail in Chap. 3.

Since 1933, there has been a great deal of work on compartmental disease transmission models, with generalizations in many directions. In particular, it is assumed in [55–57] that stays in compartments are exponentially distributed. In the generalization to age of infection models in Chap. 4, we are able to assume arbitrary distributions of stay in a compartment.

1.2.2 Stochastic Models

There are serious shortcomings in the simple Kermack–McKendrick model as a description of the beginning of a disease outbreak. Indeed, a very different kind of model is required since the Kermack–McKendrick compartmental epidemic model assumes that the sizes of the compartments are large enough that the mixing of members is homogeneous. However, at the beginning of a disease outbreak, there is a very small number of infective individuals and the transmission of infection is better captured if seen as a stochastic event that depends on the pattern of contacts between members of the population; a more satisfactory description should take such a stochastic pattern into account. We will not study stochastic models in this volume except for two sections at the start of Chap. 4 dealing with the initial stages of a disease outbreak.

The process chosen here to describe it is known as a Galton–Watson process; the result was first given in [39, 77] although there was a gap in the convergence proof. The first complete proof was given much later by Steffensen [72, 73]. The result is now a standard theorem given in many sources on branching processes, for example, [45], but did not appear in the epidemiological literature until later. To the best of the authors' knowledge, the first description in an epidemiological reference is [62] and the first epidemiological book source is the book by O. Diekmann and J.A.P. Heesterbeek [26] in 2000.

A stochastic branching process description of the beginning of a disease outbreak begins with the assumption that there is a network of contacts of individuals, which may be described by a graph with members of the population represented by vertices and with contacts between individuals represented by edges. The study of graphs originated with the abstract theory of Erdős and Rényi of the 1950s and 1960s [33–35], and has become important more recently in many areas, including in the study of social contacts, computer networks, and many other areas, as well as in the spread of communicable diseases. We will think of networks as bi-directional, with disease transmission possible in either direction along an edge. A brief taste of network

models is given at the beginning of Chap. 4. It is however important to stress the fact that this book does not get involved in the study of stochastic models or in the study of epidemics in networks—areas that deserve their own volumes.

We consider a disease outbreak that begins with a single infected individual (“patient zero”) who transmits infection to every individual to whom this individual is connected, that is, along every edge of the graph from the vertex corresponding to this individual. In other words, we assume that a disease outbreak begins when a single infective transmits infection to all of the people with whom he or she is in contact. Our development via branching processes is along the lines of that of [26]. Another approach, using a contact network perspective taken in [20, 65, 66] begins with an infected edge, corresponding to a disease outbreak started by an infective individual who passes the infection on to only one contact. This approach is the one taken more commonly in studies of epidemics on networks. It is somewhat more complicated and leads to somewhat different results, although the methods are quite similar.

In a stochastic setting, it is possible to prove that there is also a number called the basic reproduction number denoted by \mathcal{R}_0 with the properties that if $\mathcal{R}_0 < 1$ the probability that the infection will die out is 1, while if $\mathcal{R}_0 > 1$ there is a positive probability that the infection will persist leading to an epidemic. However, there is also a positive probability that the infection will increase initially but will produce only a minor outbreak dying out before triggering a major epidemic. This distinction between a minor outbreak and a major epidemic, and the result that if $\mathcal{R}_0 > 1$ there may be only a minor outbreak and not a major epidemic is intrinsic in these stochastic models and is not reflected in deterministic models, the primary theme of this book.

A possible approach to a realistic description of an epidemic would consider the use of a branching process model initially, making a transition to a compartmental model when the epidemic has become established, that is, when there are enough infectives so that the mass action mixing in the population becomes a reasonable approximation. Another approach would be to continue to use a network model throughout the course of the epidemic [63, 64, 76]. It is possible to formulate this model dynamically with the limiting case of this dynamic model, as the population size becomes very large, being the same as the compartmental model.

Past experiences and data have shown that the spread of infection in small populations is better captured in small communities as a random process. For this reason, stochastic models have an important role in disease transmission modeling. The most commonly used stochastic model includes the chain binomial model of Reed and Frost, first described in lectures in 1928 by W.H. Frost but not published until much later [1, 78]. The Reed–Frost model was actually anticipated nearly 40 years earlier by P.D. En’ko [30]. The work of En’ko was brought to public attention much later by K. Dietz [28]. E.B. Wilson and M.H. Burke have given a description of Frost’s 1928 lectures with a somewhat different derivation [78]. M. Greenwood introduced a somewhat different chain binomial model in 1931 [40]. The Reed–Frost model has been used widely as a basic stochastic model and many extensions have been formulated. The book [25] by D.J. Daley and J. Gani contains an account

of some of the more recent extensions. Also, a stochastic analogue of the Kermack–McKendrick epidemic model has been described in [6].

1.2.3 *Developments in Compartmental Models*

In the mathematical modeling of disease transmission, as in most other areas of mathematical modeling, there is always a trade-off between simple, or strategic, models, which omit most details and are designed only to highlight general qualitative behavior, and detailed, or tactical, models, usually designed for specific situations including short-term quantitative predictions. Detailed models are generally difficult or impossible to solve analytically and hence their usefulness for theoretical purposes is limited, although their strategic value may be high.

For example, very simple models for epidemics predict that an epidemic will die out after some time, leaving a part of the population untouched by disease, and this is also true of models that include control measures. This qualitative principle is not by itself very helpful in suggesting what control measures would be most effective in a given situation, but it implies that a detailed model describing the situation as accurately as possible might be useful for public health professionals. The ultimate in detailed models are agent-based models, which essentially divide the population into individuals or groups of individuals with identical behavior [32].

It is important to recognize that mathematical models to be used for making policy recommendations for management need quantitative results, and that the models needed in a public health setting require a great deal of detail in order to describe the situation accurately. For example, if the problem is to recommend what age group or groups should be the focus of attention in coping with a disease outbreak, it is essential to use a model which separates the population into a sufficient number of age groups and recognizes the interaction between different age groups. The development of high speed computing has made it possible to analyze highly detailed models rapidly.

The development of mathematical methods for the study of models for communicable diseases led to a divergence between the goals of mathematicians, who sought broad understanding, and public health professionals, who sought practical procedures for management of diseases. While mathematical modeling led to many fundamental ideas, such as the possibility of controlling smallpox by vaccination and the management of malaria by controlling the vector (mosquito) population, the practical implementation was always more difficult than the predictions of simple models. Fortunately, in recent years there have been determined efforts to encourage better communication, so that public health professionals can better understand the situations in which simple models may be useful and mathematicians can recognize that real-life public health questions are much more complicated than simple models.

In the study of compartmental disease transmission models, the population under study is divided into compartments and assumptions are made about the nature

and time rate of transfer from one compartment to another. For example, in an *SIR* model, we divide the population being studied into three classes labeled S , I , and R . We let $S(t)$ denote the number of individuals who are susceptible to the disease, that is, who are not (yet) infected at time t . $I(t)$ denotes the number of infected individuals, assumed infectious and able to spread the disease by contact with susceptible individuals. $R(t)$ denotes the number of individuals who have been infected and then removed from the possibility of being infected again or of spreading infection. In an *SIS* model, infectives recover with immunity against reinfection and the transitions are from susceptible to infective to susceptible.

The rates of transfer between compartments are expressed mathematically as derivatives with respect to time of the sizes of the compartments. Initially, we assume that the duration of stay in each compartment is exponentially distributed, and as a result models are formulated initially as differential equations. Models in which the rates of transfer depend on the sizes of compartments over the past, as well as at the instant of transfer, lead to more general types of functional equations, such as differential-difference equations or integral equations. One way in which models have expressed the idea of a reduction in contacts as an epidemic proceeds is to assume a contact rate of the form $\beta Sf(I)$ with a function $f(I)$ that grows more slowly than linearly in I . Such an assumption, while not really a mechanistic model, may give better approximation to observed data than mass action contact.

In the simple Kermack–McKendrick epidemic model there are two parameters, the rate of new infections and the recovery rate. Often, the recovery rate for a particular disease is known. In compartmental epidemic models with more compartments, the progression rate is more complicated but may also be known. It is possible to estimate the basic reproduction number if these parameters can be estimated, and there is an equation, known as the *final size relation* relating the basic reproduction number to the number of individuals infected over the course of the epidemic. There have been many presentations of this final size relation in various contexts, including models with heterogeneity of mixing [3, 12, 13, 15, 16, 26, 27, 59].

In their later work on disease transmission models [56, 57], Kermack and McKendrick did not include age of infection, and age of infection models were neglected for many years. Age of infection reappeared in the study of HIV/AIDS, in which the infectivity of infected individuals is high for a brief period after becoming infected, then quite low for an extended period, possibly several years, before increasing rapidly with the onset of full-blown AIDS. Thus the age of infection described by Kermack and McKendrick for epidemics became very important in some endemic situations; see, for example, [74, 75]. Also, HIV/AIDS has pointed to the importance of immunological ideas in the analysis on the epidemiological level.

Typically, there is a stochastic phase at the beginning of a disease outbreak, followed by an exponential increase in the number of infectives, and it may be possible to estimate this initial exponential growth rate experimentally. An estimate of the initial exponential growth rate can be used to estimate the rate of new infections, and this enables estimation of the basic reproduction number

The development and analysis of compartmental models has grown rapidly since the early models. Many of these developments are due to H.W. Hethcote [46–50]. We describe only a few of the important developments. While there are three basic compartmental disease transmission models, namely the *SIS* model, the *SIR* model without births and deaths, and the *SIR* model with births and deaths, each disease has its own properties which should be included in a model. We will describe the effects of adding heterogeneity of mixing in several chapters, including heterogeneity of mixing in Chap. 5, age structure in Chap. 13, and spatial heterogeneity in Chaps. 14 and 15. In addition, the chapters on individual diseases (Chaps. 7–12) include modeling aspects of the specific disease being studied.

For influenza, there is a significant fraction of the population which is infected but asymptomatic, with lower infectivity than symptomatic individuals. There are seasonal outbreaks which may be closely related to the strain of the previous year, that is, modified by point mutations, or it may be less closely related or it may be unrelated. The level of relation of two strains, as measured by the response of hosts that have experienced a prior influenza infection, is what we refer to as strain cross-immunity. Cross-immunity measures the level of protection earned by individuals who were infected by a related strain in a previous year. Influenza models may consider the effect that a partially efficacious vaccination may have before an outbreak, as well as the role of antiviral treatment during an outbreak. Considering multiple modes of transmission is critical when the goal is to minimize the impact of a disease in a population. Cholera may be transmitted both by direct contact and by contact with pathogen shed by infectives. In tuberculosis some individuals progress rapidly to active tuberculosis, while others progress much more slowly. Also, active-tuberculosis individuals who fail to comply with long-term treatment schedules become prime candidates for the development of a drug-resistant strain. In HIV/AIDS, the infectivity of an individual depends strongly on the time since infection, while transmission depends on modes of transmission, mixing between individuals and more. In malaria, immunity against infection is boosted by exposure to infection or by genetics (sickle cell anemia).

1.2.4 Endemic Disease Models

The analytic approaches to models for endemic diseases and epidemics are quite different. The analysis of a model for an endemic disease, carried out in Chap. 3, begins with the search for equilibria, which are, by definition, constant solutions of the model. Usually there is a disease-free equilibrium and there are one or more endemic equilibria, with a positive number of infected individuals. The next step is to linearize about each equilibrium and determine the stability of each equilibrium. Usually, if the basic reproduction number is less than 1, the only equilibrium is the disease-free equilibrium and this equilibrium is asymptotically stable. If the basic reproduction number is greater than 1, the usual situation is that the disease-free equilibrium is unstable and there is a unique endemic equilibrium which is

asymptotically stable. This approach also covers diseases in which there is vertical transmission, which is direct transmission from mother to offspring at birth [19].

However, more complicated behavior is possible. For example, if there are two strains of the disease being studied it is common to have regions in the parameter space in which there is an asymptotically stable equilibrium with only one of the strains present and a region in which there is an asymptotically stable equilibrium with both strains coexisting. Another possibility is that there is a unique endemic equilibrium but it is unstable. In this situation, there is often a Hopf bifurcation and an asymptotically stable periodic orbit around the endemic equilibrium. An example of such behavior may be found in an SIRS model, with a temporary immunity period of fixed length following recovery [51] and in an *SVIR* model [38]. If there is a periodic orbit with large amplitude and a long period, data must be gathered over a sufficiently large time interval to give an accurate picture.

Another possible behavior is a backward bifurcation. As \mathcal{R}_0 increases through 1 there is an exchange of stability between the disease-free equilibrium, which is asymptotically stable for $\mathcal{R}_0 < 1$ and unstable for $\mathcal{R}_0 > 1$, and the endemic equilibrium which exists if $\mathcal{R}_0 > 1$. The usual transition is a forward, or transcritical, bifurcation at $\mathcal{R}_0 = 1$, with an asymptotically stable endemic equilibrium and an equilibrium infective population size depending continuously on \mathcal{R}_0 .

The behavior at a bifurcation may be described graphically by the bifurcation curve, which is the graph of the infective population size I at equilibrium as a function of the basic reproduction number \mathcal{R}_0 . It has been noted [29, 42, 43, 58] that in epidemic models with multiple groups and asymmetry between groups or multiple interaction mechanisms it is possible to have a very different bifurcation behavior at $\mathcal{R}_0 = 1$. There may be multiple positive endemic equilibria for values of $\mathcal{R}_0 < 1$ and a backward bifurcation at $\mathcal{R}_0 = 1$. The qualitative behavior of a system with a backward bifurcation differs from that of a system with a forward bifurcation and the nature of these changes has been described in [11]. Since these behavioral differences are important in planning how to control a disease, it is important to determine whether a system can have a backward bifurcation. In the presence of two modes of sexually transmitted HIV, it was shown that multiple endemic equilibrium could be supported [53].

1.2.5 Diseases Transmitted by Vectors

Many diseases are transmitted from human to human indirectly, through a vector. Vectors are living organisms that can transmit infectious diseases between humans. Many vectors are bloodsucking insects that ingest disease-producing microorganisms during blood meals from an infected (human) host, and then inject it into a new host during a subsequent blood meal. The best known vectors are mosquitoes for diseases including malaria, dengue fever, chikungunya, Zika virus, Rift Valley fever, yellow fever, Japanese encephalitis, lymphatic filariasis, and West Nile fever, but ticks (for Lyme disease and tularemia), bugs (for Chagas' disease), flies (for

onchocerciasis), sandflies (for leishmaniasis), fleas (for plague, transmitted by fleas from rats to humans), and some freshwater snails (for schistosomiasis) are vectors for some diseases.

Every year there are more than a billion cases of vector-borne diseases and more than a million deaths. Vector-borne diseases account for over 17% of all infectious diseases worldwide. Malaria is the most deadly vector-borne diseases, causing an estimated 627,000 deaths in 2012. The most rapidly growing vector-borne disease is dengue, for which the number of cases has multiplied by 30 in the last 50 years. These diseases are found more commonly in tropical and sub-tropical regions where mosquitoes flourish, and in places where access to safe drinking water and sanitation systems is uncertain.

Some vector-borne diseases such as dengue, chikungunya, and West Nile virus are emerging in countries where they were unknown previously because of globalization of travel and trade and environmental challenges such as climate change. A troubling new development is the Zika virus, which has been known since 1952 but has developed a mutation in the South American outbreak of 2015 [70] which has produced very serious birth defects in babies born to infected mothers. In addition, the current Zika virus can be transmitted directly through sexual contact as well as through vectors. Chapter 6 is an introduction to the modeling of vector-borne diseases. Chapter 11 on malaria and Chap. 12 on dengue fever and the Zika virus describe modeling of specific vector-borne diseases.

Many of the important underlying ideas of mathematical epidemiology arose in the study of malaria begun by Sir R.A. Ross [69]. Malaria is one example of a disease with vector transmission, the infection being transmitted back and forth between vectors (mosquitoes) and hosts (humans). It kills hundreds of thousands of people annually, mostly children and mostly in poor countries in Africa. Among communicable diseases, only tuberculosis causes more deaths. Other vector diseases include West Nile virus, yellow fever, and dengue fever. Human diseases transmitted heterosexually may also be viewed as diseases transmitted by vectors, because males and females must be viewed as separate populations and disease is transmitted from one population to the other.

Vector-transmitted diseases require models that include both vectors and hosts. For most diseases transmitted by vectors, the vectors are insects, with a much shorter life span than the hosts, who may be humans as for malaria or animals as for West Nile virus, although there is malaria (not human malaria) in various animal populations and West Nile virus has infected humans as far as Arizona in the USA.

The compartmental structure of the disease may be different in host and vector species; for many diseases with insects as vectors an infected vector remains infected for life so that the disease may have an *SI* or *SEI* structure in the vectors and an *SIR* or *SEIR* structure in the hosts.

1.2.6 Heterogeneity of Mixing

In disease transmission models not all members of the population make contacts at the same rate. In sexually transmitted diseases there is often a “core” group of very active members who are responsible for most of the disease cases, and control measures aimed at this core group have been very effective in control [52]. In epidemics there are often “super-spreaders,” who make many contacts and are instrumental in spreading disease and in general some members of the population make more contacts than others. Chapter 5 deals with models for diseases with heterogeneous mixing and includes description of a general method (the next generation matrix) for determining the basic reproduction number for models with heterogeneous mixing. To model heterogeneity in mixing we may assume that the population is divided into subgroups with different activity levels. Formulation of models requires some assumptions about the mixing between subgroups. There have been many studies of mixing patterns in real populations, for example, [9, 10, 18, 37, 68].

It has often been observed in epidemics that most infectives do not transmit infections at all or transmit infections to very few others. This suggests that homogeneous mixing at the beginning of an epidemic may not be a good approximation.

The SARS epidemic of 2002–2003 spread much more slowly than would have been expected on the basis of the data on disease spread at the start of the epidemic. Early in the SARS epidemic of 2002–2003 it was estimated that \mathcal{R}_0 had a value between 2.2 and 3.6. At the beginning of an epidemic, the exponential rate of growth of the number of infectives is approximately $(\mathcal{R}_0 - 1)/\alpha$, where $1/\alpha$ is the generation time of the epidemic, estimated to be approximately 10 days for SARS. This would have predicted at least 30,000 cases of SARS in China during the first 4 months of the epidemic. In fact, there were fewer than 800 cases reported in this time. One explanation for this discrepancy is that the estimates were based on transmission data in hospitals and crowded apartment complexes or in the scaling of the model used to estimate parameters [24]. It was observed that there was intense activity in some locations and very little in others. This suggests that the actual reproduction number (averaged over the whole population) was much lower, perhaps in the range 1.2–1.6, and that heterogeneous mixing was a very important aspect of the epidemic.

Age is one of the most important characteristics in the modeling of populations and infectious diseases. Individuals with different ages may have different reproduction and survival capacities. Diseases may have different infection rates and mortality rates for different age groups. Individuals of different ages may also have different behaviors, and behavioral changes are crucial in control and prevention of many infectious diseases. Young individuals tend to be more active in interactions with or between populations, and in disease transmissions. Age-structured models are studied in Chap. 13.

Sexually transmitted diseases (STDs) are spread through partner interactions with pair-formations and the pair-formations process is age-dependent in most cases. For example, most HIV cases occur in the group of young adults.

Childhood diseases, such as measles, chicken pox, and rubella, are spread mainly by contacts between children of similar ages. More than half of the deaths attributed to malaria are in children under 5 years of age due to their weaker immune systems. This suggests that in models for disease transmission in an age-structured population it is necessary to allow the contact rates between two members of the population to depend on the ages of both members. Another important motivation for using age-structured models for childhood diseases is that vaccination is age-dependent (e.g., measles).

The development of age-structured models for disease transmission required development of the theory of age-structured populations. In fact, the first models for age-structured populations [61] were designed for the study of disease transmission in such populations.

1.3 Strategic Models and This Volume

This book is intended for both mathematicians and public health professionals. However, it is a book aimed primarily at practitioners who are quantitatively trained. For readers less familiar with mathematics, we are providing a website that includes notes on calculus, linear algebra, ODEs, and difference equations. We would like to repeat that it is not a book about *mathematics* per se. Hence, *mathematical rigor is not a priority* albeit we have tried to document some assertions. We wish to present the truth and nothing but the truth, but not necessarily the whole truth. Our hope is that public health people will be sufficiently interested to review mathematics if necessary and make use of the book.

We do not cover stochastic models, except for some material in Sects. 4.1 and 4.2. Some presentations of stochastic models may be found in [2, 4, 5, 25, 41].

Also, we do not cover discrete models in the main text. This is a deliberate choice that we have made to limit the size of this volume, despite the fact that a lot can be learned by the appropriate formulation of discrete models, which are unfortunately often seen as a discretization of continuous time models, that is, they are not derived directly from first principles, deviating the focus from the study of disease dynamics to the mathematical question of whether or not it is a proper discretization of a continuous time model. However, there are several projects that could give an introduction to this topic, and some references are [14, 21–23, 79].

The primary goal of this volume is to cover relatively simple models to indicate what qualitative results should be expected from more detailed tactical models. The chapters build on the work that we have done in applying epidemiological models in the context of specific diseases over the past 25 years—including some of our most recent work. It is our hope that the questions and problems highlighted in this introduction may help build collaborations between modelers, epidemiologists, and public health experts.

References

1. Abbey, H. (1952) An estimation of the Reed–Frost theory of epidemics, *Human Biol.* **24**; 201–223.
2. Allen, L.J.S. (2003) *An Introduction to Stochastic Processes with Applications to Biology*, Pearson Education Inc.
3. Andreasen, V. (2011) The final size of an epidemic and its relation to the basic reproduction number, *Bull. Math. Biol.* **73**: 2305–2321.
4. Bailey, N.T.J. (1975) *The Mathematical Theory of Infectious Diseases and its Applications*, Oxford Univ. Press.
5. Ball, F.G. (1983) The threshold behaviour of epidemic models, *J. App. Prob.* **20**: 227–241.
6. Bartlett, M.S. (1949) Some evolutionary stochastic processes *J. Roy. Stat. Soc. B* **11**: 211–229.
7. Bernoulli, D. (1760) Réflexions sur les avantages de l'inoculation, *Mercure de Paris*, p. 173.
8. Bernoulli, D. (1766) Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, *Mem. Math. Phys. Acad. Roy. Sci. Paris*, 1–45.
9. Blythe, S.P., S. Busenberg & C. Castillo-Chavez (1995) Affinity and paired-event probability, *Math. Biosc.* **128**: 265–84 .
10. Blythe, S.P., C. Castillo-Chavez, J. Palmer & M. Cheng (1991) Towards a unified theory of mixing and pair formation, *Math. Biosc.* **107**: 379–405.
11. Brauer, F. (2004) Backward bifurcations in simple vaccination models, *J. Math. Anal. & Appl.* **298**: 418–431.
12. Brauer, F. (2008) Epidemic models with treatment and heterogeneous mixing, *Bull. Math. Biol.* **70**: 1869–1885.
13. Brauer, F. (2008) Age of infection models and the final size relation, *Math. Biosc. & Eng.* **5**: 681–690.
14. Brauer, F., C. Castillo-Chavez, and Z. Feng (2010) Discrete epidemic models, *Math. Biosc. & Eng.* **7**: 1–15.
15. Brauer, F. & J. Watmough (2009) Age of infection models with heterogeneous mixing, *J. Biol. Dyn.* **3**: 324–330.
16. Breda, D., O. Diekmann, W.F. deGraaf, A. Pugliese, & R. Vermiglio (2012) On the formulation of epidemic models (an appraisal of Kermack and McKendrick, *J. Biol. Dyn.* **6**: 103–117.
17. Budd, W. (1873) *Typhoid Fever; Its Nature, Mode of Spreading, and Prevention*, Longmans, London.
18. Busenberg, S. & C. Castillo-Chavez (1989) Interaction, pair formation and force of infection terms in sexually transmitted diseases, In *Mathematical and Statistical Approaches to AIDS Epidemiology*, Lect. Notes Biomath. **83**, C. Castillo-Chavez (ed.), Springer-Verlag, Berlin-Heidelberg-New York, pp. 289–300.
19. Busenberg, S. & K.L. Cooke (1993) *Vertically Transmitted Diseases: Models and Dynamics*, Biomathematics **23**, Springer-Verlag, Berlin-Heidelberg-New York.
20. Callaway, D.S., M.E.J. Newman, S.H. Strogatz, & D.J. Watts (2000) Network robustness and fragility: Percolation on random graphs, *Phys. Rev. Letters*, **85**: 5468–5471.
21. Castillo-Chavez, C. & A. Yakubu (2001) Discrete-time *SIS* models with simple and complex dynamics, *Nonlinear Analysis, Theory. Methods, and Applications* **47**: 4753–4762.
22. Castillo-Chavez, C. & A. Yakubu (2002) Discrete-time *SIS* models with simple and complex population dynamics, in *Mathematical Approaches for Emerging and Reemerging Infectious Diseases IMA Vol.* **125**; 153–164.
23. Castillo-Chavez, C. & A. Yakubu (2002) Intraspecific competition, dispersal and disease dynamics in discrete-time patchy environments, in *Mathematical Approaches for Emerging and Reemerging Infectious Diseases IMA Vol.* **125**; 165–181.
24. Chowell, G., P.W. Fenimore, M.A. Castillo-Garsow, and C.Castillo-Chavez (2003) SARS outbreaks in Ontario, Hong Kong and Singapore: the role of diagnosis and isolation as a control mechanism. *J. Theor. Biol.* **224**: 1–8.

25. Daley, D.J. & J. Gani (1999) Epidemic Models: An Introduction, Cambridge Studies in Mathematical Biology 15, Cambridge University Press.
26. Diekmann, O. & J.A.P. Heesterbeek (2000) Mathematical epidemiology of infectious diseases: Model building, analysis and interpretation, John Wiley and Sons, ltd.
27. Diekmann,O., Heesterbeek,J.A.P., Metz,J.A.J. (1995) The legacy of Kermack and McKendrick, in Mollison,D. (ed)Epidemic Models: Their Structure and Relation to Data. Cambridge University Press, Cambridge, pp. 95–115.
28. Dietz, K, (1988) The first epidemic model: a historical note on P.D. En'ko, Australian J. Stat. **30A**: 56–65.
29. Dushoff, J., W. Huang, & C. Castillo-Chavez (1998) Backwards bifurcations and catastrophe in simple models of fatal diseases, J. Math. Biol. **36**: 227–248.
30. En'ko, P.D. (1889) On the course of epidemics of some infectious diseases, Vrach. St. Petersburg, X: 1008–1010, 1039–1042, 1061–1063. [translated from Russian by K. Dietz, Int. J. Epidemiology (1989) **18**: 749–755].
31. Dietz, K. & J.A.P. Heesterbeek (2002) Daniel Bernoulli's epidemiological model revisited, Math. Biosc. **180**: 1–221.
32. Epstein, J.M. & R.L. Axtell (1996) Growing Artificial Societies: Social Science from the Bottom Up, MIT/Brookings Institution.
33. Erdős, P. & A. Rényi (1959) On random graphs, Publicationes Mathematicae **6**: 290–297.
34. Erdős, P. & A. Rényi (1960) On the evolution of random graphs, Publ. Math. Inst. Hung. Acad. Science **5**: 17–61.
35. Erdős, P. & A. Rényi (1961) On the strengths of connectedness of a random graph, Acta Math. Scientiae Hung. **12**: 261–267.
36. Farr, W. (1840) Progress of epidemics, Second Report of the Registrar General of England and Wales, 91–98.
37. Feng, Z., Hill, A.N., Smith, P.J. Glasser, J.W. (2015) An elaboration of theory about preventing outbreaks in homogeneous populations to include heterogeneity or preferential mixing, J. Theor. Biol. **386**: 177–187.
38. Feng, Z. and H.R. Thieme (1995) Recurrent outbreaks of childhood diseases revisited: the impact of isolation, Math. Biosc. **128**: 93–130.
39. Galton, F. (1889) Natural Inheritance (2 nd ed.), App. F, pp. 241–248.
40. Greenwood, M. (1931) On the statistical measure of infectiousness, J. Hygiene **31**: 336–351.
41. Greenwood, P.E. & L.F. Gordillo (2009) Stochastic epidemic modeling, in Mathematical and Statistical Estimation Approaches in Epidemiology, G. Chowell, J.M. Hyman, L.M.A. Bettencourt, & C. Castillo-Chavez (eds.), Springer, pp. 31–52.
42. Hadeler, K.P. & C. Castillo-Chavez (1995) A core group model for disease transmission, Math Biosc. **128**: 41–55.
43. Hadeler, K.P. & P. van den Driessche (1997) Backward bifurcation in epidemic control, Math. Biosc. **146**: 15–35.
44. Hamer, W.H. (1906) Epidemic disease in England - the evidence of variability and of persistence, The Lancet **167**: 733–738.
45. Harris, T.E. (1963) The Theory of Branching Processes, Springer.
46. Hethcote, H.W. (1976) Qualitative analysis for communicable disease models, *Math. Biosc.*, **28**:335–356.
47. Hethcote, H.W. (1978) An immunization model for a heterogeneous population, *Theor. Pop. Biol.*, **14**:338–349.
48. Hethcote, H.W. (1989) Three basic epidemiological models. In *Applied Mathematical Ecology*, S. A. Levin, T.G. Hallam and L.J. Gross, eds., Biomathematics **18**: 119–144, Springer-Verlag, Berlin-Heidelberg-New York.
49. Hethcote, H.W. (1997) An age-structured model for pertussis transmission, Math. Biosci., **145**: 89–136.
50. Hethcote, H.W. (2000) The mathematics of infectious diseases, *SIAM Review*, **42**: 599–653.
51. Hethcote, H.W., H.W. Stech and P. van den Driessche (1981) Periodicity and stability in epidemic models: a survey. In: S. Busenberg & K.L. Cooke (eds.) *Differential Equations and*

- Applications in Ecology, Epidemics and Population Problems*, Academic Press, New York: 65–82.
52. Hethcote, H.W. & J.A. Yorke (1984) Gonorrhea Transmission Dynamics and Control, Lect. Notes in Biomath. 56, Springer-Verlag.
 53. Huang, W., K.L. Cooke, & C. Castillo-Chavez (1992) Stability and bifurcation for a multiple group model for the dynamics of HIV/AIDS transmission, SIAM J. App. Math. **52**: 835–853.
 54. Johnson, S. (2006) *The Ghost Map*, Riverhead Books, New York,
 55. Kermack, W.O. & A.G. McKendrick (1927) A contribution to the mathematical theory of epidemics, *Proc. Royal Soc. London*, **115**:700–721.
 56. Kermack, W.O. & A.G. McKendrick (1932) Contributions to the mathematical theory of epidemics, part. II, *Proc. Roy. Soc. London*, **138**:55–83.
 57. Kermack, W.O. & A.G. McKendrick (1933) Contributions to the mathematical theory of epidemics, part. III, *Proc. Roy. Soc. London*, **141**:94–112.
 58. Kribs-Zaleta, C.K. & J.X. Velasco-Hernandez (2000) A simple vaccination model with multiple endemic states, *Math Biosc.* **164**: 183–201.
 59. Ma, J. & D.J.D. Earn (2006) Generality of the final size formula for an epidemic of a newly invading infectious disease, *Bull. Math. Biol.* **68**: 679–702.
 60. MacDonald, G. (1957) *The Epidemiology and Control of Malaria*, Oxford University Press. Oxford University Press.
 61. McKendrick, A.G. (1926) Applications of mathematics to medical problems, *Proc. Edinburgh Math. Soc.* **44**: 98–130.
 62. Metz, J.A.J. (1978) The epidemic in a closed population with all susceptibles equally vulnerable; some results for large susceptible populations and small initial infections, *Acta Biotheoretica* **78**: 75–123.
 63. Miller, J.C. (2011) A note on a paper by Erik Volz: SIR dynamics in random networks, *J. Math. Biol.* **62**: 349–358.
 64. Miller, J.C., & E. Volz (2013) Incorporating disease and population structure into models of SIR disease in contact networks, *PLoS One* **8**: e69162, <https://doi.org/10.1371/journal.pone.0069162>.
 65. Newman, M.E.J. (2002) The spread of epidemic disease on networks, *Phys. Rev. E* **66**, 016128. ks, *Phys. Rev. E* **66**, 016128.
 66. Newman, M.E.J., S.H. Strogatz, & D.J. Watts (2001) Random graphs with arbitrary degree distributions and their applications, *Phys. Rev. E* **64**, 026118.
 67. Nowak, M., & R.M. May (1996) *Virus Dynamics: Mathematical Principles of Immunology and Virology*, Oxford Univ. Press.
 68. Nold, A. (1980) Heterogeneity in disease transmission modeling, *Math. Biosc.* **52**: 227–240.
 69. Ross, R.A.(1911) *The prevention of malaria* (2nd edition, with Addendum). John Murray, London.
 70. Schuler-Faccini, L. (2016) Possible association between Zika virus infections and microcephaly Brazil 2015, MMWR Morbidity and Mortality weekly report: 65.
 71. Snow, J. (1855) *The mode of communication of cholera* (2 nd ed.), Churchill, London.
 72. Steffensen, G.F. (1930) Om sandsynligheden for at affkommet uddør, *Matematisk Tidsskrift B* **1**: 19–23.
 73. Steffensen, G.F. (1931) Deux problèmes du calcul des probabilités, *Ann. Inst. H. Poincaré* **3**: 319–341.
 74. Thieme, H.R. and C. Castillo-Chavez (1989) On the role of variable infectivity in the dynamics of the human immunodeficiency virus. In *Mathematical and statistical approaches to AIDS epidemiology*, C. Castillo-Chavez, ed., Lect. Notes Biomath. **83**, 200–217, Springer-Verlag, Berlin-Heidelberg-New York.
 75. Thieme, H.R. and C. Castillo-Chavez (1993) How may infection-age dependent infectivity affect the dynamics of HIV/AIDS?, *SIAM J. Appl. Math.*, **53**:1447–1479.

76. Volz, E. (2008) SIR dynamics in random networks with heterogeneous connectivity, *J. Math. Biol.*, **56**: 293–310.
77. Watson, H.W. & F. Galton (1874) On the probability of the extinction of families, *J. Anthropol. Inst. Great Britain and Ireland* **4**: 138–144.
78. Wilson, E.B. & M.H. Burke (1942) The epidemic curve, *Proc. Nat. Acad. Sci.* **28**: 361–367.
79. Yakubu, A. & C. Castillo-Chavez (2002) Interplay between local dynamics and dispersal in discrete-time metapopulation models, *J. Theor. Biol.* **218**: 273–288.

Chapter 2

Simple Compartmental Models for Disease Transmission



Communicable diseases that are endemic (always present in a population) cause many deaths. For example, in 2011 tuberculosis caused an estimated 1,400,000 deaths and HIV/AIDS caused an estimated 1,200,000 deaths worldwide. According to the World Health Organization there were 627,000 deaths caused by malaria, but other estimates put the number of malaria deaths at 1,200,000. Measles, which is easily treated in the developed world, caused 160,000 deaths in 2011, but in 1980 there were 2,600,000 measles deaths. The striking reduction in measles deaths is due to the availability of a measles vaccine. Other diseases such as typhus, cholera, schistosomiasis, and sleeping sickness are endemic in many parts of the world. The effects of high disease mortality on mean life span and of disease debilitation and mortality on the economy in afflicted countries are considerable. Most of these disease deaths are in less developed countries, especially in Africa, where endemic diseases are a huge barrier to development. A reference describing the properties of many endemic diseases is [2].

For diseases that are endemic in some region, public health physicians would like to be able to estimate the number of infectives at a given time as well as the rate at which new infections arise. The effects of quarantine or vaccine in reducing the number of victims are of importance. In addition, the possibility of defeating the endemic nature of the disease and thus controlling or even eradicating the disease in a population is worthy of study.

An epidemic, which acts on a short temporal scale, may be described as a sudden outbreak of a disease that infects a substantial portion of the population in a region before disappearing. Epidemics invariably leave part of the population untouched. Often, epidemic outbreaks recur with intervals of several years between outbreaks, possibly decreasing in severity as populations develop some immunity.

The “Spanish” flu epidemic of 1919–1920 caused an estimated 50,000,000 or more deaths worldwide. The AIDS epidemic, the SARS epidemic of 2002–2003, recurring influenza pandemics such as the H1N1 influenza pandemic of 2009–2010,

and outbreaks of diseases such as the Ebola virus are events of concern and interest to many people.

The essential difference between an endemic disease and an epidemic is that in an endemic disease there is some mechanism for a flow of susceptibles into the population being studied through births of new susceptibles, immigration of susceptibles, recovery from infection without immunity against reinfection, or loss of immunity of recovered individuals. This may result in a level of infection that remains in the population. In an epidemic, there is no flow of new susceptibles into the population and the number of infectives decreases to zero because of the resulting scarcity of susceptibles.

There are many questions of interest to public health physicians confronted with a possible epidemic. For example, how severe will an epidemic be? This question may be interpreted in a variety of ways. For example, how many individuals will be affected and require treatment? What is the maximum number of people needing care at any particular time? How long will the epidemic last? Can an epidemic be averted by vaccination of enough members of the population in advance of the epidemic? How much good would quarantine of victims do in reducing the severity of the epidemic?

The idea of invisible living creatures as agents of disease goes back at least to the writings of Aristotle (384–322 BC) and was developed as a theory in the sixteenth century. The existence of microorganisms was demonstrated by Leeuwenhoek (1632–1723) with the aid of the first microscopes. The first expression of the germ theory of disease by Jacob Henle (1809–1885) came in 1840 and was developed by Robert Koch (1843–1910), Joseph Lister (1827–1912), and Louis Pasteur (1827–1875) in the latter part of the nineteenth century and the early part of the twentieth century.

The mechanism of transmission of infections is now known for most diseases. Generally, diseases transmitted by viral agents, such as influenza, measles, rubella (German measles), and chicken pox, confer immunity against reinfection, while diseases transmitted by bacteria, such as tuberculosis, meningitis, and gonorrhea, confer no immunity against reinfection. Other diseases, such as malaria, are transmitted not directly from human to human but by vectors, agents (usually insects) who are infected by humans and who then transmit the disease to humans. Heterosexual transmission of HIV/AIDS is also a vector process in which transmission goes back and forth between males and females.

In this chapter, our goal is to present and analyze simple compartmental models for disease transmission, both for endemic situations and for epidemics. Here we seek only to introduce the idea of the *basic reproduction number* and how it determines a threshold between two different outcomes. In later chapters we will study models with more detailed structure and will include the effects of different methods to try to control a disease.

2.1 Introduction to Compartmental Models

We formulate our descriptions of disease transmission as *compartmental models*, with the population under study being divided into compartments and with assumptions about the nature and time rate of transfer from one compartment to another. We divide the population being studied into three classes labeled S , I , and R . Let $S(t)$ denote the number of individuals who are susceptible to the disease, that is, who are not (yet) infected at time t . $I(t)$ denotes the number of infected individuals, assumed infective and able to spread the disease by contact with susceptibles. $R(t)$ denotes the number of individuals who have been infected and then removed from the possibility of being infected again or of spreading infection. Removal is carried out either through isolation from the rest of the population, or through immunization against infection, or through recovery from the disease with full immunity against reinfection, or through death caused by the disease. These characterizations of removed members are different from an epidemiological perspective but are often equivalent from a modeling point of view which takes into account only the state of an individual with respect to the disease.

In many diseases, infectives return to the susceptible class on recovery because the disease confers no immunity against reinfection. Such models are appropriate for most diseases transmitted by bacterial agents or helminth agents, and most sexually transmitted diseases (including gonorrhea, but not such diseases as AIDS from which there is no recovery). We use the terminology *SIS* to describe a disease with no immunity against reinfection, to indicate that the passage of individuals is from the susceptible class to the infective class and then back to the susceptible class.

We will use the terminology *SIR* to describe a disease which confers immunity against reinfection, to indicate that the passage of individuals is from the susceptible class S to the infective class I to the removed class R . Usually, diseases caused by a virus are of *SIR* type.

In addition to the basic distinction between diseases for which recovery confers immunity against reinfection and diseases for which recovered members are susceptible to reinfection, and the intermediate possibility of temporary immunity signified by a model of *SIRS* type, more complicated compartmental structure is possible. For example, there are *SEIR* and *SEIS* models, with an exposed period between being infected and becoming infective.

We are assuming that the disease transmission process is *deterministic*, that is, that the behavior of a population is determined completely by its history and by the rules which describe the model. In formulating models in terms of the derivatives of the sizes of each compartment, we are also assuming that the number of members in a compartment is a differentiable function of time. This assumption is plausible in describing an endemic state. However, in Chap. 4 we will describe the modeling of disease outbreaks and epidemics. At the beginning of a disease outbreak, there are only a few infectives and the start of a disease outbreak depends on random contacts with a small number of infectives. This will require a different approach

that we shall study in Sect. 4.1, but for the present we begin with a description of deterministic compartmental models.

The independent variable in our compartmental models is the time t , and the rates of transfer between compartments are expressed mathematically as derivatives with respect to time of the sizes of the compartments, and as a result our models are formulated initially as differential equations. Models in which the rates of transfer depend on the sizes of compartments over the past as well as at the instant of transfer lead to more general types of functional equations, such as differential-difference equations or integral equations. In this chapter, we will always assume that the duration of stay in each compartment is exponentially distributed so that our models will be systems of ordinary differential equations. In Chaps. 3 and 4 we will begin the study of more general classes of models.

In order to formulate a compartmental disease transmission model, we will need to make assumptions on the rates of flow between compartments. In the simplest models we require expressions for the rate of being infected and the rate of recovery from infection. The most common assumption about transmission of infection is *mass action incidence*. It is assumed that in a population of size N on average an individual makes βN contacts sufficient to transmit infection in unit time. Another possible assumption is *standard incidence*, in which it is assumed that on average an individual makes a constant number a of contacts sufficient to transmit infection in unit time. Standard incidence is a common assumption for sexually transmitted diseases. More realistically, one might assume that the parameters β or a are not constants, but are functions of total population size. We will always assume that contacts are *effective*. By this we mean that if there is a contact between an infective individual and a susceptible individual, infection is always passed to the susceptibles. However, it is possible that infectives in one compartment are less infectious than individuals in another compartment. In such situations, we may include a reduction factor in some contacts to model this.

With mass action incidence, (contact rate βN) since the probability that a random contact by an infective is with a susceptible is S/N , the number of new infections in unit time per infective is $(\beta N)(S/N)$, giving a rate of new infections $(\beta N)(S/N)I = \beta SI$. Alternately, we may argue that for a contact by a susceptible the probability that this contact is with an infective is I/N and thus the rate of new infections per susceptible is $(\beta N)(I/N)$, giving a rate of new infections $(\beta N)(I/N)S = \beta SI$. Note that both approaches give the same rate of new infections; in models with more complicated compartmental structure one approach may be more appropriate than the other.

With standard incidence, (contact rate a) similar reasoning leads to the result that the rate of new infections is

$$aS \frac{I}{N}.$$

If the total population size N is constant, we may use a and βN interchangeably. We will most commonly use βN , partly for historical reasons. However, for sexually

transmitted diseases the standard incidence form for contacts is more common. When we study heterogeneous mixing in Chap. 5 it will be more convenient to use the standard incidence form but it would be possible to use the mass action form instead.

A common assumption is that infectives leave the infective class at a constant rate αI per unit time. This assumption requires a fuller mathematical explanation, since the assumption of a recovery rate proportional to the number of infectives has no clear epidemiological meaning. We consider the “cohort” of members who were all infected at one time and let $u(s)$ denote the number of these who are still infective s time units after having been infected. If a fraction α of these leave the infective class in unit time then

$$u' = -\alpha u,$$

and the solution of this elementary differential equation is

$$u(s) = u(0) e^{-\alpha s}.$$

Thus, the fraction of infectives remaining infective s time units after having become infective is $e^{-\alpha s}$, so that the length of the infective period is distributed exponentially with mean $\int_0^\infty e^{-\alpha s} ds = 1/\alpha$, and this is what is really assumed. This assumption is made for simplicity as it leads to ordinary differential equation models, but in a later chapter we will study models with other distributions of the infective period.

To see that the mean infective period is $1/\alpha$, we point out since the fraction of infectives remaining infective s time units after being infected is $e^{-\alpha s}$, the fraction with infective period exactly s is

$$-\frac{d}{ds} e^{-\alpha s}$$

and the mean infective period is

$$\int_0^\infty s \frac{d}{ds} e^{-\alpha s} ds.$$

Integration by parts shows that this is

$$-\int_0^\infty e^{-\alpha s} ds = \frac{1}{\alpha}.$$

In this chapter, we will analyze *SIS* and *SIR* models for disease transmission, both with and without demographics (births and natural deaths). Our goal is to describe both endemic and epidemic situations in the simplest possible contexts. In later chapters, we will analyze more complicated models, including more

compartmental structure, more general distributions of stay in compartments, and heterogeneity of mixing. In this chapter we focus on equilibrium analysis and asymptotic behavior of models for endemic diseases and on final size relations for epidemic models.

A constant solution y_0 of a differential equation

$$y' = f(y), \quad (2.1)$$

meaning a solution of the equation $f(y) = 0$, is called an equilibrium of the differential equation. An equilibrium y_0 is said to be *stable* if every solution $y(t)$ of the differential equation (2.1) with initial value $y(0)$ sufficiently close to y_0 remains close to the equilibrium y_0 for all $t \geq 0$. An equilibrium is said to be *asymptotically stable* if it is stable and if in addition every solution with initial value sufficiently close to y_0 approaches the equilibrium as $t \rightarrow \infty$. We note that in this definition, there is no specification of the meaning of “sufficiently close.” Asymptotic stability is a local concept. An equilibrium y_0 is said to be *globally asymptotically stable* if it is stable and if solutions of (2.1) for all initial values $y(0)$ approach y_0 as $t \rightarrow \infty$. An equilibrium that is not stable is said to be *unstable*. The concepts of asymptotic stability and instability are essential for the qualitative analysis of differential equations. An important basic result is that an equilibrium y_0 of a differential equation (2.1) is asymptotically stable if $f'(y_0) < 0$ and unstable if $f'(y_0) > 0$. The case $f'(y_0) = 0$ is more difficult to analyze.

A central concept in both endemic and epidemic models is the *basic reproduction number*, to be defined in Sect. 2.2.

In the next two sections, we will discuss models for endemic situations, and in the following two sections we will discuss epidemic models.

2.2 The SIS Model

The basic compartmental models to describe the transmission of communicable diseases are contained in a sequence of three papers by W.O. Kermack and A.G. McKendrick in 1927, 1932, and 1933 [17–19].

The simplest *SIS* model, due to Kermack and McKendrick [18], is

$$\begin{aligned} S' &= -\beta SI + \alpha I \\ I' &= \beta SI - \alpha I. \end{aligned} \quad (2.2)$$

It is based on the following assumptions:

- (i) The rate of new infections is given by mass action incidence.
- (ii) Infectives leave the infective class at rate αI per unit time and return to the susceptible class.

- (iii) There is no entry into or departure from the population.
- (iv) There are no disease deaths, and the total population size is a constant N .

In an *SIS* model, the total population size N is equal to $S + I$. Later, we will allow the possibility that some infectives recover while others die of the disease to give a more general model. The hypothesis (iii) really says that the time scale of the disease is much faster than the time scale of births and deaths so that demographic effects on the population may be ignored.

Because (2.2) implies $(S + I)' = 0$, the total population $N = S + I$ is a constant. We may reduce the model (2.2) to a single differential equation by replacing S by $N - I$

$$\begin{aligned} I' &= \beta I(N - I) - \alpha I = (\beta N - \alpha)I - \beta I^2 \\ &= (\beta N - \alpha)I \left(1 - \frac{I}{N - \frac{\alpha}{\beta}}\right). \end{aligned} \quad (2.3)$$

Now (2.3) is a logistic differential equation of the form

$$I' = rI \left(1 - \frac{I}{K}\right),$$

with $r = \beta N - \alpha$ and with $K = N - \alpha/\beta$. Let us recall the analysis of the logistic equation.

The logistic equation may be solved explicitly using separation of variables; the solution satisfying the initial condition $x(0) = x_0$ is

$$x(t) = \frac{Kx_0 e^{rt}}{K - x_0 + x_0 e^{rt}} = \frac{Kx_0}{x_0 + (K - x_0)e^{-rt}}. \quad (2.4)$$

The expression (2.4) for the solution of the logistic initial value problem shows that $x(t)$ approaches the limit K as $t \rightarrow \infty$ if $x_0 > 0$ provided $r > 0, K > 0$. If $r < 0, K < 0, x(t)$ approaches the limit 0 as $t \rightarrow \infty$. This case, $K < 0$ arises in the analysis of the differential equation (2.3), but does not have any biological meaning in itself.

This qualitative result tells us that for the model (2.3), if $\beta N - \alpha < 0$ or $\beta N/\alpha < 1$, then all solutions with non-negative initial values approach the limit zero as $t \rightarrow \infty$, (the constant solution $I = N - \beta/\alpha$ is negative), while if $\beta N/\alpha > 1$, then all solutions with non-negative initial values except the constant solution $I = 0$ approach the limit $N - \alpha/\beta > 0$ as $t \rightarrow \infty$. Thus there is always a single limiting value for I but the value of the quantity $\beta N/\alpha$ determines which limiting value is approached, regardless of the initial state of the disease. In epidemiological terms this says that if the quantity $\beta N/\alpha$ is less than one the infection dies out in the sense that the number of infectives approaches zero. For this reason the constant solution $I = 0$, which corresponds to $S = N$, is called the *disease-free equilibrium*. On the other hand, if the quantity $\beta N/\alpha$ exceeds one,

the infection persists. The constant solution $I = N - \alpha/\beta$, which corresponds to $S = \alpha/\beta$, is called an *endemic equilibrium*.

The value of 1 for the quantity $\beta N/\alpha$ is a tipping point in the sense that the behavior of the solution changes if this quantity passes through 1 because of some change in the parameters of the model.

The quantity $\beta N/\alpha$ also has an epidemiological interpretation. Since βN is the number of contacts made by an average infective per unit time and $1/\alpha$ is the mean infective period, the quantity $\beta N/\alpha$ is the number of secondary infections caused by introducing a single infective into a wholly susceptible population. The basic reproduction number is defined as the number of secondary infections caused by an average infective introduced into a wholly susceptible population over the course of the disease. The basic reproduction number is usually denoted by \mathcal{R}_0 . Here, the basic reproduction number or contact number for the disease is

$$\mathcal{R}_0 = \frac{\beta N}{\alpha}. \quad (2.5)$$

In studying an infectious disease model, the basic reproduction number is a central concept and its determination is invariably an essential first step. The value one for the basic reproduction number defines a threshold at which the course of the infection changes between disappearance and persistence. It is intuitively clear that if $\mathcal{R}_0 < 1$ the infection should die out, while if $\mathcal{R}_0 > 1$ the infection should establish itself. In more highly structured models than the simple one we have developed here, the calculation of the basic reproduction number may be much more complicated, but the essential concept obtains—that of the basic reproduction number as the number of secondary infections caused by an average infective over the course of the disease.

Since an endemic equilibrium corresponds to a long-term situation, it would be more realistic to include demographic processes, that is, births and natural deaths, in our model.

A simple assumption is that there is a birth rate $\Lambda(N)$ depending on total population size, and a proportional death rate μ . Just as for disease recovery rates the assumption of a proportional death rate is equivalent to an assumption of an exponentially distributed life span with mean $1/\mu$. In the absence of disease the total population size N satisfies the differential equation

$$N' = \Lambda(N) - \mu N.$$

At this point, it is necessary to introduce some basic definitions and results to describe the qualitative behavior of solutions of this differential equation, since it is not possible to solve the differential equation analytically.

The *carrying capacity* of the population is the limiting population size K , satisfying

$$\Lambda(K) = \mu K, \quad \Lambda'(K) < \mu.$$

The condition $\Lambda'(K) < \mu$ assures the asymptotic stability of the equilibrium population size K . It is reasonable to assume that K is the only positive equilibrium, so that

$$\Lambda(N) > \mu N$$

for $0 \leq N \leq K$. For most population models,

$$\Lambda(0) = 0, \quad \Lambda''(N) \leq 0.$$

However, if $\Lambda(N)$ represents recruitment into a behavioral class, as would be natural for models of sexually transmitted diseases, it would be plausible to have $\Lambda(0) > 0$, or even to consider $\Lambda(N)$ to be a constant function. If $\Lambda(0) = 0$, we require $\Lambda'(0) > \mu$ because if this requirement is not satisfied there is no positive equilibrium and the population would die out even in the absence of disease.

A model for a disease from which infectives recover with no immunity against reinfection and that includes births and deaths is

$$\begin{aligned} S' &= \Lambda(N) - \beta SI - \mu S + \alpha I \\ I' &= \beta SI - \alpha I - \mu I, \end{aligned} \tag{2.6}$$

describing a population with a density-dependent birth rate $\Lambda(N)$ per unit time, a mass action contact rate, a proportional death rate μ in each class, and a recovery rate α .

If we add the two equations of (2.6) and use $N = S + I$, we obtain

$$N' = \Lambda(N) - \mu N.$$

Thus N approaches K as $t \rightarrow \infty$.

It is easy to verify that

$$\mathcal{R}_0 = \frac{\beta K}{\mu + \alpha}$$

because a single infective introduced into a wholly susceptible population of size K causes βK new infections in unit time and the mean infective period corrected for natural mortality is $1/(\mu + \alpha)$.

It can be shown that the endemic equilibrium of (2.6), which exists if $\mathcal{R}_0 > 1$, is always asymptotically stable. If $\mathcal{R}_0 < 1$, the system has only the disease-free equilibrium and this equilibrium is asymptotically stable. The qualitative behavior of the model (2.6) is the same as the behavior of the model (2.2) without demographics.

We were able to reduce the systems of two differential equations (2.2) to a single equation because of the assumption that the total population $S + I$ is constant or has a constant limit as $t \rightarrow \infty$. If there are deaths due to the disease this assumption is violated, and it would be necessary to use a two-dimensional system as a model.

Let us assume that infectives recover from infection at a rate αI , while infectives die of disease at a rate dI . Then the model (2.6) is replaced by the model

$$\begin{aligned} S' &= \Lambda(N) - \beta SI - \mu S + \alpha I \\ I' &= \beta SI - (\alpha + d)I - \mu I. \end{aligned} \quad (2.7)$$

Now the equation for total population size is

$$N' = \Lambda(N) - \mu N - dI.$$

The model (2.7) involves the three variables S , I , N , and we may reduce it to a two-dimensional system with variables I and N by replacing S by $(N - I)$ to give the model

$$\begin{aligned} I' &= \beta I(N - I) - (\alpha + d + \mu)I \\ N' &= \Lambda(N) - \mu N - dI. \end{aligned}$$

The analysis of this model is more difficult, and rather than going through this analysis now we will shift our attention to *SIR* models for which the general approach to models involving systems of ordinary differential equations is more readily illustrated.

2.3 The *SIR* Model with Births and Deaths

The *SIR* model of Kermack and McKendrick [18] includes births in the susceptible class proportional to total population size and a death rate in each class proportional to the number of members in the class. This model allows the total population size to grow exponentially or die out exponentially if the birth and death rates are unequal. It is applicable to such questions as whether a disease will control the size of a population that would otherwise grow exponentially. We shall return to this topic, which is important in the study of many diseases in less developed countries with high birth rates. To formulate a model in which total population size remains bounded we could follow the approach suggested by Hethcote [12] in which the total population size N is held constant by making birth and death rates equal. Such a model is

$$\begin{aligned} S' &= -\beta SI + \mu(N - S) \\ I' &= \beta SI - \alpha I - \mu I \\ R' &= \alpha I - \mu R. \end{aligned}$$

Because $S + I + R = N$ and $N' = 0$, N is constant and we can view R as determined when S and I are known. Thus we may consider the two-dimensional system

$$\begin{aligned} S' &= -\beta SI + \mu(N - S) \\ I' &= \beta SI - \alpha I - \mu I. \end{aligned} \quad (2.8)$$

We shall examine a more general *SIR* model with births and deaths for a disease that may be fatal to some infectives. For such a disease, the class R of removed members should contain only recovered members, not members removed by death from the disease. It is not possible to assume that the total population size remains constant if there are deaths due to disease; a plausible model for a disease that may be fatal to some infectives must allow the total population to vary in time.

With a mass action contact rate and a density-dependent birth rate we would have a model

$$\begin{aligned} S' &= \Lambda(N) - \beta SI - \mu S \\ I' &= \beta SI - \mu I - dI - \alpha I \\ N' &= \Lambda(N) - dI - \mu N. \end{aligned} \quad (2.9)$$

If $d = 0$, so that there are no disease deaths, the equation for N is

$$N' = \Lambda(N) - \mu N,$$

so that $N(t)$ approaches a limiting population size K provided $\Lambda'(K) < \mu$ so that the equilibrium K of the equation for N is asymptotically stable.

We shall analyze the model (2.9) with $d = 0$ qualitatively. This qualitative analysis depends on the ideas of equilibria and linearization of a system about an equilibrium, a general approach dating back to the early part of the twentieth century. In view of the remark above, our analysis will also apply to the more general model (2.9) if there are no disease deaths. Analysis of the system (2.9) with $d > 0$ is much more difficult. We will confine our study of (2.9) to a description without details.

The first stage of the analysis is to note that the model (2.9) is a properly posed problem. A properly posed problem is a problem that has a unique solution (so that solving the mathematical problem yields an expression that can have epidemiological meaning), and that the solution remains non-negative (so that it has epidemiological meaning). That is, since $S' \geq 0$ if $S = 0$ and $I' \geq 0$ if $I = 0$, we have $S \geq 0$, $I \geq 0$ for $t \geq 0$ and since $N' \leq 0$ if $N = K$ we have $N \leq K$ for $t \geq 0$. Thus the solution always remains in the biologically realistic region $S \geq 0$, $I \geq 0$, $0 \leq N \leq K$ if it starts in this region. By rights, we should verify such conditions whenever we analyze a mathematical model, but in practice this step is frequently overlooked.

Our approach will be to identify equilibria (constant solutions) and then to determine the asymptotic stability of each equilibrium. As we have defined at the

end of Sect. 2.1, asymptotic stability of an equilibrium means that a solution starting sufficiently close to the equilibrium remains close to the equilibrium and approaches the equilibrium as $t \rightarrow \infty$, while instability of the equilibrium means that there are solutions starting arbitrarily close to the equilibrium that do not approach it. Equilibrium analysis of a system of two differential equations requires the idea of linearization of a system of differential equations and some matrix algebra. To find equilibria (S_∞, I_∞) we set the right side of each of the two equations equal to zero. The second of the resulting algebraic equations factors gives two alternatives. The first alternative is $I_\infty = 0$, which will give a disease-free equilibrium, and the second alternative is $\beta S_\infty = \mu + \alpha$, which will give an endemic equilibrium, provided $\beta S_\infty = \mu + \alpha < \beta K$. If $I_\infty = 0$ the other equation gives $S_\infty = K = \Lambda/\mu$. For the endemic equilibrium the first equation gives

$$I_\infty = \frac{\Lambda}{\mu + \alpha} - \frac{\mu}{\beta}.$$

We linearize about an equilibrium (S_∞, I_∞) by letting $y = S - S_\infty$, $z = I - I_\infty$, writing the system in terms of the new variables y and z and retaining only the linear terms in a Taylor expansion. We obtain a system of two linear differential equations,

$$\begin{aligned} y' &= -(\beta I_\infty + \mu)y - \beta S_\infty z \\ z' &= \beta I_\infty y + (\beta S_\infty - \mu - \alpha)z. \end{aligned}$$

The coefficient matrix of this linear system is

$$\begin{bmatrix} -\beta I_\infty - \mu & -\beta S_\infty \\ \beta I_\infty & \beta S_\infty - \mu - \alpha \end{bmatrix}.$$

This matrix is also known as the *Jacobian* matrix. We then look for solutions whose components are constant multiples of $e^{\lambda t}$; this means that λ must be an eigenvalue of the coefficient matrix. The condition that all solutions of the linearization at an equilibrium tend to zero as $t \rightarrow \infty$ is that the real part of every eigenvalue of this coefficient matrix is negative. At the disease-free equilibrium the matrix is

$$\begin{bmatrix} -\mu & -\beta K \\ 0 & \beta K - \mu - \alpha \end{bmatrix},$$

which has eigenvalues $-\mu$ and $\beta K - \mu - \alpha$. Thus, the disease-free equilibrium is asymptotically stable if $\beta K < \mu + \alpha$ and unstable if $\beta K > \mu + \alpha$. Note that this condition for instability of the disease-free equilibrium is the same as the condition for the existence of an endemic equilibrium.

In general, the condition that the eigenvalues of a 2×2 matrix have negative real part is that the determinant be positive and the trace (the sum of the diagonal

elements) be negative. Since $\beta S_\infty = \mu + \alpha$ at an endemic equilibrium, the Jacobian matrix, or matrix of the linearization at an endemic equilibrium is

$$\begin{bmatrix} -\beta I_\infty - \mu & -\beta S_\infty \\ \beta I_\infty & 0 \end{bmatrix}$$

and this matrix has positive determinant and negative trace. Thus, the endemic equilibrium, if there is one, is always asymptotically stable. If the quantity

$$\mathcal{R}_0 = \frac{\beta K}{\mu + \alpha} = \frac{K}{S_\infty}$$

is less than one, then the system has only the disease-free equilibrium and this equilibrium is asymptotically stable. In fact, it is not difficult to prove that this asymptotic stability is *global*, that is, that every solution approaches the disease-free equilibrium. If the quantity \mathcal{R}_0 is greater than one, then the disease-free equilibrium is unstable, but there is an endemic equilibrium that is asymptotically stable. Again, the quantity \mathcal{R}_0 is the basic reproduction number. It depends on the particular disease (determining the parameter α) and on the rate of contacts, which may depend on the population density in the community being studied. The disease model exhibits a *threshold* behavior: If the basic reproduction number is less than one, then the disease will die out, but if the basic reproduction number is greater than one, then the disease will be endemic. Just as for the *SIS* model of the preceding section, the basic reproduction number is the number of secondary infections caused by a single infective introduced into a wholly susceptible population because the number of contacts per infective in unit time is βK , and the mean infective period (corrected for natural mortality) is $1/(\mu + \alpha)$.

There are two aspects of the analysis of the model (2.9) that are more complicated than the analysis of (2.8). The first is in the study of equilibria. Because of the dependence of $\Lambda(N)$ on N , it is necessary to use two of the equilibrium conditions to solve for S and I in terms of N and then substitute this into the third condition to obtain an equation for N . Then by comparing the two sides of this equation for $N = 0$ and $N = K$ it is possible to show that there must be an endemic equilibrium value of N between 0 and K if $\mathcal{R}_0 > 1$.

The second complication is in the stability analysis. Since (2.9) is a three-dimensional system that cannot be reduced to a two-dimensional system, the coefficient matrix of its linearization at an equilibrium is a 3×3 matrix and the resulting characteristic equation is a cubic polynomial equation of the form

$$\lambda^3 + a_1\lambda^2 + a_2\lambda + a_3 = 0.$$

The *Routh–Hurwitz* conditions [16, 22]

$$a_1 > 0, \quad a_1 a_2 > a_3 > 0$$

are necessary and sufficient conditions for all roots of the characteristic equation to have negative real part. A technically complicated calculation is needed to verify that these conditions are satisfied at an endemic equilibrium for the model (2.9).

The asymptotic stability of the endemic equilibrium means that the compartment sizes approach a steady state. If the equilibrium had been unstable, there would have been a possibility of sustained oscillations. Oscillations in a disease model mean fluctuations in the number of cases to be expected. If the oscillations have a long period, then this could mean that experimental data for a short period would be quite unreliable as a predictor of the future. Epidemiological models that incorporate additional factors may exhibit oscillations. A variety of such situations is described in [14, 15].

From the third equation of (2.9) we obtain

$$N' = \Lambda - \mu N - \alpha I,$$

where $N = S + I + R$. From this, we see that at the endemic equilibrium $N = \Lambda/\mu - \alpha I/\mu$, and the reduction in the population size from the carrying capacity K is

$$\frac{d}{d\mu} I_\infty = \left[\frac{\alpha K}{\mu + \alpha} - \frac{\alpha}{\beta} \right].$$

The parameter α in the *SIR* model may be considered as describing the pathogenicity of the disease. If α is large, it is less likely that $\mathcal{R}_0 > 1$. If α is small, then the total population size at the endemic equilibrium is close to the carrying capacity K of the population. Thus, the maximum population decrease caused by disease will be for diseases of intermediate pathogenicity with α not close to either 0 or 1.

Numerical simulations indicate that the approach to endemic equilibrium for an *SIR* model is like a rapid and severe epidemic if the epidemiological and demographic time scales are very different. The same happens in the *SIS* model. If there are few disease deaths, the number of infectives at endemic equilibrium may be substantial, and there may be damped oscillations of large amplitude about the endemic equilibrium. For both the *SIR* and *SIS* models we may write the differential equation for I as

$$I' = I[\beta(N)S - (\mu + \alpha)] = \beta(N)I[S - S_\infty],$$

which implies that whenever S exceeds its endemic equilibrium value S_∞ , I is increasing and epidemic-like behavior is possible. If $\mathcal{R}_0 < 1$ and $S < K$, it follows that $I' < 0$, and thus I is decreasing. Thus, if $\mathcal{R}_0 < 1$, I cannot increase and no epidemic can occur.

2.4 The Simple Kermack–McKendrick Epidemic Model

One of the early triumphs of mathematical epidemiology was the formulation of a simple model by Kermack and McKendrick in 1927 [17] whose predictions are very similar to the behavior, observed in countless epidemics, of disease that invade a population suddenly, grow in intensity, and then disappear leaving part of the population untouched. The Kermack–McKendrick model is a compartmental model based on relatively simple assumptions on the rates of flow between different classes of members of the population. The SARS epidemic of 2002–2003 revived interest in epidemic models, which had been largely ignored since the time of Kermack and McKendrick, in favor of models for endemic diseases.

The special case of the model proposed by Kermack and McKendrick in 1927 which is the starting point for our study of epidemic models is

$$\begin{aligned} S' &= -\beta SI \\ I' &= \beta SI - \alpha I \\ R' &= \alpha I. \end{aligned} \tag{2.10}$$

A flow chart is shown in Fig. 2.1.

It is based on the same assumptions that were made for the *SIS* model of Sect. 2.2, except that recovered infectives go to a removed class rather than returning to the susceptible class. We will refer to this model as the simple Kermack–McKendrick epidemic model for convenience, but we remind the reader that it is a very special case of the actual Kermack–McKendrick epidemic model. From (2.10) we see that $N = S + I + R$ is constant.

The assumptions of a constant rate of contacts and of an exponentially distributed recovery rate are unrealistically simple. More general models can be constructed and analyzed, but our goal here is to show what may be deduced from extremely simple models. It will turn out that many more realistic models exhibit very similar qualitative behaviors.

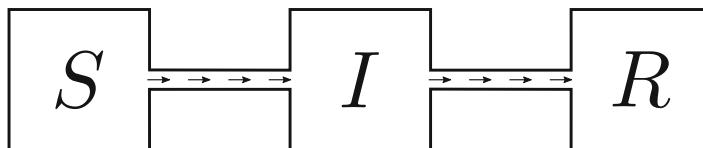


Fig. 2.1 Flow chart for the *SIR* model (2.10)

In our model, R is determined once S and I are known, and we can drop the R equation from our model, leaving the system of two equations

$$\begin{aligned} S' &= -\beta SI \\ I' &= \beta SI - \alpha I, \end{aligned} \tag{2.11}$$

together with initial conditions

$$S(0) = S_0, \quad I(0) = I_0, \quad S_0 + I_0 = N.$$

We think of introducing a small number of infectious individuals into a population of susceptibles and ask whether there will be an epidemic. We remark that the model makes sense only so long as $S(t)$ and $I(t)$ remain non-negative. Thus if either $S(t)$ or $I(t)$ reaches zero we consider the system to have terminated. We observe that $S' < 0$ for all t and $I' > 0$ if and only if $\beta S/\alpha > 1$. Thus I increases so long as $\beta S/\alpha > 1$ but since S decreases for all t , I ultimately decreases and approaches zero.

The quantity $\mathcal{R}_0 = \beta N/\alpha$ determines whether there is an epidemic. If $\mathcal{R}_0 < 1$, the infection dies out because $I'(t) < 0$ for all t , and there is no epidemic. Ordinarily, $S_0 \approx N$. If the epidemic is started by a member of the population being studied, for example by returning from travel with an infection acquired away from home, we would have $I_0 > 0$, $S_0 + I_0 = N$. A second way would be for an epidemic to be started by a visitor from outside the population. In this case, we would have $S_0 = N$. If $\mathcal{R}_0 > 1$, I increases initially and this is interpreted as saying that there is an epidemic.

Since (2.11) is a two-dimensional autonomous system of differential equations, the natural approach would be to find equilibria and linearize about each equilibrium to determine its stability. However, since every point with $I = 0$ is an equilibrium, the system (2.11) has a line of equilibria and this approach is not applicable (the linearization matrix at each equilibrium has a zero eigenvalue). The standard linearization theory for systems of ordinary differential equations is not applicable, and it is necessary to develop a new mathematical approach.

The sum of the two equations of (2.11) is

$$(S + I)' = -\alpha I.$$

Thus $S + I$ is a non-negative smooth decreasing function and therefore tends to a limit as $t \rightarrow \infty$. Also, it is not difficult to prove that the derivative of a smooth decreasing function which is bounded below must tend to zero, and this shows that

$$I_\infty = \lim_{t \rightarrow \infty} I(t) = 0.$$

Thus $S + I$ has limit S_∞ .

Integration of the sum of the two equations of (2.11) from 0 to ∞ gives

$$-\int_0^\infty (S(t) + I(t))' dt = S_0 + I_0 - S_\infty = N - S_\infty = \alpha \int_0^\infty I(t) dt.$$

Division of the first equation of (2.11) by S and integration from 0 to ∞ gives

$$\begin{aligned} \log \frac{S_0}{S_\infty} &= \beta \int_0^\infty I(t) dt \\ &= \frac{\beta}{\alpha} [N - S_\infty] \\ &= \mathcal{R}_0 \left[1 - \frac{S_\infty}{N} \right]. \end{aligned} \quad (2.12)$$

Equation (2.12) is called the *final size relation*. It gives a relation between the basic reproduction number and the size of the epidemic. Note that the final size of the epidemic, the number of members of the population who are infected over the course of the epidemic, is $N - S_\infty$. This is often described in terms of the *attack rate* $(1 - S_\infty/N)$. [Technically, the attack rate should be called an attack ratio, since it is dimensionless and is not a rate]. The attack rate is the fraction of the population that becomes infected over the course of the epidemic.

The final size relation (2.12) can be generalized to epidemic models with more complicated compartmental structure than the simple *SIR* model (2.11), including models with exposed periods, treatment models, and models including quarantine of suspected individuals and isolation of diagnosed infectives. The original Kermack–McKendrick model [17] included dependence on the time since becoming infected (age of infection), and this includes such models. We will describe this generalization in Chap. 4.

Integration of the first equation of (2.11) from 0 to t gives

$$\begin{aligned} \log \frac{S_0}{S(t)} &= \beta \int_0^t I(t) dt \\ &= \frac{\beta}{\alpha} [N - S(t) - I(t)], \end{aligned}$$

and this leads to the form

$$I(t) + S(t) - \frac{\alpha}{\beta} \log S(t) = N - \frac{\alpha}{\beta} \log S_0. \quad (2.13)$$

This implicit relation between S and I describes the orbits of solutions of (2.11) in the (S, I) plane.

In addition, since the right side of (2.12) is finite, the left side is also finite, and this shows that $S_\infty > 0$. It is not difficult to prove that there is a unique solution of the final size relation (2.12). To see this, we define the function

$$g(x) = \log \frac{S_0}{x} - \mathcal{R}_0 \left[1 - \frac{x}{N} \right].$$

Then

$$g(0+) > 0, \quad g(N) < 0,$$

and $g'(x) < 0$ if and only if

$$0 < x < \frac{N}{\mathcal{R}_0}.$$

If $\mathcal{R}_0 \leq 1$, $g(x)$ is monotone decreasing from a positive value at $x = 0+$ to a negative value at $x = N$. Thus there is a unique zero S_∞ of $g(x)$ with $S_\infty < N$. A graph of the function $g(x)$ is plotted in Fig. 2.2.

If $\mathcal{R}_0 > 1$, $g(x)$ is monotone decreasing from a positive value at $x = 0+$ to a minimum at $x = N/\mathcal{R}_0$ and then increases to a negative value at $x = N_0$. Thus there is a unique zero S_∞ of $g(x)$ with

$$S_\infty < \frac{N}{\mathcal{R}_0}.$$

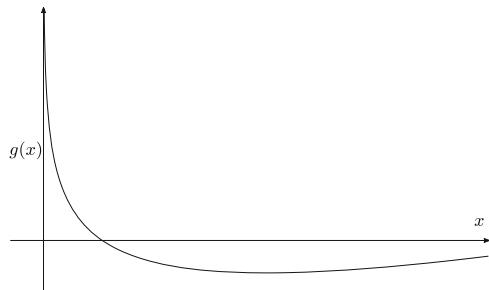
In fact,

$$\begin{aligned} g\left(\frac{S_0}{\mathcal{R}_0}\right) &= \log \mathcal{R}_0 - \mathcal{R}_0 + \frac{S_0}{N} \\ &\leq \log \mathcal{R}_0 - \mathcal{R}_0 + 1. \end{aligned}$$

Since $\log \mathcal{R}_0 < \mathcal{R}_0 - 1$ for $\mathcal{R}_0 > 0$, we actually have

$$g\left(\frac{S_0}{\mathcal{R}_0}\right) < 0,$$

Fig. 2.2 Graph of the function $g(x)$



and

$$S_\infty < \frac{S_0}{\mathcal{R}_0}. \quad (2.14)$$

An important question is how the basic reproduction number changes if a parameter of the model varies. If \mathcal{R}_0 , and therefore S_∞ , is a function of a parameter η , implicit differentiation of the final size relation (2.12) gives

$$\left(\frac{\mathcal{R}_0}{N} - \frac{1}{S_\infty} \right) \frac{dS_\infty}{d\eta} = \frac{d\mathcal{R}_0}{d\eta} \left(1 - \frac{S_\infty}{N} \right).$$

Because of (2.14), if \mathcal{R}_0 increases then S_∞ decreases.

It is generally difficult to estimate the contact rate β which depends on the particular disease being studied but may also depend on social and behavioral factors. The quantities S_0 and S_∞ may be estimated by serological studies (measurements of immune responses in blood samples) before and after an epidemic, and from these data, the basic reproduction number \mathcal{R}_0 may be estimated by using (2.12). This estimate, however, is a retrospective one which can be derived only after the epidemic has run its course.

In order to prevent the occurrence of an epidemic, if infectives are introduced into a population, it is necessary to reduce the basic reproduction number \mathcal{R}_0 below one. This may sometimes be achieved by immunization, which has the effect of transferring members of the population from the susceptible class to the removed class and thus of reducing $S(0)$. Immunization of some members of the population produces a new model. If a fraction p of the population is successfully immunized the effect is to decrease the number of susceptibles from $S(0)$ to $S(0)(1 - p)$. Originally, the basic reproduction number is $\beta N / \alpha$, but in the new situation with a decreased number of susceptibles the basic reproduction number would be $\beta N(1 - p) / \alpha$. This is less than 1 if p satisfies $\beta N(1 - p) / \alpha < 1$. This gives $1 - p < \alpha / \beta N$, or

$$p > 1 - \frac{\alpha}{\beta N} = 1 - \frac{1}{\mathcal{R}_0}.$$

Initially, the number of infectives grows exponentially because the equation for I may be approximated by

$$I' = (\beta N - \alpha)I$$

and the initial growth rate is

$$r = \beta N - \alpha = \alpha(\mathcal{R}_0 - 1).$$

This initial growth rate r may be estimated from incidence data when an epidemic begins. Since N and α may be measured, β may be calculated as

$$\beta = \frac{r + \alpha}{N}.$$

However, because of incomplete data and under-reporting of cases, this estimate may not be very accurate. This inaccuracy is even more pronounced for an outbreak of a previously unknown disease, where early cases are likely to be mis-diagnosed. Because of the final size relation, estimation of β or \mathcal{R}_0 is an important question that has been studied by a variety of approaches. Estimation of the initial growth rate from data can provide an estimate of the contact rate β . However, this relation is valid only for the model (2.11) and does not hold for models with different compartmental structure, such as an exposed period.

If $\beta S_0 > \alpha$, I increases initially to a maximum number of infectives when the derivative of I is zero, that is, when $S = \alpha/\beta$. This maximum is given by

$$I_{max} = S_0 + I_0 - \frac{\alpha}{\beta} \log S_0 - \frac{\alpha}{\beta} + \frac{\alpha}{\beta} \log \frac{\alpha}{\beta}, \quad (2.15)$$

obtained by substituting $S = \alpha/\beta$, $I = I_{max}$ into (2.13).

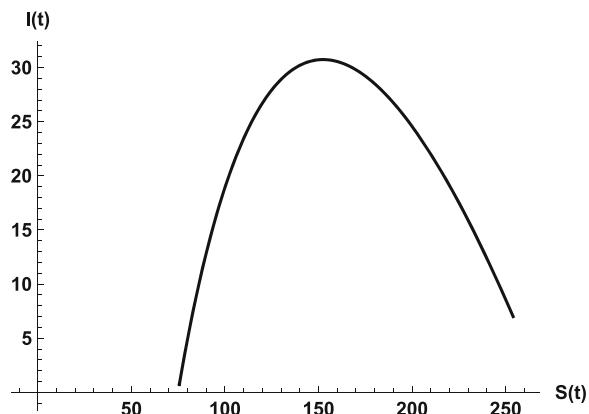
Example 1 A study of Yale University freshmen [10] reported by Hethcote [13] described an influenza epidemic with $S_0 = 0.911$, $S_\infty = 0.513$. Here we are measuring the number of susceptibles as a fraction of the total population size, or using the population size K as the unit of size. Substitution into the final size relation gives the estimate $\beta N/\alpha = 1.18$ and $\mathcal{R}_0 = 1.18$. Since we know that $1/\alpha$ is approximately 3 days for influenza, we see that βN is approximately 0.39 contacts per day per member of the population.

Example 2 (The Great Plague in Eyam) The village of Eyam near Sheffield, England suffered an outbreak of bubonic plague in 1665–1666 the source of which is generally believed to be the Great Plague of London. The Eyam plague was survived by only 83 of an initial population of 350 persons. As detailed records were preserved and as the community was persuaded to quarantine itself to try to prevent the spread of disease to other communities, the disease in Eyam has been used as a case study for modeling [21]. Detailed examination of the data indicates that there were actually two outbreaks of which the first was relatively mild. Thus we shall try to fit the model (2.11) over the period from mid-May to mid-October 1666, measuring time in months with an initial population of 7 infectives and 254 susceptibles, and a final population of 83. Raggett [21] gives values of susceptibles and infectives in Eyam on various dates, beginning with $S(0) = 254$, $I(0) = 7$, shown in Table 2.1.

The final size relation with $S_0 = 254$, $I_0 = 7$, $S_\infty = 83$ gives $\beta/\alpha = 6.54 \times 10^{-3}$, $\alpha/\beta = 153$. The infective period was 11 days, or 0.3667 month, so that $\alpha = 2.73$. Then $\beta = 0.0178$. The relation (2.15) gives an estimate of 30.4 for the maximum

Table 2.1 Eyam Plague data

Date (1666)	Susceptibles	Infectives
July 3/4	235	14.5
July 19	201	22
August 3/4	153.5	29
August 19	121	21
September 3/4	108	8
September 19	97	8
October 4/5	Unknown	Unknown
October 20	83	0

Fig. 2.3 The S - I plane

number of infectives. We use the values obtained here for the parameters β and α in the model (2.11) for simulations of the phase plane, the (S, I) plane, and for graphs of S and I as functions of t (Figs. 2.3 and 2.4). Figure 2.5 plots these data points together with the phase portrait given in Fig. 2.3 for the model (2.11).

The actual data for the Eyam epidemic are remarkably close to the predictions of this very simple model. However, the model is really too good to be true. Our model assumes that infection is transmitted directly between people. While this is possible, bubonic plague is transmitted mainly by rat fleas. When an infected rat is bitten by a flea, the flea becomes extremely hungry and bites the host rat repeatedly, spreading the infection in the rat. When the host rat dies its fleas move on to other rats, spreading the disease further. As the number of available rats decreases the fleas move to human hosts, and this is how plague starts in a human population (although the second phase of the epidemic may have been the pneumonic form of bubonic plague, which can be spread from person to person). One of the main reasons for the spread of plague from Asia into Europe was the passage of many trading ships; in medieval times ships were invariably infested with rats. An accurate model of plague transmission would have to include flea and rat populations, as well as movement in space. Such a model would be extremely complicated and its predictions might well not be any closer to observations than our simple unrealistic model. Raggett

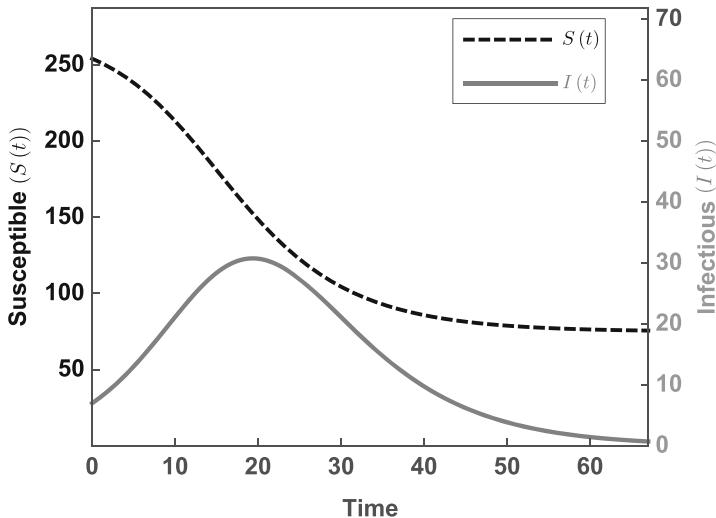


Fig. 2.4 Time plot of S and I

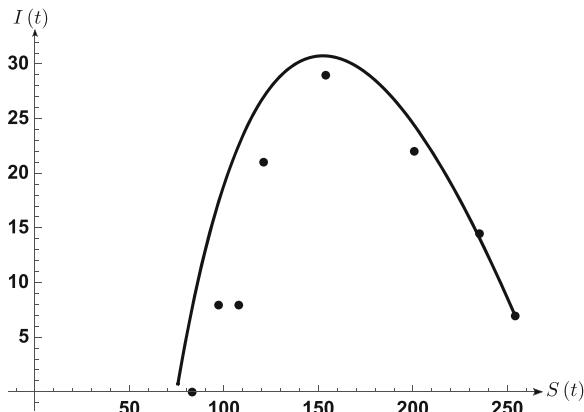


Fig. 2.5 The S - I plane, model and data

also used a stochastic model to fit the data, but the fit was rather poorer than the fit for the simple deterministic model (2.11).

In the village of Eyam the rector persuaded the entire community to quarantine itself to prevent the spread of disease to other communities. One effect of this policy was to increase the infection rate in the village by keeping fleas, rats, and people in close contact with one another, and the mortality rate from bubonic plague was much higher in Eyam than in London. Further, the quarantine could do nothing to prevent the travel of rats and thus did little to prevent the spread of disease to other communities. One message this suggests to mathematical modelers is that control strategies based on false models that do not describe the epidemiological

situation accurately may be harmful, and it is essential to distinguish between assumptions that simplify but do not alter the predicted effects substantially, and wrong assumptions which make an important difference.

2.5 Epidemic Models with Deaths due to Disease

So far, we have analyzed only models with no disease deaths, for which it may be assumed that the total population size is constant.

The assumption in the model (2.11) of a rate of contacts per infective which is proportional to population size N , called *mass action incidence* or bilinear incidence, was used in all the early epidemic models. However, it is quite unrealistic, except possibly in the early stages of an epidemic in a population of moderate size. It is more realistic to assume a contact rate which is a non-increasing function of the total population size. For example, a situation in which the number of contacts per infective in unit time is constant, called *standard incidence*, is a more accurate description for sexually transmitted diseases. If there are no disease deaths, so that the total population size remains constant, such a distinction is unnecessary.

We generalize the model (2.11) by assuming that an average member of the population makes $N\beta(N)$ contacts in unit time [4, 8]. It is reasonable to assume $\beta'(N) \leq 0$ to express the idea of saturation in the number of contacts. Then mass action incidence corresponds to the choice $a(N) = \beta N$, and standard incidence corresponds to the choice $a(N) = a$. The assumptions $a(N) = N\beta(N)$, $a'(N) \geq 0$ imply that

$$\beta(N) + N\beta'(N) \geq 0.$$

Some epidemic models [8] have used a Michaelis–Menten type of interaction of the form

$$\beta(N) = \frac{a}{1 + bN}.$$

Another form based on a mechanistic derivation for pair formation [11] leads to an expression of the form

$$\beta(N) = \frac{a}{1 + bN + \sqrt{1 + 2bN}}.$$

Data for diseases transmitted by contact in cities of moderate size [20] suggests that data fits the assumption of a form

$$\beta(N) = aN^{p-1}$$

with $p = 0.05$ quite well. All of these forms satisfy the conditions $\beta'(N) \leq 0$ and (2.5).

Because the total population size is now present in the model, we must include an equation for total population size in the model. This forces us to make a distinction between members of the population who die of the disease and members of the population who recover with immunity against reinfection. We assume, as in the previous chapter, a recovery rate αI and a disease death rate dI . We use S , I , and N as variables, with $N = S + I + R$. We now obtain a three-dimensional model

$$\begin{aligned} S' &= -\beta(N)SI \\ I' &= \beta(N)SI - (\alpha + d)I \\ N' &= -dI. \end{aligned} \tag{2.16}$$

Since N is now a decreasing function, we define $N(0) = N_0 = S_0 + I_0$. We also have the equation $R' = \alpha I$, but we do not need to include it in the model since R is determined when S , I , and N are known. We should note that if $d = 0$ the total population size remains equal to the constant N , and the model (2.16) reduces to the simpler model (2.11) with $\beta(N)$ replaced by the constant $\beta(N_0)$.

We wish to show that the model (2.16) has the same qualitative behavior as the model (2.11), namely that there is a basic reproduction number which distinguishes between disappearance of the disease and an epidemic outbreak, and that some members of the population are left untouched when the epidemic passes. These two properties are the central features of all epidemic models.

For the model (2.16) the basic reproduction number is given by

$$\mathcal{R}_0 = \frac{N_0 \beta(N_0)}{\alpha + d}$$

because a single infective introduced into a wholly susceptible population makes $c(N_0) = N_0 \beta(N_0)$ contacts in unit time, all of which are with susceptibles and thus produce new infections, and the mean infective period, corrected for mortality, is $1/(\alpha + d)$.

We assume that $\beta(0)$ is finite, thus ruling out standard incidence (standard incidence does not appear to be realistic if the total population N approaches zero, and it would be more natural to assume that $c(N)$ grows linearly with N for small N). If we let $t \rightarrow \infty$ in the sum of the first two equations of (2.16) we obtain

$$(\alpha + d) \int_0^\infty I(s)ds = S_0 + I_0 - S_\infty = N - S_\infty.$$

The first equation of (2.16) may be written as

$$-\frac{S'(t)}{S(t)} = \beta(N(t))I(t).$$

Since

$$\beta(N) \geq \beta(N_0),$$

integration from 0 to ∞ gives

$$\begin{aligned} \log \frac{S_0}{S_\infty} &= \int_0^\infty \beta(N(t))I(t)dt \\ &\geq \beta(N_0) \int_0^\infty I(t)dt \\ &= \frac{\beta(N_0)(N_0 - S_\infty)}{(\alpha + d)N_0}. \end{aligned}$$

We now obtain a final size inequality

$$\begin{aligned} \log \frac{S_0}{S_\infty} &= \int_0^\infty \beta(N(t))I(t)dt \\ &\geq \beta(N_0) \int_0^\infty I(t)dt = \mathcal{R}_0 \left[1 - \frac{S_\infty}{N_0} \right]. \end{aligned}$$

If the case fatality ratio $d/(d+\alpha)$ is small, the final size inequality is an approximate equality.

It is not difficult to show that $N(t) \geq \alpha N_0 / (\alpha + d)$ and then a similar calculation using the inequality $\beta(N) \leq \beta(\alpha N_0 / (\alpha + d)) < \infty$ shows that

$$\log \frac{S_0}{S_\infty} \leq \beta(\alpha N_0 / (\alpha + d)) \int_0^\infty I(t)dt,$$

from which we may deduce that $S_\infty > 0$.

It is important to be able to estimate the fraction of infectives who die of disease over the course of the epidemic. At every time, the rate of recovery is αI and the rate of disease deaths is dI , and thus the mortality fraction at each time is

$$\frac{d}{d + \alpha}.$$

This is what is sometimes described (incorrectly because it is not a rate) as the epidemic death rate. While the mortality fraction overall is $d/(d+\alpha)$, reports during the epidemic would underestimate the mortality fraction because there may be some infectives who would die later during the epidemic. On the other hand, for a disease like influenza, where many cases are mild enough to go unreported, reports would tend to overestimate the mortality fraction.

2.6 *Project: Discrete Epidemic Models

The discrete analogue of the continuous-time epidemic model (2.11) is

$$\begin{aligned} S_{j+1} &= S_j G_j, \\ I_{j+1} &= S_j (1 - G_j) + \sigma I_j, \\ G_j &= e^{-\beta I_j / N}, \quad j = 1, 2, \dots, \end{aligned} \tag{2.17}$$

where S_j and I_j denote the numbers of susceptible and infective individuals at time j , respectively. Here, G_j is the probability that a susceptible individual at time j will remain susceptible to time $j+1$, that is, will not be infected between time j and time $j+1$, and $\sigma = e^{-\alpha}$ is the probability that an infected individual at time j will remain infected to time $j+1$.

Assume that the initial conditions are $S(0) = S_0 > 0$, $I(0) = I_0 > 0$, and $S_0 + I_0 = N$.

Exercise 1 Consider the system (2.17).

(a) Show that the sequence $\{S_j + I_j\}$ has a limit

$$S_\infty + I_\infty = \lim_{j \rightarrow \infty} (S_j + I_j).$$

(b) Show that

$$I_\infty = \lim_{j \rightarrow \infty} I_j = 0.$$

(c) Show that

$$\log \frac{S_0}{S_\infty} = \beta \sum_{m=0}^{\infty} \frac{I_m}{N}.$$

(d) Show that

$$\log \frac{S_0}{S_\infty} = \mathcal{R}_0 \left[1 - \frac{S_\infty}{N} \right],$$

with $\mathcal{R}_0 = \frac{\beta}{1-\sigma}$.

Next, consider the case that there are k infected stages and there is treatment in some stages, with treatment rates that can be different in different stages. Assume that selection of members for treatment occurs only at the beginning of a stage. Let $I_j^{(i)}$ and $T_j^{(i)}$ denote the numbers of infected and treated individuals, respectively, in stage i ($i = 1, 2, \dots, k$) at time j . Let σ_i^I denote the probability

that an infected individual in the $I^{(i)}$ stage continues on to the next stage, either treated or untreated, and let σ_i^T denote the probability that an individual in the $T^{(i)}$ stage continues on to the next treated stage. In addition, of the members leaving an infected stage $I^{(i)}$, a fraction p_i enters treatment in $T^{(i+1)}$, while the remaining fraction q_i continues to $I^{(i+1)}$. Let m_i denote the fraction of infected members who go through the stage $I^{(i)}$, and n_i the fraction of infected members who go through the stage $T^{(i)}$. Then,

$$\begin{aligned} m_1 &= q_1, \quad m_2 = q_1 q_2, \quad \dots, \quad m_k = q_1 q_2 \cdots q_k, \\ n_1 &= p_1, \quad n_2 = p_1 + q_1 p_2, \quad \dots, \quad n_k = p_1 + q_1 p_2 + \dots + q_1 q_2 \cdots q_{k-1} p_k. \end{aligned}$$

The discrete system with treatment is

$$\begin{aligned} S_{j+1} &= S_j G_j, \\ I_{j+1}^{(1)} &= q_1 S_j (1 - G_j) + \sigma_1^I I_j^{(1)}, \\ T_{j+1}^{(1)} &= p_1 S_j (1 - G_j) + \sigma_1^T T_j^{(1)} \\ I_{j+1}^{(i)} &= q_i (1 - \sigma_{i-1}^I) I_j^{(i-1)} + \sigma_i^I \eta_i I_j^{(i)}, \\ T_{j+1}^{(i)} &= p_i (1 - \sigma_{i-1}^I) I_j^{(i-1)} + (1 - \sigma_{i-1}^T) T_j^{(i-1)} + \sigma_i^T T_j^{(i)}, \end{aligned} \tag{2.18}$$

$[i = 2, \dots, k, j \geq 0]$, with

$$G_j = e^{-\beta \sum_{i=1}^k (\epsilon_i I_j^{(i)}/N + \delta_i T_j^{(i)}/N)},$$

where ϵ_i is the relative infectivity of untreated individuals at stage i and δ_i is the relative infectivity of treated individuals at stage i .

(e) Show that

$$\mathcal{R}_c = \beta N \sum_{i=1}^k \left[\frac{\epsilon_i m_i}{1 - \sigma_i^I} + \frac{\delta_i n_i}{1 - \sigma_i^T} \right]$$

and that

$$\log \frac{S_0}{S_\infty} = \mathcal{R}_c \left[1 - \frac{S_\infty}{N} \right].$$

2.7 *Project: Pulse Vaccination

Consider an SIR model (2.9) with $\Lambda = \mu K$. For measles, typical parameter choices are $\mu = 0.02$, $\beta = 1800$, $\alpha = 100$, $K = 1$ (to normalize carrying capacity to 1) [9].

Question 1 Show that for these parameter choices $\mathcal{R}_0 \approx 18$ and to achieve herd immunity would require vaccination of about 95% of the susceptible population.

In practice, it is not possible to vaccinate 95% of a population because not all members of the population would come to be vaccinated and not all vaccinations are successful. One way to avoid recurring outbreaks of disease is “pulse vaccination” [1, 24, 25]. The basic idea behind pulse vaccination is to vaccinate a given fraction p of the susceptible population at intervals of time T with T (depending on p) chosen to ensure that the number of infectives remains small and approaches zero. In this project we will give two approaches to the calculation of a suitable function $T(p)$.

The first approach depends on the observation that I decreases so long as $S < \Gamma < (\mu + \alpha)/\beta$. We begin by vaccinating $p\Gamma$ members, beginning with $S(0) = (1 - p)\Gamma$. From (8.7),

$$S' = \mu K - \mu S - \beta SI \geq \mu K - \mu S.$$

Then $S(t)$ is greater than the solution of the initial value problem

$$S' = \mu K - \mu S, \quad S(0) = (1 - p)\Gamma.$$

Question 2 Solve this initial value problem and show that the solution obeys

$$S(t) < \Gamma, 0 \leq t < \frac{1}{\mu} \log \frac{K - (1 - p)\Gamma}{K - \Gamma}.$$

Thus a suitable choice for $T(p)$ is

$$T(p) = \frac{1}{\mu} \log \frac{K - (1 - p)\Gamma}{K - \Gamma} = \frac{1}{\mu} \log \left[1 + \frac{p\Gamma}{K - \Gamma} \right].$$

Calculate $T(p)$ for $p = m/10$ ($m = 1, 2, \dots, 10$).

The second approach is more sophisticated. Start with $I = 0$, $S' = \mu K - \mu S$. We let $t_n = nT$ ($n = 0, 1, 2, \dots$) and run the system for $0 \leq t \leq t_1 = T$. Then we let $S_1 = (1 - p)S(t_1)$. We then repeat, i.e. for $t_1 \leq t \leq t_2$, $S(t)$ is the solution of $S' = \mu K - \mu S$, $S(t_1) = S_1$, and $S_2 = (1 - p)S_1$. We obtain a sequence S_n in this way.

Question 3 Show that

$$S_{n+1} = (1 - p)K(1 - e^{-\mu T}) + (1 - p)S_n e^{-\mu T}$$

and for $t_n \leq t \leq t_{n+1}$,

$$S(t) = K \left[1 - e^{-\mu(t-t_n)} \right] + S_n e^{-\mu(t-t_n)}.$$

Question 4 Show that the solution is periodic if

$$S_{n+1} = S_n = S^* \quad (n = 0, 1, 2, \dots)$$

with

$$S^* = K \left[1 - \frac{pe^{\mu T}}{e^{\mu T} - (1-p)} \right]$$

and that the periodic solution is

$$S(t) = \begin{cases} K \left[1 - \frac{pe^{\mu T}}{e^{\mu T} - (1-p)} e^{-\mu(t-t_n)} \right] : & t_n \leq t \leq t_{n+1}, \\ S^* : & t = t_{n+1} \end{cases} \quad I(t) = 0.$$

It is possible to show by linearizing about this periodic solution that the periodic solution is asymptotically stable if

$$\frac{1}{T} \int_0^T S(t) dt < \frac{\mu + \xi}{\beta}.$$

If this condition is satisfied, the infective population will remain close to zero.

Question 5 Show that this stability condition reduces to

$$\frac{K(\mu T - p)(e^{\mu T} - 1) + pK\mu T}{\mu T [e^\mu - (1-p)]} < \frac{\mu + \xi}{\beta}.$$

Question 6 Use a computer algebra system to graph $T(p)$, where T is defined implicitly by

$$\frac{K(\mu T - p)(e^{\mu T} - 1) + pK\mu T}{\mu T [e^\mu - (1-p)]} = \frac{\mu + \xi}{\beta}.$$

Compare this expression for T with the one obtained earlier in Question 2 in this project. A larger estimate for a safe value of T would save money by allowing less frequent vaccination pulses.

2.8 *Project: A Model with Competing Disease Strains

We model a general discrete-time SIS model with two competing strains in a population with discrete and non-overlapping generations. This model arises from a particular discretization in time of the corresponding SIS continuous-time stochastic model for two competing strains.

State Variables

S_n	Population of susceptible individuals in generation n
I_n^1	Population of infected individuals with strain 1 in generation n
I_n^2	Population of infected individuals with strain 2 in generation n
T_n	Total population in generation n
f	Recruitment function

Parameters

μ	Per capita natural death rate
γ_i	Per capita recovery rate for strain i
α_i	Per capita infection rate for strain i

Construction of the model equations: The model assumes that (i) the disease is not fatal; (ii) all recruits are susceptible and the recruitment function depends only on T_n ; (iii) there are no coinfections; (iv) death, infections, and recoveries are modeled as Poisson processes with rates μ, α_i, γ_i ($i = 1, 2$); (v) the time step is measured in generations; (vi) the populations change only because of “births” (given by the recruitment function), deaths, recovery, and infection of a susceptible individual for each strain; (vii) individuals recover but do not develop permanent or temporary immunity, that is, they immediately become susceptible again.

By assumption we have that the probability of k successful encounters is a Poisson distribution, which in general has the form $p(k) = e^{-\beta} \beta^k / k!$, where β is the parameter of the Poisson distribution. In our context, only one success is necessary. Therefore, when there are no successful encounters, the expression $p(0) = e^{-\beta}$ represents the probability that a given event does not occur. For example, the probability that a susceptible individual does not become infective is $\text{Prob}(\text{not being infected by strain } i) = e^{-\alpha_i I_n^i}$, and, $\text{Prob}(\text{not recovering from strain } i) = e^{-\gamma_i I_n^i}$. Hence, $\text{Prob}(\text{not being infected}) = \text{Prob}(\text{not being infected by strain 1}) \text{Prob}(\text{not being infected by strain 2}) = e^{-\alpha_1 I_n^1} e^{-\alpha_2 I_n^2}$.

Now the probability that a susceptible does become infected is given by $1 - e^{-\alpha_i I_n^i}$. Then, $\text{Prob}(\text{infected by strain } i) = \text{Prob}(\text{infected})$. $\text{Prob}(\text{infected by strain } i | \text{infected}) = (1 - e^{-(\alpha_1 I_n^1 + \alpha_2 I_n^2)}) \frac{\alpha_i I_n^i}{\alpha_1 I_n^1 + \alpha_2 I_n^2}$.

(a) Using the above discussion, show that the dynamics are governed by the system

$$\begin{aligned} S_{n+1} &= f(T_n) + S_n e^{-\mu} e^{-(\alpha_1 I_n^1 + \alpha_2 I_n^2)} + \sum_{j=1}^2 I_n^j e^{-\mu} (1 - e^{-\gamma_j}), \\ I_{n+1}^1 &= \frac{\alpha_1 S_n I_n^1}{\alpha_1 I_n^1 + \alpha_2 I_n^2} e^{-\mu} (1 - e^{-(\alpha_1 I_n^1 + \alpha_2 I_n^2)}) + I_n^1 e^{-\mu} e^{-\gamma_1}, \\ I_{n+1}^2 &= \frac{\alpha_2 S_n I_n^2}{\alpha_1 I_n^1 + \alpha_2 I_n^2} e^{-\mu} (1 - e^{-(\alpha_1 I_n^1 + \alpha_2 I_n^2)}) + I_n^2 e^{-\mu} e^{-\gamma_2}. \end{aligned} \quad (2.19)$$

(b) Show that

$$T_{n+1} = f(T_n) + T_n e^{-\mu},$$

where

$$T_n = S_n + I_n^1 + I_n^2. \quad (2.20)$$

This equation is called the *demographic equation*. It describes the total population dynamics.

- (c) If we set $I_{n+1}^1 = I_{n+1}^2 = 0$, then model (2.19) reduces to the demographic model

$$S_n = f(S_n) + S_n e^\mu.$$

and

$$T_{n+1} = f(T_n) + T_n e^{-\mu}.$$

Check that this is the case.

- (d) Study the disease dynamics at a demographic equilibrium, that is, at a point where $T_\infty = T_\infty e^{-\mu} + f(T_\infty)$. Substitute $S_n = T_\infty - I_n^1 - I_n^2$ where T_∞ is a stable demographic equilibrium, that is, assume $T_0 = T_\infty$ to get the following equations:

$$\begin{aligned} I_{n+1}^1 &= \frac{\alpha_1 I_n^1}{\alpha_1 I_n^1 + \alpha_2 I_n^2} (T_\infty - I_n^1 - I_n^2) e^{-\mu} \left(1 - e^{-(\alpha_1 I_n^1 + \alpha_2 I_n^2)}\right) + I_n^1 e^{-\mu} e^{-\gamma_1}, \\ I_{n+1}^2 &= \frac{\alpha_2 I_n^2}{\alpha_1 I_n^1 + \alpha_2 I_n^2} (T_\infty - I_n^1 - I_n^2) e^{-\mu} \left(1 - e^{-(\alpha_1 I_n^1 + \alpha_2 I_n^2)}\right) + I_n^2 e^{-\mu} e^{-\gamma_2}. \end{aligned} \quad (2.21)$$

System (2.21) describes the dynamics of a population infected with the two strains at a demographic equilibrium.

Show that in system (2.21), if $R_1 = \frac{e^{-\mu} T_\infty \alpha_1}{1 - e^{-(\mu + \gamma_1)}} < 1$ and $R_2 = \frac{e^{-\mu} T_\infty \alpha_2}{1 - e^{-(\mu + \gamma_2)}} < 1$, then the equilibrium point $(0, 0)$ is asymptotically stable.

- (e) Interpret biologically the numbers R_i , $i = 1, 2$.
(f) Consider $f(T_n) = \Lambda$, where Λ is a constant. Show that

$$T_{n+1} = \Lambda + T_n e^{-\mu}$$

and that

$$T_\infty = \frac{\Lambda}{1 - e^{-\mu}}.$$

- (g) Consider $f(T_n) = r T_n (1 - T_n)/k$, and show that in this case the total population dynamic is given by

$$T_{n+1} = r T_n \left(1 - \frac{T_n}{k}\right) + T_n e^{-\mu}$$

and that the fixed points are

$$T_n^* = 0, \quad T_n^{**} = \frac{k(r + e^{-\mu} - 1)}{r},$$

whenever $r + e^{-\mu} > 1$.

- (h) Assume that one of the strains is missing, and determine the boundary equilibria, that is, let $I_n^i = 0$ for either $i = 1$ or 2 . Equation (2.21) reduces to

$$I_{n+1} = (T_\infty - I_n)e^{-\mu}(1 - e^{\alpha_1 I_n}) + I_n e^{-(\mu+\gamma)}.$$

Establish necessary and sufficient conditions for the stability and/or instability of boundary equilibria for the system (2.21). Compare your results with simulations of the system (2.21) and of the full system (2.19).

- (i) Does the system (2.21) have endemic ($I_1^* > 0, I_2^* > 0$) equilibria?
- (j) Simulate the full system (2.19) when the demographic equation is in the period doubling regime, where there are orbits with periods of double the length of periods for smaller parameter values. What are your conclusions?

References: [5, 6].

2.9 Project: An Epidemic Model in Two Patches

Consider the following SIS model with dispersion between two patches, Patch 1 and Patch 2, where in Patch $i \in \{1, 2\}$ at generation t , $S_i(t)$ denotes the population of susceptible individuals; $I_i(t)$ denotes the population of infected assumed infectious; $T_i(t) \equiv S_i(t) + I_i(t)$ denotes the total population size. The constant dispersion coefficients D_S and D_I measure the probability of dispersion by the susceptible and infective individuals, respectively. Observe that we are using a different notation from what we have used elsewhere, writing variables as a function of t rather than using a subscript for the independent variable in order to avoid needing double subscripts:

$$\begin{aligned} S_1(t+1) &= (1 - D_S)\tilde{S}_1(t) + D_S\tilde{S}_2(t), \\ I_1(t+1) &= (1 - D_I)\tilde{I}_1(t) + D_I\tilde{I}_2(t), \\ S_2(t+1) &= D_S\tilde{S}_1(t) + (1 - D_S)\tilde{S}_2(t), \\ I_2(t+1) &= D_I\tilde{I}_1(t) + (1 - D_I)\tilde{I}_2(t), \end{aligned}$$

where

$$\begin{aligned}\tilde{S}_i(t) &= f_i(T_i(t)) + \gamma_i S_i(t) \exp\left(\frac{-\alpha_i I_i(t)}{T_i(t)}\right) + \gamma_i I_i(t)(1 - \sigma_i), \\ \tilde{I}_i(t) &= \gamma_i(1 - \exp\left(\frac{-\alpha_i I_i(t)}{T_i(t)}\right)) S_i(t) + \gamma_i \sigma_i I_i(t)\end{aligned}$$

and

$$0 \leq \gamma_i, \sigma_i, \alpha_i, D_S, D_I \leq 1.$$

Let

$$f_i(T_i(t)) = T_i(t) \exp(r_i - T_i(t)),$$

where r_i is a positive constant.

- (a) Using computer explorations, determine whether it is possible to have a globally stable disease-free equilibrium on a patch (without dispersal) where the full system with dispersal has a stable endemic equilibrium. Do you have a conjecture?
- (b) Using computer explorations determine whether it is possible to have a globally stable endemic equilibrium on a patch (without dispersal) where the full system with dispersal has a stable disease-free equilibrium. Do you have a conjecture?

References:[3, 7, 23].

2.10 Project: Fitting Data for an Influenza Model

Consider an SIR model (2.11) with basic reproduction number 1.5.

1. Describe the qualitative changes in (S, I, R) as a function of time for different values of β and α with $\beta \in \{0.0001, 0.0002, \dots, 0.0009\}$, for the initial condition $(S, I, R) = (10^6, 1, 0)$.
2. Discuss the result of part (a) in terms of the basic reproduction number (what is β/γ ?). Use a specific disease such as influenza to provide simple interpretations for the different time courses of the disease for the different choices of β and γ .
3. Repeat the steps in part (a) for values of $\mathcal{R}_0 \in \{1.75, 2, 2.5\}$, and for each value of \mathcal{R}_0 , choose the best pair of values (β, α) that fits the slope before the first peak in the data found in Table 2.2 for reported H1N1 influenza cases in México. (Hint: normalize the data so that the peak is 1, and then multiply the data by the size of the peak in the simulations.)

Table 2.2 Reported cases for H1N1-pandemic in Mexico

Day	Cases										
75	2	95	4	115	318	135	83	155	152	175	328
76	1	96	11	116	399	136	75	156	138	176	298
77	3	97	5	117	412	137	87	157	159	177	335
78	2	98	7	118	305	138	98	158	186	178	330
79	3	99	4	119	282	139	71	159	222	179	375
80	3	100	4	120	227	140	73	160	204	180	366
81	4	101	4	121	212	141	78	161	257	181	291
82	4	102	11	122	187	142	67	162	208	182	251
83	5	103	17	123	212	143	68	163	198	183	215
84	7	104	26	124	237	144	69	164	193	184	242
85	3	105	20	125	231	145	65	165	243	185	223
86	1	106	12	126	237	146	85	166	231	186	317
87	2	107	26	127	176	147	55	167	225	187	305
88	5	108	33	128	167	148	67	168	239	188	228
89	7	109	44	129	139	149	75	169	219	189	251
90	4	110	107	130	142	150	71	170	199	190	207
91	10	111	114	131	162	151	97	171	215	191	159
92	11	112	155	132	138	152	168	172	309	192	155
93	13	113	227	133	117	153	126	173	346	193	214
94	4	114	280	134	100	154	148	174	332	194	237

2.11 Project: Social Interactions

Suppose we have a system with multiple classes of mathematical biology teachers (MBT) at time t . The classes roughly capture the MBT individual attitudes toward learning new stuff. “Reluctant” means the class of MBTs that come into the door as new hires without a disposition to learn new stuff; the positive class corresponds to those who join the MBTs with the right attitude; and the rest of the classes should be self-explanatory. The total population is divided into five classes of MBT individuals: positive (P), reluctant (R), masterful (M), unchangeable (that is, negative) (U), and inactive (I). Assume that $N(t) = R(t) + P(t) + M(t) + U(t) + I(t)$ and that the total number of MBTs is constant, that is, $N(t) = \frac{K}{\mu}$ for all t , where K is a constant. The model is

$$P' = qK - \beta P \frac{M}{K} + \delta R - \mu P$$

$$R' = (1 - q)K - (\delta + \mu)R - \alpha R$$

$$M' = \beta P \frac{M}{K} - (\gamma + \mu)M$$

$$U' = -\mu U + \alpha R$$

$$I' = \gamma M - \mu I,$$

where $q, \beta, \delta, \mu, \gamma$, and α are constants and $0 \leq q \leq 1$.

1. Interpret the parameters.
2. Look at the stability of the simplest equilibrium (your choice).
3. From \mathcal{R}_0 , discuss what would be the impact of changing parameters q, γ , and δ .
4. What are your conclusions from this model?

2.12 Exercises

1. Find the basic reproduction number and the endemic equilibrium for the *SIS* model

$$S' = -\beta SI + \alpha I$$

$$I' = \beta SI - (\alpha + d)I,$$

with a death rate dI due to disease.

2. Find the basic reproduction number for the *SIS* model (2.7) that includes births, natural deaths, and disease deaths. Show that the disease-free equilibrium is asymptotically stable if and only if $\mathcal{R}_0 < 1$, and that if $\mathcal{R}_0 > 1$ there is an asymptotically stable endemic equilibrium.
3. Modify the *SIS* model (2.2) to the situation in which there are two competing strains of the same disease, generating two infective classes I_1, I_2 under the assumption that coinfections are not possible. Does the model predict coexistence of the two strains or competitive exclusion?
- 4.* A communicable disease from which infectives do not recover may be modeled by the pair of differential equations

$$S' = -\beta SI, \quad I' = \beta SI.$$

Show that in a population of fixed size K such a disease will eventually spread to the entire population.

- 5.* Consider a disease spread by carriers who transmit the disease without exhibiting symptoms themselves. Let $C(t)$ be the number of carriers and suppose that carriers are identified and isolated from contact with others at a constant per capita rate α , so that $C' = -\alpha C$. The rate at which susceptibles become infected is proportional to the number of carriers and to the number of susceptibles, so that $S' = -\beta SC$. Let C_0 and S_0 be the numbers of carriers and susceptibles, respectively, at time $t = 0$.

- (a) Determine the number of carriers at time t from the equation for C .
- (b) Substitute the solution to part (a) into the equation for S and determine the number of susceptibles at time t .
- (c) Find $\lim_{t \rightarrow \infty} S(t)$, the number of members of the population who escape the disease.
- 6.* Consider a population of fixed size K in which a rumor is being spread by word of mouth. Let $y(t)$ be the number of people who have heard the rumor at time t and assume that everyone who has heard the rumor passes it on to r others in unit time. Thus, from time t to time $(t + h)$ the rumor is passed on $hry(t)$ times, but a fraction $y(t)/K$ of the people who hear it have already heard it, and thus there are only $hry(t) \left(\frac{K-y(t)}{K} \right)$ people who hear the rumor for the first time. Use these assumptions to obtain an expression for $y(t + h) - y(t)$, divide by h , and take the limit as $h \rightarrow 0$ to obtain a differential equation satisfied by $y(t)$.
7. At 9 AM one person in a village of 100 inhabitants starts a rumor. Suppose that everyone who hears the rumor tells one other person per hour. Using the model of the previous exercise, determine how long it will take until half the village has heard the rumor.
- 8.* If a fraction λ of the population susceptible to a disease that provides immunity against reinfection moves out of the region of an epidemic, the situation may be modeled by a system

$$S' = -\beta SI - \lambda S, \quad I' = \beta SI - \alpha I.$$

Show that both S and I approach zero as $t \rightarrow \infty$.

9. Consider the basic SIR model. We now consider a vaccination class in place of recovery:

$$\begin{aligned} S'(t) &= \mu N - \beta S \frac{I}{N} - (\mu + \phi)S \\ I'(t) &= \beta S \frac{I}{N} - (\mu + \gamma)I \\ V'(t) &= \gamma I + \phi S - \mu V. \end{aligned} \tag{2.22}$$

- (a) Show that $\frac{dN}{dt} = 0$. What does this result imply?
- (b) Discuss why it is enough to study the first two equations.
- (c) Let $\mathcal{R}_0(\phi)$ be \mathcal{R}_0 when $\phi \neq 0$, and $\mathcal{R}_0(0)$ be \mathcal{R}_0 when $\phi = 0$. Compute $\mathcal{R}_0(\phi)$. What is the value of $\mathcal{R}_0(0)$? Compare $\mathcal{R}_0(\phi)$ with $\mathcal{R}_0(0)$.
- (d) Compute the equilibria.
- (e) Do the local stability analysis of the disease-free equilibrium state and the endemic state.
10. In cases of constant recruitment, the limiting system as $t \rightarrow \infty$ and the original system usually have the same qualitative dynamics. We thus consider our previous SIR model with vaccination (refer to problem 9) changing the recruitment rate from the constant μN to the constant A . This means that a

certain fixed number of individuals join or arrive into the susceptible class per unit time. The model becomes

$$\begin{aligned} S'(t) &= \Lambda - \beta S \frac{I}{N} - \mu S \\ I'(t) &= \beta S \frac{I}{N} - (\mu + \gamma) I \\ V'(t) &= \gamma I - \mu V, \end{aligned} \quad (2.23)$$

where, again, $N(t) = S(t) + I(t) + V(t)$.

- (a) What are the units of Λ , $\beta S \frac{I}{N}$, μ , γ , β , and μS ?
- (b) Find the equation for $N'(t)$ where $N = S + I + V$, and solve this equation for $N(t)$. Observe that the population size for this model is not constant.
- (c) Show that $N(t) \rightarrow \frac{\Lambda}{\mu}$ as $t \rightarrow \infty$.
- (d) Consider the limiting system

$$\begin{aligned} S'(t) &= \mu N - \beta S I - \mu S \\ I'(t) &= \beta S I - (\mu + \gamma) I \\ V'(t) &= \gamma I - \mu V. \end{aligned} \quad (2.24)$$

Explain why it is enough to consider the first two equations as $V(t) = N - S(t) - I(t)$ when studying the dynamics of this limiting system, find \mathcal{R}_0 , and do the local stability analysis of the equilibria.

11. An epidemic model with two latent classes is described by the following system of ODE's:

$$\begin{aligned} S' &= \Lambda - \beta S \frac{I}{N} - \mu S \\ L'_1 &= p \beta S \frac{I}{N} - (\mu + k_1 + r_1) L_1 \\ L'_2 &= (1-p) \beta S \frac{I}{N} - (\mu + k_2 + r_2) L_2 \\ I' &= k_1 L_1 + k_2 L_2 - (\mu + r_3) I, \end{aligned}$$

where $N = S + L_1 + L_2 + I$. A fraction p , $0 < p < 1$, goes into the first latent class and a fraction $1 - p$ goes into the second latent class. The two classes have different rates of progression to the infectious stage. Compute the basic reproduction number \mathcal{R}_0 .

12. If vaccination strategies are incorporated for newborns, we assume that not every new birth is susceptible. Suppose that the per capita vaccination rate is p ; a newborn is vaccinated with probability p . The modified model is

$$\begin{aligned} S'(t) &= (1-p)\mu N - \beta S \frac{I}{N} - \mu S \\ I'(t) &= \beta S \frac{I}{N} - (\mu + \gamma) I \\ V'(t) &= \gamma I - \mu V + p\mu N. \end{aligned} \quad (2.25)$$

- (a) Calculate the Jacobian matrix of (2.25) at the disease-free equilibrium points?
- (b) Find the corresponding eigenvalues of the above matrix.
- (c) Find the basic reproduction numbers (\mathcal{R}_0).
- (d) Study the stability of the disease-free equilibrium points of model (2.25). Although N is constant and we could reduce this to a 2-D system, derive the stability of the full 3×3 system by using the Routh–Hurwitz criteria.
13. The same survey of Yale students described in Example 1, Sect. 2.4 reported that 91.1% were susceptible to influenza at the beginning of the year and 51.4% were susceptible at the end of the year. Estimate the basic reproduction number and decide whether there was an epidemic.
14. What fraction of Yale students in Exercise 13 would have had to be immunized to prevent an epidemic?
15. What was the maximum number of Yale students in Exercises 13 and 14 suffering from influenza at any time?
16. An influenza epidemic was reported at an English boarding school in 1978 that spread to 512 of the 763 students. Estimate the basic reproduction number.
17. What fraction of the boarding school students in Exercise 16 would have had to be immunized to prevent an epidemic?
18. What was the maximum number of boarding school students in Exercises 16 and 17 suffering from influenza at any time?
19. A disease is introduced by two visitors into a town with 1200 inhabitants. An average infective is in contact with 0.4 inhabitants per day. The average duration of the infective period is 6 days, and recovered infectives are immune against reinfection. How many inhabitants would have to be immunized to avoid an epidemic?
20. Consider a disease with $\beta N = 0.4$, $1/\alpha = 6$ days in a population of 1200 members. Suppose the disease conferred immunity on recovered infectives. How many members would have to be immunized to avoid an epidemic?
21. A disease begins to spread in a population of 800. The infective period has an average duration of 14 days and the average infective is in contact with 0.1 persons per day. What is the basic reproduction number? To what level must the average rate of contact be reduced so that the disease will die out?
22. European fox rabies is estimated to have a transmission coefficient β of $80 \text{ km}^2 \text{ years/fox}$ (assuming mass action incidence), and an average infective period of 5 days. There is a critical carrying capacity K_c measured in foxes per km^2 , such that in regions with fox density less than K_c , rabies tends to die out, while in regions with fox density greater than K_c , rabies tends to persist. Use a simple Kermack–McKendrick epidemic model to estimate K_c . [Remark: It has been suggested in Great Britain that hunting to reduce the density of foxes below the critical carrying capacity would be a way to control the spread of rabies.]
23. A large English estate has a population of foxes with a density of 1.3 foxes/km^2 . A large fox hunt is planned to reduce the fox population enough to prevent an outbreak of rabies. Assuming that the contact number β/α is $1 \text{ km}^2/\text{fox}$, find what fraction of the fox population must be caught.

24. Following a complaint from the SPCA, organizers decide to replace the fox hunt of Exercise 23 by a mass inoculation of foxes for rabies. What fraction of the fox population must be inoculated to prevent a rabies outbreak?
25. What actually occurs on the estate of these exercises is that 10% of the foxes are killed and 15% are inoculated. Is there danger of a rabies outbreak?
26. Let S , I , and R represent the densities of susceptible, infected, and recovered individuals. Suppose that recovered individuals can become susceptible again after some time. The model equations are

$$\begin{aligned}\frac{dS}{dt} &= -\beta SI + \theta R \\ \frac{dI}{dt} &= \beta SI - \alpha I \\ \frac{dR}{dt} &= \alpha I - \theta R\end{aligned}$$

where $N = S + I + R$ and β , θ , and α are the infection, loss of immunity, and recovery rates, respectively.

- (a) Reduce this model to a two-dimensional system of equations. Don't forget to show that N is constant.
- (b) Find the equilibrium points. Is there a disease-free equilibria ($I = 0$)? Is there an endemic equilibria ($I \neq 0$)? If so, when does it exist?
- (c) Determine the local stability of each of the equilibrium points you found in (b).
27. Here is another approach to the analysis of the SIR model (2.11).

- (a) Divide the two equations of the model to give

$$\frac{I'}{S'} = \frac{dI}{dS} = \frac{(\beta S - \alpha)I}{-\beta SI} = -1 + \frac{\alpha}{\beta S}.$$

- (b) Integrate to find the orbits in the (S, I) -plane,

$$I = -S + \frac{\alpha}{\beta} \log S + c,$$

with c an arbitrary constant of integration.

- (c) Define the function

$$V(S, I) = S + I - \frac{\alpha}{\beta} \log S$$

and show that each orbit is given implicitly by the equation $V(S, I) = c$ for some choice of the constant c .

- (d) Show that no orbit reaches the I -axis and deduce that $S_\infty = \lim_{t \rightarrow \infty} S(t) > 0$, which implies that part of the population escapes infection.
28. For the model (2.16) show that the final total population size is given by

$$N_\infty = \frac{\alpha}{\alpha + d} N_0 + \frac{d}{\alpha + d} S_\infty.$$

29. Consider the basic SIR model with disease deaths, but we now consider a vaccination class in place of recovery:

$$\begin{aligned} S'(t) &= \mu N - \beta S \frac{I}{N} - (\mu + \phi)S \\ I'(t) &= \beta S \frac{I}{N} - (\mu + \gamma + \delta)I \\ V'(t) &= \gamma I + \phi S - \mu V. \end{aligned} \tag{2.26}$$

- (a) Let $\mathcal{R}_0(\phi)$ be \mathcal{R}_0 when $\phi \neq 0$, and $\mathcal{R}_0(0)$ be \mathcal{R}_0 when $\phi = 0$. Compute $\mathcal{R}_0(\phi)$. What is the value of $\mathcal{R}_0(0)$? Compare $\mathcal{R}_0(\phi)$ with $\mathcal{R}_0(0)$.
- (b) Compute the equilibria.
- (c) Do the local stability analysis of the disease-free equilibrium state and the endemic state.

References

- Agur, Z.L., G. Mazor, R. Anderson, and Y. Danon (1993) Pulse mass measles vaccination across age cohorts, *Proc. Nat. Acad. Sci.*, **90**:11698–11702.
- Anderson, R.M. & R.M. May (1991) Infectious Diseases of Humans. Oxford University Press (1991)
- Arreola R., A. Crossa, and M.C. Velasco (2000) Discrete-time SEIS models with exogenous re-infection and dispersal between two patches, *Department of Biometrics, Cornell University, Technical Report Series*, BU-1533-M.
- Castillo-Chavez, C., K. Cooke, W. Huang, and S.A. Levin (1989a) The role of long incubation periods in the dynamics of HIV/AIDS. Part 1: Single Populations Models, *J. Math. Biol.*, **27**:373-98.
- Castillo-Chavez, C., W. Huang, and J. Li (1996a) Competitive exclusion in gonorrhea models and other sexually-transmitted diseases, *SIAM J. Appl. Math.*, **56**: 494–508.
- Castillo-Chavez, C., W. Huang, and J. Li (1997) The effects of females' susceptibility on the coexistence of multiple pathogen strains of sexually-transmitted diseases, *Journal of Mathematical Biology*, **35**:503–522.
- Castillo-Chavez C., and A.A. Yakubu (2000) Epidemics models on attractors, *Contemporary Mathematics*, AMS, **284**: 23–42.
John Wiley & Sons, New York, 2000.
- Dietz, K. (1982) Overall patterns in the transmission cycle of infectious disease agents. In: R.M. Anderson, R.M. May (eds) Population Biology of Infectious Diseases. Life Sciences Research Report **25**, Springer-Verlag, Berlin-Heidelberg-New York, pp. 87–102 (1982)

9. Engbert, R. and F. Drepper (1994) Chance and chaos in population biology-models of recurrent epidemics and food chain dynamics, *Chaos, Solutions & Fractals*, **4**(7):1147–1169.
10. Evans, A.S. (1982) *Viral Infections of Humans*, 2nd ed., Plenum Press, New York.
11. Heesterbeek, J.A.P. and J.A.J Metz (1993) The saturating contact rate in marriage and epidemic models. *J. Math. Biol.*, **31**: 529–539.
12. Hethcote, H.W. (1976) Qualitative analysis for communicable disease models, *Math. Biosc.*, **28**:335–356.
13. Hethcote, H.W. (1989) Three basic epidemiological models. In *Applied Mathematical Ecology*, S. A. Levin, T.G. Hallam and L.J. Gross, eds., Biomathematics **18**, 119–144, Springer-Verlag, Berlin-Heidelberg-New York.
14. Hethcote, H.W. and S.A. Levin (1989) Periodicity in epidemic models. In : S.A. Levin, T.G. Hallam & L.J. Gross (eds) *Applied Mathematical Ecology*, Biomathematics **18**, Springer-Verlag, Berlin-Heidelberg-New York: pp. 193–211.
15. Hethcote, H.W., H.W. Stech and P. van den Driessche (1981) Periodicity and stability in epidemic models: a survey. In: S. Busenberg & K.L. Cooke (eds.) *Differential Equations and Applications in Ecology, Epidemics and Population Problems*, Academic Press, New York: 65–82.
16. Hurwitz, A. (1895) Über die Bedingungen unter welchen eine Gleichung nur Wurzeln mit negativen reellen Teilen bezizt, *Math. Annalen* **46**: 273–284.
17. Kermack, W.O. & A.G. McKendrick (1927) A contribution to the mathematical theory of epidemics. *Proc. Royal Soc. London*. **115**: 700–721.
18. Kermack, W.O. & A.G. McKendrick (1932) Contributions to the mathematical theory of epidemics, part. II. *Proc. Roy. Soc. London*, **138**: 55–83.
19. Kermack, W.O. & A.G. McKendrick (1932) Contributions to the mathematical theory of epidemics, part. III. *Proc. Roy. Soc. London*, **141**: 94–112.
20. Mena-Lorca, J. & H.W. Hethcote (1992) Dynamic models of infectious diseases as regulators of population size. *J. Math. Biol.*, **30**: 693–716.
21. Raggett, G.F. (1982) Modeling the Eyam plague, *IMA Journal*, **18**:221–226.
22. Routh, E.J. (1877) *A Treatise on the Stability of a Given State of Motion: Particularly Steady Motion*, MacMillan.
23. Sánchez B.N., P.A. Gonzalez, R.A. Saenz (2000) The influence of dispersal between two patches on the dynamics of a disease, *Department of Biometrics, Cornell University, Technical Report Series*, BU-1531-M.
24. Shulgin, B., L. Stone, and Z. Agur (1998) Pulse vaccination strategy in the SIR epidemic model, *Bull. Math. Biol.*, **60**:1123–1148.
25. Stone, L., B. Shulgin, Z. Agur (2000) Theoretical examination of the pulse vaccination in the SIR epidemic model, *Math. and Computer Modeling*, **31**:207–215.

Chapter 3

Endemic Disease Models



In this chapter, we consider models for disease that may be endemic. In the preceding chapter we studied *SIS* models with and without demographics and *SIR* models with demographics. In each model, the basic reproduction number \mathcal{R}_0 determined a threshold. If $\mathcal{R}_0 < 1$ the disease dies out, while if $\mathcal{R}_0 > 1$ the disease becomes endemic. The analysis in each case involves determination of equilibria and determining the asymptotic stability of each equilibrium by linearization about the equilibrium. In each of the cases studied in the preceding chapter the disease-free equilibrium was asymptotically stable if and only if $\mathcal{R}_0 < 1$ and if $\mathcal{R}_0 > 1$ there was a unique endemic equilibrium that was asymptotically stable. In this chapter, we will see that these properties continue to hold for many more general models, but there are situations in which there may be an asymptotically stable endemic equilibrium when $\mathcal{R}_0 < 1$, and other situations in which there is an endemic equilibrium that is unstable for some values of $\mathcal{R}_0 > 1$.

In Sect. 2.3 we analyzed the *SIR* model for diseases from which infectives recover with immunity against reinfection:

$$\begin{aligned} S' &= \Lambda(N) - \beta SI - \mu S \\ I' &= \beta SI - \mu I - \alpha I - dI \\ N' &= \Lambda(N) - dI - \mu N. \end{aligned} \tag{3.1}$$

The following basic result holds for (3.1).

Theorem 3.1 *The basic reproduction number for the model (3.1) is given by*

$$\mathcal{R}_0 = \frac{\beta K}{\mu + \alpha} = \frac{K}{S_\infty}.$$

If $\mathcal{R}_0 < 1$, the system has only the disease-free equilibrium and this equilibrium is asymptotically stable.

Here, K is the population carrying capacity and S_∞ is the susceptible population size at the endemic equilibrium. The theorem says that the disease-free equilibrium is locally asymptotically stable. We recall that this means that solutions with initial values close to this equilibrium remain close to the equilibrium and approach the equilibrium as $t \rightarrow \infty$. In fact, it is not difficult to prove that this asymptotic stability is *global*, that is, that every solution approaches the disease-free equilibrium. If the quantity \mathcal{R}_0 is greater than one, then the disease-free equilibrium is unstable, but there is an endemic equilibrium that is (locally) asymptotically stable.

In fact, these properties hold for some endemic disease models with more complicated compartmental structure. We will describe some examples.

3.1 More Complicated Endemic Disease Models

3.1.1 Exposed Periods

In many infectious diseases there is an exposed period after the transmission of infection from susceptibles to potentially infective members but before these potential infectives develop symptoms and can transmit infection. To incorporate an exposed compartment with mean exposed period $1/\kappa$ we add an exposed class E and use compartments S, E, I, R and total population size $N = S + E + I + R$ to give a generalization of the epidemic model (3.1)

$$\begin{aligned} S' &= \Lambda(N) - \beta SI - \mu S \\ E' &= \beta SI - (\kappa + \mu)E \\ I' &= \kappa E - (\alpha + \mu)I. \end{aligned} \tag{3.2}$$

A flow chart is shown in Fig. 3.1.

The analysis of this model is similar to the analysis of (3.1), but with I replaced by $E + I$. That is, instead of using the number of infectives as one of the variables we use the total number of infected members, whether or not they are capable of transmitting infection.

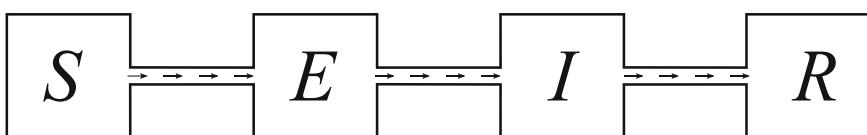


Fig. 3.1 Flow chart for the *SEIR* endemic model (3.2)

3.1.2 A Treatment Model

One form of treatment that is possible for some diseases is vaccination to protect against infection before the beginning of an epidemic. For example, this approach is commonly used for protection against annual influenza outbreaks. A simple way to model this would be to reduce the total population size by the fraction of the population protected against infection.

In reality, such inoculations are only partly effective, decreasing the rate of infection and also decreasing infectivity if a vaccinated person does become infected. This may be modeled by dividing the population into two groups with different model parameters which would require some assumptions about the mixing between the two groups. This is not difficult but we will not explore this direction until Chap. 5 on heterogeneous mixing.

If there is a treatment for infection once a person has been infected, this may be modeled by supposing that there is a rate γ proportional to the number of infectives at which infectives are selected for treatment, and that treatment reduces infectivity by a fraction δ . Suppose that the rate of removal from the treated class is η . This leads to the *SITR* model, where T is the treatment class, given by

$$\begin{aligned} S' &= \mu N - \beta S[I + \delta T] - \mu S \\ I' &= \beta S[I + \delta T] - (\alpha + \gamma + \mu)I \\ T' &= \gamma I - (\eta + \mu)T. \end{aligned} \tag{3.3}$$

A flow chart is shown in Fig. 3.2. In this model, we assume that the natural birth and death rates are equal so that the total population size remains constant.

In order to calculate the basic reproduction number, we observe that an infective in a totally susceptible population causes βN new infections in unit time, and the

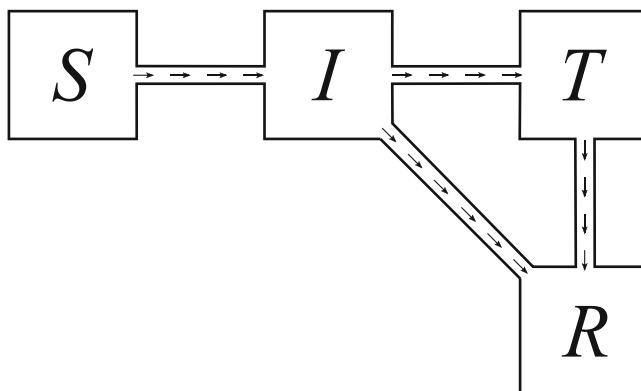


Fig. 3.2 Flow chart for the *SITR* endemic model (3.3)

mean time spent in the infective compartment is $1/(\alpha + \gamma + \mu)$. In addition, a fraction $\gamma/(\alpha + \gamma + \mu)$ of infectives are treated because $(\gamma + \alpha + \mu)$ is the rate at which the number of infectives decreases overall while γ is the rate at which these infectives are selected for treatment. While in the treatment stage the number of new infections caused in unit time is $\delta\beta N$, and the mean time in the treatment class is $1/(\eta + \mu)$. Thus

$$\mathcal{R}_0 = \frac{\beta N}{\alpha + \gamma + \mu} + \frac{\gamma}{\alpha + \gamma + \mu} \frac{\delta\beta N}{\eta + \mu}. \quad (3.4)$$

It is possible that if $\delta < 1$ and $\alpha > \eta$ the treatment may increase the reproduction number. However, since $\alpha > \eta$ would mean that treatment prolongs the infection, this is quite unlikely.

The equilibrium conditions for the model (3.3) are

$$\begin{aligned} \mu N &= \beta S[I + \delta T] + \mu S \\ \beta S[I + \delta T] &= (\alpha + \gamma + \mu)I \\ \gamma I &= (\eta + \mu)T. \end{aligned} \quad (3.5)$$

Substitution of the last of these equilibrium conditions into the second gives

$$\beta S \frac{\delta\gamma + \eta + \mu}{\eta + \mu} I = (\alpha + \gamma + \mu)I,$$

and this implies that either $I = 0$ (disease-free equilibrium) or

$$\beta S = \frac{(\alpha + \gamma + \mu)(\eta + \mu)}{\delta\gamma + \eta + \mu}$$

(endemic equilibrium). An endemic equilibrium exists if and only if the value of S given by this condition is less than N , and this is equivalent to $\mathcal{R}_0 > 1$.

The Jacobian matrix or matrix of the linearization of (3.3) at an equilibrium (S, I, T) is

$$\begin{bmatrix} -\beta(I + \delta T) - \mu & -\beta S & -\delta\beta S \\ \beta(I + \delta T) & \beta S - (\alpha + \gamma + \mu) & \delta\beta S \\ 0 & \gamma & -(\eta + \mu) \end{bmatrix},$$

and at the disease-free equilibrium $(N, 0, 0)$ this is

$$\begin{bmatrix} -\mu & -\beta N & -\delta\beta N \\ 0 & \beta N - (\alpha + \gamma + \mu) & \delta\beta N \\ 0 & \gamma & -(\eta + \mu) \end{bmatrix}.$$

The eigenvalues of this matrix are $-\mu$ and the eigenvalues of the 2×2 matrix

$$\begin{bmatrix} \beta N - (\alpha + \gamma + \mu) & \delta \beta N \\ \gamma & -(\eta + \mu) \end{bmatrix}.$$

The eigenvalues of a 2×2 matrix have negative real part if and only if the matrix has negative trace and positive determinant. The condition that the determinant is positive is

$$\beta N < \frac{(\eta + \mu)(\alpha + \gamma + \mu)}{\eta + \mu + \delta \gamma}, \quad (3.6)$$

and the condition that the trace is negative is

$$\beta N < (\alpha + \gamma + \mu) + (\eta + \mu). \quad (3.7)$$

Then, since

$$(\eta + \mu)(\alpha + \gamma + \mu) < (\alpha + \gamma + \mu)(\eta + \mu + \delta \gamma) + (\eta + \mu + \delta \gamma)(\eta + \mu),$$

if (3.6) is satisfied (3.7) is also satisfied. Thus the condition $\mathcal{R}_0 < 1$ is equivalent to (3.6) and the asymptotic stability of the disease-free equilibrium.

To show that the endemic equilibrium is asymptotically stable if it exists, that is, if $\mathcal{R}_0 > 1$, we must make use of the four conditions [26, 37] introduced in Chap. 2. A somewhat complicated calculation shows that this is indeed the case.

3.1.3 Vertical Transmission

In some diseases, notably Chagas' disease, HIV/AIDS, hepatitis B, and rinderpest (in cattle), infection may be transferred not only horizontally (by contact between individuals) but also vertically (from an infected parent to a newly born offspring) [8]. We formulate an *SIR* model with vertical transmission by assuming that a fraction q of the offspring of infective members of the population are infective at birth. For simplicity, we assume that there are no disease deaths so that the total population size N is constant, and our model is based on (3.1). The birth rate in this model is $\Lambda = \mu N$, and we assume that births are distributed proportionally among compartments. Thus the rate of births to infectives is μI , the rate of newborn infectives is $q\mu I$, and the rate of newborn susceptibles is $\mu N - q\mu I$. This leads to the model

$$\begin{aligned} S' &= \mu N - q\mu I - \beta SI - \mu S \\ I' &= q\mu I + \beta SI - \mu I - \alpha I. \end{aligned} \quad (3.8)$$

From the second equation, we see that equilibrium requires either $I = 0$ (disease-free) or $\beta S = \mu(1 - q) + \alpha$. At the disease-free equilibrium, $S = N$, $I = 0$, and the matrix of the linearization is

$$\begin{bmatrix} -\mu & -q\mu - \beta N \\ 0 & \beta N - \mu(1 - q) - \alpha \end{bmatrix}.$$

Thus the disease-free equilibrium is asymptotically stable if and only if

$$\beta N < \mu(1 - q) + \alpha.$$

This suggests that

$$\mathcal{R}_0 = \frac{\beta N + \mu q}{\mu + \alpha}.$$

To see that this is indeed correct, we note that the term $\beta N / (\mu + \alpha)$ represents horizontally transmitted infections at rate βN over a death-adjusted infective period $1 / (\mu + \alpha)$, and the term $\frac{\mu q}{\mu + \alpha}$ represents vertically transmitted infections per infective. It is not difficult to verify that the endemic equilibrium, which exists if and only if $\mathcal{R}_0 > 1$ is asymptotically stable.

3.2 Some Applications of the *SIR* Model

3.2.1 Herd Immunity

In order to prevent a disease from becoming endemic, it is necessary to reduce the basic reproduction number \mathcal{R}_0 below one. This may sometimes be achieved by immunization. If a fraction p of the $\Lambda(N)$ newborn members per unit time of the population is successfully immunized, the effect is to replace N by $N(1 - p)$, and thus to reduce the basic reproduction number to $\mathcal{R}_0(1 - p)$. The requirement $\mathcal{R}_0(1 - p) < 1$ gives $1 - p < 1/\mathcal{R}_0$, or

$$p > 1 - \frac{1}{\mathcal{R}_0}.$$

A population is said to have *herd immunity* if a large enough fraction has been immunized to assure that the disease cannot become endemic. The only disease for which this has actually been achieved worldwide is smallpox for which \mathcal{R}_0 is approximately 5, so that 80% immunization does provide herd immunity, and rinderpest, a cattle disease.

For measles, epidemiological data in the USA indicate that \mathcal{R}_0 for rural populations ranges from 5.4 to 6.3, requiring vaccination of 81.5–84.1% of the

population. In urban areas \mathcal{R}_0 ranges from 8.3 to 13.0, requiring vaccination of 88.0–92.3% of the population. In Great Britain, \mathcal{R}_0 ranges from 12.5 to 16.3, requiring vaccination of 92–94% of the population. The measles vaccine is not always effective, and vaccination campaigns are never able to reach everyone. As a result, herd immunity against measles has not been achieved (and probably never can be). An additional issue is that an anti-vaccination movement has developed, partly because of a fallacious belief that there is a link between the measles-mumps-rubella vaccine and the development of autism and partly because of a general opposition to vaccines.

Since smallpox is viewed as more serious and requires a lower percentage of the population be immunized, herd immunity was attainable for smallpox. In fact, smallpox has been eliminated; the last known case was in Somalia in 1977, and the virus is maintained now only in laboratories. The eradication of smallpox was actually more difficult than expected because high vaccination rates were achieved in some countries but not everywhere, and the disease persisted in some countries. The eradication of smallpox was possible only after an intensive campaign for worldwide vaccination [22].

3.2.2 Age at Infection

In order to calculate the basic reproduction number \mathcal{R}_0 for a disease modeled by a system (3.1), we need to know the values of the contact rate β and the parameters μ , K , and α . The parameters μ , K , and α can usually be measured experimentally but the contact rate β is difficult to determine directly. There is an indirect method of estimating \mathcal{R}_0 in terms of the life expectancy and the mean age at infection which enables us to avoid having to estimate the contact rate. In this calculation, we will assume that β is constant, but we will also indicate the modifications needed when β is a function of total population size N . The calculation assumes exponentially distributed life spans and infective periods. The result is valid so long as the life span is exponentially distributed, but if the life span is not exponentially distributed the result could be quite different.

Consider the “age cohort” of members of a population born at some time t_0 and let a be the age of members of this cohort. If $y(a)$ represents the fraction of members of the cohort who survive to age (at least) a , then the assumption that a fraction μ of the population dies per unit time means that $y'(a) = -\mu y(a)$. Since $y(t_0) = 1$, we may solve this first order initial value problem to obtain $y(a) = e^{-\mu a}$. The fraction dying at (exactly) age a is $-y'(a) = \mu y(a)$. The mean life span is the average age at death, which is $\int_0^\infty a[-y'(a)]da$, and if we integrate by parts we find that this life expectancy is

$$\int_{t_0}^\infty [-ay'(a)]da = [-ay(a)]_{t_0}^\infty + \int_{t_0}^\infty y(a)da = \int_{t_0}^\infty y(a)da.$$

Since $y(a) = e^{-\mu a}$, this reduces to $1/\mu$. The life expectancy is often denoted by L , so that we may write

$$L = \frac{1}{\mu}.$$

The rate at which surviving susceptible members of this cohort become infected at age a and time $t_0 + a$ is $\beta I(t_0 + a)$. Thus, if $z(a)$ is the fraction of the age cohort alive and still susceptible at age a , $z'(a) = -[\mu + \beta I(t_0 + a)]z(a)$. Solution of this first linear order differential equation gives

$$z(a) = e^{-[\mu a + \int_0^a \beta I(t_0 + b) db]} = y(a)e^{-\int_0^a \beta I(t_0 + b) db}.$$

The mean length of time in the susceptible class for members who may become infected, as opposed to dying while still susceptible, is

$$\int_0^\infty e^{-\int_0^a \beta I(t_0 + b) db} da,$$

and this is the mean age at which members become infected. If the system is at an equilibrium I_∞ , this integral may be evaluated, and the mean age at infection, denoted by A , is given by

$$A = \int_0^\infty e^{-\beta I_\infty a} da = \frac{1}{\beta I_\infty}.$$

For our model the endemic equilibrium is

$$I_\infty = \frac{\mu K}{\mu + \alpha} - \frac{\mu}{\beta},$$

and this implies

$$\frac{L}{A} = \frac{\beta I_\infty}{\mu} = \mathcal{R}_0 - 1. \quad (3.9)$$

This relation is very useful in estimating basic reproduction numbers. For example, in some urban communities in England and Wales between 1956 and 1969 the average age of contracting measles was 4.8 years. If life expectancy is assumed to be 70 years, this indicates $\mathcal{R}_0 = 15.6$.

If β is a function $\beta(N)$ of total population size and K is the carrying capacity, the relation (3.9) becomes

$$\mathcal{R}_0 = \frac{\beta(K)}{\beta(N_0)} \left[1 + \frac{L}{A} \right].$$

If disease mortality does not have a large effect on total population size, in particular if there is no disease mortality, this relation is very close to (3.9).

The relation between age at infection and basic reproduction number indicates that measures such as inoculations, which reduce \mathcal{R}_0 , will increase the average age at infection. For diseases such as rubella (German measles), whose effects may be much more serious in adults than in children, this indicates a danger that must be taken into account: While inoculation of children will decrease the number of cases of illness, it will tend to increase the danger to those who are not inoculated or for whom the inoculation is not successful. Nevertheless, the number of infections in older people will be reduced, although the fraction of cases which are in older people will increase.

3.2.3 The Inter-Epidemic Period

Many common childhood diseases, such as measles, whooping cough, chicken pox, diphtheria, and rubella, exhibit variations from year to year in the number of cases. These fluctuations are frequently regular oscillations, suggesting that the solutions of a model might be periodic. This does not agree with the predictions of the model we have been using in this section; however, it would not be inconsistent with solutions of the characteristic equation, which are complex conjugate with small negative real part corresponding to lightly damped oscillations approaching the endemic equilibrium. Such behavior would look like recurring epidemics. If the eigenvalues of the matrix of the linearization at an endemic equilibrium are $-u \pm iv$, where $i^2 = -1$, then the solutions of the linearization are of the form $B e^{-ut} \cos(vt + c)$, with decreasing “amplitude” $B e^{-ut}$ and “period” $\frac{2\pi}{v}$.

For the model (3.1) we recall that at the endemic equilibrium we have

$$\beta I_\infty + \mu = \mu \mathcal{R}_0, \quad \beta S_\infty = \mu + \alpha$$

and the matrix of the linearization is

$$\begin{bmatrix} -\mu \mathcal{R}_0 & -(\mu + \alpha) \\ \mu(\mathcal{R}_0 - 1) & 0 \end{bmatrix}.$$

The eigenvalues are the roots of the quadratic equation

$$\lambda^2 + \mu \mathcal{R}_0 \lambda + \mu(\mathcal{R}_0 - 1)(\mu + \alpha) = 0,$$

which are

$$\lambda = \frac{-\mu \mathcal{R}_0 \pm \sqrt{\mu^2 \mathcal{R}_0^2 - 4\mu(\mathcal{R}_0 - 1)(\mu + \alpha)}}{2}.$$

If the mean infective period $1/\alpha$ is much shorter than the mean life span $1/\mu$, we may neglect the terms that are quadratic in μ . Thus, the eigenvalues are approximately

$$\frac{-\mu \mathcal{R}_0 \pm \sqrt{-4\mu(\mathcal{R}_0 - 1)\alpha}}{2},$$

and these are complex with imaginary part $\sqrt{\mu(\mathcal{R}_0 - 1)\alpha}$. This indicates oscillations with period approximately

$$\frac{2\pi}{\sqrt{\mu(\mathcal{R}_0 - 1)\alpha}}.$$

We use the relation $\mu(\mathcal{R}_0 - 1) = \mu L/A$ and the mean infective period $\tau = 1/\alpha$ to see that the interepidemic period T is approximately $2\pi\sqrt{A\tau}$. Thus, for example, for recurring outbreaks of measles with an infective period of 2 weeks or $1/26$ year in a population with a life expectancy of 70 years with R_0 estimated as 15, we would expect outbreaks spaced 2.76 years apart. Also, as the “amplitude” at time t is $e^{-\mu R_0 t/2}$, the maximum displacement from equilibrium is multiplied by a factor $e^{-(15)(2.76)/140} = 0.744$ over each cycle. In fact, many observations of measles outbreaks indicate less damping of the oscillations, suggesting that there may be additional influences that are not included in our simple model. To explain oscillations about the endemic equilibrium a more complicated model is needed. One possible generalization would be to assume seasonal variations in the contact rate [13, 27]. This is a reasonable supposition for a childhood disease most commonly transmitted through school contacts, especially in winter in cold climates. Note, however, that data from observations are never as smooth as model predictions and models are inevitably gross simplifications of reality which cannot account for random variations in the variables. It may be difficult to judge from experimental data whether an oscillation is damped or persistent.

3.2.4 “Epidemic” Approach to Endemic Equilibrium

In the model (3.1) the demographic time scale described by the birth and natural death rates μK and μ and the epidemiological time scale described by the rate α of departure from the infective class may differ substantially. Think, for example, of a natural death rate $\mu = 1/75$, corresponding to a human life expectancy of 75 years, and epidemiological parameter $\alpha = 25$, describing a disease from which all infectives recover after a mean infective period of $1/25$ year, or 2 weeks. Suppose we consider a carrying capacity $K = 1000$ and take $\beta = 0.1$, indicating that an average infective makes $(0.1)(1000) = 100$ contacts per year. Then $\mathcal{R}_0 = 4.00$, and at the endemic equilibrium we have $S_\infty = 250.13$, $I_\infty = 0.40$, $R_\infty = 749.47$. This

equilibrium is globally asymptotically stable and is approached from every initial state.

However, if we take $S(0) = 999$, $I(0) = 1$, $R(0) = 0$, simulating the introduction of a single infective into a susceptible population and solve the system numerically we find that the number of infectives rises sharply to a maximum of 400 and then decreases to almost zero in a period of 0.4 year, or about 5 months. In this time interval the susceptible population decreases to 22 and then begins to increase, while the removed (recovered and immune against reinfection) population increases to almost 1000 and then begins a gradual decrease. The size of this initial “epidemic” could not have been predicted from our qualitative analysis of the system (3.1). On the other hand, since μ is so small compared to the other parameters of the model, we might consider neglecting μ , replacing it by zero in the model. If we do this, the model reduces to the simple Kermack–McKendrick epidemic model (without births and deaths) of Sect. 2.4.

If we follow the model (3.1) over a longer time interval we find that the susceptible population grows to 450 after 46 years, then drops to 120 during a small epidemic with a maximum of 18 infectives, and exhibits widely spaced epidemics decreasing in size. It takes a very long time before the system comes close to the endemic equilibrium and remains close to it. The large initial epidemic conforms to what has often been observed in practice when an infection is introduced into a population with no immunity, such as the smallpox inflicted on the Aztecs by the invasion of Cortez.

If we use the model (3.1) with the same values of β , K , and μ , but take $\alpha = 0$, $d = 25$ to describe a disease fatal to all infectives, we obtain very similar results. Now the total population is $S + I$, which decreases from an initial size of 1000 to a minimum of 22 and then gradually increases and eventually approaches its equilibrium size of 250.53. Thus, the disease reduces the total population size to one-fourth of its original value, suggesting that infectious diseases may have large effects on population size. This is true even for populations which would grow rapidly in the absence of infection, as we shall see in a later section (Sect. 3.7).

3.3 Temporary Immunity

In the *SIR* models that we have studied, it has been assumed that the immunity received by recovery from the disease is permanent. This is not always true, as there may be a gradual loss of immunity with time. In addition, there are often mutations in a virus, and as a result the active disease strain is sufficiently different from the strain from which an individual has recovered and the immunity received may wane.

Temporary immunity may be described by an *SIRS* model in which a rate of transfer from R to S is added to an *SIR* model. For simplicity, we confine our attention to epidemic models, without including births, natural deaths, and disease deaths, but the analysis of models including births and deaths would lead to the same conclusions. Thus we begin with a model

$$\begin{aligned} S' &= -\beta SI + \theta R \\ I' &= \beta SI - \alpha I \\ R' &= \alpha I - \theta R, \end{aligned}$$

with a proportional rate θ of loss of immunity.

Since $N' = (S + I + R)' = 0$, the total population size N is constant, and we may replace R by $N - S - I$ and reduce the model to a two-dimensional system

$$\begin{aligned} S' &= -\beta SI + \theta(N - S - I) \\ I' &= \beta SI - \alpha I. \end{aligned} \tag{3.10}$$

Equilibria are solutions of the system

$$\beta SI + \theta S + \theta I = \theta N$$

$$\alpha I + \theta S + \theta I = \theta N,$$

and there is a disease-free equilibrium $S = \alpha/\beta$, $I = 0$. If $\mathcal{R}_0 = \beta N/\alpha > 1$, there is also an endemic equilibrium with

$$\beta S = \alpha, \quad (\alpha + \theta)I = \theta(N - S).$$

The matrix of the linearization of (3.10) at an equilibrium (S, I) is

$$A = \begin{bmatrix} -(\beta I + \theta) & -(\beta S + \theta) \\ \beta I & \beta S - \alpha \end{bmatrix}.$$

At the disease-free equilibrium A has the sign structure

$$\begin{bmatrix} - & - \\ 0 & \beta N - \alpha \end{bmatrix}.$$

This matrix has negative trace and positive determinant if and only if $\beta N < \alpha$, or $\mathcal{R}_0 < 1$. At an endemic equilibrium, the matrix has sign structure

$$\begin{bmatrix} - & - \\ + & 0 \end{bmatrix}.$$

and thus always has negative trace and positive determinant. We see from this that, as in other models studied in this chapter, the disease-free equilibrium is asymptotically stable if and only if the basic reproduction number is less than 1 and the endemic equilibrium, which exists if and only if the basic reproduction number

exceeds 1, is always asymptotically stable. However, it is possible for a different *SIRS* model to have quite different behavior.

3.3.1 *Delay in an *SIRS* Model

We consider an *SIRS* model, which assumes a constant period of temporary immunity following recovery from the infection in place of an exponentially distributed period of temporary immunity [24]. We assume that there is a temporary immunity period of fixed length ω , after which recovered infectives revert to the susceptible class. The resulting model is described by the system of differential-difference equations

$$\begin{aligned} S'(t) &= -\beta S(t)I(t) + \alpha I(t - \omega) \\ I'(t) &= \beta S(t)I(t) - \alpha I(t) \\ R'(t) &= \alpha I(t) - \alpha I(t - \omega). \end{aligned} \tag{3.11}$$

The equilibrium analysis of a system of differential-difference equations with a delay ω is analogous to the equilibrium analysis of a system of ordinary differential equations, but there are important variations. Instead of assigning an initial condition at $t = 0$ it is necessary to assign initial data on the interval $-\omega \leq t \leq 0$. Equilibria of a system of differential-difference equations are constant solutions, just as for systems of differential equations, and the process of linearization about an equilibrium is the same.

The characteristic equation at an equilibrium is the condition that the linearization at the equilibrium has a solution whose components are constant multiples of $e^{\lambda t}$. In the ordinary differential equation case, this is just the equation that determines the eigenvalues of the coefficient matrix, a polynomial equation, but in the general case, it is a transcendental equation. The result on which our analysis depends, which we state without proof, is that an equilibrium is asymptotically stable if all roots of the characteristic equation have negative real part, or equivalently that the characteristic equation have no roots with real part greater than or equal to zero [5].

In (3.11), since $N = S + I + R$ is constant, we may discard the equation for R and use a two-dimensional model

$$\begin{aligned} S'(t) &= -\beta S(t)I(t) + \alpha I(t - \omega) \\ I'(t) &= \beta S(t)I(t) - \alpha I(t). \end{aligned} \tag{3.12}$$

Equilibria are given by $I = 0$ or $\beta S = \alpha$. There is a disease-free equilibrium $S = N, I = 0$. There is also an endemic equilibrium for which $\beta S = \alpha$. However, the two equations for S and I give only a single equilibrium condition. To determine the

endemic equilibrium (S_∞, I_∞) , we must write the equation for R in the integrated form

$$R(t) = \int_{t-\omega}^t \alpha I(x) dx$$

to give $R_\infty = \omega \alpha I_\infty$. We also have $\beta S_\infty = \alpha$, and from $S_\infty + I_\infty + R_\infty = N$ we obtain

$$\beta I_\infty = \frac{(\beta N - \alpha)}{1 + \omega \alpha}.$$

To linearize about an equilibrium (S_∞, I_∞) of (3.12) we substitute

$$S(t) = S_\infty + u(t), \quad I(t) = I_\infty + v(t),$$

and neglect the quadratic term, giving the linearization

$$\begin{aligned} u'(t) &= -\beta I_\infty u(t) - \beta S_\infty v(t) + \alpha v(t - \omega) \\ v'(t) &= \beta I_\infty u(t) + \beta S_\infty v(t) - \alpha v(t). \end{aligned}$$

The characteristic equation is the condition on λ that this linearization has a solution

$$u(t) = u_0 e^{\lambda t}, \quad v(t) = v_0 e^{\lambda t},$$

and this is

$$\begin{aligned} (\beta I_\infty + \lambda)u_0 + (\beta S_\infty - \alpha e^{-\lambda \omega})v_0 &= 0 \\ \beta I_\infty u_0 + (\beta S_\infty - \alpha - \lambda)v_0 &= 0, \end{aligned}$$

or

$$\det \begin{bmatrix} \lambda + \beta I_\infty & \beta S_\infty - \alpha e^{\lambda \omega} \\ \beta I_\infty & \beta S_\infty - \alpha - \lambda \end{bmatrix}.$$

This reduces to

$$\alpha \beta I_\infty \frac{1 - e^{-\lambda \omega}}{\lambda} = -[\lambda + \alpha + \beta S_\infty + \beta I_\infty]. \quad (3.13)$$

At the disease-free equilibrium $S_\infty = N$, $I_\infty = 0$, this reduces to a linear equation with a single root $\lambda = -\beta N - \alpha$, which is negative if and only if $\mathcal{R}_0 = \beta N / \alpha < 1$.

We think of ω and N as fixed and consider β and α as parameters. If $\alpha = 0$ the Eq. (3.13) is linear and its only root is $-\beta S_\infty - \beta I_\infty < 0$. Thus, there is a region in

the (α, β) parameter space containing the β -axis, in which all roots of (3.13) have negative real part. In order to find how large this stability region is, we make use of the fact that the roots of (3.13) depend continuously on β and α . A root can move into the right half-plane only by passing through the value zero or by crossing the imaginary axis as βN and α vary. Thus, the stability region contains the β -axis and extends into the plane until there is a root $\lambda = 0$ or until there is a pair of pure imaginary roots $\lambda = \pm iy$ with $y > 0$. Since the left side and right side of (3.13) have opposite sign for real $\lambda \geq 0$, there cannot be a root $\lambda = 0$.

The condition that there is a root $\lambda = iy$ is

$$\alpha\beta I_\infty \frac{1 - e^{-i\omega\alpha}}{iy} = -(iy + \alpha + \beta S_\infty + \beta I_\infty) \quad (3.14)$$

and separation into real and imaginary parts gives the pair of equations

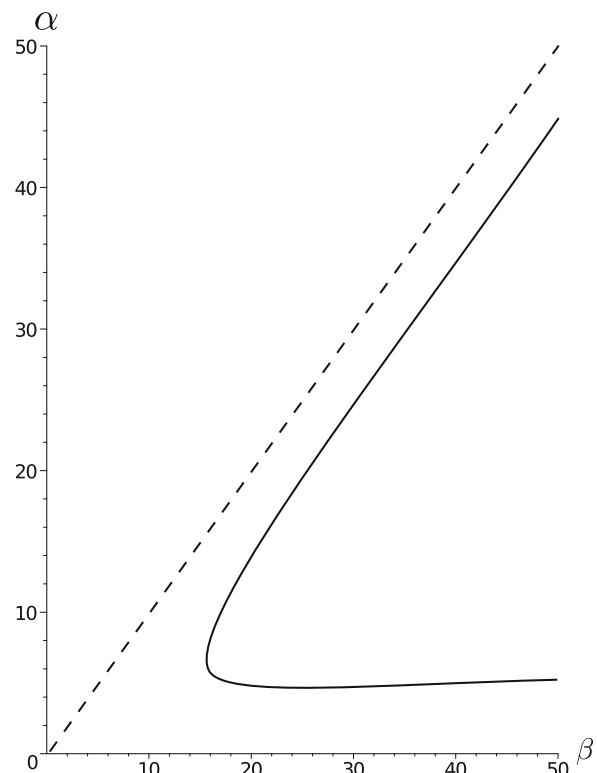
$$\alpha\beta \frac{\sin \omega y}{y} = -[\alpha + \beta S_\infty + \beta I_\infty], \quad \alpha\beta I_\infty \frac{1 - \cos \omega y}{y} = y. \quad (3.15)$$

To satisfy the first condition, it is necessary to have $\omega\alpha > 1$ since $|\sin \omega y| \leq |\omega y|$ for all y . This implies, in particular, that the endemic equilibrium is asymptotically stable if $\omega\alpha < 1$. In addition, it is necessary to have $\sin \omega y < 0$. There is an infinite sequence of intervals on which $\sin \omega y < 0$, the first being $\pi < \omega y < 2\pi$. For each of these intervals, the equations (3.15) define a curve in the (β, α) plane parametrically with y as parameter. The region in the plane below the first of these curves is the region of asymptotic stability, that is, the set of values of β and α for which the endemic equilibrium is asymptotically stable. This curve is shown for $\omega = 1$, $N = 1$ in Fig. 3.3. Since $\mathcal{R}_0 = \beta N / \alpha > 1$, only the portion of the (β, α) plane below the line $\alpha = \beta N$ is relevant.

The new feature of the model of this section is that the endemic equilibrium is not asymptotically stable for all parameter values. What is the behavior of the model if the parameters are such that the endemic equilibrium is unstable? A plausible suggestion is that since the loss of stability corresponds to a root $\lambda = iy$ of the characteristic equation there are solutions of the model behaving like the real part of e^{iyt} , that is, that there are periodic solutions. This is exactly what does happen according to a very general result called the Hopf bifurcation theorem [25], which says that when roots of the characteristic equation cross the imaginary axis a stable periodic orbit arises.

From an epidemiological point of view periodic behavior is unpleasant. It implies fluctuations in the number of infectives which makes it difficult to allocate resources for treatment. It is also possible for oscillations to have a long period. This means that if data are measured over only a small time interval the actual behavior may not be displayed. Thus, the identification of situations in which an endemic equilibrium is unstable is an important problem.

Fig. 3.3 Region of asymptotic stability for endemic equilibria ($\omega = 1$, $N = 1$)



3.4 A Simple Model with Multiple Endemic Equilibria

In compartmental models for the transmission of communicable diseases there is usually a basic reproduction number \mathcal{R}_0 , representing the mean number of secondary infections caused by a single infective introduced into a susceptible population. If $\mathcal{R}_0 < 1$ there is a disease-free equilibrium which is asymptotically stable, and the infection dies out. If $\mathcal{R}_0 > 1$ the usual situation is that there is a unique endemic equilibrium which is asymptotically stable, and the infection persists. Even if the endemic equilibrium is unstable, the instability commonly arises from a Hopf bifurcation [25], described in Sect. 3.3, and the infection still persists but in an oscillatory manner. More precisely, as \mathcal{R}_0 increases through 1 there is an exchange of stability between the disease-free equilibrium and the endemic equilibrium (which is negative as well as unstable and thus biologically meaningless if $\mathcal{R}_0 < 1$).

There are, however, situations in which there may be more than one endemic equilibrium even in very simple epidemic models, and we describe such a model suggested in [42, 43]. We consider an *SIS* model in a population of constant total size N with treatment of infectives, assuming that the treatment cures the infection but that there is a maximum capacity for treatment. Thus we assume a model

$$\begin{aligned} S' &= -\beta SI + h(I) \\ I' &= \beta SI - \alpha I - h(I), \end{aligned} \tag{3.16}$$

assuming a treatment function $h(I)$ of the form

$$h(I) = \begin{cases} rI, & (I < I^*) \\ rI^*, & (I \geq I^*), \end{cases}$$

in which r is a constant representing the treatment rate up to a maximum capacity rI^* . Since the total population size $S + I$ is a constant N , we may replace S by $N - I$ and reduce the model to a single equation

$$I' = \beta I(N - I) - \alpha I - h(I) = g(I). \tag{3.17}$$

There is a disease-free equilibrium $I = 0$, and it is easily verified that the disease-free equilibrium is asymptotically stable if and only if $\mathcal{R}_0 = \beta N / (\alpha + r) < 1$.

For $I \leq I^*$,

$$g(I) = \beta I(N - I) - (\alpha + r)I,$$

and an endemic equilibrium with $I \leq I^*$ is a positive solution I_∞ of $g(I) = 0$, namely

$$I_\infty = N - \frac{\alpha + r}{\beta} = N \left(1 - \frac{1}{\mathcal{R}_0} \right),$$

and there is such an equilibrium if and only if

$$I^* \geq N - \frac{\alpha + r}{\beta} = N \left(1 - \frac{1}{\mathcal{R}_0} \right). \tag{3.18}$$

For $I \leq I^*$,

$$g'(I) = \beta(N - 2\beta I - (\alpha + r)),$$

and $g'(I_\infty) < 0$ if and only if

$$N - \frac{\alpha + r}{\beta} < 2I_\infty = 2 \left(N - \frac{\alpha + r}{\beta} \right),$$

and this is equivalent to $\mathcal{R}_0 > 1$. Thus, the equilibrium $I_\infty \leq I^*$ exists and is asymptotically stable if and only if (3.18) is satisfied.

Equilibria $I > I^*$ are solutions of the quadratic equation

$$g(I) = -\beta I^2 + (\beta N - \alpha)I - rI^* = 0,$$

which are

$$I = \frac{(\beta N - \alpha) + \sqrt{(\beta N - \alpha)^2 - 4r\beta I^*}}{2\beta}, \quad J = \frac{(\beta N - \alpha) - \sqrt{(\beta N - \alpha)^2 - 4r\beta I^*}}{2\beta}.$$

Then $J < (\beta N - \alpha)/2\beta$ and $I > (\beta N - \alpha)/2\beta$. For these to qualify as equilibria, they must also be greater than I^* and less than N , but it is possible to choose parameter values such that the model (3.16) has more than one endemic equilibrium. For example, the choices

$$\alpha = 0.5, \quad r = 0.5, \quad N = 1, \quad I^* = 0.05,$$

so that $\mathcal{R}_0 = \beta$, give two equilibria I, J for some values of β , including some values with $\mathcal{R}_0 < 1$. With these parameter values, $I = J = 0.279$ when $\beta = 0.779$.

An equilibrium I_∞ of the differential equation $I' = g(I)$ is asymptotically stable if $g'(I_\infty) < 0$, and unstable if $g'(I_\infty) > 0$. From this, it is easy to deduce that the equilibrium J is unstable, while the equilibrium I is asymptotically stable. If we plot the equilibrium values as functions of β , the curve I begins at the point $(0.779, 0.279)$ and goes upwards to the right, while the curve J goes downward to the right from the same starting point. Because of the choice $I^* = 0.05$, only the portion of the J curve above the line $I = 0.05$ is relevant. For $0.779 \leq \mathcal{R}_0 \leq 1$ there are two asymptotically stable equilibria, namely 0 and I separated by an unstable equilibrium J . For this reason, we have drawn the J curve as a dotted curve in Fig. 3.4.

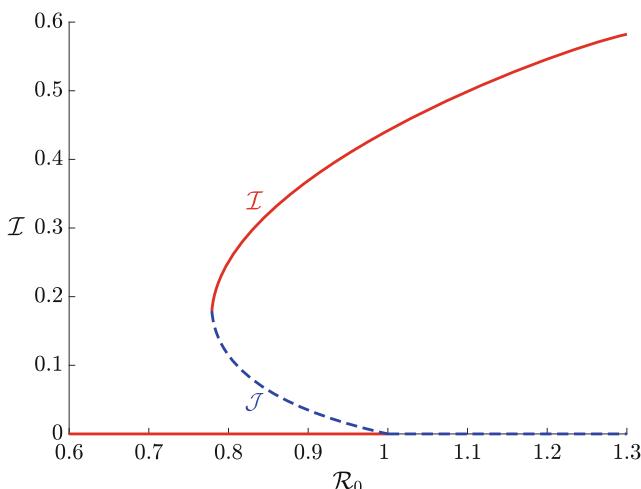


Fig. 3.4 Multiple endemic equilibria

A *bifurcation curve*, a graph of equilibria as a function of the basic reproduction number, as in Fig. 3.4, gives a good deal of information about the behavior of endemic equilibria. We observe, for example, that in Fig. 3.4, there are endemic equilibria for some values of the basic reproduction number less than 1, and that there is a discontinuity in the endemic equilibria at $\mathcal{R}_0 = 1$.

3.5 A Vaccination Model: Backward Bifurcations

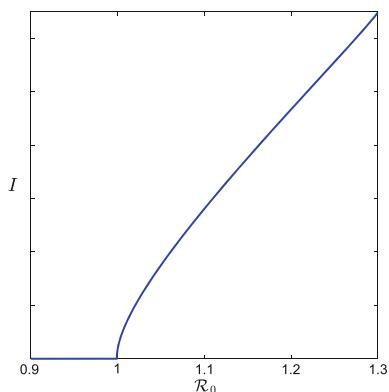
In a compartmental model, there is a bifurcation, or change in equilibrium behavior, at $\mathcal{R}_0 = 1$ but the equilibrium infective population size depends continuously on \mathcal{R}_0 . Such a transition is called a forward, or transcritical, bifurcation.

The behavior at a bifurcation may be described graphically by the bifurcation curve, which is the graph of equilibrium infective population size I as a function of the basic reproduction number \mathcal{R}_0 . For a forward bifurcation, the bifurcation curve is as shown in Fig. 3.5.

It has been noted [14, 20, 21, 29] that in epidemic models with multiple groups and asymmetry between groups or multiple interaction mechanisms it is possible to have a very different bifurcation behavior at $\mathcal{R}_0 = 1$. There may be multiple positive endemic equilibria for values of $\mathcal{R}_0 < 1$ and a backward bifurcation at $\mathcal{R}_0 = 1$. This means that the bifurcation curve has the form shown in Fig. 3.4 with a broken curve denoting an unstable endemic equilibrium that separates the domains of attraction of asymptotically stable equilibria.

The qualitative behavior of an epidemic system with a backward bifurcation differs from that of a system with a forward bifurcation in at least three important ways. If there is a forward bifurcation at $\mathcal{R}_0 = 1$ it is not possible for a disease to invade a population if $\mathcal{R}_0 < 1$ because the system will return to the disease-free equilibrium $I = 0$ if some infectives are introduced into the population. On the other hand, if there is a backward bifurcation at $\mathcal{R}_0 = 1$ and enough infectives are

Fig. 3.5 Forward bifurcation



introduced into the population to put the initial state of the system above the unstable endemic equilibrium with $\mathcal{R}_0 < 1$, the system will approach the asymptotically stable endemic equilibrium.

Other differences are observed if the parameters of the system change to produce a change in \mathcal{R}_0 . With a forward bifurcation at $\mathcal{R}_0 = 1$ the equilibrium infective population remains zero so long as $\mathcal{R}_0 < 1$ and then increases continuously as \mathcal{R}_0 increases. With a backward bifurcation at $\mathcal{R}_0 = 1$, there is an asymptotically stable disease-free equilibrium so long as $\mathcal{R}_0 < 1$ but there is also an asymptotically stable endemic equilibrium for some values of $\mathcal{R}_0 < 1$ and as \mathcal{R}_0 increases through 1 the infective population size jumps to the positive endemic equilibrium. In the other direction, if a disease is being controlled by means that decrease \mathcal{R}_0 it is sufficient to decrease \mathcal{R}_0 to 1 if there is a forward bifurcation at $\mathcal{R}_0 = 1$ but it is necessary to bring \mathcal{R}_0 well below 1 if there is a backward bifurcation.

These behavior differences are important in planning how to control a disease; a backward bifurcation at $\mathcal{R}_0 = 1$ makes control more difficult. One control measure often used is the reduction of susceptibility to infection produced by vaccination. By vaccination, we mean either an inoculation that reduces susceptibility to infection or an education program such as encouragement of better hygiene or avoidance of risky behavior for sexually transmitted diseases. Whether vaccination is inoculation or education, typically it reaches only a fraction of the susceptible population and is not perfectly effective. In an apparent paradox, models with vaccination may exhibit backward bifurcations, making the behavior of the model more complicated than the corresponding model without vaccination. It has been argued [6] that a partially effective vaccination program applied to only part of the population at risk may increase the severity of outbreaks of such diseases as HIV/AIDS.

We will give a qualitative analysis of a model which may have a variable total population size $N \leq K$ for which there is a possibility of a backward bifurcation. The model we will study adds vaccination to the simple SIS model with births and natural deaths but with no disease deaths studied in Sect. 2.2. We have considered the model

$$\begin{aligned} S' &= \Lambda(N) - \beta(N)SI - \mu S + \alpha I \\ I' &= \beta SI - (\mu + \alpha)I, \end{aligned} \tag{3.19}$$

where the population carrying capacity K is defined by $\Lambda(K) = \mu K$, $\Lambda'(K) < \mu$ and the contact rate $\beta(N)$ is a function of total population size with $N\beta(N)$ non-decreasing and $\beta(N)$ non-increasing. We have seen that there is a disease-free equilibrium $I = 0$ that is asymptotically stable if

$$\mathcal{R}_0 = \frac{K\beta(K)}{\mu + \alpha} < 1.$$

If $\mathcal{R}_0 > 1$ the disease-free equilibrium is unstable but there is an endemic equilibrium that is asymptotically stable.

To the model (3.19) we add the assumption that in unit time a fraction φ of the susceptible class is vaccinated. The vaccination may reduce but not completely eliminate susceptibility to infection. We model this by including a factor σ , $0 \leq \sigma \leq 1$, in the infection rate of vaccinated members with $\sigma = 0$ meaning that the vaccine is perfectly effective and $\sigma = 1$ meaning that the vaccine has no effect. We describe the new model by including a vaccinated class V , with

$$\begin{aligned} S' &= \mu N - \beta(N)SI - (\mu + \varphi)S + \alpha I \\ I' &= \beta(N)SI + \sigma\beta(N)VI - (\mu + \alpha)I \\ V' &= \varphi S - \sigma\beta(N)VI - \mu V \end{aligned} \quad (3.20)$$

and $N = S + I + V$. Since N is constant, we can replace S by $N - I - V$ to give the equivalent system

$$\begin{aligned} I' &= \beta[N - I - (1 - \sigma)V]I - (\mu + \alpha)I \\ V' &= \varphi[N - I] - \sigma\beta VI - (\mu + \varphi)V \end{aligned} \quad (3.21)$$

with $\beta = \beta(N)$. The system (3.21) is the basic vaccination model which we will analyze. We remark that if the vaccine is completely ineffective, $\sigma = 1$, then (3.21) is equivalent to an *SIS* model. If all susceptibles are vaccinated immediately (formally, $\varphi \rightarrow \infty$), the model (3.21) is equivalent to

$$I' = \sigma\beta I(K - I) - (\mu + \alpha)I$$

which is an *SIS* model with basic reproduction number

$$\mathcal{R}_0^* = \frac{\sigma\beta K}{\mu + \alpha} = \sigma\mathcal{R}_0 \leq \mathcal{R}_0.$$

We will think of the parameters μ , α , φ , and σ as fixed and will view β as variable. In practice, the parameter φ is the one most easily controlled, and later we will express our results in terms of an uncontrolled model with parameters β , μ , α , and σ fixed and examine the effect of varying φ . With this interpretation in mind, we will use $\mathcal{R}(\varphi)$ to denote the basic reproduction number of the model (3.21), and we will see that

$$\mathcal{R}_0^* \leq \mathcal{R}(\varphi) \leq \mathcal{R}_0.$$

Equilibria of the model (3.21) are solutions of

$$\begin{aligned} \beta I [K - I - (1 - \sigma)V] &= (\mu + \alpha)I \\ \varphi[K - I] &= \sigma\beta VI + (\mu + \varphi)V. \end{aligned} \quad (3.22)$$

If $I = 0$, then the first of these equations is satisfied and the second leads to

$$V = \frac{\varphi}{\mu + \varphi} K.$$

This is the disease-free equilibrium.

The matrix of the linearization of (3.21) at an equilibrium (I, V) is

$$\begin{bmatrix} -2\beta I - (1 - \sigma)\beta V - (\mu + \alpha) + \beta K & -(1 - \sigma)\beta I \\ -(\varphi + \sigma\beta V) & -(\mu + \varphi + \sigma\beta I) \end{bmatrix}.$$

At the disease-free equilibrium this matrix is

$$\begin{bmatrix} -(1 - \sigma)\beta V - (\mu + \alpha) + \beta K & 0 \\ -(\varphi + \sigma\beta V) & -(\mu + \varphi) \end{bmatrix}$$

which has negative eigenvalues, implying the asymptotic stability of the disease-free equilibrium, if and only if

$$-(1 - \sigma)\beta V - (\mu + \alpha) + \beta K < 0.$$

Using the value of V at the disease-free equilibrium this condition is equivalent to

$$\mathcal{R}(\varphi) = \frac{\beta K}{\mu + \alpha} \cdot \frac{\mu + \sigma\varphi}{\mu + \varphi} = \mathcal{R}_0 \frac{\mu + \sigma\varphi}{\mu + \varphi} < 1.$$

The case $\varphi = 0$ is that of no vaccination with $\mathcal{R}(0) = \mathcal{R}_0$, and $\mathcal{R}(\varphi) < \mathcal{R}_0$ if $\varphi > 0$. We note that $\mathcal{R}_0^* = \sigma\mathcal{R}_0 = \lim_{\varphi \rightarrow \infty} \mathcal{R}(\varphi) < \mathcal{R}_0$.

If $0 \leq \sigma < 1$ endemic equilibria are solutions of the pair of equations

$$\begin{aligned} \beta [K - I - (1 - \sigma)V] &= \mu + \alpha \\ \varphi[K - I] &= \sigma\beta VI + (\mu + \varphi)V. \end{aligned} \tag{3.23}$$

We eliminate V using the first equation of (3.23) and substitute into the second equation to give an equation of the form

$$AI^2 + BI + C = 0 \tag{3.24}$$

with

$$\begin{aligned} A &= \sigma\beta \\ B &= (\mu + \theta + \sigma\varphi) + \sigma(\mu + \alpha) - \sigma\beta K \\ C &= \frac{(\mu + \alpha)(\mu + \theta + \varphi)}{\beta} - (\mu + \theta + \sigma\varphi)K. \end{aligned} \tag{3.25}$$

If $\sigma = 0$ (3.24) is a linear equation with unique solution.

$$I = K - \frac{(\mu + \alpha)(\mu + \varphi)}{\beta\mu} = K \left[1 - \frac{1}{\mathcal{R}(\varphi)} \right]$$

which is positive if and only if $\mathcal{R}(\varphi) > 1$. Thus if $\sigma = 0$ there is a unique endemic equilibrium if $\mathcal{R}(\varphi) > 1$ that approaches zero as $\mathcal{R}(\varphi) \rightarrow 1+$ and there cannot be an endemic equilibrium if $\mathcal{R}(\varphi) < 1$. In this case, it is not possible to have a backward bifurcation at $\mathcal{R}(\varphi) = 1$.

We note that $C < 0$ if $\mathcal{R}(\varphi) > 1$, $C = 0$ if $\mathcal{R}(\varphi) = 1$, and $C > 0$ if $\mathcal{R}(\varphi) < 1$. If $\sigma > 0$, so that (3.24) is quadratic and if $\mathcal{R}(\varphi) > 1$ then there is a unique positive root of (3.24) and thus there is a unique endemic equilibrium. If $\mathcal{R}(\varphi) = 1$, then $C = 0$ and there is a unique non-zero solution of (3.24) $I = -B/A$ which is positive if and only if $B < 0$. If $B < 0$ when $C = 0$ there is a positive endemic equilibrium for $\mathcal{R}(\varphi) = 1$. Since equilibria depend continuously on φ there must then be an interval to the left of $\mathcal{R}(\varphi) = 1$ on which there are two positive equilibria

$$I = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}.$$

This establishes that the system (3.21) has a backward bifurcation at $\mathcal{R}(\varphi) = 1$ if and only if $B < 0$ when β is chosen to make $C = 0$.

We can give an explicit criterion in terms of the parameters μ , φ , σ for the existence of a backward bifurcation at $\mathcal{R}(\varphi) = 1$. When $\mathcal{R}(\varphi) = 1$, $C = 0$ so that

$$(\mu + \sigma\varphi)\beta K = (\mu + \alpha)(\mu + \varphi). \quad (3.26)$$

The condition $B < 0$ is

$$(\mu + \sigma\varphi) + \sigma(\mu + \alpha) < \sigma\beta K$$

with βK determined by (3.26), or

$$\sigma(\mu + \alpha)(\mu + \varphi) > (\mu + \sigma\varphi)[(\mu + \sigma\varphi) + \sigma(\mu + \alpha)]$$

which reduces to

$$\sigma(1 - \sigma)(\mu + \alpha)\varphi > (\mu + \sigma\varphi)^2. \quad (3.27)$$

A backward bifurcation occurs at $\mathcal{R}(\varphi) = 1$, with βK given by (3.26) if and only if (3.27) is satisfied. We point out that for an *SI* model, where $\alpha = 0$, the condition (3.27) becomes

$$\sigma(1 - \sigma)\mu\varphi > (\mu + \sigma\varphi)^2.$$

But

$$\begin{aligned} (\mu + \sigma\varphi)^2 &= \mu^2 + \sigma^2\varphi^2 + 2\mu\sigma\varphi \\ &> 2\mu\sigma\varphi > \sigma(1 - \sigma)\mu\varphi \end{aligned}$$

because $\sigma < 1$. Thus a backward bifurcation is not possible if $\alpha = 0$, that is, for an *SI* model. Likewise, (3.27) cannot be satisfied if $\sigma = 0$.

If $C > 0$ and either $B \geq 0$ or $B^2 < 4AC$, there are no positive solutions of (3.24) and thus there are no endemic equilibria. Equation (3.24) has two positive solutions, corresponding to two endemic equilibria, if and only if $C > 0$, or $\mathcal{R}(\varphi) < 1$, and $B < 0$, $B^2 > 4AC$, or $B < -2\sqrt{AC} < 0$. If $B = -2\sqrt{AC}$, there is one positive solution $I = -B/2A$ of (3.24).

If (3.27) is satisfied, so that there is a backward bifurcation at $\mathcal{R}(\varphi) = 1$, there are two endemic equilibria for an interval of values of β from

$$\beta K = \frac{(\mu + \alpha)(\mu + \varphi)}{\mu + \sigma\varphi}$$

corresponding to $\mathcal{R}(\varphi) = 1$ to a value β_c defined by $B = -2\sqrt{AC}$. To calculate β_c , we let $x = \mu + \alpha - \beta K$, $U = \mu + \sigma\varphi$ to give $B = \sigma x + U$, $\beta C = \beta K U + (\mu + \alpha)(\mu + \varphi)$. Then $B^2 = 4AC$ becomes

$$(\sigma x + U)^2 + 4\beta\sigma KU - 4\sigma(\mu + \alpha)(\mu + \varphi) = 0$$

which reduces to

$$(\sigma x)^2 - 2U(\sigma x) + \left[U^2 + 4\sigma(1 - \sigma)(\mu + \alpha)\varphi \right] = 0$$

with roots

$$\sigma x = U \pm 2\sqrt{\sigma(1 - \sigma)(\mu + \alpha)\varphi}.$$

For the positive root $B = \sigma x + U > 0$, and since we require $B < 0$ as well as $B^2 - 4AC = 0$, we obtain β_c from $\sigma x = U - 2\sqrt{\sigma(1 - \sigma)(\mu + \alpha)\varphi}$ so that

$$\sigma\beta_c K = \sigma(\mu + \alpha) + 2\sqrt{\sigma(1 - \sigma)(\mu + \alpha)\varphi} - (\mu + \sigma\varphi). \quad (3.28)$$

Then the critical basic reproduction number \mathcal{R}_c is given by

$$\mathcal{R}_c = \frac{\mu + \sigma\varphi}{\mu + \varphi} \cdot \frac{\sigma(\mu + \alpha) + 2\sqrt{\sigma(1 - \sigma)(\mu + \alpha)\varphi} - (\mu + \sigma\varphi)}{\sigma(\mu + \alpha)\varphi}$$

and it is possible to verify with the aid of (3.28) that $\mathcal{R}_c < 1$.

3.5.1 The Bifurcation Curve

In drawing the bifurcation curve (the graph of I as a function of $\mathcal{R}(\varphi)$), we think of β as variable with the other parameters $\mu, \alpha, \sigma, Q, \varphi$ as constant. Then $\mathcal{R}(\varphi)$ is a constant multiple of β and we can think of β as the independent variable in the bifurcation curve.

Implicit differentiation of the equilibrium condition (3.24) with respect to β gives

$$(2AI + B) \frac{dI}{d\beta} = \sigma I(K - I) + \frac{(\mu + \alpha)(\mu + \varphi)}{\beta^2}.$$

It is clear from the first equilibrium condition in (3.23) that $I \leq K$ and this implies that the bifurcation curve has positive slope at equilibrium values with $2AI + B > 0$ and negative slope at equilibrium values with $2AI + B < 0$. If there is not a backward bifurcation at $\mathcal{R}(\varphi) = 1$, then the unique endemic equilibrium for $\mathcal{R}(\varphi) > 1$ satisfies

$$2AI + B = \sqrt{B^2 - 4AC} > 0$$

and the bifurcation curve has positive slope at all points where $I > 0$. Thus the bifurcation curve is as shown in Fig. 3.5.

If there is a backward bifurcation at $\mathcal{R}(\varphi) = 1$, then there is an interval on which there are two endemic equilibria given by

$$2AI + B = \pm \sqrt{B^2 - 4AC}.$$

The bifurcation curve has negative slope at the smaller of these and positive slope at the larger of these. Thus the bifurcation curve is as shown in Fig. 3.4.

The condition $2AI + B > 0$ is also significant in the local stability analysis of endemic equilibria. An endemic equilibrium of (3.21) is (locally) asymptotically stable if and only if it corresponds to a point on the bifurcation curve at which the curve is increasing. To prove this, we observe that the matrix of the linearization of (3.21) at an equilibrium (I, V) is

$$\begin{bmatrix} -2\beta I - (1 - \sigma)\beta V - (\mu + \alpha) + \beta K & -(1 - \sigma)\beta I \\ -(\varphi + \sigma\beta V) & -(\mu + \varphi + \sigma\beta I) \end{bmatrix}.$$

Because of the equilibrium conditions (3.23), the matrix at an endemic equilibrium (I, V) is

$$\begin{bmatrix} -\beta I & -(1 - \sigma)\beta I \\ -(\varphi + \sigma\beta V) & -(\mu + \varphi + \sigma\beta I) \end{bmatrix}.$$

This has negative trace, and its determinant is

$$\begin{aligned} & \sigma(\beta I)^2 + \beta I(\mu + \varphi) - (1 - \sigma)\varphi\beta I - (1 - \sigma)\beta V \cdot \sigma\beta I \\ &= \beta I[2\sigma\beta I + (\mu + \sigma\varphi) + \sigma(\mu + \alpha) - \sigma\beta K] \\ &= \beta I[2AI + B]. \end{aligned}$$

If $2AI + B > 0$, that is, if the bifurcation curve has positive slope, then the determinant is positive and the equilibrium is asymptotically stable. If $2AI + B < 0$ the determinant is negative and the equilibrium is unstable. In fact, it is a saddle point. A saddle point in the plane is an equilibrium at which the linearization has one positive eigenvalue and one negative eigenvalue. This means that there are two orbits approaching the saddle point called stable separatrices and two orbits going out from the saddle point, called unstable separatrices. Because orbits cannot cross the separatrices, the stable separatrices divide the plane into two regions and divide the plane into two domains of attraction. The stable separatrices in the (I, V) plane separate the domains of attraction of the other (asymptotically stable) endemic equilibrium and the disease-free equilibrium.

3.6 *An SEIR Model with General Disease Stage Distributions

The ODE models considered in earlier parts of the chapter, except the SIRS model in Sect. 3.3, assume exclusively that the durations of disease stages (e.g., latent and infectious stages) are exponentially distributed. This assumption, while making the models and their analyses easier, is not biologically realistic for most infectious diseases. A more appropriate distribution is the gamma distribution, for which the probability of remaining in the stage is given by

$$p_n(s) = \sum_{k=0}^{n-1} \frac{(n\theta s)^k s^{-n\theta s}}{k!} \quad (3.29)$$

where $1/\theta$ is the mean of the distribution and n is the shape parameter. The exponential distribution is the special case when $n = 1$. The other extreme case is when $n \rightarrow \infty$, which corresponds to a fixed duration. Figure 3.6 illustrates the gamma distribution for various n values. Let $1/\kappa$ and $1/\alpha$ denote the mean latent and infectious periods, and let m and n denote the shape parameter for the latent and infectious stages, respectively. It has been reported [44] that for measles

$$1/\kappa = 8, \quad 1/\alpha = 5, \quad m = n = 20,$$

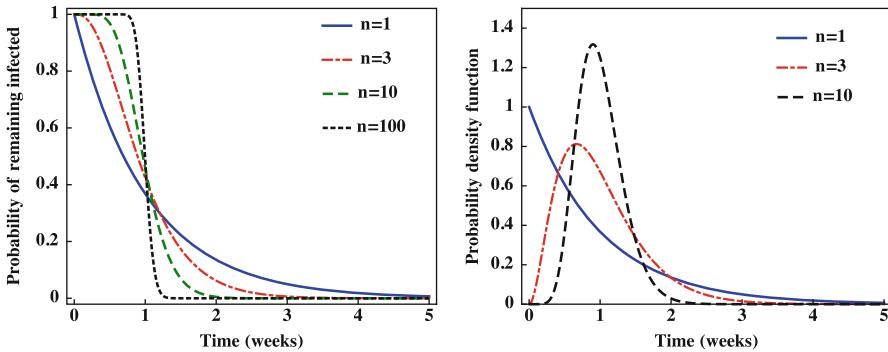


Fig. 3.6 Depictions of the survival probability (3.29) for the infectious period (left) and the probability density function (right) with different shape parameter values (n). The special case of $n = 1$ gives the exponential distribution. The mean infectious period ($1/\theta$) is chosen to be 1 week

and for smallpox

$$1/\kappa = 14, \quad 1/\alpha = 8.6, \quad m = 40, \quad n = 4.$$

There are other cases where the disease stage durations do not fit well by the standard family distributions. Epidemiological models with non-exponential distributions such as the gamma distribution have been previously studied (see, for example, [23, 30, 31]). In these studies, the authors discussed various drawbacks associated with the exponential distribution assumption. For example, it is pointed out that constant recovery is a poor description of real-world infections, and they show that in models with more realistic distributions of disease stages less stable behavior may be expected and disease persistence may be diminished [23, 30]. In [18] it was demonstrated that when control measures such as quarantine and isolation are considered, models with exponential and gamma distributions may generate contradictory evaluations on control strategies. Thus, it will be helpful to have mathematical results for models that allow arbitrary distributions. This is the goal of this section.

Let $P_E, P_I : [0, \infty) \rightarrow [0, 1]$ describe the durations of the exposed (latent) and infective stages, respectively. That is, $P_i(s)$ ($i = E, I$) gives the probability that the disease stage i lasts longer than s time units (or the probability of being still in the same stage at stage age s). Then, the derivative $-\dot{P}_i(s)$ ($i = E, I$) gives the rate of removal from the stage i at stage age s by the natural progression of the disease. These duration functions have the following properties:

$$P_i(0) = 1, \quad \dot{P}_i(s) \leq 0, \quad \int_0^\infty P_i(s) ds < \infty, \quad i = E, I.$$

For the vital dynamics, we use the simplest function $e^{-\mu t}$ for the probability of survival (because our focus is on the effect of arbitrary distribution for disease stages). Let the numbers of initial susceptible and removed individuals be $S_0 > 0$ and $R_0 > 0$ respectively. Let $E_0(t)e^{-\mu t}$ and $I_0(t)e^{-\mu t}$ be the non-increasing functions that represent the numbers of individuals that were initially exposed and infective, respectively, and are still alive and in the respective classes at time t . $E(0)$ and $I(0)$ are constants representing the number of individuals in the E and I classes, respectively, at time $t = 0$. Let $\tilde{I}_0(t)$ denote those initially infected who have moved into the I , and are still alive at time t . Consider the force of infection $\lambda(t)$ that takes the form

$$\lambda(t) = c \frac{I(t)}{N}. \quad (3.30)$$

Then the number of individuals who became exposed at some time $s \in (0, t)$ and are still alive and in the E class at time t is

$$E(t) = \int_0^t \lambda(s)S(s)P_E(t-s)e^{-\mu(t-s)}ds + E_0(t)e^{-\mu t},$$

and the number of infectious individuals at time t is

$$I(t) = \int_0^t \int_0^\tau \lambda(s)S(s)[-P_E(\tau-s)]P_I(t-\tau)e^{-\mu(t-s)}dsd\tau + I_0(t)e^{-\mu t} + \tilde{I}_0(t).$$

Assume that the recruitment rate is μN and they all enter the S class. Then the $SEIR$ model reads

$$\begin{aligned} S(t) &= \int_0^t \mu Ne^{-\mu(t-s)}ds - \int_0^t \lambda(s)S(s)e^{-\mu(t-s)}ds + S_0e^{-\mu t}, \\ E(t) &= \int_0^t \lambda(s)S(s)P_E(t-s)e^{-\mu(t-s)}ds + E_0(t)e^{-\mu t}, \\ I(t) &= \int_0^t \int_0^\tau \lambda(s)S(s)[-P_E(\tau-s)]P_I(t-\tau)e^{-\mu(t-s)}dsd\tau + \tilde{I}(t), \end{aligned} \quad (3.31)$$

where $\lambda(t)$ is given in (3.30), and $\tilde{X}(t) = X_0(t)e^{-\mu t} + \tilde{X}_0(t)$ ($X = Q, I, H, R$). It can be shown that under standard assumptions on initial data and parameter functions the system (3.31) has a unique non-negative solution defined for all positive time.

When specific distributions are assumed for functions P_E and P_I the system (3.31) might be simplified. In particular, under the exponential distribution assumption (EDA) and gamma distribution assumption (GDA) the system can be reduced to be ODE systems, which will be referred to as the exponential distribution model (EDM) and gamma distribution model (GDM), respectively. This allows the examination of how the distribution assumptions may affect the model predictions.

Let

$$a(\tau) = e^{-\mu\tau} \int_0^\tau [-\dot{P}_E(\tau - u)] P_I(u) du. \quad (3.32)$$

The reproduction number is given by

$$\mathcal{R}_c = \int_0^\infty c a(\tau) d\tau. \quad (3.33)$$

To see the biological meaning of the expression (3.33) and to simplify the notation in later sections we introduce the following quantities:

$$\begin{aligned} \mathcal{T}_E &= \int_0^\infty [-\dot{P}_E(s)] e^{-\mu s} ds, \quad \mathcal{T}_I = \int_0^\infty [-\dot{P}_I(s)] e^{-\mu s} ds, \\ \mathcal{D}_E &= \int_0^\infty P_E(s) e^{-\mu s} ds, \quad \mathcal{D}_I = \int_0^\infty P_I(s) e^{-\mu s} ds. \end{aligned} \quad (3.34)$$

\mathcal{T}_E , \mathcal{T}_I represent, respectively, the probability that exposed individuals survive and become infectious, and the probability that infectious individuals survive and become recovered. \mathcal{D}_E represents the mean sojourn time (death-adjusted) in the exposed stage, and \mathcal{D}_I represents the mean sojourn time (death-adjusted) in the infectious stage. Using (3.34) we can rewrite \mathcal{R}_c in (3.33) as

$$\mathcal{R}_c = c \int_0^\infty a(\tau) d\tau = c \mathcal{T}_{E_k} \mathcal{D}_{I_l}. \quad (3.35)$$

System (3.31) always has the disease-free equilibrium (DFE), and an endemic equilibrium may exist depending on the value of \mathcal{R}_0 as described below. The proof of the result can be found in [18].

Result For System (3.31), the DFE is a global attractor if $\mathcal{R}_c < 1$ and unstable if $\mathcal{R}_c > 1$, in which case an endemic equilibrium exists and is stable.

To compare the model behavior under different stage distributions, let P_E and P_I be the gamma distributions with the duration functions $P_E(s) = p_m(s, \kappa)$ and $P_I(s) = p_n(s, \alpha)$, with mean $1/\kappa$ and $1/\alpha$, respectively. When $m = n = 1$, where m and n are the shape parameters, P_E and P_I are exponential distributions the general model (3.31) has the usual form of the standard SEIR model:

$$\begin{aligned} S' &= \mu N - cS \frac{I}{N} - \mu S, \\ E' &= cS \frac{I}{N} - (\kappa + \mu)E, \\ I' &= \kappa E - (\alpha + \mu)I. \end{aligned} \quad (3.36)$$

For other integers m and n , then the integral equation model (3.31) can be reduced to an ordinary differential equation model. It has been noted that the use

of the gamma distribution $p_n(s, \theta)$ for a disease stage, e.g., the exposed stage, is equivalent to assuming that the entire stage is replaced by a series of n sub-stages, and each of the sub-stages is exponentially distributed with the removal rate $n\theta$ and the mean sojourn time T/n , where $T = 1/\theta$ is the mean sojourn time of the entire stage (see, for example, [23, 30, 32]). This approach of converting a gamma distribution to a sequence of exponential distributions is known as the “linear chain trick”. In this case, the general model (3.31) reduces to the following ODEs:

$$\begin{aligned} S' &= \mu N - cS \frac{I}{N} - \mu S, \\ E'_1 &= cS \frac{I}{N} - (m\kappa + \mu)E_1, \\ E'_j &= m\kappa E_{j-1} - (m\kappa + \mu)E_j, \quad j = 2, \dots, m, \\ I'_1 &= m\kappa E_m - (n\alpha + \mu)I_1, \\ I'_j &= n\alpha I_{j-1} - (n\alpha + \mu)I_j, \quad j = 2, \dots, n, \\ \text{with } I &= \sum_{j=1}^n I_j. \end{aligned} \quad (3.37)$$

From the formula (3.35) we get the reproduction number for system (3.37):

$$\mathcal{R}_c = \frac{(m\kappa)^m}{(\mu + m\kappa)^m} \frac{c}{\mu + n\alpha} \sum_{j=0}^{n-1} \frac{(n\alpha)^j}{(\mu + n\alpha)^j}. \quad (3.38)$$

The qualitative behavior of the two systems (3.36) and (3.37) is the same due to the results stated above. For the quantitative behavior, some differences exist. For example, Fig. 3.7 illustrates that the model with gamma distributions tend to generate a lower frequency in oscillations than the model with exponential

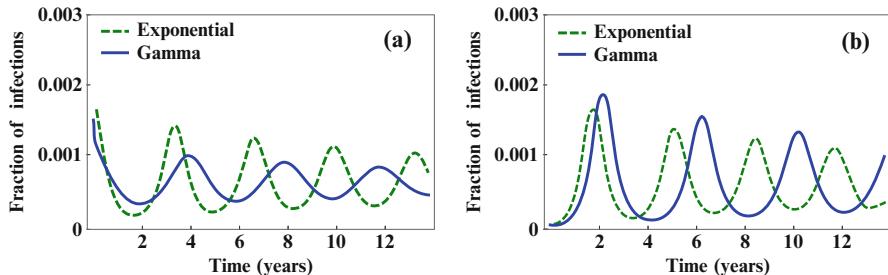


Fig. 3.7 Comparison of simulations of the SEIR model with exponential and gamma distributions (3.36) and (3.37). Plot in (a) is for the case when the initial fraction of infectious individuals is higher (0.0015) while in (b) it is lower (0.0001). We observe in (a) that the solution of the model with gamma distribution has much lower magnitude in the oscillation than the solution of the model with exponential distribution, whereas in (b) it is opposite. However, in both (a) and (b) the frequency of the oscillation is lower for the gamma distribution model than for the exponential distribution model. The parameter values used are $1/\kappa = 2$ (day), $1/\alpha = 7$ (day), $1/\mu = 15$ (year). This is more appropriate for modeling school entry and exit. The values of c is chosen so that $\mathcal{R}_0 = 2$

distributions. As for the magnitude of oscillations, either model can have a higher magnitude than the other depending on the initial conditions.

3.6.1 *Incorporation of Quarantine and Isolation

As pointed out earlier, the model with gamma distributions for disease stages (3.37) has similar qualitative behavior as the model with exponential distributions (3.36). As will be shown in this section, when control measures with quarantine and isolation are considered, models with exponential and gamma distributions can generate very different quantitative outcomes, including contradictory evaluations regarding control strategies.

Let ρ denote the isolation efficiency with $\rho = 1$ representing complete effectiveness. Thus, when $\rho < 1$ the isolated individuals can transmit the infection with a reduced infectivity $1 - \rho$. The force of infection is

$$\lambda(t) = c \frac{I(t) + (1 - \rho)H(t)}{N}. \quad (3.39)$$

Let $k(s), l(s): [0, \infty) \rightarrow [0, 1]$ denote, respectively, the probabilities that exposed, infective individuals have not been quarantined, isolated at stage age s . Hence, $1 - k(s) =: \bar{k}(s)$, $1 - l(s) =: \bar{l}(s)$ give the respective probabilities of being quarantined, isolated before reaching stage age s . Assume that $k(0) = l(0) = 1$, $\dot{k}(s) \leq 0$ and $\dot{l}(s) \leq 0$. Consider the simpler case when the survivals from quarantine and isolation are described by the exponential functions

$$k(s) = e^{-\chi s}, \quad l(s) = e^{-\phi s} \quad (3.40)$$

with χ and ϕ being constants, we have

$$E_0(t) = E(0)e^{-(\chi+\alpha)t}, \quad I_0(t) = I(0)e^{-(\phi+\delta)t}, \quad \text{etc.} \quad (3.41)$$

The model with general distributions as well as quarantine and isolation is

$$\begin{aligned} S(t) &= \int_0^t \mu N e^{-\mu(t-s)} ds - \int_0^t \lambda(s) S(s) e^{-\mu(t-s)} ds + S_0 e^{-\mu t}, \\ E(t) &= \int_0^t \lambda(s) S(s) P_E(t-s) k(t-s) e^{-\mu(t-s)} ds + E_0(t) e^{-\mu t}, \\ Q(t) &= \int_0^t \int_0^\tau \lambda(s) S(s) [-P_E(\tau-s) \dot{k}(\tau-s)] P_E(t-\tau | \tau-s) e^{-\mu(t-s)} ds d\tau + \tilde{Q}(t), \\ I(t) &= \int_0^t \int_0^\tau \lambda(s) S(s) [-\dot{P}_E(\tau-s) k(\tau-s)] P_I(t-\tau) l(t-\tau) e^{-\mu(t-s)} ds d\tau + \tilde{I}(t), \end{aligned}$$

$$\begin{aligned}
H(t) &= \int_0^t \int_0^u \int_0^\tau \lambda(s) S(s) [-\dot{P}_E(\tau-s)k(\tau-s)][-P_I(u-\tau)\dot{l}(u-\tau)] \\
&\quad \times P_I(t-u|u-\tau)e^{-\mu(t-s)}dsd\tau du \\
&+ \int_0^t \int_0^\tau \lambda(s) S(s) [-\dot{P}_E(\tau-s)\bar{k}(\tau-s)]P_I(t-\tau)e^{-\mu(t-s)}dsd\tau + \tilde{H}(t), \\
R(t) &= \int_0^t \int_0^\tau \lambda(s) S(s) [-\dot{P}_E(\tau-s)][1-P_I(t-\tau)]e^{-\mu(t-s)}dsd\tau + \tilde{R}(t),
\end{aligned} \tag{3.42}$$

where $\lambda(t)$ is given in (3.39) and $\tilde{X}(t) = X_0(t)e^{-\mu t} + \tilde{X}_0(t)$ ($X = Q, I, H, R$). Again $\tilde{X}(t) \rightarrow 0$ as $t \rightarrow \infty$.

Let

$$\begin{aligned}
a_1(\tau) &= e^{-\mu\tau} \int_0^\tau [-\dot{P}_E(\tau-u)k(\tau-u)]P_I(u)\bar{l}(u)du, \\
a_2(\tau) &= e^{-\mu\tau} \int_0^\tau [-\dot{P}_E(\tau-u)k(\tau-u)]P_I(u)\bar{l}(u)du, \\
a_3(\tau) &= e^{-\mu\tau} \int_0^\tau [-\dot{P}_E(\tau-u)\bar{k}(\tau-u)]P_I(u)du,
\end{aligned} \tag{3.43}$$

where $\bar{k}(s) = 1 - k(s)$, $\bar{l}(s) = 1 - l(s)$. Let

$$A(\tau) = a_1(\tau) + (1-\rho)[a_2(\tau) + a_3(\tau)].$$

Then, the reproduction number is

$$\mathcal{R}_c = c \int_0^\infty A(\tau)d\tau. \tag{3.44}$$

We can also rewrite \mathcal{R}_0 in the following form:

$$\mathcal{R}_c = \mathcal{R}_I + \mathcal{R}_{IH} + \mathcal{R}_{QH}, \tag{3.45}$$

where

$$\begin{aligned}
\mathcal{R}_I &= c \int_0^\infty a_1(\tau)d\tau = c \mathcal{T}_{E_k} \mathcal{D}_{I_l}, \\
\mathcal{R}_{IH} &= (1-\rho)c \int_0^\infty a_2(\tau)d\tau = (1-\rho)c \mathcal{T}_{E_k} (\mathcal{D}_I - \mathcal{D}_{I_l}), \\
\mathcal{R}_{QH} &= (1-\rho)c \int_0^\infty a_3(\tau)d\tau = (1-\rho)c (\mathcal{T}_E - \mathcal{T}_{E_k}) \mathcal{D}_I.
\end{aligned}$$

and

$$\begin{aligned}\mathcal{T}_E &= \int_0^\infty [-\dot{P}_E(s)]e^{-\mu s}ds, \quad \mathcal{T}_{E_k} = \int_0^\infty [-\dot{P}_E(s)k(s)]e^{-\mu s}ds, \\ \mathcal{T}_I &= \int_0^\infty [-\dot{P}_I(s)]e^{-\mu s}ds, \quad \mathcal{T}_{I_l} = \int_0^\infty [-\dot{P}_I(s)l(s)]e^{-\mu s}ds, \\ \mathcal{D}_E &= \int_0^\infty P_E(s)e^{-\mu s}ds, \quad \mathcal{D}_{E_k} = \int_0^\infty P_E(s)k(s)e^{-\mu s}ds, \\ \mathcal{D}_I &= \int_0^\infty P_I(s)e^{-\mu s}ds, \quad \mathcal{D}_{I_l} = \int_0^\infty P_I(s)l(s)e^{-\mu s}ds.\end{aligned}\tag{3.46}$$

The three components, $\mathcal{R}_I, \mathcal{R}_{IH}, \mathcal{R}_{QH}$ in \mathcal{R}_c represent contributions from the I class and from the H class through isolation and quarantine, respectively. \mathcal{T}_E and \mathcal{T}_{E_k} represent, respectively, the probability and the “quarantine-adjusted” probability that exposed individuals survive and become infectious. \mathcal{T}_I and \mathcal{T}_{I_l} represent, respectively, the probability and the “isolation-adjusted” probability that infectious individuals survive and become recovered. \mathcal{D}_E and \mathcal{D}_{E_k} represent, respectively, the mean sojourn time (death-adjusted) and the “quarantine-adjusted” mean sojourn time (death-adjusted as well) in the exposed stage. \mathcal{D}_I and \mathcal{D}_{I_l} represent, respectively, the mean sojourn time (death-adjusted) and the “isolation-adjusted” mean sojourn time (death-adjusted as well) in the infectious stage.

For system (3.42), the same results as for system (3.31) holds, i.e., the DFE is a global attractor if $\mathcal{R}_c < 1$ and unstable if $\mathcal{R}_c > 1$, in which case an endemic equilibrium exists and is stable.

3.6.2 *The Reduced Model of (3.42) Under GDA

Again let P_E and P_I be the gamma distributions with the duration functions $P_E(s) = p_m(s, \kappa)$ and $P_I(s) = p_n(s, \alpha)$, with mean $1/\kappa$ and $1/\alpha$, respectively. Then using the functions $k(s)$ and $l(s)$ given in (3.40) we can differentiate the equations in system (3.42) and obtain the following system of ordinary differential equations

$$\begin{aligned}S' &= \mu N - cS \frac{I + (1 - \rho)H}{N} - \mu S, \\ E'_1 &= cS \frac{I + (1 - \rho)H}{N} - (\chi + m\kappa + \mu)E_1, \\ E'_j &= m\kappa E_{j-1} - (\chi + m\kappa + \mu)E_j, \quad j = 2, \dots, m, \\ Q'_1 &= \chi E_1 - (m\kappa + \mu)Q_1, \\ Q'_j &= \chi E_j + m\kappa Q_{j-1} - (m\kappa + \mu)Q_j, \quad j = 2, \dots, m, \\ I'_1 &= m\kappa E_m - (\phi + n\alpha + \mu)I_1,\end{aligned}\tag{3.47}$$

$$\begin{aligned}
I'_j &= n\alpha I_{j-1} - (\phi + n\alpha + \mu)I_j, \quad j = 2, \dots, n, \\
H'_1 &= m\kappa Q_m + \phi I_1 - (n\alpha + \mu)H_1, \\
H'_j &= n\alpha H_{j-1} + \phi I_j - (n\alpha + \mu)H_j, \quad j = 2, \dots, n, \\
R' &= n\alpha I_n + n\alpha H_n - \mu R, \\
\text{with } I &= \sum_{j=1}^n I_j, \quad H = \sum_{j=1}^n H_j.
\end{aligned}$$

In the special case when $m = n = 1$, the system (3.47) reduces to:

$$\begin{aligned}
S' &= \mu N - cS \frac{I+(1-\rho)H}{N} - \mu S, \\
E' &= cS \frac{I+(1-\rho)H}{N} - (\chi + \kappa + \mu)E, \\
Q' &= \chi E - (\kappa + \mu)Q, \\
I' &= \kappa E - (\phi + \alpha + \mu)I, \\
H' &= \kappa Q + \phi I - (\alpha + \mu)H.
\end{aligned} \tag{3.48}$$

From the formula (3.45) we get the reproduction number for system (3.47):

$$\begin{aligned}
\mathcal{R}_c &= \frac{(m\kappa)^m}{(\mu + m\kappa)^m} \frac{c}{\mu + n\delta} \sum_{j=0}^{n-1} \frac{(n\alpha)^j}{(\mu + n\alpha)^j} \\
&\quad \left[1 - \rho \left(1 - \frac{(\mu + m\kappa)^m}{(\mu + m\kappa + \chi)^m} \frac{\mu + n\alpha}{\mu + n\alpha + \phi} \frac{\sum_{j=0}^{n-1} \frac{(n\alpha)^j}{(\mu + n\alpha + \phi)^j}}{\sum_{j=0}^{n-1} \frac{(n\alpha)^j}{(\mu + n\alpha)^j}} \right) \right],
\end{aligned} \tag{3.49}$$

with the derivatives

$$\frac{\partial \mathcal{R}_c}{\partial \chi} = -c\rho \frac{m(m\kappa)^m}{(\mu + m\kappa + \chi)^{m+1}} \sum_{j=0}^{n-1} \frac{(n\alpha)^j}{(\mu + n\alpha + \phi)^{j+1}} < 0, \tag{3.50}$$

$$\frac{\partial \mathcal{R}_c}{\partial \phi} = -c\rho \frac{(m\kappa)^m}{(\mu + m\kappa + \chi)^m} \sum_{j=0}^{n-1} \frac{(j+1)(n\alpha)^j}{(\mu + n\alpha + \phi)^{j+2}} < 0. \tag{3.51}$$

3.6.3 *Comparison of EDM and GDM

In this section, we show that when the GDA is used to replace the EDA, model predictions regarding the effectiveness of disease intervention policies may be different both quantitatively and qualitatively. We illustrate this by comparing the two models, GDM (3.47) and EDM (3.48). Two criteria are used in the comparison.

One is the impact of control measures described by χ and ϕ on the reduction in the magnitude of \mathcal{R}_c and the other one is the reduction in the number of cumulative infections C at the end of an epidemic (the final epidemic size).

From (3.49) to (3.51) we know that the reproduction number \mathcal{R}_c for GDM decreases with increasing χ and ϕ . Similarly, using the formula (3.45) we get the reproduction number for the EDM, $\mathcal{R}_c = \mathcal{R}_I + \mathcal{R}_{IH} + \mathcal{R}_{QH}$, where

$$\begin{aligned}\mathcal{R}_I &= \frac{ck}{(\mu + \kappa + \chi)(\mu + \alpha + \phi)}, \\ \mathcal{R}_{IH} &= \frac{(1 - \rho)ck}{\mu + \kappa + \chi} \left(\frac{1}{\mu + \alpha} - \frac{1}{\mu + \alpha + \phi} \right), \\ \mathcal{R}_{QH} &= (1 - \rho)c \left(\frac{\kappa}{\mu + \kappa} - \frac{\kappa}{\mu + \kappa + \chi} \right) \frac{1}{\mu + \alpha},\end{aligned}$$

which can be written in a simpler form as:

$$\mathcal{R}_c = \frac{\kappa}{\mu + \kappa} \frac{c}{\mu + \alpha} \left[1 - \rho \left(1 - \frac{\mu + \kappa}{\mu + \kappa + \chi} \frac{\mu + \alpha}{\mu + \alpha + \phi} \right) \right]. \quad (3.52)$$

The derivatives of \mathcal{R}_c with respect to the control parameters are

$$\begin{aligned}\frac{\partial \mathcal{R}_c}{\partial \chi} &= -c\rho \frac{\kappa}{(\mu + \kappa + \chi)^2} \frac{1}{\mu + \alpha + \phi} < 0, \\ \frac{\partial \mathcal{R}_c}{\partial \phi} &= -c\rho \frac{\kappa}{\mu + \kappa + \chi} \frac{1}{(\mu + \alpha + \phi)^2} < 0.\end{aligned}$$

Hence, the reproduction number \mathcal{R}_c for EDM also decreases as the control parameters χ and ϕ increase. Therefore, both models seem to work well when the impact of each individual control measure is considered. When we try to compare model predictions of combined control strategies, however, inconsistent predictions by the two models are observed. For example, in Fig. 3.8a, b, \mathcal{R}_c for both models is plotted either as a function of ϕ for a fixed value of $\chi = 0.05$, or as a function of χ for a fixed value $\phi = 0.05$, or as a function of both χ and ϕ with $\chi = \phi$. For any vertical line except the one at 0.1, the three curves intersect the vertical line at three points that represent three control strategies. The order of these points (from top to bottom) determines the order of effectiveness (from low to high) of the corresponding control strategies since a larger \mathcal{R}_c value will most likely lead to a higher disease prevalence. The order of these three points (labeled by a circle, a triangle, and a square) predicted by the EDM and the GDM is clearly different for the selected parameter sets, suggesting conflict assessments of interventions between the two models. These conflict assessments are also shown when we compare the C values. For example, Fig. 3.8a shows that the strategy corresponding to $\chi = 0.3, \phi = 0.05$ (indicated by the triangle) is more effective

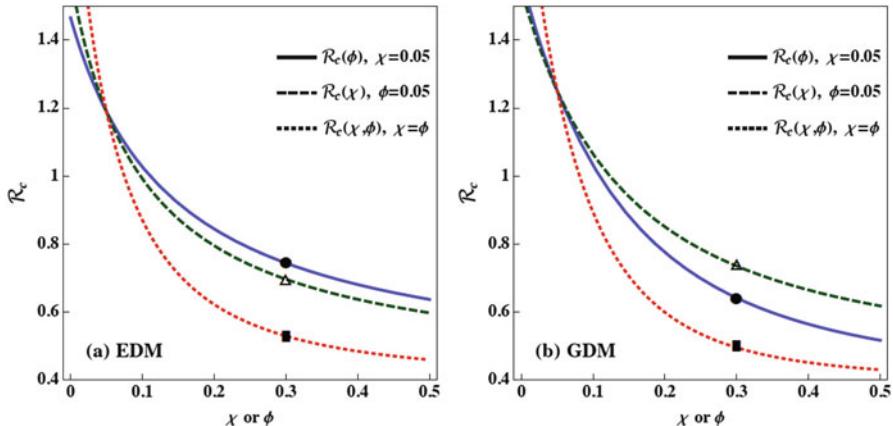


Fig. 3.8 Comparison of the EDM and the GDM on the impact of various control measures. (a) and (b) are plots of the reproduction number \mathcal{R}_c as functions of control measures (χ and ϕ) for the EDM and GDM, respectively

than the strategy corresponding to $\chi = 0.05, \phi = 0.3$ (indicated by the solid circle). However, Fig. 3.8b shows the opposite, i.e., the strategy corresponding to $\chi = 0.3, \phi = 0.05$ (indicated by the triangle) is less effective than the strategy corresponding to $\chi = 0.05, \phi = 0.3$ (indicated by the solid circle). The parameter values used in Fig. 3.8 are $c = 0.2$, $\rho = 0.8$, $\kappa = 1/7$, and $\alpha = 1/10$, corresponding to a disease with a latency period of $1/\kappa = 7$ days and an infectious period of $1/\alpha = 10$ days (e.g., SARS).

To examine in more detail the quantitative differences between the two models we conducted intensive simulations of the EDM and the GDM for various control measures, some of which are illustrated in Fig. 3.9. In this figure, the parameters for gamma distributions are $m = n = 3$, E and I represent the fraction of latent and infectious fractions $E = (E_1 + E_2 + E_3)/N$ and $I = (I_1 + I_2 + I_3)/N$, respectively. The latent and infectious periods are $\kappa = 1/7$ and $\alpha = 1/10$. The cumulative infection is calculated by integrating the incidence function, i.e.,

$$C(t) = \int_0^t cS(s)[I(s) + (1 - \rho)H(s)]/Nds$$

where $H = H_1 + H_2 + H_3$. Figure 3.9a, b is for Strategy I which implements quarantine alone with $\chi = 0.07$, and Fig. 3.9c, d is for Strategy II which implements isolation alone with $\phi = 0.06$. The effectiveness of these control measures is reflected by the corresponding $C(t)$ values. According to Fig. 3.9a, c, the EDM predicts that Strategy II is *more* effective than Strategy I as the number C of cumulative infections (fractions) under Strategy II is 30% lower than the C value under Strategy I (notice that $C \approx 0.3$ and $C \approx 0.2$ under strategies I and II, respectively). However, according to Fig. 3.9b, d, the GDM predicts that Strategy

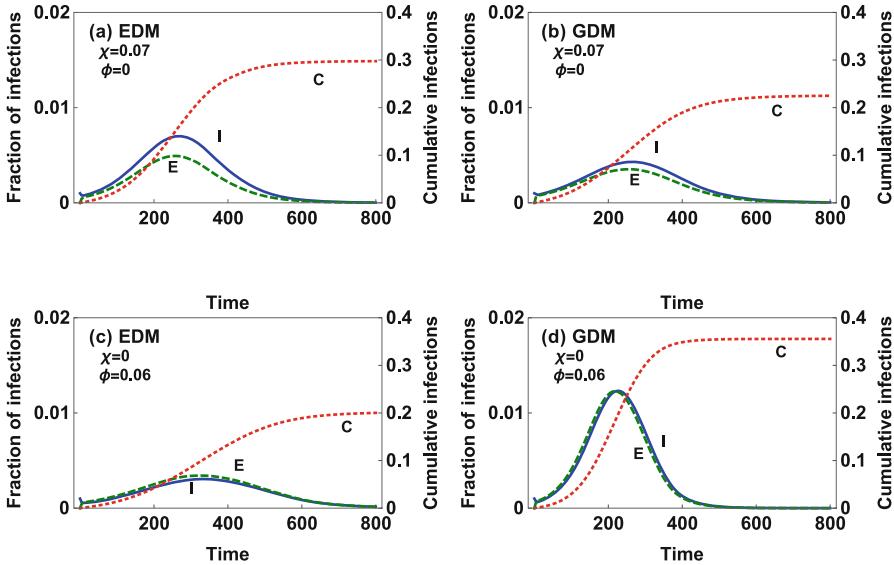


Fig. 3.9 Comparison of control strategies and evaluations given by the exponential distribution model (EDM) and the gamma distribution model (GDM) with $m = n = 3$. Two strategies represented by χ and ϕ are compared: Strategy I involves quarantine alone ($\chi = 0.07$ and $\phi = 0$) while Strategy II involves isolation alone ($\chi = 0$ and $\phi = 0.06$). Other parameter values used are $1/\kappa = 7$ (day), $1/\alpha = 10$ (day), $1/\mu = 75$ (year), and $c = 0.2$. The time unit is day

II is less effective than Strategy II as the number C of cumulative infections under Strategy I is 30% lower than the C value under Strategy II (notice that $C \approx 0.23$ and $C \approx 0.36$ under strategies I and II, respectively). Obviously, in this example, the predictions by the EDM and by the GDM are inconsistent.

One of the main reasons for the discrepancy between models with exponential and gamma distributions is the memoryless property of the exponential distribution. This can be made more transparent by examining the expected remaining sojourns from the distributions. Under the gamma distribution $p_n(s, \theta)$ (or simply denoted by $p_n(s)$) with $n \geq 2$, the expected remaining sojourn at stage age s is

$$\mathcal{M}_n(s) = \int_0^\infty \frac{p_n(t+s)}{p_n(s)} dt = \frac{1}{p_n(s)} \int_s^\infty p_n(t) dt = \frac{1}{n\theta} \frac{\sum_{k=0}^{n-1} \sum_{j=0}^k \frac{(n\theta s)^j}{j!}}{\sum_{k=0}^{n-1} \frac{(n\theta s)^k}{k!}}.$$

After checking $\mathcal{M}'_n(s) < 0$ and $\lim_{s \rightarrow \infty} \mathcal{M}_n(s) \rightarrow T/n$ where $T = 1/\theta$, we know that $\mathcal{M}_n(s)$ strictly decreases with stage age s , and that when s is large the expected remaining sojourn can be as small as T/n . Hence, the expected remaining sojourn in a stage is indeed dependent on the time already spent in the stage. Therefore, the gamma distribution $p_n(s)$ for $n \geq 2$ provides a more realistic description than the exponential distribution $p_1(s)$ for which $\mathcal{M}_1(s) = T$ for all s .

3.7 Diseases in Exponentially Growing Populations

Many parts of the world experienced very rapid population growth in the eighteenth century. The population of Europe increased from 118 million in 1700 to 187 million in 1800. In the same time period the population of Great Britain increased from 5.8 to 9.15 million, and the population of China increased from 150 to 313 million [33]. The population of English colonies in North America grew much more rapidly than this, aided by substantial immigration from England, but the native population, which had been reduced to one tenth of their previous size by disease following the early encounters with Europeans and European diseases, grew even more rapidly. While some of these population increases may be explained by improvements in agriculture and food production, it appears that an even more important factor was the decrease in the death rate due to diseases. Disease death rates dropped sharply in the eighteenth century, partly from better understanding of the links between illness and sanitation and partly because the recurring invasions of bubonic plague subsided, perhaps due to reduced susceptibility. One plausible explanation for these population increases is that the bubonic plague invasions served to control the population size, and when this control was removed the population size increased rapidly.

In developing countries it is quite common to have high birth rates and high disease death rates. In fact, when disease death rates are reduced by improvements in health care and sanitation it is common for birth rates to decline as well, since families no longer need to have as many children to ensure that enough children survive to take care of the older generations. Again, it is plausible to assume that population size would grow exponentially in the absence of disease but is controlled by disease mortality.

The *SIR* model with births and deaths of Kermack and McKendrick [28] includes births in the susceptible class proportional to population size and a natural death rate in each class proportional to the size of the class. Let us analyze a model of this type with birth rate r and a natural death rate $\mu < r$. For simplicity we assume the disease is fatal to all infectives with disease death rate α , so that there is no removed class and the total population size is $N = S + I$. Our model is

$$\begin{aligned} S' &= r(S + I) - \beta SI - \mu S \\ I' &= \beta SI - (\mu + \alpha)I. \end{aligned} \tag{3.53}$$

From the second equation we see that equilibria are given by either $I = 0$ or $\beta S = \mu + \alpha$. If $I = 0$, the first equilibrium equation is $rS = \mu S$, which implies $S = 0$ since $r > \mu$. It is easy to see that the equilibrium $(0,0)$ is unstable. What actually would happen if $I = 0$ is that the susceptible population would grow exponentially with exponent $r - \mu > 0$. If $\beta S = \mu + \alpha$, the first equilibrium condition gives

$$r \frac{\mu + \alpha}{\beta} + rI - (\mu + \alpha)I - \frac{\mu(\mu + \alpha)}{\beta} = 0,$$

which leads to

$$(\alpha + \mu - r)I = \frac{(r - \mu)(\mu + \alpha)}{\beta}.$$

Thus, there is an endemic equilibrium provided $r < \alpha + \mu$, and it is possible to show by linearizing about this equilibrium that it is asymptotically stable. On the other hand, if $r > \alpha + \mu$ there is no positive equilibrium value for I . In this case we may add the two differential equations of the model to give

$$N' = (r - \mu)N - \alpha I \geq (r - \mu)N - \alpha N = (r - \mu - \alpha)N$$

and from this we may deduce that N grows exponentially. For this model, either we have an asymptotically stable endemic equilibrium or population size grows exponentially. In the case of exponential population growth we may have either vanishing of the infection or an exponentially growing number of infectives.

If only susceptibles contribute to the birth rate, as may be expected if the disease is sufficiently debilitating, the behavior of the model is quite different. Let us consider the model

$$\begin{aligned} S' &= rS - \beta SI - \mu S = S(r - \mu - \beta I) \\ I' &= \beta SI - (\mu + \alpha)I = I(\beta S - \mu - \alpha) \end{aligned} \quad (3.54)$$

which has the same form as the Lotka–Volterra predator–prey model of population dynamics. This system has two equilibria, obtained by setting the right sides of each of the equations equal to zero, namely $(0, 0)$ and an endemic equilibrium $((\mu + \alpha)/\beta, (r - \mu)/\beta)$. It turns out that the qualitative analysis approach we have been using is not helpful as the equilibrium $(0, 0)$ is unstable and the eigenvalues of the coefficient matrix at the endemic equilibrium have real part zero. In this case the behavior of the linearization does not necessarily carry over to the full system. However, we can obtain information about the behavior of the system by a method that begins with the elementary approach of separation of variables for first order differential equations. We begin by taking the quotient of the two differential equations and using the relation

$$\frac{I'}{S'} = \frac{dI}{dS}$$

to obtain the separable first order differential equation

$$\frac{dI}{dS} = \frac{I(\beta S - \mu - \alpha)}{S(r - \beta I)}.$$

Separation of variables gives

$$\int \left(\frac{r}{I} - \beta \right) dI = \int \left(\beta - \frac{\mu + \alpha}{S} \right) dS.$$

Integration gives the relation

$$\beta(S + I) - r \log I - (\mu + \alpha) \log S = c$$

where c is a constant of integration. This relation shows that the quantity

$$V(S, I) = \beta(S + I) - r \log I - (\mu + \alpha) \log S$$

is constant on each orbit (path of a solution in the (S, I) plane). Each of these orbits is a closed curve corresponding to a periodic solution.

We may view the model as describing an epidemic initially, leaving a susceptible population small enough that infection cannot establish itself. Then there is a steady population growth until the number of susceptibles is large enough for an epidemic to recur. During this growth stage the infective population is very small and random effects may wipe out the infection, but the immigration of a small number of infectives will eventually restart the process. As a result, we would expect recurrent epidemics. In fact, bubonic plague epidemics did recur in Europe for several hundred years. If we modify the demographic part of the model to assume limited population growth rather than exponential growth in the absence of disease, the effect would be to give behavior like that of the model studied in the previous section, with an endemic equilibrium that is approached slowly in an oscillatory manner if $\mathcal{R}_0 > 1$.

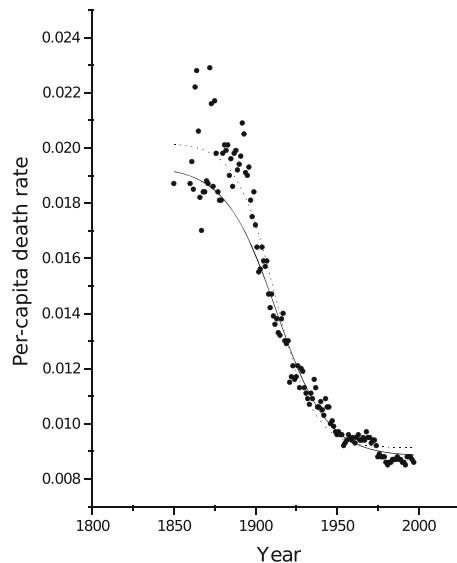
3.8 Project: Population Growth and Epidemics

When one tries to fit epidemiological data over a long time interval to a model, it is necessary to include births and deaths in the population. Throughout the book we have considered population models with birth and death rates that are constant in time. However, population growth often may be fit better by assuming a linear population model with a time-dependent growth rate, even though this does not have a model-based interpretation. There could be many reasons for variations in birth and death rates; we could not quantify the variations even if we knew all of the reasons. Let $r(t) = \frac{dN}{dt}/N$ denote the time-dependent per capita growth rate. To estimate $r(t)$ from linear interpolation of census data, proceed as follows:

1. Let N_i and N_{i+1} be the consecutive census measurements of population size taken at times t_i and t_{i+1} , respectively. Let $\Delta N = N_{i+1} - N_i$, $\Delta t = t_{i+1} - t_i$, and $\delta N = N(t + \delta t) - N(t)$.
2. If $t_i \leq t \leq t_{i+1}$, $\frac{\Delta N}{\Delta t} = \frac{\delta N}{\delta t}$, then we make the estimate $r(t) \approx \frac{\Delta N}{\Delta t N(t)}$.
3. A better approximation is obtained by replacing $N(t)$ by $N(t + \delta t/2)$. Why? Show that in this case, $r(t) \approx \left(\frac{\delta t}{2} + \frac{N(t)\Delta t}{\Delta N}\right)^{-1}$.

Table 3.1 Population data growth for the USA

Year	Population size	Year	Population size	Year	Population size
1700	250,888	1800	5,308,483	1900	75,994,575
1710	331,711	1810	7,239,881	1910	91,972,266
1720	466,185	1820	9,638,453	1920	105,710,620
1730	629,445	1830	12,866,020	1930	122,775,046
1740	905,563	1840	17,069,453	1940	131,669,275
1750	1,170,760	1850	23,192,876	1950	151,325,798
1760	1,593,625	1860	31,443,321	1960	179,323,175
1770	2,148,076	1870	39,818,449	1970	203,302,031
1780	2,780,369	1880	50,155,783	1980	226,542,199
1790	3,929,214	1890	62,947,714	1990	248,718,301
–	–	–	–	2000	274,634,000

Fig. 3.10 Observed death rate (filled circle) and the best fit obtained with the function (3.55)

Question 1 Use the data of Table 3.1 to estimate the growth rate $r(t)$ for the population of the USA.

Figure 3.10 shows the time evolution of the USA mortality rate. This mortality rate is fit well by

$$\mu = \mu_0 + \frac{\mu_0 - \mu_f}{1 + e^{(t-t'_{1/2})/\Delta'}} \quad (3.55)$$

with $\mu_0 = 0.01948$, $\mu_f = 0.008771$, $t'_{1/2} = 1912$, and $\Delta' = 16.61$. Then the “effective birth rate” $b(t)$ is defined as the real birth rate plus the immigration rate.

Question 2 Estimate $b(t)$ using $r(t) = b(t) - \mu(t)$, with $r(t)$ found in Question 1.

Consider an SEIR disease transmission model. We assume that:

- An average infective individual produces β new infections per unit of time when all contacts are with susceptibles but that otherwise, this rate is reduced by the ratio S/N .
- Individuals in the exposed class E progress to the infective class at the per capita rate k .
- There is no disease-induced mortality or permanent immunity, and there is a mean infective period of $1/\gamma$.

We define $\gamma = r + \mu$. The model becomes:

$$\begin{aligned}\frac{dS}{dt} &= bN - \mu S - \beta S \frac{I}{N}, \\ \frac{dE}{dt} &= \beta S \frac{I}{N} - (k + \mu) E, \\ \frac{dI}{dt} &= k E - (r + \mu) I, \\ \frac{dR}{dt} &= r I - \mu R.\end{aligned}\tag{3.56}$$

Question 3

- Show that the mean number of secondary infections (belonging to the exposed class) produced by one infective individual in a population of susceptibles is $Q_0 = \beta/\gamma$.
- Assuming that k and μ are time-independent, show that \mathcal{R}_0 is given by $Q_0 f$, where $f = k/(k + \mu)$. What is the epidemiological interpretation of $Q_0 f$?

The usual measure of the severity of an epidemic is the incidence of infective cases. The incidence of infective cases is defined as the number of new infective individuals per year. If we take 1 year as the unit of time, the incidence of infective cases is given approximately by kE . The incidence rate of infective cases per 100,000 population is given approximately by $10^5 kE/N$.

Tuberculosis (TB) is an example of a disease with an exposed (noninfective) stage. Infective individuals are called active TB cases. Estimated incidence of active TB in the USA was in a growing phase until around 1900 and then experienced a subsequent decline. The incidence rate of active TB exhibited a declining trend from 1850 (see Table 3.2 and Fig. 3.11). The proportion of exposed individuals who survive the latency period and become infective is $f = \frac{k}{k+\mu}$. The number f will be used as a measure of the risk of developing active TB by exposed individuals.

Question 4 Assume that the mortality rate varies according to the expression (3.55), and that the value of b found in Question 2 is used. Set $\gamma = 1 \text{ years}^{-1}$ and $\beta = 10 \text{ years}^{-1}$, both constant through time. Simulate TB epidemics starting in 1700 assuming constant values for f . Can you reproduce the observed trends (Table 3.2)?

Table 3.2 Reported incidence and incidence rate (per 100,000 population) of active TB

Year	Incidence rate	Incidence	Year	Incidence rate	Incidence
1953	53	84,304	1976	15	32,105
1954	49.3	79,775	1977	13.9	30,145
1955	46.9	77,368	1978	13.1	28,521
1956	41.6	69,895	1979	12.6	27,769
1957	39.2	67,149	1980	12.3	27,749
1958	36.5	63,534	1981	11.9	27,337
1959	32.5	57,535	1982	11	25,520
1960	30.8	55,494	1983	10.2	23,846
1961	29.4	53,726	1984	9.4	22,255
1962	28.7	53,315	1985	9.3	22,201
1963	28.7	54,042	1986	9.4	22,768
1964	26.6	50,874	1987	9.3	22,517
1965	25.3	49,016	1988	9.1	22,436
1966	24.4	47,767	1989	9.5	23,495
1967	23.1	45,647	1990	10.3	25,701
1969	19.4	39,120	1992	10.5	26,673
1970	18.3	37,137	1993	9.8	25,287
1971	17.1	35,217	1994	9.4	24,361
1972	15.8	32,882	1995	8.7	22,860
1973	14.8	30,998	1996	8	21,337
1974	14.2	30,122	1997	7.4	19,885
1975	15.9	33,989	1998	6.8	18,361

It is not possible to obtain a good fit of the data of Table 3.2 to the model (3.56). It is necessary to use a refinement of the model that includes time-dependence in the parameters, and the next step is to describe such a model. The risk of progression to active TB depends strongly on the standard of living. An indirect measure of the standard of living can be obtained from the life expectancy at birth. The observed life expectancy for the USA is approximated well by the sigmoid shape function

$$\tau = \tau_f + \frac{(\tau_0 - \tau_f)}{1 + \exp[(t - t_{1/2})/\Delta]}, \quad (3.57)$$

shown in Fig. 3.12. Here τ_0 and τ_f are asymptotic values for life expectancy; $t_{1/2} = 1921.3$ is the time by which life expectancy reaches the value $(\tau_0 + \tau_f)/2$; and $\Delta = 18.445$ determines the width of the sigmoid.

Assume that the risk f varies exactly like life expectancy, that is, assume that f is given by

$$f(t) = f_f + \frac{(f_i - f_f)}{1 + \exp[(t - t_{1/2})/\Delta]}. \quad (3.58)$$

Fig. 3.11 Incidence of active TB

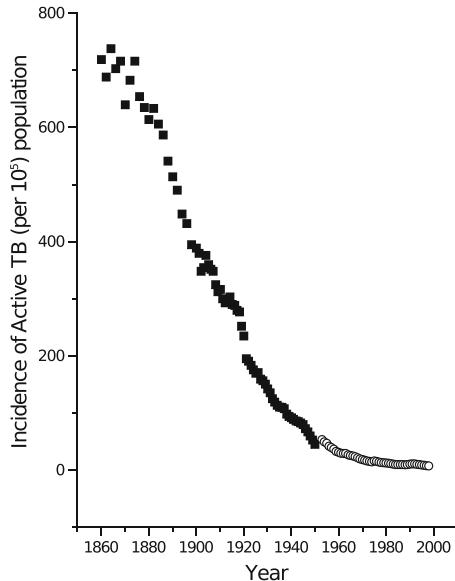
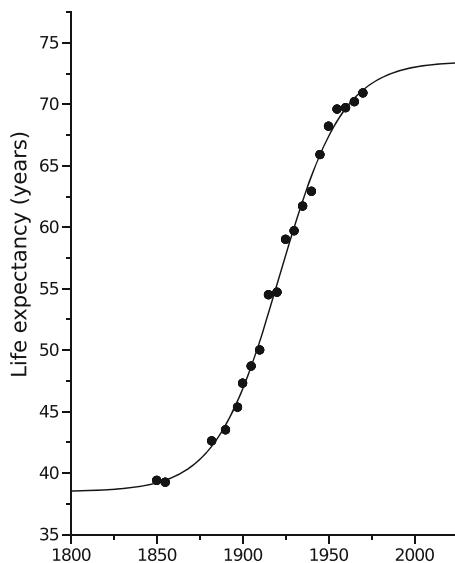


Fig. 3.12 Observed average life expectancy at birth (filled circle) and its best fit (continuous line) using expression (3.57)



We refine the model (3.56) by replacing the parameter k by the variable expression $\mu f(t)/(1 - f(t))$ and $k + \mu$ by $\mu/(1 - f(t))$, obtained from the relation $f = k/(k + \mu)$. Since the time scale of the disease is much faster than the demographic time scale, the recovery rate r is approximately equal to γ . This gives the model

$$\begin{aligned}\frac{dS}{dt} &= b(t)N - \mu(t)S - \beta S \frac{I}{N}, \\ \frac{dE}{dt} &= \beta S \frac{I}{N} - \frac{\mu(t)}{1-f(t)} E, \\ \frac{dI}{dt} &= \frac{\mu(t)f(t)}{1-f(t)} E - \gamma I, \\ \frac{dR}{dt} &= \gamma I - \mu(t)R.\end{aligned}\tag{3.59}$$

Question 5 Simulate TB epidemics starting in 1700 using the model (3.59) with $\gamma = 1 \text{ years}^{-1}$ and $\beta = 10 \text{ years}^{-1}$, both constant, and with $\mu(t)$ given by (3.55) and $f(t)$ given by (3.58). Find values of f_0 and f_f for which an accurate reproduction of the observed TB trends (Table 3.2) is achieved.

References: [1–4, 9–11, 15, 16, 38–41].

3.9 *Project: An Environmentally Driven Infectious Disease

Consider an environmentally driven infectious disease such as cholera and toxoplasmosis (a parasite disease caused by *T. gondii*). For this type of disease, the transmission occurs when susceptible hosts have contacts with a contaminated environment, and the rate of environment contamination is dependent on both the number of infected hosts and the average pathogen load within an infected host. One way to model the transmission dynamics for such a disease is to consider both the disease transmission at the population level and the infection process within the hosts. The following model couples a simple within-host system for cell–parasite interactions (e.g., see [34–36]) and an endemic SI model with an interaction with a contaminated environment:

$$\begin{aligned}\dot{T} &= \Lambda - kVT - mT, \\ \dot{T}^* &= kVT - (m + d)T^*, \\ \dot{V} &= g(E) + pT^* - cV, \\ \dot{S} &= \mu(S + I) - \lambda ES - \mu S, \\ \dot{I} &= \lambda ES - \mu I, \\ \dot{E} &= \theta(V)I(1 - E) - \gamma E.\end{aligned}\tag{3.60}$$

Here, the variables for the within-host system $T = T(t)$, $T^* = T^*(t)$, and $V = V(t)$ are the densities of healthy cells, infected cells, and parasite load, respectively. $S = S(t)$ and $I = I(t)$ denote the numbers of susceptible and infective individuals at time t , respectively. Λ denotes the recruitment rate of cells; k is the per-capita infection rate of cells; m and d are the per-capita background and infection-induced

cell mortalities, respectively; p denotes the parasite production rate by an infected cell and c is the within-host clearance rate of pathogens.

The variables $S(t)$ and $I(t)$ denote the numbers of susceptible and infective hosts at time t , and $E(t)$ ($0 \leq E \leq 1$) represents the level of environmental contamination at time t , or the concentration of the pathogen per unit area of a region being considered. The parameter λ denotes the per-capita infection rate of hosts in a contaminated environment; μ denotes per-capita birth and natural death rate of hosts; and γ denotes the rate of pathogen clearance in the environment.

The function $g(E)$ in the V equation represents an added rate in the change of parasite load due to the continuous ingestion of parasites by the host from a contaminated environment, and is assumed to have the following properties:

$$g(0) = 0, \quad g(E) \geq 0, \quad g'(E) > 0, \quad g''(E) \leq 0. \quad (3.61)$$

One of the simplest forms for $g(E)$ is the linear function $g(E) = aE$, where a is a positive constant. Other forms of $g(E)$ include $g_1(E) = aE/(1+bE)$ with a and b being positive constants and $g_2(E) = aE^q$ ($q < 1$).

For the analysis of the coupled model (3.60), a commonly used approach is to consider that the within-host system (consisting of the T , T^* , and V equations) occurs on a much faster time scale than the between-host system (consisting of S , I , and E equations), which allows the substitution of a stable equilibrium of the fast-system (treating the slow-variables as constant) into the slow-system and study the lower-dimensional slow system (see, e.g., [7, 12, 17, 19]). The system for the fast variables is

$$\begin{aligned} \dot{T} &= \Lambda - kVT - mT \\ \dot{T}^* &= kVT - (m + d)T^* \\ \dot{V} &= g(E) + pT^* - cV, \end{aligned} \quad (3.62)$$

and the system for the slow variables is

$$\begin{aligned} \dot{S} &= \mu(S + I) - \lambda ES - \mu S, \\ \dot{I} &= \lambda ES - \mu I, \\ \dot{E} &= \theta IV(1 - E) - \gamma E. \end{aligned} \quad (3.63)$$

Question 1 Consider the fast system (3.62). The within-host reproduction number \mathcal{R}_w (w for within) is given by

$$\mathcal{R}_w = \frac{kpT_0}{c(m + d)} \quad (3.64)$$

where $T_0 = \Lambda/\mu$.

- (a) Let $E > 0$ be a constant. Show that (3.62) has a unique biologically feasible equilibrium (which depends on E) $\tilde{U}(E) = (\tilde{T}(E), \tilde{T}^*(E), \tilde{V}(E))$.
- (b) Show that the unique equilibrium $\tilde{U}(E) = (\tilde{T}(E), \tilde{T}^*(E), \tilde{V}(E))$ of (3.62) is globally, asymptotically stable.

Hint: Consider the following Lyapunov function

$$\begin{aligned}\mathcal{L}(T, T^*, V) = & \tilde{T} \left(\frac{T}{\tilde{T}} - \log \frac{T}{\tilde{T}} - 1 \right) + \tilde{T}^* \left(\frac{T^*}{\tilde{T}^*} - \log \frac{T^*}{\tilde{T}^*} - 1 \right) \\ & + \frac{m+d}{p} \tilde{V} \left(\frac{V}{\tilde{V}} - \log \frac{V}{\tilde{V}} - 1 \right).\end{aligned}$$

Consider the case when $\mathcal{R}_w > 1$. It can be verified that $\tilde{V}(0) = \lim_{E \rightarrow 0} \tilde{V}(E) > 0$. Note that the total population of hosts $N = S + I$ remains constant for all $t > 0$. Thus, the fast system (3.62) can be reduced to a two-dimensional system (by ignoring the S equations). Note also that \tilde{U} is g.a.s. in the fast system. We can replace the fast variable V in (3.62) by $\tilde{V}(E)$ and study the following fast system

$$\begin{aligned}\dot{I} &= \lambda E(N - I) - \mu I, \\ \dot{E} &= \theta I \tilde{V}(E)(1 - E) - \gamma E.\end{aligned}\tag{3.65}$$

The reproduction number for the between-host system, which is denoted by \mathcal{R}_b (b for between) and defined as

$$\mathcal{R}_b = \frac{\theta \tilde{V}(0)}{\mu} \frac{\lambda N}{\gamma}.\tag{3.66}$$

Therefore, \mathcal{R}_b represents the number of secondary infections through the environment by one infected individual during the entire infectious period in a completely susceptible host population and environment.

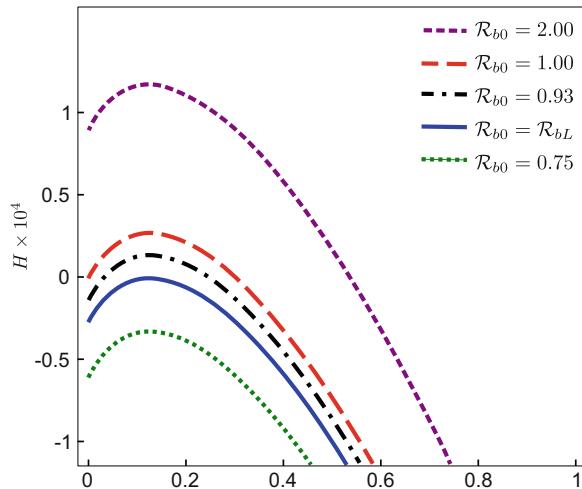
Let $\hat{W} = (\hat{I}, \hat{E})$ denote a biologically feasible equilibrium for (3.65). Show that $\hat{I} = \lambda \hat{E} N / (\lambda \hat{E} + \mu)$ and \hat{E} is a solution of the equation $F(E) = G(E)$, $0 < E < 1$, where

$$\begin{aligned}F(E) &= \frac{1-E}{c} \left[g(E) + \frac{pm}{m+d} (T_0 - \tilde{T}(E)) \right], \\ G(E) &= \frac{\gamma E}{\theta N} + \frac{\mu \gamma}{\theta \lambda N}.\end{aligned}\tag{3.67}$$

Equivalently, \hat{E} is a zero of the function $H(E) = F(E) - G(E)$ with $0 < E < 1$.

- (a) Let $\hat{W}_0 = (0, 0)$ denote the infection-free equilibrium of (3.65). Show that \hat{W}_0 locally asymptotically stable when $\mathcal{R}_b < 1$ and unstable when $\mathcal{R}_b > 1$.

Fig. 3.13 Plot of the function $H(E)$ for different \mathcal{R}_b value between 0.75 and 2. A zero of $H(E)$ in $(0, 1)$ corresponds to positive equilibrium of the slow system (3.65)



- (b) Show that it is possible for the equation $H(E) = F(E) - G(E) = 0$ to have 0, 1, or 2 solutions in $(0, 1)$.

Hint: Show first that $H''(E) > 0$ for $0 < E < 1$.

- (c) Figure 3.13 illustrates a numerical plot of the function $H(E)$ for various \mathcal{R}_b (by varying λ) values between 0.75 and 2. Other parameter values used are: $\Lambda = 6 \times 10^3$, $k = 1.5 \times 10^{-6}$, $m = 0.3$, $d = 0.2$, $a = 4 \times 10^5$, $c = 50$, $\mathcal{R}_{w0} = 1.09$ ($p = 908$), $N = 10^4$, $\mu = 4 \times 10^{-4}$, $\theta = 1 \times 10^{-10}$, and $\gamma = 0.02$. What do you observe? How does the number of solutions of $H(E) = 0$ depend on \mathcal{R}_b ?
- (d) From the plot in part (c) we can observe that there exists a lower bound $\mathcal{R}_{bL} \in (0, 1)$ such that for all $\mathcal{R}_b \in (\mathcal{R}_{bL}, 1)$, the equation $H(E) = F(E) - G(E) = 0$ has two solutions in $(0, 1)$, which correspond to two positive equilibria $\hat{W}_i = (\hat{I}_i, \hat{E}_i)$ ($i = 1, 2$) with $\hat{I}_2 > \hat{I}_1$. In this case, prove analytically that \hat{W}_2 is locally asymptotically stable and \hat{W}_1 is unstable. (**Hint:** Check the sign of the eigenvalues of the Jacobian matrix at \hat{W}).
- (e) For the full system (3.60), conduct numerical simulations to confirm the results stated in parts (a)–(d), which are obtained by separating the fast and slow systems.
- (i) Reproduce the Fig. 3.14 (left) by plotting the fraction of infected $I(t)/N$ vs. time with several sets of initial conditions. Use the same parameter values as in Part (c) except that $p = 850$ (corresponding to $\mathcal{R}_b = 0.37 < 1$), $a = 5 \times 10^5$, $\lambda = 5.5 \times 10^{-4}$, and $\gamma = 0.015$.
- (ii) Reproduce the phase portrait shown in Fig. 3.14 (right), which illustrates one fast variable (V) and one slow variable (E). Use the same parameter values as in Part (c) except that $p = 10^3$ (corresponding to $\mathcal{R}_b > 1$) and $a = 4 \times 10^4$.

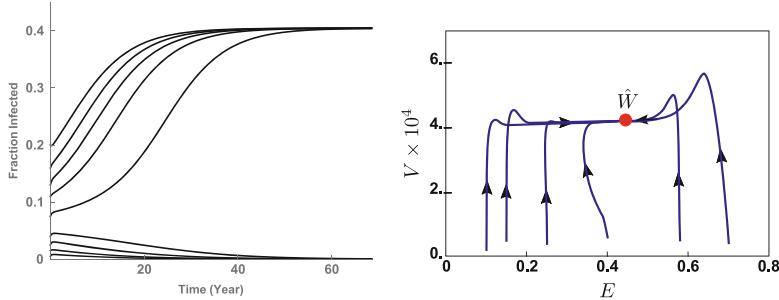


Fig. 3.14 Left: Time plot of the full system (3.60) for $\mathcal{R}_b \in (\mathcal{R}_{bL}, 1)$, in which case there are two stable equilibria, one with the infection-free and one with positive infection level. Right: Phase portrait of the full system (3.60) for $\mathcal{R}_b > 1$, in which case there is a unique stable equilibrium

3.10 *Project: A Two-Strain Model with Cross Immunity

This project concerns a two-strain model with cross immunity. Divide the population into ten different classes: susceptibles (S), infected with strain i (I_i), primary infection), isolated with strain i (Q_i), recovered from strain i (R_i , as a result of primary infection), infected with strain i (V_i , secondary infection), given that the population had recovered from strains $j \neq i$, and recovered from both strains (W). Let A denote the population of non-isolated individuals and let $\frac{\beta_i S(I_i + V_i)}{A}$ be the rate at which susceptibles become infected with strain i . That is, the i th ($i \neq j$) incidence rate is assumed to be proportional to both the number of susceptibles and the available modified proportion of i -infectious individuals, $\frac{(I_i + V_i)}{A}$. Let σ_{ij} denote a measure of the cross-immunity provided by a prior infection with strain i to exposure with strain j ($i \neq j$). Consider the following model

$$\begin{aligned} \frac{dS}{dt} &= A - \sum_{i=1}^2 \beta_i S \frac{(I_i + V_i)}{A} - \mu S, \\ \frac{dI_i}{dt} &= \beta_i S \frac{(I_i + V_i)}{A} - (\mu + \gamma_i + \delta_i) I_i, \\ \frac{dQ_i}{dt} &= \delta_i I_i - (\mu + \alpha_i) Q_i, \\ \frac{dR_i}{dt} &= \gamma_i I_i + \alpha_i Q_i - \beta_j \sigma_{ij} R_i \frac{(I_j + V_j)}{A} - \mu R_i, \quad j \neq i \\ \frac{dV_i}{dt} &= \beta_i \sigma_{ij} R_j \frac{(I_i + V_i)}{A} - (\mu + \gamma_i) V_i, \quad j \neq i \\ \frac{dW}{dt} &= \sum_{i=1}^2 \gamma_i V_i - \mu W, \\ A &= S + W + \sum_{i=1}^2 (I_i + V_i + R_i). \end{aligned} \tag{3.68}$$

The basic reproduction number for strain i is

$$\mathcal{R}_i = \frac{\beta_i}{\mu + \gamma_i + \delta_i}, \quad i = 1, 2.$$

Assume that $\sigma_{12} = \sigma_{21} = \sigma$. The values of reproduction numbers \mathcal{R}_i and the cross-immunity levels σ determine the existence and stability of equilibrium points of the system (3.68). Let E_i denote the boundary equilibria where only strain i is present ($i = 1, 2$).

Question 1 Consider the case when changes in \mathcal{R}_i are due to changes in β_i . Let $f(\mathcal{R}_1)$ and $g(\mathcal{R}_2)$ be the two functions given by

$$f(\mathcal{R}_1) = \frac{\mathcal{R}_1}{1 + \sigma(\mathcal{R}_1 - 1) \left(1 + \frac{\delta_2}{\mu + \gamma_2} \right) \left(1 - \frac{\mu(\mu + \alpha_1)}{(\mu + \gamma_1)(\mu + \alpha_1) + \alpha_1 \delta_1} \right)} \quad (3.69)$$

and

$$g(\mathcal{R}_2) = \frac{\mathcal{R}_2}{1 + \sigma(\mathcal{R}_2 - 1) \left(1 + \frac{\delta_1}{\mu + \gamma_1} \right) \left(1 - \frac{\mu(\mu + \alpha_2)}{(\mu + \gamma_2)(\mu + \alpha_2) + \alpha_2 \delta_2} \right)}, \quad (3.70)$$

and let σ_1^* and σ_2^* be the critical values such that

$$f'(\mathcal{R}_1) \equiv \frac{\partial f(\mathcal{R}_1, \sigma)}{\partial \mathcal{R}_1} \Big|_{\sigma_1^*} = 0, \quad g'(\mathcal{R}_2) \equiv \frac{\partial g(\mathcal{R}_2, \sigma)}{\partial \mathcal{R}_2} \Big|_{\sigma_2^*} = 0. \quad (3.71)$$

Determine the properties of f and g and sketch these functions in the $\mathcal{R}_1 - \mathcal{R}_2$ plane.

Question 2

- (a) Determine the region(s) in the $\mathcal{R}_1 - \mathcal{R}_2$ plane for the existence of E_1 and E_2 .
- (b) Determine the conditions for the stabilities of E_1 and E_2 .

3.11 Exercises

1. Consider the following SEIR model with disease-induced mortality:

$$\begin{aligned} \frac{dS}{dt} &= \mu N - \beta S \frac{I}{N} - \mu S, \\ \frac{dE}{dt} &= \beta S \frac{I}{N} - (\kappa + \mu) E, \end{aligned}$$

$$\frac{dI}{dt} = \kappa E - (\gamma + \mu + \delta)I,$$

$$\frac{dR}{dt} = \gamma I - \mu R,$$

$$N = S + E + I + R,$$

where δ denotes the per capita rate of disease related death.

- (a) Compute the basic reproduction number \mathcal{R}_0 .
 - (b) Does the system have an endemic equilibrium? If yes, find the condition in terms of \mathcal{R}_0 .
 - (c) Show that the endemic equilibrium is locally asymptotically stable whenever it exists.
 - (d) Reduce the system to a three-dimensional system by introducing fractions $u = S/N, x = E/N, y = I/N, z = R/N$.
2. Show that the endemic equilibrium of (3.3) is asymptotically stable if $\mathcal{R}_0 > 1$.
3. Consider a population in which a fraction $p \in (0, 1)$ of newborns are successfully vaccinated and assume permanent immunity after infection and vaccination. Assume that infectious individuals are treated at a per capita rate r . Let \mathcal{R}_c denote the control reproduction number such that the disease-free equilibrium is locally asymptotically stable when $\mathcal{R}_c < 1$. Consider a disease for which $\beta = 0.86$, $\gamma = 1/14 \text{ days}^{-1}$, $\mu = 1/75 \text{ years}^{-1}$. Use the following SIR model to calculate the threshold immunity level p_c such that $\mathcal{R}_c < 1$ for $p > p_c$.

$$\begin{aligned}\frac{dS}{dt} &= \mu N(1-p) - \beta S \frac{I}{N} - \mu S, \\ \frac{dI}{dt} &= \beta S \frac{I}{N} - (\gamma + \mu + r)I, \\ \frac{dR}{dt} &= \mu Np + (\gamma + r)I - \mu R, \\ N &= S + I + R.\end{aligned}$$

- (a) Find p_c in the absent of treatment (i.e., $r = 0$).
 - (b) Find p_c when $r = 0.2$.
 - (c) Plot p_c as a function of r .
 - (d) Plot \mathcal{R}_c as a function of p and r
 - (e) Plot several contour curves of \mathcal{R}_c in the (p, r) plane including the curve for $\mathcal{R}_c = 1$.
4. Consider the SIRS model (3.10).
- (a) Find the expression for the fraction I^*/N of the infected individuals at the endemic equilibrium.
 - (b) Explore the dependence of I^*/N on the immunity loss (θ), particularly in the two extreme cases when the immunity period is very short or very long.

- 5.* Consider the SIR model with delay (3.12).
- Find the endemic equilibrium.
 - Let $\beta = 0.86$ and $\alpha = 1/14$. Determine the threshold value ω_c such that the stability of the endemic equilibrium switches its stability.
6. Consider the vaccination model (3.21).
- Verify that $\mathcal{R}_c < 1$ whenever there is a backward bifurcation.
 - Show how to choose φ to make $\mathcal{R}_c < \mathcal{R}_0$, assuming that all parameters other than φ are kept fixed.
 - Is it possible to improve the vaccine (decrease σ) enough to make $\mathcal{R}_c < \mathcal{R}_0$, assuming all parameters other than σ are kept fixed?
- 7.* Consider the model with a gamma distribution (3.47) and the exponential distribution (3.48). Compare the behavior of the two models under the scenarios specified below. Assume that all parameters have the same values as in Fig. 3.10 except the control parameters χ and φ .
- $\chi = \varphi = 0$. Do you observe any differences in the disease prevalence between the two models? Explain why or why not.
 - Compare the two models under two strategies. Strategy I: $\chi = 0.08$ and $\varphi = 0$, Strategy II: $\chi = 0$ and $\varphi = 0.08$. Do you observe any differences in the disease prevalence between the two models? Explain why or why not.

References

- Aparicio, J.P., A. Capurro, and C. Castillo-Chavez (2000) Markers of disease evolution: the case of tuberculosis, *J. Theor. Biol.*, **215**: 227–238.
- Aparicio J. P.,A.F. Capurro, and C. Castillo-Chavez (2000) On the fall and rise of tuberculosis. Department of Biometrics, Cornell University, Technical Report Series, BU-1477-M.
- Aparicio J.P., A. Capurro, and C. Castillo-Chavez (2002) Frequency Dependent Risk of Infection and the Spread of Infectious Diseases. In: *Mathematical Approaches for Emerging and Reemerging Infectious Diseases : Models, Methods and Theory*, Edited by Castillo-Chavez, C. with S. Blower, P. van den Driessche, D. Kirschner, and A.A. Yakubu, Springer-Verlag (2002), pp. 341–350.
- Aparicio, J.P., A. Capurro, and C. Castillo-Chavez (2002) On the long-term dynamics and re-emergence of tuberculosis. In: *Mathematical Approaches for Emerging and Reemerging Infectious Diseases: Models, Methods and Theory*, Edited by Castillo-Chavez, C. with S. Blower, P. van den Driessche, D. Kirschner, and A.A. Yakubu, Springer-Verlag, (2002) pp. 351–360.
- Bellman, R.E. and K.L. Cooke (1963) Differential-Difference Equations, Academic Press, New York.
- Blower, S.M. & A.R. Mclean (1994) Prophylactic vaccines, risk behavior change, and the probability of eradicating HIV in San Francisco, *Science* **265**: 1451–1454.
- Boldin, B. and O. Diekmann (2008) Superinfections can induce evolutionarily stable coexistence of pathogens, *J. Math. Biol.* **56**: 635–672.
- Busenberg, S. and K.L. Cooke (1993) Vertically Transmitted Diseases: Models and Dynamics, Biomathematics **23**, Springer-Verlag, Berlin-Heidelberg-New York.

9. Castillo-Chavez, C., and Z. Feng (1997) To treat or not to treat; the case of tuberculosis, *J. Math. Biol.*, **35**: 629–656.
10. Castillo-Chavez, C., and Z. Feng (1998) Global stability of an age-structure model for TB and its applications to optimal vaccination strategies, *Math. Biosc.*, **151**: 135–154.
11. Castillo-Chavez, C. and Z. Feng (1998) Mathematical models for the disease dynamics of tuberculosis, *Advances in Mathematical Population Dynamics - Molecules, Cells, and Man* (O. Arino, D. Axelrod, M. Kimmel, (eds)): 629–656, World Scientific Press, Singapore.
12. Cen, X., Z. Feng and Y. Zhao (2014) Emerging disease dynamics in a model coupling within-host and between-host systems, *J. Theor. Biol.*, **361**: 141–151.
13. Conlan, A.J.K. and B.T. Grenfell (2007) Seasonality and the persistence and invasion of measles, *Proc. Roy. Soc. B.*, **274**: 1133–1141.
14. Dushoff, J., W. Huang, & C. Castillo-Chavez (1998) Backwards bifurcations and catastrophe in simple models of fatal diseases, *J. Math. Biol.* **36**: 227–248.
15. Feng, Z., C. Castillo-Chavez, and A. Capurro (2000) A model for TB with exogenous re-infection, *J. Theor. Biol.* **5**: 235–247.
16. Feng, Z., W. Huang and C. Castillo-Chavez (2001) On the role of variable latent periods in mathematical models for tuberculosis, *J. Dynamics and Differential Equations* **13**: 425–452.
17. Feng, Z., J. Velasco-Hernandez and B. Tapia-Santos (2013) A mathematical model for coupling within-host and between-host dynamics in an environmentally-driven infectious disease, *Math. Biosc.*, **241**: 49–55.
18. Feng, Z., D. Xu, & H. Zhao (2007) Epidemiological models with non-exponentially distributed disease stages and applications to disease control, *Bull. Math. biol.* **69**: 1511–1536.
19. Gilchrist, M.A. and D. Coombs (2006) Evolution of virulence: interdependence, constraints, and selection using nested models, *Theor. Pop. Biol.* **69**: 145–153.
20. Hadeler, K.P. & C. Castillo-Chavez (1995) A core group model for disease transmission, *Math Biosc.* **128**: 41–55.
21. Hadeler, K.P. and P. van den Driessche (1997) Backward bifurcation in epidemic control, *Math. Biosc.* **146**: 15–35.
22. Hethcote, H.W. (1978) An immunization model for a heterogeneous population, *Theor. Pop. Biol.* **14**: 338–349.
23. Hethcote, H.W. and D.W. Tudor (1980) Integral equation models for endemic infectious diseases, *J. Math. Biol.* **9**: 37–47.
24. Hethcote, H.W., H.W. Stech, and P. van den Driessche (1981) Periodicity and stability in epidemic models: A survey, in *Differential Equations and Applications in Ecology, Epidemics, and Population Problems*, S. Busenberg and K.L. Cooke (eds.), Academic Press, New York, pages 65–82.
25. Hopf, E. (1942) Abzweigung einer periodischen Lösungen von einer stationären Lösung eines Differentialsystems, Berlin Math-Phys. Sachsische Akademie der Wissenschaften, Leipzig, **94**: 1–22.
26. Hurwitz, A. (1895) Über die Bedingungen unter welchen eine Gleichung nur Wurzeln mit negativen reellen Teilen bezügt, *Math. Annalen* **46**: 273–284.
27. Keeling, M.J. and B.T. Grenfell (1997) Disease extinction and community size: Modeling the persistence of measles, *Science* **275**: 65–67.
28. Kermack, W.O. and A.G. McKendrick (1932) Contributions to the mathematical theory of epidemics, part. II, *Proc. Roy. Soc. London*, **138**: 55–83.
29. Kribs-Zaleta, C.M. and J.X. Velasco-Hernandez (2000) A simple vaccination model with multiple endemic states, *Math Biosc.* **164**: 183–201.
30. Lloyd, A. L. (2001) Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods, *Proc. Royal Soc. London. Series B: Biological Sciences* **268**: 985–993.
31. Lloyd, A. L. (2001b) Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and **60** (1), 59–71.
32. MacDonald, N. (1978) *Time lags in biological models* (Vol. 27), Heidelberg, Springer-Verlag.
33. McNeill, W.H. (1976) *Plagues and Peoples*, Doubleday, New York.

34. Nowak, M.A., S. Bonhoeffer, G. M. Shaw and R. M. May (1997) Anti-viral drug treatment: dynamics of resistance in free virus and infected cell populations, *J. Theor. Biol.* **184**: 203–217.
35. Perelson, A.S., D. E. Kirschner and R. De Boer (1993) Dynamics of HIV infection of CD4⁺ T cells, *Math. Biosc.* **114**: 81–125.
36. Perelson, A.S. and P. W. Nelson (2002) Modelling viral and immune system dynamics, *Nature Rev. Immunol.* **2**: 28–36.
37. Routh, E.J. (1877) *A Treatise on the Stability of a Given State of Motion: Particularly Steady Motion*, MacMillan.
38. U. S. Bureau of the Census (1975) *Historical statistics of the United States: colonial times to 1970*, Washington, D. C. Government Printing Office.
39. U.S. Bureau of the Census (1980) *Statistical Abstracts of the United States*, 101st edition.
40. U.S. Bureau of the Census (1991) *Statistical Abstracts of the United States*, 111th edition.
41. U.S. Bureau of the Census (1999) *Statistical Abstracts of the United States*, 119th edition.
42. Wang, W. (2006) Backward bifurcations of an epidemic model with treatment, *Math. Biosc.* **201**: 58–71.
43. Wang, W. and S. Ruan (2004) Bifurcations in an epidemic model with constant removal rate of the infectives, *J. Math. Anal. & Appl.* **291**: 775–793.
44. Wearing, H. J., P. Rohani, & M. J. Keeling (2005) Appropriate models for the management of infectious diseases, *PLoS Medicine* **2** (7), e174.

Chapter 4

Epidemic Models



In this chapter we describe models for epidemics, acting on a sufficiently rapid time scale that demographic effects, such as births, natural deaths, immigration into and emigration out of a population may be ignored. The prototype epidemic model is the simple Kermack–McKendrick model studied in Sect. 2.4.

We have established that the simple Kermack–McKendrick epidemic model

$$\begin{aligned} S' &= -\beta SI \\ I' &= \beta SI - \alpha I \end{aligned} \tag{4.1}$$

has the basic properties:

1. There is a basic reproduction number \mathcal{R}_0 such that if $\mathcal{R}_0 < 1$, the disease dies out while if $\mathcal{R}_0 > 1$ there is an epidemic.
2. The number of infectives always approaches zero and the number of susceptibles always approaches a positive limit S_∞ as $t \rightarrow \infty$.
3. There is a relation

$$\log \frac{S_0}{S_\infty} = \mathcal{R}_0 \left[1 - \frac{S_\infty}{N} \right] \tag{4.2}$$

between the reproduction number and the final size of the epidemic which is an equality if there are no disease deaths.

We will extend this model to models with more compartments and general distributions of stay in a compartment, for which these properties also hold, but we begin this chapter by presenting a more realistic description of the beginning of a disease outbreak than the compartmental approach. We will assume throughout this chapter that there are no disease deaths; the effects of including disease deaths in an epidemic model are the same as those described for the Kermack–McKendrick model with disease deaths in Sect. 2.5.

4.1 A Branching Process Disease Outbreak Model

The Kermack–McKendrick compartmental epidemic model assumes that the sizes of the compartments are large enough that the mixing of members is homogeneous, or at least that there is homogeneous mixing in each subgroup if the population is stratified by activity levels. However, at the beginning of a disease outbreak, there is a very small number of infective individuals and the transmission of infection is a stochastic event depending on the pattern of contacts between members of the population; a description should take this pattern into account.

Our approach will be to give a stochastic branching process description of the beginning of a disease outbreak to be applied so long as the number of infectives remains small, distinguishing a (minor) disease outbreak confined to this stage from a (major) epidemic, which occurs if the number of infectives begins to grow at an exponential rate. Once an epidemic has started we may switch to a deterministic compartmental model, arguing that in a major epidemic, contacts would tend to be more homogeneously distributed. Implicitly, we are thinking of an infinite population, and by a major epidemic, we mean a situation in which a nonzero fraction of the population is infected, and by a minor outbreak, we mean a situation in which the infected population may grow but remains a negligible fraction of the population.

There is an important difference between the behavior of branching process models and the behavior of models of Kermack–McKendrick type, namely that, as we shall see in this section, for a stochastic disease outbreak model if $\mathcal{R}_0 < 1$ the probability that the infection will die out is 1, but if $\mathcal{R}_0 > 1$ there is a positive probability that the infection will increase initially but will produce only a minor outbreak and will die out before triggering a major epidemic.

We describe the network of contacts between individuals by a graph with members of the population represented by vertices and with contacts between individuals represented by edges. The study of graphs originated with the abstract theory of Erdős and Rényi of the 1950s and 1960s [13–15]. It has become important in many areas of application, including social contacts and computer networks, as well as the spread of communicable diseases. We will think of networks as bi-directional, with disease transmission possible in either direction along an edge.

An edge is a contact between vertices that can transmit infection. The number of edges of a graph at a vertex is called the *degree* of the vertex. The degree distribution of a graph is $\{p_k\}$, where p_k is the fraction of vertices having degree k . The degree distribution is fundamental in the description of the spread of disease.

We think of a small number of infectives in a population of susceptibles large enough that in the initial stage we may neglect the decrease in the size of the susceptible population. Our development begins along the lines of that of [12] and then develops along the lines of [10, 32, 34]. We assume that the infectives make contacts independently of one another and let p_k denote the probability that the number of contacts by a randomly chosen individual is exactly k , with $\sum_{k=0}^{\infty} p_k = 1$. In other words, $\{p_k\}$ is the degree distribution of the vertices of

the graph corresponding to the population network. For the moment, we assume that every contact leads to an infection, but we will relax this assumption later.

It is convenient to define the *generating function*

$$G_0(z) = \sum_{k=0}^{\infty} p_k z^k.$$

Since $\sum_{k=0}^{\infty} p_k = 1$, this power series converges for $0 \leq z \leq 1$, and may be differentiated term by term. Thus

$$p_k = \frac{G_0^{(k)}(0)}{k!}, \quad k = 0, 1, 2, \dots.$$

It is easy to verify that the generating function has the properties

$$G_0(0) = p_0, \quad G_0(1) = 1, \quad G_0'(z) > 0, \quad G_0''(z) > 0.$$

The mean degree, which we denote by $\langle k \rangle$ or z_1 , is

$$\langle k \rangle = \sum_{k=1}^{\infty} kp_k = G_0'(1).$$

More generally, we define the moments

$$\langle k^j \rangle = \sum_{k=1}^{\infty} k^j p_k, \quad j = 1, 2, \dots \infty.$$

When a disease is introduced into a network, we think of it as starting at a vertex (patient zero) who transmits infection to every individual to whom this individual is connected, that is, along every edge of the graph from the vertex corresponding to this individual. We may think of this individual as being inside the population, as when a member of a population returns from travel after being infected, or as being outside the population, as when someone visits another population and brings back an infection. For transmission of disease after this initial contact, we need to use the *excess degree* of a vertex. If we follow an edge to a vertex, the excess degree of this vertex is one less than the degree. We use the excess degree because infection cannot be transmitted back along the edge whence it came. The probability of reaching a vertex of degree k , or excess degree $(k - 1)$, by following a random edge is proportional to k , and thus the probability that a vertex at the end of a random edge has excess degree $(k - 1)$ is a constant multiple of $k p_k$ with the constant chosen to make the sum over k of the probabilities equal to 1. Then the probability that a

vertex has excess degree $(k - 1)$ is

$$q_{k-1} = \frac{kp_k}{\langle k \rangle}.$$

This leads to a generating function $G_1(z)$ for the excess degree

$$G_1(z) = \sum_{k=1}^{\infty} q_{k-1} z^{k-1} = \sum_{k=1}^{\infty} \frac{kp_k}{\langle k \rangle} z^{k-1} = \frac{1}{\langle k \rangle} G'_0(z),$$

and the mean excess degree, which we denote by $\langle k_e \rangle$, is

$$\begin{aligned} \langle k_e \rangle &= \frac{1}{\langle k \rangle} \sum_{k=1}^{\infty} k(k-1)p_k \\ &= \frac{1}{\langle k \rangle} \sum_{k=1}^{\infty} k^2 p_k - \frac{1}{\langle k \rangle} \sum_{k=1}^{\infty} kp_k \\ &= \frac{\langle k^2 \rangle}{\langle k \rangle} - 1 = G'_1(1). \end{aligned}$$

We let $\mathcal{R}_0 = G'_1(1)$, the mean excess degree. This is the mean number of secondary cases caused by patient zero over the course of the outbreak and is the basic reproduction number as usually defined; the threshold for an epidemic is determined by \mathcal{R}_0 . The quantity $\langle k_e \rangle = G'_1(1)$ is sometimes written in the form

$$\langle k_e \rangle = G'_1(1) = \frac{z_2}{z_1},$$

where $z_2 = \sum_{k=1}^{\infty} k(k-1)p_k = \langle k^2 \rangle - \langle k \rangle$ is the mean number of second neighbors of a random vertex. We note that \mathcal{R}_0 depends not only on the mean degree but also on its variance.

Our next goal is to calculate the probability that the infection will die out and will not develop into a major epidemic, proceeding in two steps. First we find the probability that a secondary infected vertex (a vertex which has been infected by another vertex in the population) will not spark a major epidemic. The computations are performed on a branching approximation to an actual epidemic valid only at the beginning of the epidemic. In an actual epidemic some contacts are “wasted” on individuals who have already been infected.

Suppose that the secondary infected vertex has excess degree j . We let z_n denote the probability that this infection dies out within the next n generations. For the infection to die out in n generations each of the j secondary infections coming from the initial secondary infected vertex must die out in $(n-1)$ generations. The probability of this is z_{n-1} for each secondary infection, and the probability that

all secondary infections will die out in $(n - 1)$ generations is z_{n-1}^j . Now z_n is the sum over j of these probabilities, weighted by the probability q_j of j secondary infections. Thus

$$z_n = \sum_{j=0}^{\infty} q_j z_{n-1}^j = G_1(z_{n-1}), \quad z_0 = 0.$$

We assume $G_1(0) \geq 0$ so that there is an interval $0 < z \leq w$ on which $G_1(z) \geq z$; w is the smallest positive solution of $z = G_1(z)$ and $z_n < w$. If $G_1(0) = 0$, then $q_0 = 0$, and $p_1 = 0$. Thus the assumption $G_1(0) > 0$ is that the contact graph has no vertices with only one edge.

The sequence z_n is an increasing sequence and has a limit z_∞ , which is the probability that this infection will die out eventually. Then z_∞ is the limit as $n \rightarrow \infty$ of the solution of the difference equation

$$z_n = G_1(z_{n-1}), \quad z_0 = 0.$$

Thus z_∞ must be an equilibrium of this difference equation, that is, a solution of $z = G_1(z)$. Since w is the smallest positive solution of $z = G_1(z)$, $z \leq G_1(z) \leq G_1(w) = w$ for $0 \leq z \leq w$. It follows by induction that

$$z_n \leq w, \quad n = 0, 1, \dots, \infty.$$

From this we deduce that

$$z_\infty = w.$$

The equation $G_1(z) = z$ has a root $z = 1$ since $G_1(1) = 1$. Because the function $G_1(z) - z$ has a positive second derivative, its derivative $G'_1(z) - 1$ is increasing and can have at most one zero. This implies that the equation $G_1(z) = z$ has at most two roots in $0 \leq z \leq 1$. If $\mathcal{R}_0 < 1$ the function $G_1(z) - z$ has a negative first derivative

$$G'_1(z) - 1 \leq G'_1(1) - 1 = \mathcal{R}_0 - 1 < 0$$

and the equation $G_1(z) = z$ has only one root, namely $z = 1$. On the other hand, if $\mathcal{R}_0 > 1$ the function $G_1(z) - z$ is positive for $z = 0$ and negative near $z = 1$ since it is zero at $z = 1$ and its derivative is positive for $z < 1$ and z near 1. Thus in this case the equation $G_1(z) = z$ has a second root $z_\infty < 1$.

This root z_∞ is the probability that an infection transmitted along one of the edges at the initial secondary vertex will die out, and this probability is independent of the excess degree of the initial secondary vertex. It is also the probability that an infection originating outside the population, such as an infection brought into the population under study from outside, will die out.

Next, we calculate the probability that an infection originating at a primary infected vertex, such as an infection introduced by a visitor from outside the population under study, will die out. The probability that the disease outbreak will die out eventually is the sum over k of the probabilities that the initial infection in a vertex of degree k will die out, weighted by the degree distribution $\{p_k\}$ of the original infection, and this is

$$\sum_{k=0}^{\infty} p_k z_{\infty}^k = G_0(z_{\infty}).$$

To summarize this analysis, we see that if $\mathcal{R}_0 < 1$, the probability that the infection will die out is 1. On the other hand, if $\mathcal{R}_0 > 1$, there is a solution $z_{\infty} < 1$ of

$$G_1(z) = z$$

and there is a probability $1 - G_0(z_{\infty}) > 0$ that the infection will persist, and will lead to an epidemic. However, there is a positive probability $G_0(z_{\infty})$ that the infection will increase initially but will produce only a minor outbreak and will die out before triggering a major epidemic. This distinction between a minor outbreak and a major epidemic, and the result that if $\mathcal{R}_0 > 1$, there may be only a minor outbreak and not a major epidemic are aspects of stochastic models not reflected in deterministic models.

If contacts between members of the population are random, corresponding to the assumption of mass action in the transmission of disease, then the probabilities p_k are given by the *Poisson distribution*

$$p_k = \frac{e^{-c} c^k}{k!}$$

for some constant c [9, pp. 142–143]. The generating function for the Poisson distribution is $e^{c(z-1)}$. Then $G_1(z) = G_0(z)$, and $\mathcal{R}_0 = c$, so that

$$G_1(z) = G_0(z) = e^{\mathcal{R}_0(z-1)}.$$

The commonly observed situation that most infectives do not pass on infection but there are a few “super-spreading events” [35] corresponds to a probability distribution that is quite different from a Poisson distribution, and could give a quite different probability that an epidemic will occur. For example, if $\mathcal{R}_0 = 2.5$ the assumption of a Poisson distribution gives $z_{\infty} = 0.107$ and $G_0(z_{\infty}) = 0.107$, so that the probability of an epidemic is 0.893. The assumption that nine out of ten infectives do not transmit infection while the tenth transmits 25 infections gives

$$G_0(z) = (z^{25} + 9)/10, \quad G_1(z) = z^{24}, \quad z_{\infty} = 0, \quad G_0(z_{\infty}) = 0.9,$$

from which we see that the probability of an epidemic is 0.1. Another example, possibly more realistic, is to assume that a fraction $(1 - p)$ of the population follows a Poisson distribution with constant r while the remaining fraction p consists of super-spreaders each of whom makes L contacts. This would give a generating function

$$G_0(z) = (1 - p)e^{r(z-1)} + pz^L$$

$$G_1(z) = \frac{r(1 - p)e^{r(z-1)} + pLz^{L-1}}{r(1 - p) + pL},$$

and

$$\mathcal{R}_0 = \frac{r^2(1 - p) + pL(L - 1)}{r(1 - p) + pL}.$$

For example, if $r = 2.2$, $L = 10$, $p = 0.01$, numerical simulation gives

$$\mathcal{R}_0 = 2.5, \quad z_\infty = 0.146,$$

so that the probability of an epidemic is 0.849.

These examples demonstrate that the probability of a major epidemic depends strongly on the nature of the contact network. Simulations suggest that for a given value of the basic reproduction number the Poisson distribution is the one with the maximum probability of a major epidemic.

It has been observed that in many situations there is a small number of long range connections in the graph, allowing rapid spread of infection. There is a high degree of clustering (some vertices with many edges) and there are short path lengths. Such a situation may arise if a disease is spread to a distant location by an air traveler. This type of network is called a *small world* network. Long range connections in a network can increase the likelihood of an epidemic dramatically.

4.1.1 Transmissibility

Contacts do not necessarily transmit infection. For each contact between individual of whom one has been infected and the other is susceptible, there is a probability that infection will actually be transmitted. This probability depends on such factors as the closeness of the contact, the infectivity of the member who has been infected, and the susceptibility of the susceptible member. We assume that there is a mean probability T , called the *transmissibility*, of transmission of infection. The transmissibility depends on the rate of contacts, the probability that a contact will transmit infection, the duration time of the infection, and the susceptibility. Until now we have assumed that all contacts transmit infection, that is, that $T = 1$.

In this section, we will continue to assume that there is a network describing the contacts between members of the population whose degree distribution is given by the generating function $G_0(z)$, but we will assume in addition that there is a mean transmissibility T .

When disease begins in a network, it spreads to some of the vertices of the network. Edges that are infected during a disease outbreak are called *occupied*, and the size of the disease outbreak is the cluster of vertices connected to the initial vertex by a continuous chain of occupied edges.

The probability that exactly m infections are transmitted by an infective vertex of degree k is

$$\binom{k}{m} T^m (1 - T)^{k-m}.$$

We define $\Gamma_0(z, T)$ be the generating function for the distribution of the number of occupied edges attached to a randomly chosen vertex, which is the same as the distribution of the infections transmitted by a randomly chosen individual for any (fixed) transmissibility T . Then

$$\begin{aligned}\Gamma_0(z, T) &= \sum_{m=0}^{\infty} \left[\sum_{k=m}^{\infty} p_k \binom{k}{m} T^m (1 - T)^{(k-m)} \right] z^m \\ &= \sum_{k=0}^{\infty} p_k \left[\sum_{m=0}^k \binom{k}{m} (zT)^m (1 - T)^{(k-m)} \right] \\ &= \sum_{k=0}^{\infty} p_k [zT + (1 - T)]^k = G_0(1 + (z - 1)T).\end{aligned}\tag{4.3}$$

In this calculation we have used the binomial theorem to see that

$$\sum_{m=0}^k \binom{k}{m} (zT)^m (1 - T)^{(k-m)} = [zT + (1 - T)]^k.$$

Note that

$$\Gamma_0(0, T) = G_0(1 - T), \quad \Gamma_0(1, T) = G_0(1) = 1, \quad \Gamma'_0(z, T) = TG'_0(1 + (z - 1)T).$$

For secondary infections we need the generating function $\Gamma_1(z, T)$ for the distribution of occupied edges leaving a vertex reached by following a randomly chosen edge. This is obtained from the excess degree distribution in the same way,

$$\Gamma_1(z, T) = G_1(1 + (z - 1)T)$$

and

$$\Gamma_1(0, T) = G_1(1-T), \quad \Gamma_1(1, T) = G_1(1) = 1, \quad \Gamma'_1(z, T) = TG'_1(1+(z-1)T).$$

The basic reproduction number is now

$$\mathcal{R}_0 = \Gamma'_1(1, T) = TG'_1(1).$$

The calculation of the probability that the infection will die out and will not develop into a major epidemic follows the same lines as the argument for $T = 1$. The result is that if $\mathcal{R}_0 = TG'_1(1) < 1$, the probability that the infection will die out is 1. If $\mathcal{R}_0 > 1$ there is a solution $z_\infty(T) < 1$ of

$$\Gamma_1(z, T) = z,$$

and a probability $1 - \Gamma_0(z_\infty(T), T) > 0$ that the infection will persist, and will lead to an epidemic. However, there is a positive probability $\Gamma_1(z_\infty(T), T)$ that the infection will increase initially but will produce only a minor outbreak and will die out before triggering a major epidemic.

Another interpretation of the basic reproduction number is that there is a *critical transmissibility* T_c defined by

$$T_c G'_1(1) = 1.$$

In other words, the critical transmissibility is the transmissibility that makes the basic reproduction number equal to 1. If the mean transmissibility can be decreased below the critical transmissibility, then an epidemic can be prevented.

The measures used to try to control an epidemic may include contact interventions, that is, measures affecting the network such as avoidance of public gatherings and rearrangement of the patterns of interaction between caregivers and patients in a hospital, and transmission interventions such as careful hand washing or face masks to decrease the probability that a contact will lead to disease transmission.

4.2 Network and Compartmental Epidemic Models

Compartmental models for epidemics are not suitable for describing the beginning of a disease outbreak because they assume that all members of a population are equally likely to make contact with a very small number of infectives. Thus, as we have seen in the preceding section, stochastic branching process models are better descriptions of the beginning of an epidemic. They allow the possibility that even if a disease outbreak has a reproduction number greater than 1, it may be only a minor outbreak and may not develop into a major epidemic. One possible approach to a more realistic description of an epidemic would be to use a branching process model

initially and then make a transition to a compartmental model when the epidemic has become established and there are enough infectives that mass action mixing in the population is a reasonable approximation. Another approach would be to continue to use a network model throughout the course of the epidemic. In this section we shall indicate how a compartmental approach and a network approach are related. The development is taken from [30, 31, 43].

We assume that there is a known static *configuration model* (CM) network in which the probability that a node u has degree k_u is $P(k_u)$. We let $G_0(z)$ denote the probability generating function of the degree distribution,

$$G_0(z) = \sum_{k=0}^{\infty} p_k z^k,$$

with mean degree $\langle k \rangle = G'_0(1)$.

The per-edge contact rate from an infected node is assumed to be β , and it is assumed that infected nodes recover at a rate α . We use an edge-based compartmental model because the probability that a random neighbor is infected is not necessarily the same as the probability that a random individual is infected. We let $S(t)$ denote the fraction of nodes that are susceptible at time t , $I(t)$ the fraction of nodes that are infective at time t , and $R(t)$ the fraction of nodes that are recovered at time t . It is easy to write an equation for R' , the rate at which infectives recover. If we know $S(t)$, we can find $I(t)$, because a decrease in S gives a corresponding increase in I . Since

$$S(t) + I(t) + R(t) = 1,$$

we need only find the probability that a randomly selected node is susceptible.

We assume that the hazard of infection for a susceptible node u is proportional to the degree k_u of the node. Each contact is represented by an edge of the network joining u to a neighboring node. We let φ_I denote the probability that this neighbor is infective. Then the per-edge hazard of infection is

$$\lambda_E = \beta \varphi_I.$$

Assuming that edges are independent, u 's hazard of infection at time t is

$$\lambda_u(t) = k_u \lambda_E(t) = k_u \beta \varphi_I(t).$$

Consider a randomly selected node u and let $\theta(t)$ be the probability that a random neighbor has not transmitted infection to u at time t . Then the probability that u is susceptible is θ^{k_u} . Averaging over all nodes, we see that the probability that a random node u is susceptible is

$$S(t) = \sum_{k=0}^{\infty} P(k) [\theta(t)]^k = G_0(\theta(t)). \quad (4.4)$$

We break θ into three parts,

$$\theta = \varphi_S + \varphi_I + \varphi_R,$$

with φ_S the probability that a random neighbor v of u is susceptible, φ_I the probability that a random neighbor v of u is infective but has not transmitted infection to u , and φ_R the probability that a random neighbor v has recovered without transmitting infection to u . Then the probability that v has transmitted infection to u is $1 - \theta$.

Since infected neighbors recover at rate α , the flux from φ_I to φ_R is $\alpha\varphi_I$. Thus

$$\varphi'_R = \alpha\varphi_I.$$

It is easy to see from this that

$$R' = \alpha I. \quad (4.5)$$

Since edges from infected neighbors transmit infection at rate β , the flux from φ_I to $(1 - \theta)$ is $\beta\varphi_I$. Thus

$$\theta' = -\beta\varphi_I. \quad (4.6)$$

To obtain φ'_I we need the flux into and out of the φ_I compartment. The incoming flux from φ_S results from infection of the neighbor. The outgoing flux to φ_R corresponds to recovery of the neighbor without having transmitted infection, and the outgoing flux to $(1 - \theta)$ corresponds to transmission without recovery. The total outgoing flux is $(\alpha + \beta)\varphi_I$.

To determine the flux from φ_S to φ_I , we need the rate at which a neighbor changes from susceptible to infective. Consider a random neighbor v of u ; the probability that v has degree k is $kp(k)/\langle k \rangle$. Since there are $(k - 1)$ neighbors of v that could have infected v , the probability that v is susceptible is θ^{k-1} . Averaging over all k , we see that the probability that a random neighbor v of u is susceptible is

$$\varphi_S = \sum_{k=0}^{\infty} \frac{kp(k)}{\langle k \rangle} \theta^{k-1} = \frac{G'_0(\theta)}{G'_0(1)}. \quad (4.7)$$

To calculate φ_R , we note that the flux from φ_I to φ_R and the flux from φ_I to $(1 - \theta)$ are proportional with proportionality constant α/β . Since both φ_R and $(1 - \theta)$ start at zero,

$$\varphi_R = \frac{\alpha}{\beta}(1 - \theta). \quad (4.8)$$

Now, using (4.6)–(4.8), and

$$\varphi_I = \theta - \varphi_S - \varphi_R,$$

we obtain

$$\theta' = -\beta\varphi_I = -\beta\theta + \beta\varphi_S + \beta\varphi_R = -\beta\theta + \beta \frac{G'_0(\theta)}{G'_0(1)} + \alpha(1 - \theta). \quad (4.9)$$

We now have a dynamic model consisting of Eqs. (4.4), (4.5), (4.9), and $S + I + R = 1$. We wish to show a relationship between this set of equations and the simple Kermack–McKendrick compartmental model (4.1). In order to accomplish this, we need only show under what conditions we would have $S' = -\beta SI$.

Differentiating (4.4) and using (4.6), we obtain

$$S' = G'_0(\theta)\theta' = -G'_0(\theta)\beta\varphi_I.$$

Consider a large population with N members, each making $C \leq N - 1$ contacts, so that

$$S = \theta^C, \quad G'_0(\theta) = \frac{CS}{\theta} S'(\theta),$$

and

$$S' = -\beta CS \frac{\varphi_I}{\theta}.$$

We now let $C \rightarrow \infty$ (which implies $N \rightarrow \infty$) in such a way that

$$\hat{\beta} = \beta C$$

remains constant. Then

$$S' = -\hat{\beta} \frac{\varphi_I}{\theta}.$$

We will now show that

$$\frac{\varphi_I}{\theta} \approx 1,$$

and this will yield the desired approximation

$$S' = -\hat{\beta} SI. \quad (4.10)$$

The probability that an edge to a randomly chosen node has not transmitted infection is θ (assuming that the given target node cannot transmit infection), and

the probability that in addition it is connected to an infected node is φ_I . Because $\hat{\beta} = \beta C$ is constant and therefore bounded as C grows, only a fraction no greater than a constant multiple of I/C of edges to the target node may have transmitted infection from a node that is still infected. For large values of C , φ_I is approximately I . Similarly, θ is approximately 1 as $C \rightarrow \infty$. Thus $\varphi_I/\theta \approx I$ as $C \rightarrow \infty$. This gives the desired approximate equation for S . The result remains valid if all degrees are close to the average degree as the average degree grows.

The edge-based compartmental modeling approach that we have used can be generalized in several ways. For example, heterogeneity of mixing can be included. In general, one would expect that early infections would be in individuals having more contacts, and thus that an epidemic would develop more rapidly than a mass action compartmental model would predict. When contact duration is significant, as would be the case in sexually transmitted diseases, an individual with a contact would play no further role in disease transmission until a new contact is made, and this can be incorporated in a network model.

The network approach to disease modeling is a rapidly developing field of study, and there will undoubtedly be fundamental developments in our understanding of the modeling of disease transmission. Some useful references are [6, 27–29, 32–34, 39].

In the remainder of this chapter, we assume that we are in an epidemic situation following a disease outbreak that has been modeled initially by a branching process. Thus we return to the study of compartmental models. We will study models having more compartmental structure than the simple Kermack–McKendrick *SIR* epidemic model (4.1).

4.3 More Complicated Epidemic Models

4.3.1 Exposed Periods

In many infectious diseases there is an exposed period after the transmission of infection from susceptibles to potentially infective members but before these potential infectives develop symptoms and can transmit infection. To incorporate an exposed period with mean exposed period $1/\kappa$ we add an exposed class E and use compartments S, E, I, R and total population size $N = S + E + I + R$ to give a generalization of the epidemic model (4.1)

$$\begin{aligned} S' &= -\beta SI \\ E' &= \beta SI - \kappa E \\ I' &= \kappa E - \alpha I. \end{aligned} \tag{4.11}$$

A flow chart is shown in Fig. 4.1.

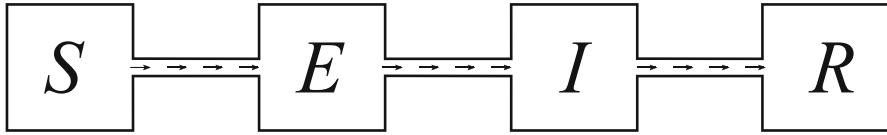


Fig. 4.1 Flow chart for the SEIR model (4.11)

In this model there are two infected compartments, namely E and I with different infectivities. The parameter β represents the rate of effective contacts. The infectivity of a member of E is zero, and therefore contacts between members of S and members of E do not produce new infections. The analysis of this model is similar to the analysis of (4.1), but with I replaced by $E + I$. That is, instead of using the number of infectives as one of the variables we use the total number of infected members, whether or not they are capable of transmitting infection.

For the model (4.11) it is no longer possible to distinguish whether there is an epidemic or not by determining whether the number of infectives grows or decreases initially. A more general characterization is given by whether the equilibrium with all members of the population susceptible is unstable (epidemic) or asymptotically stable (no epidemic). We will use a situation in which the disease-free equilibrium is unstable as our definition of an epidemic.

For the model (4.11), the matrix of the linearization at the equilibrium $S = N$, $E = 0$, $I = 0$ is

$$\begin{bmatrix} 0 & 0 & -\beta N \\ 0 & -\kappa & \beta N \\ 0 & \kappa & -\alpha \end{bmatrix}.$$

The eigenvalues of this matrix are zero and the eigenvalues of the 2×2 matrix are

$$\begin{bmatrix} -\kappa & \beta N \\ \kappa & -\alpha \end{bmatrix}.$$

The zero eigenvalue corresponds to movement along the line of equilibria. The other two eigenvalues have negative real part, corresponding to stability of the equilibrium with respect to solutions starting near the equilibrium but not on the line of equilibria, and failure of an epidemic to develop (since the trace of the matrix is negative) if and only if the determinant of the matrix is positive. The determinant is $\kappa(\alpha - \beta N)$, and this is positive if and only if $\mathcal{R}_0 < 1$.

If there is an epidemic, the initial exponential growth rate is the largest eigenvalue of the matrix, and this is the largest root of the quadratic characteristic equation

$$\lambda^2 + (\alpha + \kappa)\lambda - \kappa(\beta N - \alpha) = 0.$$

Since the constant term in this equation is negative if $\mathcal{R}_0 > 1$, there is one negative and one positive root, and the positive root is

$$\lambda = \frac{-(\alpha + \kappa) + \sqrt{(\alpha - \kappa)^2 + 4\kappa\beta N}}{2}.$$

This is the initial exponential growth rate; note that it is not the same as the initial exponential growth rate for the *SIR* model (4.1). The effect of an exposed period is to decrease the initial exponential growth rate.

4.3.2 A Treatment Model

One form of treatment that is possible for some diseases is vaccination to protect against infection before the beginning of an epidemic. For example, this approach is commonly used for protection against annual influenza outbreaks. A simple way to model this would be to reduce the total population size by the fraction of the population protected against infection.

In reality such inoculations are only partly effective, decreasing the rate of infection and also decreasing infectivity if a vaccinated person does become infected. This may be modeled by dividing the population into two groups with different model parameters which would require some assumptions about the mixing between the two groups. This is not difficult but we will not explore this direction until Chap. 5 on heterogeneous mixing.

If there is a treatment for infection once a person has been infected, this may be modeled by supposing that a fraction γ per unit time of infectives is selected for treatment, and that treatment reduces infectivity by a fraction δ . Thus effective contacts between members of S and members of T would produce only δ new infections per contact. Suppose that the rate of removal from the treated class is η . This leads to the *SITR* model, where T is the treatment class, given by

$$\begin{aligned} S' &= -\beta S[I + \delta T] \\ I' &= \beta S[I + \delta T] - (\alpha + \gamma)I \\ T' &= \gamma I - \eta T. \end{aligned} \tag{4.12}$$

A flow chart is shown in Fig. 4.2.

It is reasonable to assume that treatment does not slow recovery, so that $\eta \geq \alpha$. Also, since the total time in the infectious and treatment stages should not be greater than the mean infective period, we should expect that

$$\frac{1}{\alpha} \geq \frac{1}{\gamma} + \frac{1}{\eta}.$$

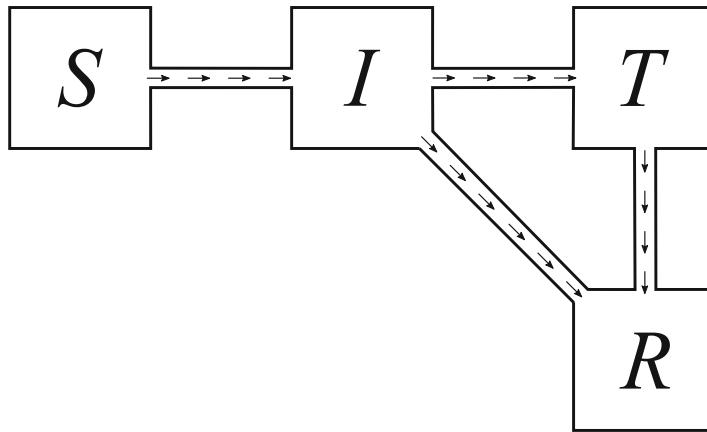


Fig. 4.2 Flow chart for the $SITR$ model (4.12)

It is not difficult to prove, much as was done for the model (4.1) that

$$S_\infty = \lim_{t \rightarrow \infty} S(t) > 0, \quad \lim_{t \rightarrow \infty} I(t) = \lim_{t \rightarrow \infty} T(t) = 0.$$

In order to calculate the basic reproduction number, we may argue that an infective in a totally susceptible population causes βN new infections in unit time, and the mean time spent in the infective compartment is $1/(\alpha + \gamma)$. In addition, a fraction $\gamma/(\alpha + \gamma)$ of infectives are treated. While in the treatment stage the number of new infections caused in unit time is $\delta\beta N$, and the mean time in the treatment class is $1/\eta$. Thus \mathcal{R}_0 is

$$\mathcal{R}_0 = \frac{\beta N}{\alpha + \gamma} + \frac{\gamma}{\alpha + \gamma} \frac{\delta\beta N}{\eta}. \quad (4.13)$$

It is also possible to establish the final size relation (4.2) by means very similar to those used for the simple model (4.1). We integrate the first equation of (4.12) to obtain

$$\log \frac{S_0}{S_\infty} = \int_0^\infty \beta[I(t) + \delta T(t)]dt.$$

Integration of the third equation of (4.12) gives

$$\gamma \int_0^\infty I(t)dt = \eta \int_0^\infty T(t)dt.$$

Integration of the sum of the first two equations of (4.12) gives

$$N - S_\infty = (\alpha + \gamma) \int_0^\infty I(t) dt.$$

Combination of these three equations and (4.13) gives (4.2).

For the model (4.12), the matrix of the linearization at the equilibrium $S = N$, $I = 0$, $T = 0$ is

$$\begin{bmatrix} 0 & -\beta N & -\delta\beta N \\ 0 & \beta N - (\alpha + \gamma) & \delta\beta N \\ 0 & \gamma & -\eta \end{bmatrix}.$$

The eigenvalues of this matrix are zero and the eigenvalues of the 2×2 matrix are

$$\begin{bmatrix} \beta N - (\alpha + \gamma) & \delta\beta N \\ \gamma & -\eta \end{bmatrix}.$$

The nonzero eigenvalues have negative real part, corresponding to stability of the equilibrium and failure of an epidemic to develop if and only if the determinant of the matrix is positive and the trace is negative. The determinant is

$$-\beta N(\eta + \delta\gamma) + \eta(\alpha + \gamma),$$

which is positive if and only if $\mathcal{R}_0 < 1$, and this condition implies that the trace is negative.

If there is an epidemic, the initial exponential growth rate is the largest eigenvalue of the matrix, and this is the largest root of the quadratic characteristic equation

$$\lambda^2 + [\beta N - (\alpha + \gamma + \eta)]\lambda + \beta N(\eta + \delta\gamma) - \eta(\alpha + \gamma) = 0.$$

Since the constant term in this equation is negative if $\mathcal{R}_0 > 1$, there is one negative and one positive root, and the positive root is the initial exponential growth rate. Notice that this rate depends on the treatment rate.

4.3.3 An Influenza Model

In some diseases, such as influenza, at the end of a stage individuals may proceed to one of two stages. There is a latent period after which a fraction p of latent individuals L proceeds to an infective stage I , while the remaining fraction $(1 - p)$ proceeds to an asymptomatic stage A , with infectivity reduced by a factor δ and a

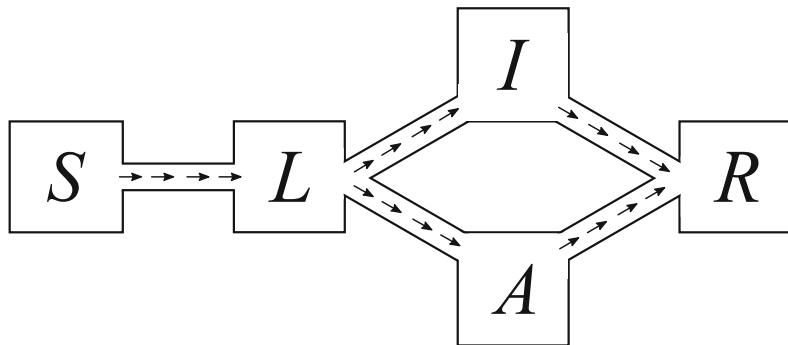


Fig. 4.3 Flowchart for the influenza model (4.14)

different period $1/\eta$. The influenza model of [3, 5] is

$$\begin{aligned} S' &= -\beta S[I + \delta A] \\ L' &= \beta S[I + \delta A] - \kappa L \\ I' &= p\kappa L - \alpha I \\ A' &= (1-p)\kappa L - \eta A \end{aligned} \tag{4.14}$$

and

$$\mathcal{R}_0 = \beta N \left[\frac{p}{\alpha} + \frac{\delta(1-p)}{\eta} \right].$$

A flow chart is shown in Fig. 4.3.

The same approach used in earlier examples leads to the same final size relation (4.2).

The model (4.14) is an example of a differential infectivity model. In such models, also used in the study of HIV/AIDS [23], individuals enter a specific group when they become infected and stay in that group over the course of the infection. Different groups may have different parameter values. For example, for influenza infective and asymptomatic members may have different infectivities and different periods of stay in the respective stages.

4.3.4 A Quarantine–Isolation Model

For an outbreak of a new disease, where no vaccine is available, isolation of diagnosed infectives and quarantine of people who are suspected of having been infected (usually by tracing of contacts of diagnosed infectives) are the only control measures available. We formulate a model to describe the course of an epidemic,

originally introduced for modeling the SARS epidemic of 2002–2003 [19], when control measures are begun under the assumptions:

1. Exposed members may be infective with infectivity reduced by a factor ε_E , $0 \leq \varepsilon_E < 1$.
2. Exposed members who are not isolated become infective at rate κ_E .
3. We introduce a class Q of quarantined members and a class J of isolated (hospitalized) members and exposed members are quarantined at a proportional rate γ_Q in unit time (in practice, a quarantine will also be applied to many susceptibles, but we ignore this in the model). Quarantine is not perfect, but reduces the contact rate by a factor ε_Q . The effect of this assumption is that some susceptibles make fewer contacts than the model assumes.
4. Infectives are diagnosed at a proportional rate γ_J per unit time and isolated. Isolation is imperfect, and there may be transmission of disease by isolated members, with an infectivity factor of ε_J .
5. Quarantined members are monitored and when they develop symptoms at rate κ_Q they are isolated immediately.
6. Infectives leave the infective class at rate α_I and isolated members leave the isolated class at rate α_J .

These assumptions lead to the $SEQIJR$ model [19]

$$\begin{aligned} S' &= -\beta S[\varepsilon_E E + \varepsilon_E \varepsilon_Q Q + I + \varepsilon_J J] \\ E' &= \beta S[\varepsilon_E E + \varepsilon_E \varepsilon_Q Q + I + \varepsilon_J J] - (\kappa_E + \gamma_Q)E \\ Q' &= \gamma_Q E - \kappa_J Q \\ I' &= \kappa_E E - (\alpha_I + \gamma_J)I \\ J' &= \kappa_Q Q + \gamma_J I - \alpha_J J. \end{aligned} \tag{4.15}$$

The model before control measures are begun is the special case

$$\gamma_Q = \gamma_J = \kappa_Q = \alpha_J = 0, \quad Q = J = 0$$

of (4.15). It is the same as (4.11).

A flow chart is shown in Fig. 4.4.

We define the *control reproduction number* \mathcal{R}_c to be the number of secondary infections caused by a single infective in a population consisting only of susceptibles with the control measures in place. It is analogous to the basic reproduction number but instead of describing the very beginning of the disease outbreak it describes the beginning of the recognition of the epidemic. The basic reproduction number is the value of the control reproduction number with

$$\gamma_Q = \gamma_J = \kappa_Q = \alpha_J = 0.$$

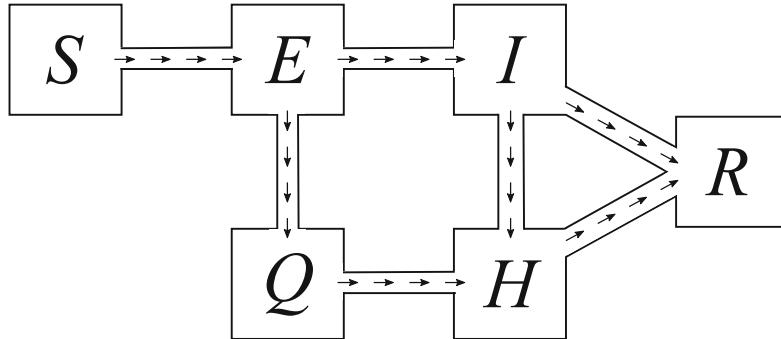


Fig. 4.4 Flowchart for the $SEQIJR$ model (4.15)

We have already calculated \mathcal{R}_0 for (4.11) and we may calculate \mathcal{R}_c in the same way but using the full model with quarantined and isolated classes. We obtain

$$\mathcal{R}_c = \frac{\beta N \varepsilon_E}{D_1} + \frac{\beta N \kappa_E}{D_1 D_2} + \frac{\varepsilon_Q \beta N \varepsilon_E \gamma_Q}{D_1 \kappa_Q} + \frac{\varepsilon_J \beta N \kappa_E \gamma_J}{\alpha_J D_1 D_2} + \frac{\varepsilon_J \beta N \gamma_Q}{\alpha_J D_1},$$

where $D_1 = \gamma_Q + \kappa_E$ and $D_2 = \gamma_J + \alpha_I$.

Each term of \mathcal{R}_c has an epidemiological interpretation. The mean duration in E is $1/D_1$ with contact rate $\beta \varepsilon_E$, giving a contribution to \mathcal{R}_c of $\beta N \varepsilon_E / D_1$. A fraction κ_E / D_1 goes from E to I , with contact rate β and mean duration $1/D_2$, giving a contribution of $\beta N \kappa_E / D_1 D_2$. A fraction γ_Q / D_1 goes from E to Q , with contact rate $\varepsilon_E \varepsilon_Q \beta$ and mean duration $1/\kappa_Q$, giving a contribution of $\varepsilon_E \varepsilon_Q \beta N \gamma_Q / D_1 \kappa_Q$. A fraction $\kappa_E \gamma_J / D_1 D_2$ goes from E to I to J , with a contact rate of $\varepsilon_J \beta$ and a mean duration of $1/\alpha_J$, giving a contribution of $\varepsilon_J \beta N \kappa_E \gamma_J / \alpha_J D_1 D_2$. Finally, a fraction γ_Q / D_1 goes from E to Q to J with a contact rate of $\varepsilon_J \beta$ and a mean duration of $1/\alpha_J$ giving a contribution of $\varepsilon_J \beta N \gamma_Q / D_1 \alpha_J$. The sum of these individual contributions gives \mathcal{R}_c .

In the model (4.15) the parameters γ_Q and γ_J are *control* parameters which may be chosen in the attempt to manage the epidemic. The parameters ε_Q and ε_J depend on the strictness of the quarantine and isolation processes and are thus also control measures in a sense. The other parameters of the model are specific to the disease being studied. While they are not variable, their measurements are subject to experimental error.

The linearization of (4.15) at the disease-free equilibrium $(N, 0, 0, 0, 0)$ has matrix

$$\begin{bmatrix} \varepsilon_E \beta N - (\kappa_E + \gamma_Q) & \varepsilon_E \varepsilon_Q \beta & \beta N & \varepsilon_J \beta N \\ \gamma_Q & -\kappa_Q & 0 & 0 \\ \kappa_E & 0 & -(\alpha_I + \gamma_J) & 0 \\ 0 & \kappa_Q & \gamma_J & -\alpha_J \end{bmatrix}.$$

The corresponding characteristic equation is a fourth degree polynomial equation whose leading coefficient is 1 and whose constant term is a positive constant multiple of $1 - \mathcal{R}_c$, thus positive if $\mathcal{R}_c < 1$ and negative if $\mathcal{R}_c > 1$. If $\mathcal{R}_c > 1$ there is a positive eigenvalue, corresponding to an initial exponential growth rate of solutions of (4.15). If $\mathcal{R}_c < 1$ it is possible to show that all eigenvalues of the coefficient matrix have negative real part, and thus solutions of (4.15) die out exponentially [42].

In order to show that analogues of the relation (4.2) and $S_\infty > 0$ derived for the model (4.1) are valid for the management model (4.15), we begin by integrating the equations for $S + E$, Q , I , J , of (4.15) with respect to t from $t = 0$ to $t = \infty$, using the initial conditions

$$S(0) + E(0) = N(0) = N, \quad Q(0) = I(0) = J(0) = 0.$$

We continue by integrating the equation for S and then an argument similar to the one used for (4.1) but technically more complicated may be used to show that $S_\infty > 0$ for the treatment model (4.15) and also to establish the final size relation

$$\log \frac{S_0}{S_\infty} = \mathcal{R}_c \left[1 - \frac{S_\infty}{N} \right].$$

Thus the asymptotic behavior of the management model (4.15) is the same as that of the simpler model (4.1).

In the various compartmental models that we have studied, there are significant common features. This suggests that compartmental models can be put into a more general framework. In fact, this general framework is the age of infection epidemic model originally introduced by Kermack and McKendrick in [24]. We will explore this generalization in Sect. 4.5.

4.4 An SIR Model with a General Infectious Period Distribution

In the simple model (4.1) studied in Sect. 2.4 we have assumed that the infective period is exponentially distributed. Now let us consider an *SIR* epidemic model in a population of constant size N with mass action incidence in which $P(s)$ is the fraction of individuals who are still infective at time s after having become infected. The model is

$$\begin{aligned} S' &= -\beta S(t)I(t) \\ I(t) &= I_0(t) + \int_0^t [-S'(t-s)]P(s)ds. \end{aligned} \tag{4.16}$$

Here, $I_0(t)$ is the number of individuals infective initially at $t = 0$ who are still infective at time t . Then

$$I_0(t) \leq (N - S_0)P(t),$$

because if all initial infectives were newly infected we would have equality in this relation, and if some initial infectives had been infected before the starting time $t = 0$, they would recover earlier.

We assume that $P(s)$ is a non-negative, non-increasing function with $P(0) = 1$. We assume also that the mean infective period $\int_0^\infty P(s)ds$ is finite. Since a single infective causes βN new infections in unit time and $\int_0^\infty P(s)ds$ is the mean infective period, it is easy to calculate

$$\mathcal{R}_0 = \beta N \int_0^\infty P(s) ds.$$

In this model it would be possible to assume that not all contacts by infectives with susceptibles lead to new infections. This could be achieved by incorporating a reduction factor δ in the equation for S and in the distribution P . Since S is a non-negative decreasing function, it follows as for (4.1) that $S(t)$ decreases to a limit S_∞ as $t \rightarrow \infty$, but we must proceed differently to show that $I(t) \rightarrow 0$. This will follow if we can prove that $\int_0^t I(s) ds$ is bounded as $t \rightarrow \infty$. We have

$$\begin{aligned} \int_0^t I(s) ds &= \int_0^t I_0(s) ds + \int_0^t \int_0^s [-S'(s-u)]P(u)duds \\ &\leq (N - S_0) \int_0^t P(s)ds + \int_0^t \int_u^t [-S'(s-u)]ds P(u)du \\ &\leq (N - S_0) \int_0^t P(s)ds + \int_0^t [S_0 - S(t-u)]P(s)ds \\ &\leq N \int_0^t P(s)ds. \end{aligned}$$

Since $\int_0^\infty P(s)ds$ is assumed to be finite, it follows that $\int_0^t I(s) ds$ is bounded, and thence that $I(t) \rightarrow 0$.

Now integration of the first equation in (4.16) from 0 to ∞ gives

$$\log \frac{S_0}{S_\infty} = \beta \int_0^\infty I(s)ds < \infty,$$

and this shows that $S_\infty > 0$.

If all initially infected individuals are newly infected, so that $I_0(t) = (N - S_0)P(t)$, integration of the second equation of (4.16) gives

$$\begin{aligned} \int_0^\infty I(s) ds &= \int_0^\infty I_0(s) ds + \int_0^\infty \int_0^s [-S'(s-u)]P(u)duds \\ &= (N - S_0) \int_0^\infty P(u)du + \int_0^\infty \int_u^\infty [-S'(s-u)]ds P(u)du \\ &= (N - S_0) \int_0^\infty P(u)du + \int_0^\infty [S_0 - S_\infty]P(u)du \\ &= (N - S_\infty) \int_0^\infty P(u)du \\ &= \mathcal{R}_0 \left[1 - \frac{S_\infty}{N} \right], \end{aligned}$$

and this is the final size relation, identical to (4.2). If there are individuals who were infected before time $t = 0$, a positive term

$$(N - S_0) \int_0^\infty P(t)dt - \int_0^\infty I_0(t)dt$$

must be subtracted from the right side of this equation.

The generalization to arbitrary infective periods in this section is a component of the age of infection epidemic model of [24], which also incorporates general compartmental structures. The examples of this section and the previous section are all special cases of the age of infection model.

4.5 The Age of Infection Epidemic Model

The general epidemic model described by Kermack and McKendrick [24] included a dependence of infectivity on the time since becoming infected (age of infection). We let $S(t)$ denote the number of susceptibles at time t and let $\varphi(t)$ be the total infectivity at time t , defined as the sum of products of the number of infected members with each infection age and the mean infectivity for that infection age. We assume that on average, members of the population make a constant number βN of contacts in unit time. We let $B(s)$ be the fraction of infected members remaining infected at infection age s and let $\pi(s)$ with $0 \leq \pi(s) \leq 1$ be the mean infectivity (if infected) at infection age s . Then we let

$$A(s) = \pi(s)B(s),$$

the mean infectivity of members of the population with infection age s including those who are no longer infectious. Here, the factor π must include the probability that a contact will produce a new infection. We assume that there are no disease deaths, so that the total population size is a constant N .

Since $-S'(t-s)$ is the number of new infections at time $(t-s)$ and $A(s)$ is the remaining infectivity of these new infections at time t (infection age (s)), the total infectivity at time t is

$$\varphi(t) = \int_0^\infty [-S'(t-s)]A(s)ds.$$

It is assumed that the disease outbreak begins at time $t = 0$, so that $S(u) = N$ for $u < 0$. There may be a discontinuity in $S(u)$ at $u = 0$ corresponding to an initial infective distribution.

We may write the age of infection epidemic model as

$$\begin{aligned} S' &= -\beta S\varphi \\ \varphi(t) &= \int_0^\infty [-S'(t-s)]A(s)ds = \int_0^\infty \beta S(t-s)\varphi(t-s)A(s)ds. \end{aligned} \quad (4.17)$$

The parameter β continues to represent the rate of effective contacts, and the transmission probability is contained in φ . The basic reproduction number is

$$\mathcal{R}_0 = \beta N \int_0^\infty A(s)ds. \quad (4.18)$$

We may write the model in a single equation

$$S'(t) = \beta S(t) \int_0^\infty A(s)S'(t-s)ds.$$

There is an invasion criterion. Initially, in a fully susceptible population when $S(t)$ is close to $S_0 = N$, we may replace $S(t)$ by its initial value N in this single equation, giving a linear equation. The condition that this linear equation have a solution $S(t) = Ne^{rt}$ is

$$1 = \beta N \int_0^\infty A(s)e^{-rs}ds. \quad (4.19)$$

The result that the initial exponential growth rate in an epidemic is given by the solution of (4.19) was given in [20, 36] and later in [44]. It allows us to estimate the basic reproduction number for an epidemic, provided we know not only the mean infective period but also the distribution of infective periods. This result is valid for age of infection epidemic models, and thus is applicable to compartmental models

that can be interpreted in an age of infection context, that is, provided we are able to calculate the infectivity function $A(s)$.

Combination of the relations (4.18) and (4.19) gives a relation between the initial exponential growth rate r and the basic reproduction number \mathcal{R}_0 , namely

$$\mathcal{R}_0 = \frac{\int_0^\infty A(s)ds}{\int_0^\infty e^{-rs} A(s)ds}.$$

Then $\mathcal{R}_0 > 1$ if and only if $r > 0$.

This relation provides a means to estimate the basic reproduction number from measurements of the initial exponential growth rate provided the infectivity distribution is known. We define an epidemic as a situation in which for the model we have $r > 0$, so that initially the solution grows exponentially.

Division of the model equation (4.17) by $S(t)$ and integration with respect to t from 0 to ∞ gives, with an interchange of order of integration gives

$$\log \frac{S_0}{S_\infty} = \int_0^\infty [S(-s) - S_\infty]A(s)ds.$$

Since we are assuming $S(-s) = N$ if $s > 0$, we have

$$\log \frac{S_0}{S_\infty} = \int_0^\infty [N - S_\infty]A(s)ds,$$

and then we obtain the final size relation

$$\log \frac{S_0}{S_\infty} = \mathcal{R}_0 \left[1 - \frac{S_\infty}{N} \right].$$

The final size relation is sometimes presented in the form

$$\log \frac{S_0}{S_\infty} = \mathcal{R}_0 \left[1 - \frac{S_\infty}{S_0} \right], \quad (4.20)$$

see for example [4, 12, 22], since normally $S(0) \approx N$. This form would also represent the final size relation for an epidemic started by someone outside the population under study, so that $S_0 = N$, $I_0 = 0$.

4.5.1 A General SEIR Model

We consider an *SEIR* model with general distributions of stay in both the exposed and infective period. The model is related to the model (3.31), which also includes births and natural deaths. Suppose the fraction of exposed individuals who are still

in the exposed class s time units after being exposed is $P_E(s)$ and the fraction of individuals who are still in the infectious class s time units after entering the infectious class is $P_I(s)$, with $P_E(s)$, $P_I(s)$ non-negative, non-increasing functions such that

$$P_E(0) = P_I(0) = 1, \quad \int_0^\infty P_E(s)ds < \infty, \quad \int_0^\infty P_I(s)ds < \infty.$$

Then P_E and P_I represent survival probabilities in the classes E and I , respectively. The probability density function for E , which will appear in the model, is

$$q(\tau) = -P'_E(\tau).$$

We assume that E_0 newly exposed members enter the exposed class at time $t = 0$. Then

$$\begin{aligned} S' &= -\beta SI \\ E(t) &= E_0 P_E(t) + \int_0^t [-S'(u)] P_E(t-u) du. \end{aligned}$$

In this equation we are assuming that all effective contacts transmit infection. Differentiation of the equation for $E(t)$ gives

$$E'(t) = E_0 P'_E(t) - S'(t) + \int_0^t [-S'(u)] P'_E(t-u) du,$$

and this shows that the input to the infectious stage at time t is

$$-E_0 P'_E(t) - \int_0^t [-S'(u)] P'_E(t-u) du.$$

If we assume that $S'(u) = 0$ for $u < 0$ and that $S'(u)$ has a jump of $-E_0$ at $u = 0$, then we may write the equation for $E(t)$ as

$$E(t) = \int_0^\infty [-S'(s)] P_E(t-s) ds.$$

From the equation for $E(t)$, we see that the output from E to I is

$$-E_0 P'_E(t) - \int_0^t [-S'(s)] P'_E(t-s) ds = - \int_0^\infty [-S'(s)] P'_E(t-s) ds.$$

Then

$$\begin{aligned} I(t) &= - \int_0^\infty \int_0^\infty S'(s) P'_E(t-s-u) P_I(u) du ds \\ &\quad - \int_0^\infty \int_0^\infty [-S'(s)] P'_E(t-s-u) P_I(u) du S'(s) ds. \end{aligned}$$

We now have

$$I(t) = \int_0^\infty [-S'(s)] A_I(t-s) ds,$$

with

$$A_I(z) = - \int_0^\infty P'_E(z-v) P_I(v) dv.$$

Then the model is

$$\begin{aligned} S' &= -\beta SI \\ E(t) &= \int_0^\infty [-S'(s)] P_E(t-s) ds \\ I(t) &= \int_0^\infty [-S'(s)] A_I(t-s) ds, \end{aligned} \tag{4.21}$$

which is in age of infection form with $\Phi = I$ and $A(z) = A_I(z)$. Then

$$\begin{aligned} \mathcal{R}_0 &= \beta N \int_0^\infty A(z) dz \\ &= -\beta N \int_0^\infty \int_0^z P'_E(z-u) P_I(u) du dz \\ &= \beta N \int_0^\infty P_I(u) du, \end{aligned} \tag{4.22}$$

using $-\int_0^\infty P'_E(v) dv = P_E(0) - P_E(\infty) = 1$.

The initial exponential growth rate of the general SEIR model (4.21) satisfies

$$\beta N \int_0^\infty e^{-rs} \int_0^s [-P'_E(s-u)] P_I(u) du ds = 1,$$

which reduces to

$$\begin{aligned} 1 &= \beta N \int_0^\infty q(v) e^{-rv} dv \int_0^\infty e^{-ru} P_I(u) du \\ &= \beta N \left[1 - r \int_0^\infty e^{-rv} P_E(v) dv \right] \int_0^\infty e^{-ru} P_I(u) du, \end{aligned} \quad (4.23)$$

with the aid of integration by parts.

It is important to remember that in this example we are assuming that effective contacts transmit new infections. We have not allowed for reductions in infectivity or susceptibility, but it would be possible to include these factors into the model. From (4.23) we see that the effect of an exposed period is to decrease the initial exponential growth rate. It also indicates that the result of using the relation (4.19) depends on knowledge of the compartmental structure of the model.

4.5.2 A General Treatment Model

Consider the treatment model (4.12) of Sect. 4.3. We now extend this to an age of infection model with general infective and treatment staged distributions. Assume that the distribution of infective periods is given by $P_I(\tau)$, and the distribution of periods in treatment is given by $P_T(\tau)$. Then the SITR model becomes

$$\begin{aligned} S'(t) &= -\beta S(t)[I(t) + \delta T(t)] \\ I(t) &= I_0 P_I(t) + \int_0^t [-S'(t-\sigma)] e^{-\gamma\sigma} P_I(\sigma) d\sigma \\ T(t) &= \int_0^\infty \gamma I(t-\sigma) P_T(\sigma) d\sigma, \end{aligned} \quad (4.24)$$

and

$$\varphi(t) = I(t) + \delta T(t).$$

Assuming that $S(u)$ has a jump of $-I_0$ at $u = 0$, we may write the equation for I as

$$I(t) = - \int_0^\infty [-S'(t-s)] e^{-\gamma s} P_I(s) ds.$$

Substituting the expression for I into the equation for T , we obtain

$$\begin{aligned} T(t) &= \int_0^\infty \gamma \int_0^\infty [-S'(t-s-\sigma)] e^{-\gamma s} P_I(s) P_T(\sigma) ds d\sigma \\ &= \int_0^\infty \gamma \int_0^\infty [-S'(t-s-\sigma)] P_T(\sigma) d\sigma e^{-\gamma s} P_I(s) ds \end{aligned}$$

$$\begin{aligned}
&= \int_0^\infty \gamma \int_s^\infty [-S'(t-v)] P_T(v-s) dv e^{-\gamma s} P_I(s) ds \\
&= \int_0^\infty [-S'(t-v)] \gamma \int_0^v P_T(v-s) e^{-\gamma s} P_I(s) ds dv \\
&= \int_0^\infty [-S'(t-v)] A(v) dv,
\end{aligned}$$

with

$$A(v) = \int_0^v \gamma P_T(v-s) e^{-\gamma s} P_I(s) ds.$$

Writing

$$\varphi(t) = I(t) + \delta T(t),$$

we now have the model (4.24) in age of infection form (without transmission probabilities),

$$\begin{aligned}
S'(t) &= -\beta S(t)\varphi(t) \\
\varphi(t) &= - \int_0^\infty [-S'(t-s)] [e^{-\gamma s} P_I(s) + \delta A(s)] ds.
\end{aligned} \tag{4.25}$$

From this we see that

$$\begin{aligned}
\mathcal{R}_0 &= \beta N \int_0^\infty [e^{-\gamma s} P_I(s) + \delta A(s)] ds \\
&= \beta N \int_0^\infty e^{-\gamma s} P_I(s) ds + \beta N \delta \gamma \int_0^\infty \int_0^s P_T(s-u) du ds].
\end{aligned}$$

With exponentially distributed infective and treatment periods, $P_I(s) = e^{-\alpha s}$, $P_T(s) = e^{-\eta s}$ we calculate \mathcal{R}_0 , obtaining

$$\begin{aligned}
\mathcal{R}_0 &= \beta N \int_0^\infty e^{-(\alpha+\gamma)\tau} d\tau \left[1 + \delta \gamma \int_0^\infty e^{-\eta\tau} d\tau \right] \\
&= \frac{\beta N}{\alpha + \gamma} \left[1 + \frac{\delta \gamma}{\eta} \right],
\end{aligned}$$

the same result as (4.13).

An arbitrary choice of treatment period distribution with mean $1/\eta$ does not affect the quantity \mathcal{R}_0 , but different infective period distributions may have a significant effect. For example, let us take $\gamma = 1$ and assume the mean infective

period is 1. Then, with an exponential distribution, $P_I(\tau) = e^{-\tau}$,

$$\int_0^\infty e^{-\tau} P_I(\tau) d\tau = \int_0^\infty e^{-2\tau} d\tau = \frac{1}{2}.$$

With an infective period of fixed length 1,

$$\int_0^\infty e^{-\tau} P_I(\tau) d\tau = \int_0^1 e^{-\tau} d\tau = (1 - e^{-1}) = 0.632.$$

Thus a model with an infective period of fixed length would lead to a basic reproduction number more than 25% higher than a model with an exponentially distributed infective period that has the same mean.

4.5.3 A General Quarantine/Isolation Epidemic Model

To cope with a disease outbreak for which there is no (as yet) known treatment, the only methods available are isolation of diagnosed infective and quarantine of suspected exposed members of the population. This approach was used during the *SARS* epidemic of 2003, and modeled in [19]. An *SEIR* model with general exposed and infective periods as well as quarantine and isolation has been analyzed in [47]. While the model in [47] was deterministic, the analysis was from a probabilistic point of view. Here we add quarantine and isolation to the model (4.21) and derive some of the results of [47] from a compartmental approach. The model is related to the model (3.42), which includes births and natural deaths, and is more general because it does not assume that quarantine and isolation are perfectly effective.

We move members from the exposed class to a quarantine class Q at rate ψ and from the infective class to an isolated (hospitalized) class H at rate φ . For simplicity, we assume that both quarantine and isolation are perfectly effective, so that no infections are transmitted from either quarantined or isolated members. Then we need not include Q or H in the model unless we wish to track the number of individuals quarantined and isolated. We need only adjust the model (4.21) to include the removals. With quarantine but not yet isolation, the new equation for E is

$$E(t) = E_0 e^{-\psi t} P_E(t) + \int_0^t [-S'(s)] e^{-\psi(t-s)} P_E(t-s) ds.$$

The input to I at time u becomes

$$E_0 q_E(u) + \int_0^u [-S'(\tau)] e^{-\psi(u-\tau)} q_E(u-\tau) d\tau.$$

Now, $I(t)$ is given by

$$\begin{aligned} I(t) &= E_0 \int_0^t q_E(u) e^{-\psi(t-u)} P_I(t-u) du \\ &\quad + E_0 \int_0^t [-S'(s)] q_E(u-s) e^{-\psi(u-s)} ds P_I(t-u) du. \end{aligned}$$

The first term in this expression may be written as $I_0(t)$, and the second term may be simplified, using interchange of the order of integration in the iterated integral much as in the previous section, to yield

$$\int_0^t [-S'(s)] T(t-s) ds$$

with

$$T(v) = \int_0^v q_E(y) e^{-\psi y} P_I(v-y) dy.$$

Then the model is

$$\begin{aligned} S' &= -\beta SI \\ E(t) &= E_0 e^{-\psi t} P_E(t) + \int_0^t [-S'(s)] e^{-\psi(t-s)} P_E(t-s) ds \\ I(t) &= I_0(t) + \int_0^t [-S'(s)] T(t-s) ds. \end{aligned}$$

If we add isolation at a rate φ of infectives, we obtain

$$I(t) = e^{-\varphi t} I_0(t) + \int_0^t [-S'(s)] e^{-\varphi(t-s)} T(t-s) ds,$$

and the quarantine/isolation model is

$$\begin{aligned} S' &= -\beta SI \\ E(t) &= E_0 e^{-\psi t} P_E(t) + \int_0^t [-S'(s)] e^{-\psi(t-s)} P_E(t-s) ds \\ I(t) &= E_0 e^{-\varphi t} I_0(t) + \int_0^t [-S'(s)] e^{-\varphi(t-s)} T(t-s) ds. \end{aligned} \tag{4.26}$$

The reproduction number is now a control reproduction number $\mathcal{R}_c(\psi, \varphi)$, depending on the control parameters ψ and φ ,

$$\begin{aligned}\mathcal{R}_c(\psi, \varphi) &= \beta N \int_0^\infty e^{-\varphi\tau} T(\tau) d\tau \\ &= \beta N \int_0^\infty e^{-\varphi\tau} E_0 \int_0^\tau q(y) e^{-\psi y} P_I(\tau - y) dy d\tau \\ &= \beta N \int_0^\infty e^{-\psi y} q(y) \left[\int_y^\infty e^{-\varphi(\tau-y)} P_I(\tau - y) d\tau \right] dy \\ &= \beta N \int_0^\infty e^{-\psi y} q(y) dy \int_0^\infty e^{-\varphi u} P_I(u) du.\end{aligned}$$

If we know the functions P_I and q , we may calculate the sensitivity of \mathcal{R}_0 to the quarantine and isolation rates and thus compare quarantine and isolation as management strategies. In [47], this expression is written in terms of the Laplace transforms of q and P_I ,

$$\mathcal{R}_c(\psi, \varphi) = \beta N E_0 \mathcal{L}_q(\psi) \mathcal{L}_{P_I}(\varphi).$$

By comparison of the mean infectivity functions of the models (4.21) and (4.26), we see that the effect of quarantine and/or isolation on an SEIR model is to decrease the initial exponential growth rate.

As suggested by these examples, there are general methods for calculation of integrals involving $A(\tau)$ without the necessity of calculating the function A explicitly [7, 48].

A naive view of epidemic models suggests that if we know the basic reproduction number, we can determine the size of an epidemic. There are several quite different problems with this view. We have seen that estimation of the basic reproduction number from early observed data depends on the structure of the model being assumed. For example, if there is an exposed period, the initial exponential growth rate is less than if infected individuals become infectious immediately but this does not affect the basic reproduction number. This implies that an estimate of the basic reproduction number from the initial exponential growth rate will be too small. A second problem is that in an epidemic, early data may be incomplete and inaccurate, and its use may lead to a poor estimate of the reproduction number.

4.6 The Gamma Distribution

There is ample evidence that exponential distributions of stay in compartments are much less realistic than gamma distributions [16, 17, 26, 46]. This has already been suggested in Sect. 3.6. A gamma distribution $P(\tau)$ with parameter n and period $1/\alpha$

can be represented as a sequence of n exponential distributions $P_i(\tau)$ with period $1/n\alpha$. Thus an *SIR* epidemic model with a gamma distribution for the infectious period may be represented by the system

$$\begin{aligned} S' &= -\beta SI \\ I'_1 &= \beta SI - n\alpha I_1 \\ I'_j &= n\alpha I_{j-1} - n\alpha I_j, \quad j = 2, 3, \dots, n. \end{aligned} \tag{4.27}$$

We may view this as an age of infection model

$$\begin{aligned} S' &= -\beta SI \\ I(t) &= I_0(t) + \int_0^t [-S'(\tau)] P(\tau) d\tau, \end{aligned} \tag{4.28}$$

in which $P(\tau)$ represents the fraction of infectives with infection age τ .

To use the age of infection interpretation, we need to determine the kernel $P(\tau)$ in order to calculate its integral. We let $u_k(\tau)$ be the fraction of infected members with infection age τ in the k -th subinterval. Then

$$\begin{aligned} u'_1 &= -n\alpha u_1, \quad u_1(0) = 1 \\ u'_j &= n\alpha u_{j-1} - n\alpha u_j, \quad u_j(0) = 0, \quad j = 2, 3, \dots, n. \end{aligned} \tag{4.29}$$

It is easy to solve this system of differential equations recursively, and we obtain

$$u_k(\tau) = \frac{(n\alpha)^{k-1}}{(k-1)!} \tau^{k-1} e^{-n\alpha\tau}, \quad k = 1 \dots n.$$

This gives the desired kernel

$$P(\tau) = \sum_{k=1}^n u_k(\tau) = \sum_{k=1}^n \frac{(n\alpha)^{(k-1)}}{(k-1)!} \tau^{(k-1)} e^{-n\alpha\tau}. \tag{4.30}$$

In order to evaluate the integral $\int_0^\infty P(\tau) d\tau$, we start with the integration formula

$$\int t^{k-1} e^{-ct} dt = -\frac{1}{c} t^{k-1} e^{-ct} + \frac{k-1}{c} \int t^{k-2} e^{-ct} dt, \tag{4.31}$$

obtained using integration by parts; then with limits of integration 0 to ∞ and $c = n\alpha$, we obtain

$$\int_0^\infty t^{k-1} e^{-n\alpha t} dt = \frac{k-1}{n\alpha} \int_0^\infty t^{k-2} e^{-n\alpha t} dt,$$

which leads by induction to the formula

$$\int_0^\infty t^{k-1} e^{-n\alpha t} dt = \frac{(k-1)!}{(n\alpha)^{k-1}} \cdot \frac{1}{n\alpha}. \quad (4.32)$$

Combination of (4.30), (4.32) gives

$$\int_0^\infty P(\tau) d\tau = \sum_{k=1}^n \frac{1}{n\alpha} = \frac{1}{\alpha}.$$

The mean infective period is independent of n but a larger value of n gives a smaller variance. The limiting case as $n \rightarrow \infty$ is an infective period of fixed length $1/\alpha$. In practice, infective periods often are bunched closer together than would be predicted by an exponential distribution and it is common to choose a gamma distribution with a value of n corresponding to the observed distribution.

We will need to use the value of $\int_0^\infty e^{-\lambda\tau} P(\tau) d\tau$. To evaluate this integral, we have

$$\int_0^\infty e^{-\lambda\tau} u_k(\tau) d\tau = \frac{(n\alpha)^{k-1}}{(k-1)!} \int_0^\infty \tau^{k-1} e^{-(\lambda+n\alpha)\tau} d\tau = \frac{(n\alpha)^{k-1}}{(\lambda+n\alpha)^k},$$

using (4.32). From this we obtain

$$\int_0^\infty e^{-\lambda\tau} P(\tau) d\tau = \sum_{k=0}^{n-1} \frac{(n\alpha)^{k-1}}{(\lambda+n\alpha)^k} = \frac{1}{(\lambda+n\alpha)} \sum_{k=0}^{n-1} \frac{n\alpha}{\lambda+n\alpha}. \quad (4.33)$$

But this is a geometric series, whose sum is

$$\frac{1}{\lambda+n\alpha} \frac{1 - \left(\frac{n\alpha}{\lambda+n\alpha}\right)^n}{1 - \frac{n\alpha}{\lambda+n\alpha}} = \frac{1 - \left(\frac{n\alpha}{\lambda+n\alpha}\right)^n}{\lambda}.$$

We now have the formula

$$\int_0^\infty e^{-\lambda\tau} P(\tau) d\tau = \frac{1 - \left(\frac{n\alpha}{\lambda+n\alpha}\right)^n}{\lambda}. \quad (4.34)$$

The limiting case as $n \rightarrow \infty$ of a gamma distribution is a distribution

$$P(t) = 1 \quad (0 \leq t \leq \frac{1}{\alpha}), \quad P(t) = 0 \quad (t > \frac{1}{\alpha}),$$

for which

$$\int_0^\infty e^{-\lambda t} P(t) dt = \frac{1 - e^{-\lambda/\alpha}}{\gamma}.$$

Then

$$\frac{1}{\alpha + \frac{\gamma}{n}} < \frac{1 - e^{-\gamma/\alpha}}{\lambda}. \quad (4.35)$$

It is possible to show, by use of L'Hôpital's rule that

$$\lim_{n \rightarrow \infty} \left(\frac{n\alpha}{\lambda + n\alpha} \right)^n = e^{-\lambda/\alpha},$$

and this implies that the limiting value as $n \rightarrow \infty$ of the integral $\int_0^\infty e^{-\lambda\tau} P(\tau) d\tau$ for a gamma distribution with parameter n is the integral corresponding to the limiting distribution.

4.7 Interpretation of Data and Parametrization

We have carried out qualitative analysis of various disease transmission models in terms of the parameters of the model. In modeling a specific disease outbreak it is necessary to assign values to the parameters of the model in order to estimate the effects of possible control measures. For an epidemic disease, we ignore demographic quantities. If we have information about the duration of infection in an individual, we need to estimate the contact rate and the rates of transitions between compartments in order to be able to estimate the basic reproduction number and then to analyze the qualitative behavior of the model. However, different assumptions about the compartmental structure of the model and about the transition rates between compartments would lead to different results.

In an epidemic situation we may be able to estimate the contact rate from the initial exponential growth rate. For example, in the simplest possible epidemic model, the Kermack–McKendrick model with no disease deaths

$$\begin{aligned} S' &= -\beta SI \\ I' &= \beta SI - \alpha I \end{aligned} \quad (4.36)$$

initially (or as long as S remains close to the total population size N), $I(t)$ may be approximated by $e^{(\beta N - \alpha)t}$. Thus the initial exponential growth rate is approximately $\beta N - \alpha$. If we plot observed values of $\log I(t)$ as a function of t , we would expect

that the graph, after an initial stochastic phase would be a straight line, until S/N becomes significantly less than 1 and the graph bends downward.

Let us suppose that we can estimate the slope r of this line. Then the value of r gives an estimate of $\beta N - \alpha$. From this, we obtain an estimate for the basic reproduction number

$$\mathcal{R}_0 = \frac{\beta N}{\alpha} \approx \frac{r + \alpha}{\alpha} = \frac{r}{\alpha} + 1. \quad (4.37)$$

4.7.1 Models of SIR Type

In the simple Kermack–McKendrick epidemic model (4.36) we assume that the infectious period is exponentially distributed with mean $1/\alpha$. In practice, the period of stay in the infectious class usually has a smaller variance than an exponential distribution.

As we have suggested in Sect. 4.6, a gamma distribution is likely to be much more realistic than an exponential distribution for stay in a compartment [16, 17, 26, 46]. A gamma distribution $P(\tau)$ with parameter n and period $1/\alpha$ can be represented as a sequence of n exponential distributions $P_i(\tau)$ with period $1/n\alpha$. Then an SIR epidemic model with a gamma distribution for the infectious period may be represented by the system (4.27). We used the formula (4.30) to evaluate the integrals

$$\int_0^\infty P(\tau)d\tau = \sum_{k=1}^n \frac{1}{n\alpha} = \frac{1}{\alpha},$$

and

$$\int_0^\infty e^{-r\tau} P(\tau)d\tau = \frac{1 - \left(\frac{n\alpha}{r+n\alpha}\right)^n}{r}. \quad (4.38)$$

The limiting case as $n \rightarrow \infty$ of a gamma distribution is a distribution

$$P(t) = 1 \quad (0 \leq t \leq \frac{1}{\alpha}), \quad P(t) = 0 \quad (t > \frac{1}{\alpha}),$$

for which

$$\int_0^\infty e^{-rt} P(t)dt = \frac{1 - e^{-r/\alpha}}{\alpha}.$$

Then

$$\frac{1}{\alpha + \frac{\gamma}{n}} < \frac{1 - e^{-r/\alpha}}{r}. \quad (4.39)$$

It is possible to show, by use of L'Hôpital's rule that

$$\lim_{n \rightarrow \infty} \left(\frac{n\alpha}{r + n\alpha} \right)^n = e^{-r/\alpha},$$

and this implies that the limiting value as $n \rightarrow \infty$ of the integral $\int_0^\infty e^{-r\tau} P(\tau) d\tau$ for a gamma distribution with parameter n is the integral corresponding to the limiting distribution.

As we have seen in Sect. 4.4, for an age of infection *SIR* model with an infectious period distribution given by a function $P(\tau)$, the basic reproduction number is given by

$$\mathcal{R}_0 = \beta N \int_0^\infty P(\tau) d\tau, \quad (4.40)$$

while the initial exponential growth rate λ satisfies

$$\beta N \int_0^\infty e^{-r\tau} P(\tau) d\tau = 1. \quad (4.41)$$

Elimination of βN between the Eqs. (4.40) and (4.41) gives a formula for the basic reproduction number determined by the initial exponential growth rate r and the distribution $P(\tau)$,

$$\mathcal{R}_0 = \frac{\int_0^\infty P(\tau) d\tau}{\int_0^\infty e^{-r\tau} P(\tau) d\tau}. \quad (4.42)$$

If we assume that the mean infective period $\int_0^\infty P(\tau) d\tau$ is known and the initial exponential growth rate can be measured, we may use the relation (4.42) to calculate the basic reproduction number for various choices of the infectious period distribution. As more information about the infectious period distribution becomes known, it is possible to obtain more refined estimates of the reproduction number.

Thus, for example, the choice of an exponentially distributed infective period $P(\tau) = e^{-\alpha\tau}$ gives

$$\int_0^\infty e^{-\lambda\tau} d\tau = \frac{1}{\lambda + \alpha}, \quad \mathcal{R}_0 = 1 + \frac{r}{\alpha}.$$

The choice of a fixed length infective period gives

$$r \int_0^\infty e^{-r\tau} d\tau = \frac{1 - e^{-r/\alpha}}{r}, \quad \mathcal{R}_0 = \beta/\alpha(1 - e^{-r/\alpha}).$$

For a gamma distribution with parameter n ,

$$\int_0^\infty e^{-\tau} P(\tau) d\tau = 1 - \left(\frac{n\alpha}{r + n\alpha} \right)^n, \quad \mathcal{R}_0 = \frac{1}{1 - (\frac{n\alpha}{r + n\alpha})^n}.$$

4.7.2 Models of SEIR Type

Now suppose we assume a simple *SEIR* model,

$$\begin{aligned} S' &= -\beta SI \\ E' &= \beta SI - \kappa E \\ I' &= \kappa E - \alpha I. \end{aligned} \tag{4.43}$$

The exponential growth rate for this model is

$$r = \frac{-(\alpha + \kappa) + \sqrt{(\alpha - \kappa)^2 + 4\kappa\beta N}}{2},$$

and this yields the estimate

$$\beta N = \alpha + \frac{r^2}{\kappa} + r \frac{\alpha + \kappa}{\kappa}. \tag{4.44}$$

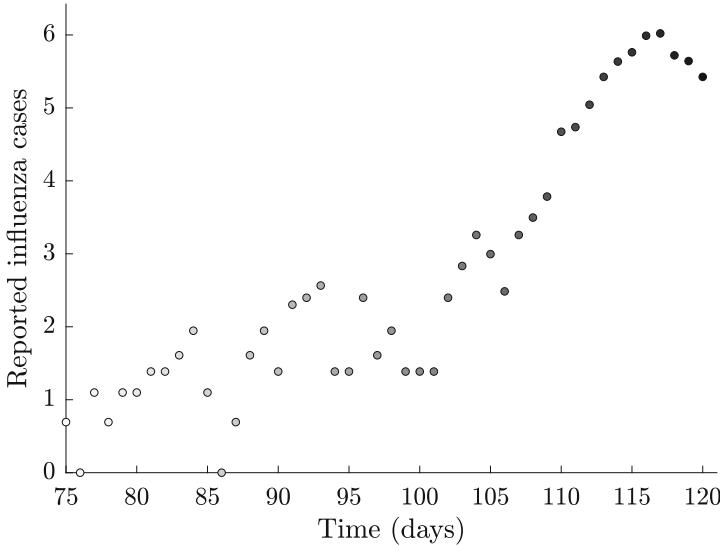
Observe that this estimate for βN is larger than the estimate obtained for the *SIR* model and therefore the assumption of an *SEIR* model leads to a larger estimate than the *SIR* model for $\mathcal{R}_0 = \beta N / \alpha$. A given exponential growth rate in an *SEIR* model corresponds to a larger reproduction number than in an *SIR* model.

Example 1

Consider the data given in Table 2.2 in Sect. 2.10 for reported H1N1 influenza cases in México. We plot $\log I(t)$ as a function of t , obtaining the graph shown in Fig. 4.5. This graph, after an initial stochastic phase, appears to exhibit linear growth from day 97 to day 117, with a slope of 0.22. The data is for influenza, which we assume to be described by an *SEIR* model with an exposed period of 1.9 days followed by an infectious period of 4.1 days. The relation (4.45) with the values

$$\alpha = 1/4.1, \quad \kappa = 1/1.9, \quad r = 0.22$$

yields $\beta = 0.65$, and then $\mathcal{R}_0 = \beta/\alpha = 2.67$.

**Fig. 4.5** Influenza cases in Mexico

It is important to note that this estimate depends not only on the initial exponential growth rate but also on the compartmental structure assumed for the model and on the transition rates between compartments.

We may write a general *SEIR* model in age of infection form,

$$\begin{aligned} S' &= -\beta S \varphi \\ \varphi(t) &= \varphi_0(t) + \int_0^t \beta S(t-\tau) \varphi(t-\tau) A(\tau) d\tau, \end{aligned}$$

where $A(\tau)$ is the mean infectivity of an infected individual, whether exposed or infective, at infection age τ .

If we separate infected individuals into exposed and infective, and think of an exposed period distribution P_E and an infectious period distribution P_I , the model is the *SEIR* model of Sect. 4.5.1

$$\begin{aligned} S' &= -\beta SI \\ E(t) &= E_0 P_E(t) + \int_0^t [-S'(u)] P_E(t-u) du \\ I(t) &= I_0(t) - \int_0^t [-S'(u)] \left[\int_u^t P'_E(v-u) P_I(t-v) dv \right] du. \end{aligned} \tag{4.45}$$

This is of age of infection form with mean infectivity

$$A(t-u) = - \int_u^t P'_E(v-u) P_E(t-v) dv$$

and

$$\int_0^\infty A(\tau) d\tau = \int_0^\infty P_I(\tau) d\tau.$$

Also,

$$\begin{aligned} \int_0^\infty e^{-r\tau} A(\tau) d\tau &= - \int_0^\infty P'_E(v) e^{-rv} dv \int_0^\infty e^{-ru} P_I(u) du \\ &= \left[1 - r \int_0^\infty e^{-rv} P_E(v) dv \right] \int_0^\infty e^{-ru} P_I(u) du, \end{aligned}$$

and we may use this relation to calculate \mathcal{R}_0 for a variety of choices of P_E and P_I , including gamma distributions and periods of fixed length.

With exponentially distributed exposed and infectious periods, we obtain

$$\mathcal{R}_0 = 1 + \frac{r}{\alpha} + \frac{r}{\kappa} + \frac{r^2}{\alpha\kappa},$$

while for exposed and infective periods of fixed lengths we obtain

$$\mathcal{R}_0 = \frac{re^{r/\kappa}}{\alpha(1 - e^{-r/\alpha})} > e^r/\kappa.$$

It is possible to show that the basic reproduction number corresponding to fixed exposed and infective periods is always larger than the basic reproduction number corresponding to exponentially distributed exposed and infectious periods, and the basic reproduction number corresponding to exposed and infectious periods both having gamma distributions is between these two values.

4.7.3 Mean Generation Time

In a simple epidemic model, if we know the contact rate and the infective period, then we may be able to determine the basic reproduction number. As we have seen in the previous section, knowledge of the initial exponential growth rate, which may sometimes be observed experimentally leads to information about the contact rate. Another quantity which may sometimes be observed experimentally, is the mean generation time, meaning roughly the mean time between a primary case and a

secondary case caused by this primary case. This quantity has been given various definitions, not all equivalent. Another term often used to describe this idea is the serial interval.

The mean generation time is defined [18, 40] as the mean time from infection to the onset of infectiousness. It may be estimated from data early in a disease outbreak [40, 44, 45]. If we assume that the rate of secondary infections caused by an individual is proportional to the infectiousness of the individual [37], then the rate of secondary infections caused at infection age τ is proportional to $A(\tau)$. This assumption would be valid only early in an epidemic when the depletion of the susceptible population would be negligible and the number of infectives is a sufficiently small fraction of the total population size that a susceptible is at risk of infection only from a single infective. Thus the distribution of new infections caused at infection age τ is

$$\frac{A(s)}{\int_0^\infty A(s)ds},$$

whose mean is

$$T_G = \frac{\int_0^\infty sA(s)ds}{\int_0^\infty A(s)ds}, \quad (4.46)$$

and this is the mean generation time. We should note that this describes the time interval between successive infections, but what may be observed is the time interval between development of clinical symptoms, which is not the same thing, but we may hope that it is a reasonable approximation.

In order to calculate the mean generation time for the general *SEIR* model (4.45), we use (4.46) with

$$A(s) = A_I(s) = - \int_0^s P'_E(s-u)P_I(u)du.$$

As we have seen in the calculations for the model (4.45), $\int_0^\infty A(s)ds = \int_0^\infty P_I(s)ds$. We have also,

$$\begin{aligned} \int_0^\infty sA(s)ds &= - \int_0^\infty \int_0^s sP'_E(s-u)P_I(u)dsdu \\ &= - \int_0^\infty \left[\int_u^\infty sP'_E(s-u)ds \right] P_I(u)du. \end{aligned} \quad (4.47)$$

But

$$-\int_u^\infty sE'_Q(s-u)ds = - \int_0^\infty (u+v)P'_E(v)dv = u + \int_0^\infty P_E(v)dv,$$

using integration by parts. Thus (4.47) becomes

$$\int_0^\infty s A(s) ds = \int_0^\infty u P_I(u) du + \int_0^\infty P_I(u) du \int_0^\infty P_E(v) dv.$$

This shows that the mean generation time for the *SEIR* model (4.45) is

$$T_G = \frac{\int_0^\infty u P_I(u) du}{\int_0^\infty P_I(u) du} + \int_0^\infty P_E(u) du.$$

Note that this is the sum of the mean period of stay in the exposed stage and a quantity depending on the infective period distribution. It is easy to calculate that in the case of an exponential distribution with mean $1/\alpha$ this quantity is $1/\alpha$, the mean infective period, while in the case of a constant infectious period of length $1/\alpha$ this quantity is $1/2\alpha$, half the mean infective period. In general, for an *SEIR* model the mean generation time is equal to the sum of the mean exposed period and a quantity depending on the distribution of the infective period.

It is convenient to define the mean exposed time T_E and the mean infective time T_I . For the general *SEIR* model (4.45) that we have been studying, we have

$$T_E = \frac{1}{\kappa}, \quad T_I = \frac{1}{\alpha}.$$

For exponentially distributed exposed and infective periods,

$$T_G = T_E + T_I,$$

and we may write

$$\mathcal{R}_0 = 1 + r \frac{1}{\kappa} + r \frac{1}{\alpha} + r^2 \frac{1}{\kappa \alpha} = 1 + r T_E + r T_I + r^2 T_E (T_G - T_E).$$

For exposed and infective periods of fixed length,

$$T_G = T_E + \frac{1}{2} T_I,$$

and we have

$$\mathcal{R}_0 = \frac{\frac{r}{\alpha} e^{r/\kappa}}{1 - e^{-r/\alpha}} = \frac{\frac{r}{\alpha} e^{r/\kappa + r/2\alpha}}{e^{r/2\alpha} - e^{-r/2\alpha}} = \frac{\frac{r}{\alpha} e^{r T_G}}{2 \sinh(r/2\alpha)}.$$

Since $\sinh x > x$ for $x > 0$, we obtain the bound

$$\mathcal{R}_0 < e^{r T_G}. \tag{4.48}$$

For a model with an exposed period having a gamma distribution with parameter m and an infective period having a gamma distribution with parameter n it is possible to show that the corresponding reproduction number, which we denote by $\mathcal{R}_0^{(m,n)}$, is

$$\mathcal{R}_0^{(m,n)} = \frac{r \left(\frac{r}{\kappa m} + 1 \right)^m}{\alpha \left[1 - \left(\frac{r}{\alpha n} + 1 \right)^{-n} \right]}. \quad (4.49)$$

It is possible to show that $\mathcal{R}_0^{(m,n)}$ is an increasing function of both m and n , and that as $m, n \rightarrow \infty$, $\mathcal{R}_0^{(m,n)}$ approaches the reproduction number corresponding to exposed and infective periods of fixed length.

4.8 *Effect of Timing of Control Programs on Epidemic Final Size

To explore the influence of seasonally forced transmission in disease control and the epidemic final size of influenza, we consider an extension of the standard SIR epidemic model by incorporating a periodic function $c(t)$ for the transmission rate as well as drug treatment and/or vaccination.

When model parameters vary with time, analytic results on final size will be very difficult to obtain. Most results are based on numerical simulations. The example presented here with periodic transmission rate examines the potential negative effect of vaccination and/or treatment uses depending on the timing of implementation. These scenarios are based on the consideration that vaccines and antiviral drugs may not be available at the beginning of an epidemic. We focus on three important measures when assessing the effect of a control program: (a) peak size of the epidemic curve (the maximum number of infections during the course of a pandemic); (b) peak time (the time at which the peak occurs); and (c) final size (the total number of infections at the end of a pandemic). Main objectives of effective control strategies should include lowering the peak size to keep demand for facilities below available supply, lowering the final size to reduce morbidity, and delaying the peak to provide more time for response.

Unlike in the case of constant parameters, where vaccination and treatment will always help reduce morbidity, the case of a periodic transmission rate $c(t)$ may generate non-intuitive results. That is, the model can exhibit outcomes in which an increased use of vaccination or antiviral drugs will lead to a higher morbidity. To demonstrate this, we first consider the following epidemic model with time-dependent infection rate

$$\begin{aligned} \frac{dS}{dt} &= -c(t) \frac{SI}{N} \\ \frac{dI}{dt} &= c(t) \frac{SI}{N} - \alpha I, \end{aligned} \quad (4.50)$$

where

$$c(t) = c_0[1 + \epsilon \cos(2\pi t/365)]$$

with the initial conditions

$$S(t_0) = (1 - \varphi_0)(N - I_0), \quad I(t_0) = I_0, \quad R(t_0) = 0. \quad (4.51)$$

The parameter ϵ denotes the magnitude of the seasonal forcing and the period is assumed to be 1 year. We do not include vaccinated individuals in the R class because the value of $R(t)$ at the end of the disease outbreak will be used to measure the final epidemic size. Notice that the initial time t_0 represents the time at which the vaccination program starts. Some simulation results are presented in Fig. 4.6.

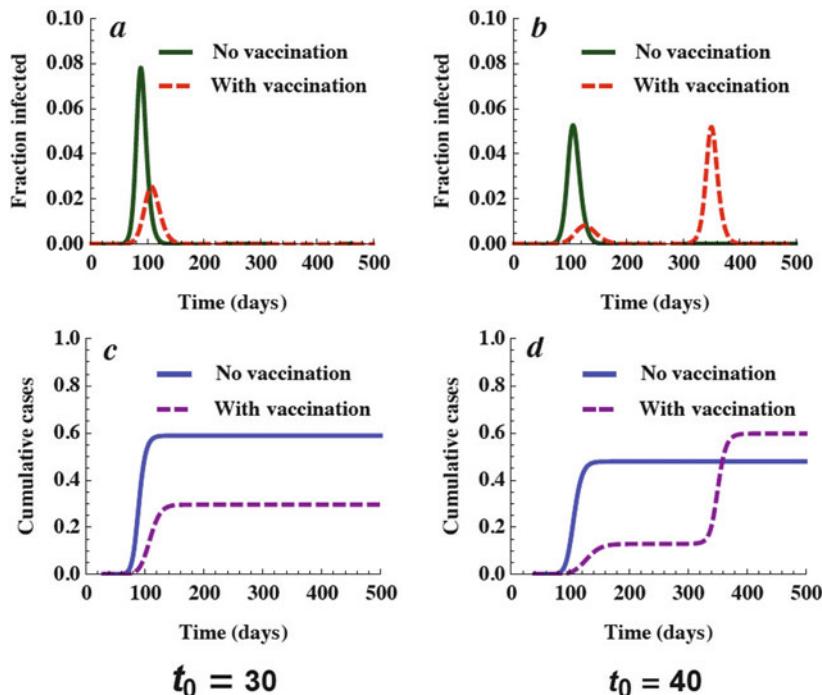


Fig. 4.6 Plots of epidemic curves and cumulative infections for different t_0 values. These plots show the dramatic effect that the start time of an epidemic (t_0) can have on its progression in a periodic environment given by the SIR model (4.50) with initial conditions (4.51). From the plots in (a) and (c) we observe that when $t_0 = 30$, vaccination reduced both the peak size and the peak time although there is not much delay in the peak time. However, plots in (b) and (d) show that, although the size of the first epidemic peak is reduced, a second (and higher) peak is also generated. More significantly, the final epidemic size is increased, showing the sensitivity of the behavior of the epidemic to t_0 .

In Fig. 4.6, both the epidemic curve and curve representing the cumulative infections are plotted. For the transmission function $c(t)$, the parameter values used are those of an H1N1-like disease with $c_0 = 0.5$ and $\epsilon = 0.35$. We assume $1/\alpha = 3$ days. These values are also used in other figures except when specified differently. The figure shows two sets of simulations corresponding to two different times t_0 of initial introduction of pathogens. One case is for $t_0 = 30$ and the other is for $t_0 = 40$, and for demonstration purposes the vaccination level is chosen to be $\varphi_0 = 0.1$.

Some interesting observations can be made from Fig. 4.6. We can first compare the outcomes between the model without vaccination (the solid curves) and the model with vaccination (the dashed curves). It demonstrates the dependence of the model outcomes on the time of introduction (t_0) of initial infections. Particularly, we observe that although the peak and final sizes are both decreased for the case of $t_0 = 30$, the case of $t_0 = 40$ differs significantly. Although the first wave is reduced, the second wave is also generated; and moreover, the final epidemic size is even higher than that without vaccination, demonstrating a scenario in which the use of vaccine may have an adverse effect.

4.9 Directions for Generalization

A fundamental assumption in the model (4.1) is homogeneous mixing, that all individuals are equivalent in contacts. A more realistic approach would include separation of the population into subgroups with differences in behavior. For example, in many childhood diseases the contacts that transmit infection depend on the ages of the individuals, and a model should include a description of the rate of contact between individuals of different ages. Other heterogeneities that may be important include activity levels of different groups and spatial distribution of populations. Network models may be formulated to include heterogeneity of mixing, or more complicated compartmental models can be developed.

An important question which should be kept in mind in the formulation of epidemic models is the extent to which the fundamental properties of the simple model (4.1) carry over to more elaborate models.

4.10 Some Warnings

An actual epidemic differs considerably from the idealized model (4.1) or (4.17). Some notable differences are:

1. When it is realized that an epidemic has begun, individuals are likely to modify their behavior by avoiding crowds to reduce their contacts and by being more careful about hygiene to reduce the risk that a contact will produce infection.
2. If a vaccine is available for the disease which has broken out, public health measures will include vaccination of part of the population. Various vaccination

strategies are possible, including vaccination of health care workers and other first line responders to the epidemic, vaccination of members of the population who have been in contact with diagnosed infectives, or vaccination of members of the population who live in close proximity to diagnosed infectives.

3. Isolation may be imperfect; in-hospital transmission of infection was a major problem in the SARS epidemic.
4. In the SARS epidemic of 2002–2003, in-hospital transmission of disease from patients to health care workers because of imperfect isolation accounted for many of the cases. This points to an essential heterogeneity in disease transmission which must be included whenever there is any risk of such transmission.

4.11 *Project: A Discrete Model with Quarantine and Isolation

Project 1 This project concerns a general single-outbreak discrete epidemic model involving arbitrarily distributed stage-duration distributions for the latent and the infectious stages. Clarity is added to model derivation by making use of a probabilistic perspective. Hence, we let X and Y represent the time an individual spends in latent (E, Q) and infectious (I, H) classes, respectively. Similarly, we denote by Z the time at which an exposed individual is quarantined (from E to Q), and W the time at which an infected individual is isolated (from I to H). X, Y, Z , and W must take values on $\{1, 2, 3, \dots\}$ and their dynamics are assumed to be governed by underlying probabilistic processes. It is understood that the distribution of waiting time in each of the above classes can be described by probability distributions associated with the random variables X, Y, Z , and W :

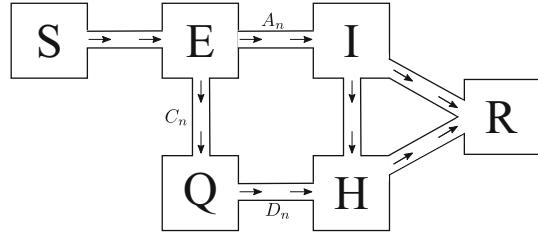
$$\begin{aligned} p_i &= \mathbb{P}(X > i) \text{ probability of remaining latent } i \text{ steps after entering } E, \\ q_i &= \mathbb{P}(Y > i) \text{ probability of remaining infectious } i \text{ steps after entering } I, \\ k_i &= \mathbb{P}(Z > i) \text{ probability of not quarantined } i \text{ steps after entering } E, \\ l_i &= \mathbb{P}(W > i) \text{ probability of not isolated } i \text{ steps after entering } I. \end{aligned} \tag{4.52}$$

Assume that $p_0 = q_0 = k_0 = l_0 = 1$, meaning that the latency, infectious, quarantine , and isolation periods last at least one time step. For ease of description, we introduce the following notation:

- A_n the input to E at time n (new infections),
- B_n the input to I at time n ,
- C_n the input to Q at time n ,
- D_n the input to H from Q at time n ,
- F_n ($= B_n + D_n$) the total input to infectious class at time n .

A transition diagram using the above notation is shown in Fig. 4.7.

Fig. 4.7 Transition diagram for the model with arbitrarily distributed stage durations



Using the above notation the model with arbitrarily distributed stage durations can be written as:

$$\begin{aligned}
 S_{n+1} &= S_n G_n, \quad G_n = e^{-\frac{\beta}{N}[I_n + (1-\rho)H_n]}, \\
 E_{n+1} &= A_{n+1} + A_n p_1 k_1 + \cdots + A_1 p_n k_n + A_0 p_{n+1} k_{n+1}, \\
 Q_{n+1} &= A_n p_1 (1 - k_1) + \cdots + A_1 p_n (1 - k_n) + A_0 p_{n+1} (1 - k_{n+1}), \\
 I_{n+1} &= B_{n+1} + B_n q_1 l_1 + \cdots + B_1 q_n l_n, \\
 H_{n+1} &= F_{n+1} + F_n q_1 + \cdots + F_1 q_n - I_{n+1}, \quad n = 0, 1, 2, \dots,
 \end{aligned} \tag{4.53}$$

where A_n , B_n , C_n , and D_n are given by

$$\begin{aligned}
 A_{n+1} &= S_n(1 - G_n) = S_n - S_{n+1}, \quad n \geq 0, \text{ with } A_0 = E_0, \\
 B_{n+1} &= A_n(1 - p_1) + A_{n-1}(p_1 - p_2)k_1 + \cdots + A_1(p_{n-1} - p_n)k_{n-1} \\
 &\quad + A_0(p_n - p_{n-1})k_n, \quad \text{with } B_0 = 0, k_0 = 1, \\
 C_{n+1} &= E_n - E_{n+1} + A_{n+1} - B_{n+1}, \quad \text{with } C_0 = 0, \\
 D_{n+1} &= Q_n - (Q_{n+1} - C_{n+1}).
 \end{aligned} \tag{4.54}$$

The initial conditions in model (4.53) are S_0 , $E_0 > 0$ and $I_0 = Q_0 = H_0 = R_0 = 0$.

Question 1 Show that the control reproduction number \mathcal{R}_c is given by

$$\mathcal{R}_c = \mathcal{R}_I + \mathcal{R}_{IH} + \mathcal{R}_{QH}, \tag{4.55}$$

where

$$\begin{aligned}
 \mathcal{R}_I &= \beta \mathcal{T}_{E_k} \mathcal{D}_{I_l}, \\
 \mathcal{R}_{IH} &= \beta(1 - \rho) \mathcal{T}_{E_k} (\mathcal{D}_I - \mathcal{D}_{I_l}), \\
 \mathcal{R}_{QH} &= \beta(1 - \rho)(1 - \mathcal{T}_{E_k}) \mathcal{D}_I.
 \end{aligned} \tag{4.56}$$

Here, $\mathcal{D}_E = \mathbb{E}(X)$ and $\mathcal{D}_I = \mathbb{E}(Y)$ are the mean sojourn times in the exposed and infectious stages, respectively. \mathcal{D}_{E_k} and \mathcal{D}_{I_l} denote the “quarantine-adjusted”

and “isolation-adjusted” mean sojourn times in the exposed and infective stages, respectively, given by

$$\begin{aligned}\mathcal{D}_{E_k} &= \mathbb{E}(X \wedge Z) = \sum_{j=0}^{\infty} \mathbb{P}(X \wedge Z > j) = 1 + \sum_{j=1}^{\infty} p_j k_j, \\ \mathcal{D}_l &= \mathbb{E}(Y \wedge W) = \sum_{j=0}^{\infty} \mathbb{P}(Y \wedge W > j) = 1 + \sum_{j=1}^{\infty} q_j l_j.\end{aligned}$$

\mathcal{T}_{E_k} denotes the proportion of individuals in the E class who become infective before going to the Q class, i.e., the event $E \rightarrow I$ occurs before $E \rightarrow Q$. This quantity is given by

$$\mathcal{T}_{E_k} = \mathbb{P}(X \leq Z) = \sum_{j=1}^{\infty} (p_{j-1} - p_j) k_{j-1}.$$

The three quantities in (4.56) represent the stage-specific reproduction numbers representing, respectively, the average number of secondary infections produced by individuals in the I stage, in the H stage with quarantine (QH) and without (IH).

Question 2 Let \mathcal{R}_c be as given in (4.55). Show that the epidemic final size generated by the dynamics of model (4.53) satisfies the following final size relationship:

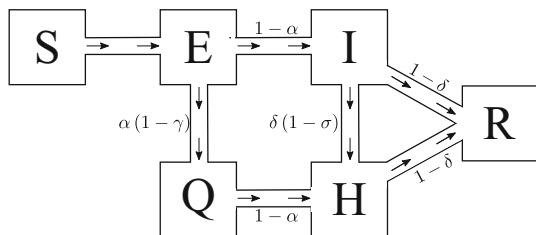
$$\log \frac{S_0}{S_\infty} = \left(1 - \frac{S_\infty}{N}\right) \mathcal{R}_c. \quad (4.57)$$

Question 3 Consider the case where the distributions in (4.52) are geometric. Assume that $X \sim \text{Geom}(\alpha)$, $Y \sim \text{Geom}(\delta)$, $Z \sim \text{Geom}(\gamma)$, and $W \sim \text{Geom}(\sigma)$. That is,

$$p_i = \alpha^i, \quad q_i = \delta^i, \quad k_i = \gamma^i, \quad l_i = \sigma^i, \quad i = 0, 1, 2, \dots.$$

In this case, the transition diagram 4.7 can be replaced by that shown in Fig. 4.8

Fig. 4.8 Transition diagram for the model with geometric stage distributions



Show that under this assumption, the general model (4.53) simplifies to

$$\begin{aligned} S_{n+1} &= S_n G_n \\ E_{n+1} &= (1 - G_n)S_n + \alpha\gamma E_n \\ Q_{n+1} &= \alpha(1 - \gamma)E_n + \alpha Q_n \\ I_{n+1} &= (1 - \alpha)E_n + \delta\sigma I_n \\ H_{n+1} &= (1 - \alpha)Q_n + \delta(1 - \sigma)I_n + \delta H_n, \quad n = 0, 1, 2, 3, \dots \end{aligned} \tag{4.58}$$

Note that the system (4.58) has the same form as the usually considered model with constant probabilities for transitions between stages. We refer to this model as the geometric distribution model (GDM). The questions below concern the GDM (4.58).

- (a) Determine the quantities \mathcal{D}_E , \mathcal{D}_I , \mathcal{D}_{E_k} , \mathcal{D}_{I_l} , and \mathcal{T}_{E_k} .
- (b) Compute \mathcal{R}_I , \mathcal{R}_{IH} , \mathcal{R}_{QH} , and the control reproduction number \mathcal{R}_c .

Question 4 We can examine how different distribution assumptions may influence the values of the reproduction number \mathcal{R}_c given in (4.55) and (4.56). Particularly, the reduction in \mathcal{R}_c by quarantine and isolation can be different under different distribution assumptions. To examine this, we can consider two specific distributions, geometric distribution assumption (GDA) and Poisson distribution assumption (PDA). For the purpose of comparison, we take X_g and X_p with mean μ_1 and Y_g and Y_p with mean μ_2 (the subscripts g and p represent geometric and Poisson, respectively). The quarantine (Z) and isolation (W) are assumed to have geometric distributions with parameters γ or σ , respectively. Consider the parameter values $\beta = 0.75$, $\rho = 0.95$, $\mu_1 = 5$, $\mu_2 = 10$. Compare two strategies: Strategy I corresponds to $\gamma = 0.5$ and $\sigma = 0.8$ and Strategy II corresponds to $\gamma = 0.8$ and $\sigma = 0.5$. Figure 4.9 plots the control reproduction number \mathcal{R}_c vs γ or σ . The left figure is for $\mathcal{R}_{c,g}$ (i.e., when X and Y are geometric distributions) and the right figure is for $\mathcal{R}_{c,p}$ (i.e., when X and Y are Poisson distributions). Figure 4.9 shows reduction in \mathcal{R}_c under Strategy I (represented by a dot on the dashed curve) and Strategy II (represented by a solid square on the solid curve).

- (a) Reproduce Fig. 4.9.
- (b) Under the GDA, which of the strategies I and II is more effective in reducing the control reproduction number?
- (c) Under the PDA, which of the strategies I and II is more effective in reducing the control reproduction number?
- (d) Do the two distribution assumptions generate the same assessment? Why or why not?
- (e) The formula (4.57) holds for arbitrarily distributed latent and infectious periods. Because reductions of \mathcal{R}_c by quarantine and isolation depend on distribution assumptions, compare the effects of these control measures on the epidemic final size under the GDA and PDA.

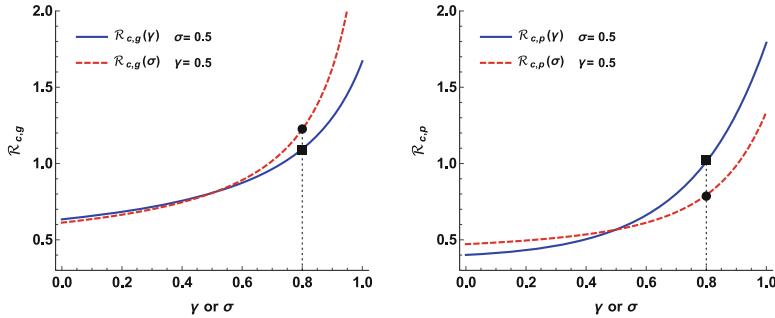


Fig. 4.9 Comparison of the GDM and PDM in terms of their evaluations on various control measures when the corresponding control reproduction numbers, $\mathcal{R}_{c,g}$ and $\mathcal{R}_{c,p}$ are used. The two control strategies considered are represented by $\gamma = 0.5$ and $\sigma = 0.8$ (labeled with a dot on the dashed curve) and by $\gamma = 0.8$ and $\sigma = 0.5$ (labeled with a solid square on the solid curve)

Project 2 Consider the epidemic model (4.26), which includes general distributions for the latent and infectious stages with survival function $P_i(s)$ ($i = E, I$), as well as a control measures quarantine and isolation represented by parameters ψ and φ . The model also assumes that the isolation is 100% effective, i.e., the isolated individuals do not contribute to the disease transmission. Let $q_i(s) = -P'_i(s)$ ($i = E, I$). The corresponding reproduction number is

$$\mathcal{R}_c(\psi, \varphi) = \beta N \int_0^\infty e^{-\psi y} q_E(y) dy \int_0^\infty e^{-\varphi u} P_I(u) du, \quad (4.59)$$

or in terms of the Laplace transforms of q_E and P_I

$$\mathcal{R}_c(\psi, \varphi) = \beta N \mathcal{L}_{q_E}(\psi) \mathcal{L}_{P_I}(\varphi). \quad (4.60)$$

In the more general case of imperfect isolation, the isolation may not be 100% effective. Let σ denote the isolation efficiency for isolated individuals with $0 \leq \sigma \leq 1$. Then the effective transmission rate for the isolated (or hospitalized) individuals (H) is reduced to $(1 - \sigma)\beta$. In this case, the S equation in (4.26) becomes

$$S' = -\beta S[I + (1 - \sigma)H]. \quad (4.61)$$

Question 1 Let μ_I denote the mean infective period, i.e., $\mu_I = \int_0^\infty P_I(s)ds$. Show that the formula (4.60) can be extended to be a function of ψ , φ , and σ as below:

$$\mathcal{R}_c(\psi, \varphi, \sigma) = (1 - \sigma)\beta N \mathcal{L}_{q_E}(\psi) \mu_I + \sigma \beta N \mathcal{L}_{q_E}(\psi) \mathcal{L}_{P_I}(\varphi). \quad (4.62)$$

Question 2 If the stage durations are both gamma distributed with shape parameters $k_E \geq 1$ and $k_I \geq 1$, i.e.,

$$q_i(s, \mu_i, k_i) = \frac{1}{\Gamma(k_i)} \frac{k_i}{\mu_i} \left(\frac{k_i}{\mu_i} s \right)^{k_i-1} e^{-\frac{k_i}{\mu_i} s}, \quad s > 0, \quad i = E, I.$$

- (a) Show that in the absence of quarantine (i.e., $\phi = 0$) the control reproduction number is

$$\mathcal{R}_c(0, \varphi, \sigma) = (1 - \sigma)\beta N \mu_I + \frac{\sigma \beta N}{\varphi} \left[1 - \left(1 + \frac{\mu_I}{k_I} \varphi \right)^{-k_I} \right]. \quad (4.63)$$

- (b) Let $\beta N = 0.5$. Plot $\mathcal{R}_c(0, \varphi, \sigma)$ given in (4.63) for various sets of values of $k_E = k_I = k$, μ_I , and σ_I . For example, $k = 1, 2, 4$, $\mu_I = 3, 5, 8$, and $\sigma = 0.1, 0.5, 0.8, 1$. Summarize the observations from these plots in terms of the dependence of \mathcal{R}_c on the shape parameter, mean period of infection, and isolation efficiency.
- (c) Let $k_E = k_I = k$. Note that $k = 1$ corresponds to an exponentially distributed stage, which is the most commonly assumed distribution. For the two values $k = 1$ and 4 , draw the contour plots of $\mathcal{R}_c(0, \varphi, \sigma)$ in the (φ, σ) plane showing a few contour curves including the one corresponding to $\mathcal{R}_c(0, \varphi, \sigma) = 1$. Other parameter values are $\beta N = 0.5$, $\mu_I = 3$, $\sigma = 1$. Discuss the observed difference from the contour plots between the cases of $k = 1$ and $k = 4$ in terms of the threshold values of φ for which $\mathcal{R}_c(0, \varphi, \sigma) = 1$.

4.12 Project: Epidemic Models with Direct and Indirect Transmission

Some diseases may be spread in more than one manner. For example, cholera may be spread by person to person transmission but may also be transmitted indirectly through a pathogen released by infectives through a medium such as contaminated water [38, 41].

Consider an epidemic model with direct (person to person) and indirect (through a medium such as contaminated water) transmission. To a simple *SIR* model we add a pathogen *B* shed by infectives. We assume that the infectivity of the pathogen is proportional to its concentration, suggesting mass action transmission. The resulting model is

$$\begin{aligned} S' &= -\beta_1 SI - \beta_2 SB \\ I' &= \beta_1 SI + \beta_2 SB - \gamma I \\ R' &= \gamma I \\ B' &= rI - \delta B, \end{aligned} \quad (4.64)$$

with initial conditions

$$S(0) = S_0, \quad I(0) = I_0, \quad B(0) = B_0,$$

in a population of constant total size $N = S_0 + I_0$, with $R(0) = 0$. In general, $N = S + I + R$. In this model r represents the rate at which an infectious individual sheds pathogen and δ represents the rate at which the pathogen loses infectivity.

Question 1 Show that the basic reproduction number

$$\mathcal{R}_0 = \frac{\beta_1 N}{\gamma} + \frac{r\beta_2 N}{\gamma\delta}.$$

In this expression, the first term represents secondary infections caused directly by a single infective introduced into a wholly susceptible population, infecting βN susceptibles in unit time for a time period $1/\gamma$. The second term represents secondary infections caused indirectly through the pathogen since a single infective sheds a quantity r of pathogen in unit time for a time period $1/\gamma$ and this pathogen infects βN susceptibles in unit time for a time period $1/\delta$.

Question 2 Derive the final size relation

$$\begin{aligned} \log \frac{S_0}{S_\infty} &= \left(\beta_1 + \frac{\beta_1 N r}{\gamma \delta} \right) \left[1 - \frac{S_\infty}{N} \right] + \beta_1 \frac{B_0}{\delta} \\ &= \mathcal{R}_0 \left[1 - \frac{S_\infty}{N} \right] + \beta_1 \frac{B_0}{\delta}. \end{aligned} \tag{4.65}$$

This implies $S_\infty > 0$.

In order to cover such generalizations of the model (4.1) as multiple infective stages and arbitrary distributions of stay in a stage, we give an age of infection model

$$\begin{aligned} S'(t) &= -S(t)[\beta_1 \varphi(t) + \beta_2 B(t)] \\ \varphi(t) &= \varphi_0(t) + \int_0^t [-S'(\tau)] P(\tau) d\tau \\ B(t) &= B_0(t) + \int_0^t r \varphi(\tau) Q(\tau) d\tau. \end{aligned} \tag{4.66}$$

In this model, $\varphi(t)$ represents the total infectivity of individuals with age of infection t , $\varphi_0(t)$ represents the total infectivity at time t of individuals who were already infected at time $t = 0$, $B_0(t)$ represents the pathogen concentration at time t remaining from pathogen concentration that was already present at time $t = 0$, $P(\tau)$ represents the mean infectivity of individuals at age of infection τ , normally the product of the fraction of infectives still infective at age of infection τ and the

relative infectivity at that infection age, and $Q(\tau)$ represents the fraction of pathogen remaining τ time units after having been shed by an infective. The function Q is monotone non-increasing with $Q(0) = 1$, $\int_0^\infty Q(\tau)d\tau < \infty$. Since infectivity of an individual may depend on the age of infection of the individual, the function P is not necessarily non-increasing, but we assume $\int_0^\infty P(\tau)d\tau < \infty$.

Question 3 Show that for the model (4.66), the basic reproduction number is

$$\mathcal{R}_0 = \beta_1 N \int_0^\infty P(\tau)d\tau + r\beta_2 N \int_0^\infty P(\tau)d\tau \int_0^\infty Q(\tau)d\tau.$$

In this expression, the first term represents new infection transmitted directly by a single infectious individual inserted into a totally susceptible population, while the second term represents secondary infections caused by this individual indirectly through shedding of pathogen.

Question 4 Obtain the final size relation

$$\begin{aligned} \log \frac{S_0}{S_\infty} = \mathcal{R}_0 \left[\frac{S_0 - S_\infty}{N} \right] + \beta_1 \int_0^\infty \varphi_0(t)dt \\ + r\beta_2 \int_0^\infty Q(t)dt \int_0^\infty \varphi_0(t)dt + \beta_2 \int_0^\infty B_0(t)dt. \end{aligned} \quad (4.67)$$

If all infections at time zero have infection age zero, then

$$\varphi_0(t) = [N - S_0]P(t), \quad \int_0^\infty \varphi_0(t)dt = [N - S_0] \int_0^\infty P(t)dt,$$

and if the entire pathogen concentration at time zero has infection age zero, then

$$B_0(t) = B_0 Q(t), \quad \int_0^\infty B_0(t)dt = B_0 \int_0^\infty Q(t)dt$$

with some constant B_0 . In this case, the final size relation (9.10) takes the form

$$\log \frac{S_0}{S_\infty} = \mathcal{R}_0 \left[1 - \frac{S_\infty}{N} \right] + \beta_2 B_0 \int_0^\infty Q(t)dt. \quad (4.68)$$

The final size relation has a term arising from an initial pathogen concentration that tends to decrease S_∞ .

In general, because Q is monotone non-increasing,

$$\int_0^\infty B_0(t)dt \leq B_0 \int_0^\infty Q(t)dt.$$

If P is monotone non-increasing,

$$\int_0^\infty \varphi_0(t)dt \leq [N - S_0] \int_0^\infty P(t)dt.$$

If P is not monotone, which may occur, for example, if there is an exposed stage followed by an infective stage with higher infectivity, this is not necessarily true. However, if there are no infectives initially, so that the epidemic is started by the pathogen, then $\varphi_0(t) = 0$ and $S_0 = N$. Then (4.67) remains valid without the need to assume that P is monotone.

These results have been established only for a constant rate of pathogen shedding. If the rate of pathogen depends on the age of infection, the equation for B in the model (4.66) should be replaced by an equation

$$B(t) = B_0(t) + \int_0^t r(t-\tau)\varphi(t-\tau)Q(\tau)d\tau.$$

It is not possible to treat the corresponding model as an age of infection model, but we can view it as a staged progression model.

We consider an epidemic with progression from S through k infected stages I_1, I_2, \dots, I_k , as analyzed in [7], but with the addition of a pathogen. We assume that in stage j the relative infectivity is ε_j , the distribution of stay in the stage is given by P_j with $P_j(0) = 1$, $\int_0^\infty P_j(t)dt < \infty$, and P_j monotone non-increasing, so that the infectivity of an individual in stage j is $A_j(\tau) = \varepsilon_j P_j(\tau)$. There are no disease deaths and the total population size N is constant. We assume initial conditions

$$S(0) = 0, \quad I_1(0) = I_0, \quad I_2(0) = I_3(0) = \dots = I_k(0) = 0, \quad R(0) = 0.$$

The total infectivity is given by

$$\varphi(t) = \sum_{j=1}^k \varepsilon_j I_j(t).$$

We let $B_j(t)$ be the quantity of pathogen shed by infectives in the stage I_j and let Q_j denote the distribution of stay of pathogen shed by infectives in this stage, with $Q_j(0) = 1$, $\int_0^\infty Q_j(t)dt < \infty$, and Q_j monotone non-increasing. We let r_j be the shedding rate in this stage. We also define the total quantity of pathogen,

$$B(t) = \sum_{j=1}^k B_j(t).$$

Then

$$B_j(t) = B_j^0(t) + \int_0^t r_j I_j(t-\tau)Q_j(\tau)d\tau. \quad (4.69)$$

Question 5 Show that the basic reproduction number is

$$\mathcal{R}_0 = \beta_1 N \sum_{j=1}^k \varepsilon_j \int_0^\infty P_j(t) dt + \beta_2 N \sum_{j=1}^k r_j \int_0^\infty P_j(t) dt \int_0^\infty Q_j(t) dt.$$

Question 6 For simplicity, we assume that all individuals infected at time zero have infection age zero for $t = 0$, and also that there is a new quantity of pathogen B_0 introduced at time zero, so that $B_0(t) = B_0 Q(t)$. Obtain the final size relation

$$\begin{aligned} \log \frac{S_0}{S_\infty} &= \beta_1 \sum_{j=1}^k \varepsilon_j \int_0^\infty P_j(t) dt [N - S_\infty] \\ &\quad + \beta_2 \sum_{j=1}^k r_j \int_0^\infty P_j(t) dt \int_0^\infty Q_j(t) dt [N - S_\infty] \\ &\quad + \beta_2 B_0 \int_0^\infty Q(t) dt \\ &= \mathcal{R}_0 \left[1 - \frac{S_\infty}{N} \right] + \beta_2 B_0 \int_0^\infty Q(t) dt. \end{aligned} \tag{4.70}$$

References [8]. Other sources of information about cholera include [1, 2, 11, 21, 25, 38, 41].

4.13 Exercises

1. Show that it is not possible for a major epidemic to develop unless at least one member of the contact network has degree at least 3.
2. What is the probability of a major epidemic if every member of the contact network has degree 3?
3. Show that if $G_1(0) = 0$, then
 - (a) $G'_1(0) \leq 0$
 - (b) $G_1(z) < z$ for $0 \leq z \leq 1$
 - (c) $z_\infty = 0$
 - (d) $\mathcal{R}_0 > 1$
4. Consider a truncated Poisson distribution, with

$$p_k = \begin{cases} \frac{e^{-c} c^k}{k!}, & k \leq 10, \\ 0, & k > 10. \end{cases}$$

Estimate (numerically) the probability of a major epidemic if $c = 1.5$.

5. Show that the probability generating function for an exponential distribution, given by

$$p_k = (1 - e^{-1/r})e^{-k/r},$$

is

$$G_0(z) = \frac{1 - e^{-1/r}}{1 - ze^{-1/r}}.$$

6. A power law distribution is given by

$$p_k = Ck^{-\alpha}.$$

For what values of α is it possible to normalize this (i.e., choose C to make $\sum p_k = 1$)?

7. Consider a network in which the contacts between members of the population are random. Assume that the fraction of vertices having degree k , or the probabilities p_k are given by the Poisson distribution $p_k = e^c c^k / k!$, where $c = \mathcal{R}_0 = 3$. Let \mathbf{P} denote the probability of a minor outbreak.
- (a) Determine the equation that can be used to solve for \mathbf{P} .
 - (b) Find \mathbf{P} .
 - (b) Find the probability of a major epidemic.
8. Compare the qualitative behaviors of the models

$$S' = -\beta SI, \quad I' = \beta SI - \alpha I,$$

and

$$S' = -\beta SI, \quad E' = \beta SI - \kappa E, \quad I' = \kappa E - \alpha I,$$

with

$$\beta N = 1/3, \quad \alpha = 1/6, \quad \kappa = 1/2, \quad S(0) = 999, \quad I(0) = 1.$$

These models represent an SIR epidemic model and an SEIR epidemic model respectively with a mean infective period of 6 days and a mean exposed period of 2 days. Do numerical simulations to decide whether the exposed period affects the behavior of the model noticeably.

9. Consider three basic epidemic models—the simple SIR model,

$$S' = -\beta SI,$$

$$I' = \beta SI - \alpha I,$$

the SEIR model with some infectivity in the exposed period,

$$\begin{aligned} S' &= -\beta[S(I + \varepsilon E)], \\ E' &= \beta[S(I + \varepsilon E) - \kappa E], \\ I' &= \kappa E - \alpha I, \end{aligned}$$

and the SIR model with treatment,

$$\begin{aligned} S' &= -\beta S(I + \delta T), \\ I' &= \beta S(I + \delta T) - (\alpha + \varphi)I, \\ T' &= \varphi I - \eta T. \end{aligned}$$

Use the parameter values

$$\beta N = \frac{1}{3}, \quad \alpha = \frac{1}{4}, \quad \varepsilon = \frac{1}{2}, \quad \kappa = \frac{1}{2}, \quad \delta = \frac{1}{2}, \quad \eta = \frac{1}{4}, \quad \varphi = 1,$$

and the initial values

$$S(0) = 995, \quad E(0) = 0, \quad I(0) = 5, \quad T(0) = 0.$$

For each model,

- (a) Calculate the reproduction number and the epidemic final size.
 - (b) Do some numerical simulations to obtain the epidemic size by determining the change in S , the maximum number of infectives by measuring I , and the duration of the epidemic.
10. Consider the treatment model (4.12). Assume that $\beta N = 860$, $\delta = 0.5$, $\eta = 1/10$, and $S_0 = N - 1$.
- (a) Explore numerically how the final epidemic size can be affected by the treatment effort (measured by the parameter γ). Plot the relationship.
 - (b) Determine the final epidemic size without treatment.
 - (c) Is it possible to reduce the final size by 50% by increasing γ ? If yes, find that γ value.
11. For the model (4.15), what should be the constraints on parameter values so that passage from E to Q to J to R is at least as rapid as the passage from E to I to R , and the passage from I to J to R is at least as rapid as the passage from I to R ?
12. Consider the model (4.15) with quarantine and isolation. Assume that $\beta N = 860$, $\kappa_E = 1/4$, $\alpha_I = 1/10$, $\varepsilon_E = 0.4$, $\varepsilon_Q = 0.1$, $\varepsilon_J = 0.6$.

- (a) In the absence of control (i.e., $\gamma_Q = \gamma_J = \kappa_Q = \alpha_J = 0$), what is the final epidemic size?
 (b) Plot the final size as a function of γ_Q and γ_J .
 (c) Is it possible to reduce the final size by 50% by increasing γ_Q and/or γ_J ? If yes, identify one pair of values of γ_Q and γ_J to achieve the goal.
13. Formulate an *SEIERT* model, that is, a model with both an exposed class and a treatment class
- (a) Draw a flow chart and calculate the basic reproduction number
 (b) Determine the final size relation
14. Consider an SIR model in which a fraction θ of infectives is isolated in a perfectly quarantined class Q with standard incidence (meaning that individuals make a contacts in unit time of which a fraction $I/(N - Q)$ is infective), given by the system

$$\begin{aligned}S' &= -\beta NS \frac{I}{N - Q}, \\I' &= \beta NS \frac{I}{N - Q} - (\theta + \alpha)I, \\Q' &= \theta I - \gamma Q, \\R' &= \alpha I + \gamma Q.\end{aligned}$$

- (a) Find the equilibria.
 (b) Find the basic reproduction number \mathcal{R}_0 .
 (c) For parameters, take $\alpha = 0.1$, $\theta = 1, 2, 4$, $\gamma = 0.2$. Sketch the phase plane of the system and observe what is happening.
15. Isolation/quarantine is a complicated process because we don't live in a perfect world. In hospitals, patients may inadvertently or deliberately break from isolation and in the process have casual contacts with others including medical personnel and visitors. Taking this into account, we are led to the model

$$\begin{aligned}S' &= -\beta NS \frac{[I + \rho \tau Q]}{N - \sigma Q}, \\I' &= \beta NS \frac{[I + \rho \tau Q]}{N - \sigma Q} - (\theta + \alpha)I, \\Q' &= \theta I - \gamma Q, \\R' &= \alpha I + \gamma Q.\end{aligned}$$

- (a) Determine all the parameters in the system and define each parameter.
 (b) Show that the population is constant.

- (c) Find all equilibria.
 (d) Find the reproduction number \mathcal{R}_0 .
 (e) Describe the asymptotic behavior of the model, including its dependence on the basic reproduction number.
16. Formulate a model analogous to (4.12) for which treatment is not started immediately, but begins at time $\tau > 0$. Can you say anything about the dependence of the reproduction number on τ ?
 17. In the simple *SIR* model (4.1) it is assumed that all effective contacts between susceptibles and infectives lead to new infections. Suppose, however, that only a fraction δ of contacts by infectives actually transmit disease and only a fraction σ of contacts of susceptibles actually lead to new infections. This would lead to a model

$$\begin{aligned} S' &= -\delta\sigma\beta SI \\ I' &= \delta\sigma\beta SI - \alpha I. \end{aligned}$$

Find the basic reproduction number and the final size relation for this model.

18. Consider the model in Sect. 4.4 and assume that $P(\tau)$ has the value 1 if $0 \leq \tau \leq T$ and 0 otherwise.
 (a) Calculate the basic reproduction number
 (b) Plot the final epidemic size for several different values of T .
19. Consider the model in Sect. 4.4 and assume that $P(\tau)$ follows a gamma distribution with mean equal to 10 and shape parameter equal to n .
 (a) Plot final epidemic size for $n = 1, 3, 5, 7, 9$.
 (b) Does the final size change with n ? If yes, what is the observed pattern of the change?
20. The quarantine–isolation model (4.26) is not quite in the age of infection form that we have been studying in this section because of the initial terms in the equations. Transform it to age of infection form by incorporating the initial terms into infinite integrals and calculate the control reproduction number from this form.
21. Determine the initial exponential growth rate for the model (4.11) by interpreting it as an age of infection model.
22. Determine the initial exponential growth rate for the model (4.12) by interpreting it as an age of infection model.
23. Determine the initial exponential growth rate for the general treatment model (4.24) of Sect. 4.5.
24. Consider the model (4.23) with $A(\tau) = e^{-0.1\tau}$.
 (a) Determine the critical value of a_c such that there is an epidemic if $a > a_c$ and no epidemic if $a < a_c$.
 (b) Is there an epidemic for $a = 1.5$? If yes, find the exponential growth rate of the epidemic.

25. Find the solution of the system (4.29).
26. Verify the relation (4.30).
27. Verify the Eq. (4.32).
28. Determine the mean generation time for an *SEIR* model with an infectious period having a gamma distribution with parameter n .
29. Establish the relation (4.49).
30. Consider the model (4.50) with initial conditions given in (4.51) with $t_0 = 30$. The parameter values c_0 , ε , and α are the same as those used in Fig. 4.6. Make plots similar to those in Fig. 4.6 under the following scenarios.
 - (a) Plot the epidemic curves for three values of vaccination fraction p : 0.05, 0.1, and 0.15. Compare the shape of each epidemic curve with that for the case of no vaccination (i.e., $p = 0$).
 - (b) Plot the cumulative cases for $p = 0.05$, 0.1, and 0.15. Compare each case with the case of no vaccination.
 - (c) How many epidemic waves do you observe in each of the three cases?
 - (d) Is it true that the higher the fraction of people vaccinated the lower the final size will be?
31. Repeat Exercise 30 but for $t_0 = 40$. Describe the qualitative differences in epidemic curves and cumulative cases between the case $t_0 = 30$ and the case $t_0 = 40$? What are the implications of these differences for public health policy decisions?

References

1. Alexanderian, A., M.K. Gobbert, K.R. Fister, H. Gaff, S. Lenhart, and E. Schaefer (2011) An age-structured model for the spread of epidemic cholera: Analysis and simulation, Nonlin. Anal. Real World Appl. **12**: 3483–3498.
2. Andrews, J.R. & and S. Basu (2011) Transmission dynamics and control of cholera in Haiti: an epidemic model, Lancet **377**: 1248–1255.
3. Arino, J., F. Brauer, P. van den Driessche, J. Watmough & J. Wu (2006) Simple models for containment of a pandemic, J. Roy. Soc. Interface, **3**: 453–457.
4. Arino, J. F. Brauer, P. van den Driessche, J. Watmough & J. Wu (2007) A final size relation for epidemic models, Math. Biosc. & Eng. **4**: 159–176.
5. Arino, J., F. Brauer, P. van den Driessche, J. Watmough & J. Wu (2008) A model for influenza with vaccination and antiviral treatment, Theor. Pop. Biol. **253**: 118–130.
6. Bansal, S., J. Read, B. Pourbohloul, and L.A. Meyers (2010) The dynamic nature of contact networks in infectious disease epidemiology, J. Biol. Dyn., **4**: 478–489.
7. Brauer, F., C. Castillo-Chavez, and Z. Feng (2010) Discrete epidemic models, Math. Biosc. & Eng. **7**: 1–15.
8. Brauer, F., Z. Shuai, & P. van den Driessche (2013) Dynamics of an age-of-infection cholera model, Math. Biosc. & Eng. **10**: 1335–1349.
9. Brauer, F., P. van den Driessche and J. Wu, eds. (2008) Mathematical Epidemiology, Lecture Notes in Mathematics, Mathematical Biosciences subseries 1945, Springer, Berlin - Heidelberg - New York.

10. Callaway, D.S., M.E.J. Newman, S.H. Strogatz, D.J. Watts (2000) Network robustness and fragility: Percolation on random graphs, *Phys. Rev. Letters*, **85**: 5468–5471.
11. Codeço, C.T. (2001) Endemic and epidemic dynamics of cholera: the role of the aquatic reservoir, *BMC Infectious Diseases* **1**: 1.
12. Diekmann, O. & J.A.P. Heesterbeek (2000) *Mathematical Epidemiology of Infectious Diseases*. Wiley, Chichester.
13. Erdős, P. & A. Rényi (1959) On random graphs, *Publicationes Mathematicae* **6**: 290–297.
14. Erdős, P. & A. Rényi (1960) On the evolution of random Pub. Math. Inst. Hung. Acad. Science **5**: 17–61.
15. Erdős, P. & A. Rényi (1961) On the strengths of connectedness of a random graph, *Acta Math. Scientiae Hung.* **12**: 261–267.
16. Feng, Z. (2007) Final and peak epidemic sizes for *SEIR* models with quarantine and isolation, *Math. Biosc. & Eng.* **4**: 675–686.
17. Feng, Z., D. Xu & W. Zhao (2007) Epidemiological models with non-exponentially distributed disease stages and applications to disease control, *Bull. Math. Biol.* **69**: 1511–1536.
18. Fine, P.E.M. (2003) The interval between successive cases of an infectious disease, *Am. J. Epid.* **158**: 1039–1047.
19. Gumel, A., S. Ruan, T. Day, J. Watmough, P. van den Driessche, F. Brauer, D. Gabrielson, C. Bowman, M.E. Alexander, S. Ardal, J. Wu & B.M. Sahai (2004) Modeling strategies for controlling SARS outbreaks based on Toronto, Hong Kong, Singapore and Beijing experience, *Proc. Roy. Soc. London* **271**: 2223–2232.
20. Diekmann, O., J.A.P. Heesterbeek, & J.A.J. Metz (1995) The legacy of Kermack and McKendrick, in *Epidemic Models: Their Structure and Relation to Data*, D. Mollison, ed., Cambridge University Press, pp. 95–115.
21. Hartley, D.M., J.G. Morris Jr. & D.L. Smith (2006) Hyperinfectivity: a critical element in the ability of *V. cholerae* to cause epidemics? *PLOS Med.* **3**:63–69.
22. Heffernan, J.M., R.J. Smith? & L.M. Wahl (2005) Perspectives on the basic reproductive ratio, *J. Roy. Soc. Interface* **2**: 281–293.
23. Hyman, J.M., J. Li & E. A. Stanley (1999) The differential infectivity and staged progression models for the transmission of HIV, *Math. Biosc.* **155**: 77–109.
24. Kermack, W.O. & A.G. McKendrick (1927) A contribution to the mathematical theory of epidemics. *Proc. Royal Soc. London*. **115**: 700–721.
25. King, A.A. E.L. Ionides, M. Pascual, & M.J. Bouma (2008) Inapparent infectious and cholera dynamics, *Nature* **454**: 877–890.
26. Lloyd, A.L. (2001) Realistic distributions of infectious periods in epidemic models: Changing patterns of persistence and dynamics, *Theor. Pop. Biol.*, **60**: 59–71.
27. Meyers, L.A. (2007) Contact network epidemiology: Bond percolation applied to infectious disease prediction and control, *Bull. Am. Math. Soc.*, **44**: 63–86.
28. Meyers, L.A., M.E.J. Newman and B. Pourbohloul(2006) Predicting epidemics on directed contact networks, *J. Theor. Biol.*, **240**: 400–418.
29. Meyers, L.A., B. Pourbohloul, M.E.J. Newman, D.M. Skowronski, and R.C. Brunham (2005) Network theory and SARS: predicting outbreak diversity, *J. Theor. Biol.*, **232**: 71–81.
30. Miller, J.C. (2010) A note on a paper by Erik Volz: SIR dynamics in random networks. *J. Math. Biol.* <https://doi.org/10.1007/s00285-010-0337-9>.
31. Miller, J.C. and E. Volz (2011) Simple rules govern epidemic dynamics in complex networks, to appear.
32. Newman, M.E.J. (2002) The spread of epidemic disease on networks, *Phys. Rev. E* **66**, 016128.
33. Newman, M.E.J. (2003) The structure and function of complex networks. *SIAM Review*, **45**: 167–256.
34. Newman, M.E.J., S.H. Strogatz & D.J. Watts (2001) Random graphs with arbitrary degree distributions and their applications, *Phys. Rev. E* **64**.

35. Riley, S., C. Fraser, C.A. Donnelly, A.C. Ghani, L.J. Abu-Raddad, A.J. Hedley, G.M. Leung, L-M Ho, T-H Lam, T.Q. Thach, P. Chau, K-P Chan, S-V Lo, P-Y Leung, T. Tsang, W. Ho, K-H Lee, E.M.C. Lau, N.M. Ferguson, & R.M. Anderson (2003) Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions, *Science* **300**: 1961–1966.
36. Roberts, M.G. and J.A.P. Heesterbeek (2003) A new method for estimating the effort required to control an infectious disease, *Proc. Roy. Soc. London B* **270**: 1359–1364.
37. Scalia-Tomba, G., A. Svensson, T. Asikainen, and J. Giesecke (2010) Some model based considerations on observing generation times for communicable diseases, *Math. Biosc.* **223**: 24–31.
38. Shuai, Z. & P. van den Driessche (2011) Global dynamics of cholera models with differential infectivity, *Math. Biosc.* **234**: 118–126.
39. Strogatz, S.H. (2001) Exploring complex networks, *Nature*, **410**: 268–276.
40. Svensson, A. (2007) A note on generation time in epidemic models, *Math. Biosc.* **208**: 300–311.
41. Tien, J.H. & D.J.D. Earn (2010) Multiple transmission pathways and disease dynamics in a waterborne pathogen model, *Bull. Math. Biol.* **72**: 1506–1533.
42. van den Driessche, P. & J. Watmough (2002) Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math. Biosc.*, **180**: 29–48.
43. Volz, E. (2008) SIR dynamics in random networks with heterogeneous connectivity, *J. Math. Biol.*, **56**: 293–310.
44. Wallinga, J. & M. Lipsitch (2007) How generation intervals shape the relationship between growth rates and reproductive numbers, *Proc. Royal Soc. B* **274**: 599–604.
45. Wallinga, J. & P. Teunis (2004) Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures, *Am. J. Epidemiol.* **160**: 509–516,
46. Wearing, H.J., P. Rohani & M. J. Keeling (2005) Appropriate models for the management of infectious diseases, *PLOS Medicine*, **2**: 621–627.
47. Yan, P. and Z. Feng (2010) Variability order of the latent and the infectious periods in a deterministic *SEIR* epidemic model and evaluation of control effectiveness *Math. Biosc.* **224**: 43–52.
48. Yang, C.K. and F. Brauer (2009) Calculation of \mathcal{R}_0 for age-of-infection models, *Math. Biosc. & Eng.* **5**: 585–599

Chapter 5

Models with Heterogeneous Mixing



5.1 A Vaccination Model

To cope with annual seasonal influenza epidemics there is a program of vaccination before the “flu” season begins. Each year a vaccine is produced aimed at protecting against the three influenza strains considered most dangerous for the coming season. We formulate a model to add vaccination to the simple *SIR* model under the assumption that vaccination reduces susceptibility (the probability of infection if a contact with an infected member of the population is made).

We consider a population of total size N and assume that a fraction γ of this population is vaccinated prior to a disease outbreak. Thus we have unvaccinated and vaccinated sub-populations of sizes $N_U = (1 - \gamma)N$ and $N_V = \gamma N$, respectively. We assume that vaccinated members have susceptibility to infection reduced by a factor σ , $0 \leq \sigma \leq 1$, with $\sigma = 0$ describing a perfectly effective vaccine and $\sigma = 1$ describing a vaccine that has no effect. We assume also that vaccinated individuals who are infected have infectivity reduced by a factor δ and both vaccinated and unvaccinated individuals have a recovery rate α . The number of contacts in unit time per member is a_U for unvaccinated individuals and a_V for vaccinated individuals. These may be equal.

In this chapter we will study models in which there is more than one susceptible or infective compartment, and it is convenient to formulate such models using the number of contacts in unit time instead of the number of contacts multiplied by the total population size. Thus, for example, instead of writing the simple epidemic model as

$$S' = \beta SI, \quad I' = \beta SI - \alpha I,$$

we would write it as

$$S' = -aS \frac{I}{N}, \quad I' = aS \frac{I}{N} - \alpha I.$$

We let S_U, S_V, I_U, I_V denote the number of unvaccinated susceptibles, the number of vaccinated susceptibles, the number of unvaccinated infectives, and the number of vaccinated infectives, respectively. The resulting model is

$$\begin{aligned} S'_U &= -a_U S_U \left[\frac{I_U}{N_U} + \delta \frac{I_V}{N_V} \right] \\ S'_V &= -\sigma a_V S_V \left[\frac{I_U}{N_U} + \delta \frac{I_V}{N_V} \right] \\ I'_U &= a_U S_U \left[\frac{I_U}{N_U} + \delta \frac{I_V}{N_V} \right] - \alpha I_U \\ I'_V &= \sigma a_V S_V \left[\frac{I_U}{N_U} + \delta \frac{I_V}{N_V} \right] - \alpha I_V. \end{aligned} \tag{5.1}$$

The initial conditions prescribe $S_U(0), S_V(0), I_U(0), I_V(0)$, with

$$S_U(0) + I_U(0) = N_U, \quad S_V(0) + I_V(0) = N_V.$$

Since the infection now is beginning in a population which is not fully susceptible, we speak of the *control reproduction number* \mathcal{R}_c rather than the basic reproduction number. However, as we will soon see, calculation of the control reproduction number will require a more general definition and a considerable amount of technical computation. The computation method is applicable to both basic and control reproduction numbers. We will use the term reproduction number to denote either a basic reproduction number or a control reproduction number. We are able to obtain final size relations without knowledge of the reproduction number but these final size relations do contain information about the reproduction number, and more.

Since S_U and S_V are decreasing non-negative functions they have limits $S_U(\infty)$ and $S_V(\infty)$, respectively, as $t \rightarrow \infty$. The sum of the equations for S_U and I_U in (5.1) is

$$(S_U + I_U)' = -\alpha I_U,$$

from which we conclude, just as in the analysis of the simple *SIR* model in Sect. 2.4, that $I_U(t) \rightarrow 0$ as $t \rightarrow \infty$, and that

$$\alpha \int_0^\infty I_U(t) dt = N_U - S_U(\infty). \tag{5.2}$$

Similarly, using the sum of the equations for S_V and I_V , we see that $I_V(t) \rightarrow 0$ as $t \rightarrow \infty$, and that

$$\alpha \int_0^\infty I_V(t)dt = N_V - S_V(\infty). \quad (5.3)$$

Integration of the equation for S_U in (5.1) and use of (5.2), (5.3) gives

$$\begin{aligned} \log \frac{S_U(0)}{S_U(\infty)} &= a_U \left[\int_0^\infty I_U(t)dt + \delta \int_0^\infty I_V(t)dt \right] \\ &= \frac{a_U}{\alpha} \left[1 - \frac{S_U(\infty)}{N_U} \right] + \frac{\delta a_U}{\alpha} \left[1 - \frac{S_V(\infty)}{N_V} \right]. \end{aligned} \quad (5.4)$$

A similar calculation using the equation for S_V gives

$$\log \frac{S_V(0)}{S_V(\infty)} = \frac{\sigma a_V}{\alpha} \left[1 - \frac{S_U(\infty)}{N_U} \right] + \frac{\delta \sigma a_V}{\alpha} \left[1 - \frac{S_V(\infty)}{N_V} \right]. \quad (5.5)$$

This pair of Eqs. (5.4) and (5.5) are the final size relations. They make it possible to calculate $S_U(\infty)$, $S_V(\infty)$ if the parameters of the model are known.

It is convenient to define the matrix

$$\mathcal{R} = \begin{bmatrix} \mathcal{R}_{11} & \mathcal{R}_{12} \\ \mathcal{R}_{21} & \mathcal{R}_{22} \end{bmatrix} = \begin{bmatrix} \frac{a_U}{\alpha} & \frac{\delta a_U}{\alpha} \\ \frac{\sigma a_V}{\alpha} & \frac{\delta \sigma a_V}{\alpha} \end{bmatrix}.$$

The element \mathcal{R}_{ij} can be interpreted as the average number of susceptibles of group i infected by an infective of type j over its infectious period.

Then the final size relations (5.4), (5.5) may be written as

$$\begin{aligned} \log \frac{S_U(0)}{S_U(\infty)} &= \mathcal{R}_{11} \left[1 - \frac{S_U(\infty)}{N_U} \right] + \mathcal{R}_{12} \left[1 - \frac{S_V(\infty)}{N_V} \right] \\ \log \frac{S_V(0)}{S_V(\infty)} &= \mathcal{R}_{21} \left[1 - \frac{S_U(\infty)}{N_U} \right] + \mathcal{R}_{22} \left[1 - \frac{S_V(\infty)}{N_V} \right]. \end{aligned} \quad (5.6)$$

The matrix \mathcal{R} is closely related to the reproduction number. In the next section we describe a general method for calculating reproduction numbers that will involve a matrix which is similar to this matrix.

5.2 The Next Generation Matrix and the Basic Reproduction Number

Up to this point, we have calculated reproduction numbers by following the secondary cases caused by a single infective introduced into a population. However, if there are sub-populations with different susceptibility to infection, as in the vaccination model introduced in Sect. 5.1, it is necessary to follow the secondary infections in the sub-populations separately, and this approach will not yield the reproduction number. It is necessary to give a more general approach to the meaning of the reproduction number, and this is done through the *next generation matrix* [18, 19, 45]. The underlying idea is that we must calculate the matrix whose (i, j) entry is the number of secondary infections caused in compartment i by an infected individual in compartment j . We will follow the development in [45, 46] for ordinary differential equation models, even though the approach of [18, 19] is more general.

In a compartmental disease transmission model, we sort individuals into compartments based on a single, discrete state variable. A compartment is called a *disease compartment* if the individuals therein are infected. Note that this use of the term disease is broader than the clinical definition and includes stages of infection such as exposed stages in which infected individuals are not necessarily infective. Suppose there are n disease compartments and m non-disease compartments, and let $x \in R^n$ and $y \in R^m$ be the sub-populations in each of these compartments. Further, we denote by \mathcal{F}_i the rate at which secondary infections increase the i^{th} disease compartment and by \mathcal{V}_i the rate at which disease progression, death, and recovery decrease the i^{th} compartment, that is, \mathcal{V}_i is the net outflow from the i^{th} compartment due to disease progression, death, and recovery, with inflow from other compartments yielding a negative contribution. The compartmental model can then be written in the form

$$\begin{aligned} x'_i &= \mathcal{F}_i(x, y) - \mathcal{V}_i(x, y), \quad i = 1, \dots, n, \\ y'_j &= g_j(x, y), \quad j = 1, \dots, m. \end{aligned} \tag{5.7}$$

Note that the decomposition of the dynamics into \mathcal{F} and \mathcal{V} and the designation of compartments as infected or uninfected may not be unique; different decompositions correspond to different epidemiological interpretations of the model. The definitions of \mathcal{F} and \mathcal{V} used here differ slightly from those in [45].

The derivation of the basic reproduction number is based on the linearization of the ODE model about a disease-free equilibrium. We make the following assumptions:

- $\mathcal{F}_i(0, y) = 0$ and $\mathcal{V}_i(0, y) = 0$ for all $y \geq 0$ and $i = 1, \dots, n$.
- The disease-free system $y' = g(0, y)$ has a unique equilibrium that is asymptotically stable, that is, all solutions with initial conditions of the form $(0, y)$ approach a point $(0, y_0)$ as $t \rightarrow \infty$. We refer to this point as the disease-free equilibrium.

The first assumption says that all new infections are secondary infections arising from infected hosts; there is no immigration of individuals into the disease compartments. It ensures that the disease-free set, which consists of all points of the form $(0, y)$, is invariant. That is, any solution with no infected individuals at some point in time will be free of infection for all time. The second assumption ensures that the disease-free equilibrium is also an equilibrium of the full system.

Next, we further assume that

- $\mathcal{F}_i(x, y) \geq 0$ for all non-negative x and y and $i = 1, \dots, n$.
- $\mathcal{V}_i(x, y) \leq 0$ whenever $x_i = 0$, $i = 1, \dots, n$.
- $\sum_{i=1}^n \mathcal{V}_i(x, y) \geq 0$ for all non-negative x and y .

The reasons for these assumptions are that the function \mathcal{F} represents new infections and cannot be negative, each component, \mathcal{V}_i , represents a net outflow from compartment i and must be negative (inflow only) whenever the compartment is empty, and the sum $\sum_{i=1}^n \mathcal{V}_i(x, y)$ represents the total outflow from all infected compartments. Terms in the model leading to increases in $\sum_{i=1}^n x_i$ are assumed to represent secondary infections and therefore belong in \mathcal{F} .

Suppose that a single infected person is introduced into a population originally free of disease. The initial ability of the disease to spread through the population is determined by an examination of the linearization of (5.7) about the disease-free equilibrium $(0, y_o)$. It is easy to see that the assumption $\mathcal{F}_i(0, y) = 0$, $\mathcal{V}_i(0, y) = 0$ implies

$$\frac{\partial \mathcal{F}_i}{\partial y_j}(0, y_o) = \frac{\partial \mathcal{V}_i}{\partial y_j}(0, y_o) = 0$$

for every pair (i, j) . This implies that the linearized equations for the disease compartments, x , are decoupled from the remaining equations and can be written as

$$x' = (F - V)x, \quad (5.8)$$

where F and V are the $n \times n$ matrices with entries

$$F = \frac{\partial \mathcal{F}_i}{\partial x_j}(0, y_o) \quad \text{and} \quad V = \frac{\partial \mathcal{V}_i}{\partial x_j}(0, y_o).$$

Because of the assumption that the disease-free system $y' = g(0, y)$ has a unique asymptotically stable equilibrium, the linear stability of the system (5.7) is completely determined by the linear stability of the matrix $(F - V)$ in (5.8).

The number of secondary infections produced by a single infected individual can be expressed as the product of the expected duration of the infective period and the rate at which secondary infections occur. For the general model with n disease compartments, these are computed for each compartment for a hypothetical index case. The expected time the index case spends in each compartment is given by the integral $\int_0^\infty \phi(t, x_0) dt$, where $\phi(t, x_0)$ is the solution of (5.8) with $F = 0$

(no secondary infections) and non-negative initial conditions, x_0 , representing an infected index case:

$$x' = -Vx, \quad x(0) = x_0. \quad (5.9)$$

In effect, this solution shows the path of the index case through the disease compartments from the initial exposure through death or recovery with the i^{th} component of $\phi(t, x_0)$ interpreted as the probability that the index case (introduced at time $t = 0$) is in disease state i at time t . The solution of (5.9) is $\phi(t, x_0) = e^{-Vt}x_0$, where the exponential of a matrix is defined by the Taylor series

$$e^A = I + A + \frac{A^2}{2} + \frac{A^3}{3!} + \cdots + \frac{A^k}{k!} + \cdots.$$

This series converges for all t and $\int_0^\infty \phi(t, x_0) dt = V^{-1}x_0$ (see, for example, [29]). The (i, j) entry of the matrix V^{-1} can be interpreted as the expected time an individual initially introduced into disease compartment j spends in disease compartment i .

The (i, j) entry of the matrix F is the rate at which secondary infections are produced in compartment i by an index case in compartment j . Hence, the expected number of secondary infections produced by the index case is given by

$$\int_0^\infty Fe^{-Vt}x_0 dt = FV^{-1}x_0.$$

Following Diekmann and Heesterbeek [18], the matrix $K_L = FV^{-1}$ is referred to as the next generation matrix with large domain for the system at the disease-free equilibrium. The (i, j) entry of K is the expected number of secondary infections in compartment i produced by individuals initially in compartment j , assuming, of course, that the environment seen by the individual remains homogeneous for the duration of its infection.

Shortly, we will describe some results from matrix theory which imply that the matrix, $K_L = FV^{-1}$ is non-negative and therefore has a non-negative eigenvalue, $\mathcal{R}_0 = \rho(FV^{-1})$, such that there are no other eigenvalues of K with modulus greater than \mathcal{R}_0 and there is a non-negative eigenvector ω associated with \mathcal{R}_0 [7, Theorem 1.3.2]. This eigenvector is in a sense the distribution of infected individuals that produces the greatest number, \mathcal{R}_0 , of secondary infections per generation. Thus, \mathcal{R}_0 and the associated eigenvector ω suitably define a “typical” infective and the basic reproduction number can be rigorously defined as the spectral radius of the matrix, K_L . The spectral radius of a matrix K_L , denoted by $\rho(K_L)$, is the maximum of the moduli of the eigenvalues of K_L . If K_L is irreducible, then \mathcal{R}_0 is a simple eigenvalue of K_L and is strictly larger in modulus than all other eigenvalues of K_L . However, if K_L is reducible, which is often the case for diseases with

multiple strains, then K_L may have several positive real eigenvectors corresponding to reproduction numbers for each competing strain of the disease.

We have interpreted the reproduction number for a disease as the number of secondary infections produced by an infected individual in a population of susceptible individuals. If the reproduction number $\mathcal{R}_0 = \rho(FV^{-1})$ is consistent with the differential equation model, then it should follow that the disease-free equilibrium is asymptotically stable if $\mathcal{R}_0 < 1$ and unstable if $\mathcal{R}_0 > 1$.

This is shown through a sequence of lemmas.

The spectral bound (or abscissa) of a matrix A is the maximum real part of all eigenvalues of A . If each entry of a matrix T is non-negative we write $T \geq 0$ and refer to T as a non-negative matrix. A matrix of the form $A = sI - B$, with $B \geq 0$, is said to have the Z sign pattern. These are matrices whose off-diagonal entries are negative or zero. If in addition, $s \geq \rho(B)$, then A is called an M-matrix. Note that in this section, I denotes an identity matrix, not a population of infectious individuals. The following lemma is a standard result from [7].

Lemma 5.1 *If A has the Z sign pattern, then $A^{-1} \geq 0$ if and only if A is a non-singular M-matrix.*

The assumptions we have made imply that each entry of F is non-negative and that the off-diagonal entries of V are negative or zero. Thus V has the Z sign pattern. Also, the column sums of V are positive or zero, which, together with the Z sign pattern, implies that V is a (possibly singular) M-matrix [7, condition M_{35} of Theorem 6.2.3]. In what follows, it is assumed that V is non-singular. In this case, $V^{-1} \geq 0$, by Lemma 5.1. Hence, $K_L = FV^{-1}$ is also non-negative.

Lemma 5.2 *If F is non-negative and V is a non-singular M-matrix, then $\mathcal{R}_0 = \rho(FV^{-1}) < 1$ if and only if all eigenvalues of $(F - V)$ have negative real parts.*

Proof Suppose $F \geq 0$ and V is a non-singular M-matrix. By the proof of Lemma 5.1, $V^{-1} \geq 0$. Thus, $(I - FV^{-1})$ has the Z sign pattern, and by Lemma 5.1, $(I - FV^{-1})^{-1} \geq 0$ if and only if $\rho(FV^{-1}) < 1$. From the equalities $(V - F)^{-1} = V^{-1}(I - FV^{-1})^{-1}$ and $V(V - F)^{-1} = I + F(V - F)^{-1}$, it follows that $(V - F)^{-1} \geq 0$ if and only if $(I - FV^{-1})^{-1} \geq 0$. Finally, $(V - F)$ has the Z sign pattern, so by Lemma 5.1, $(V - F)^{-1} \geq 0$ if and only if $(V - F)$ is a non-singular M-matrix. Since the eigenvalues of a non-singular M-matrix all have positive real parts, this completes the proof. \square

Theorem 5.1 *Consider the disease transmission model given by (5.7). The disease-free equilibrium of (5.7) is locally asymptotically stable if $\mathcal{R}_0 < 1$, but unstable if $\mathcal{R}_0 > 1$.*

Proof Let F and V be defined as above, and let J_{21} and J_{22} be the matrices of partial derivatives of g with respect to x and y evaluated at the disease-free equilibrium. The Jacobian matrix for the linearization of the system about the disease-free equilibrium has the block structure

$$J = \begin{bmatrix} F - V & 0 \\ J_{21} & J_{22} \end{bmatrix}.$$

The disease-free equilibrium is locally asymptotically stable if the eigenvalues of the Jacobian matrix all have negative real parts. Since the eigenvalues of J are those of $(F - V)$ and J_{22} , and the latter all have negative real parts by assumption, the disease-free equilibrium is locally asymptotically stable if all eigenvalues of $(F - V)$ have negative real parts. By the assumptions on \mathcal{F} and \mathcal{V} , F is non-negative and V is a non-singular M-matrix. Hence, by Lemma 5.2 all eigenvalues of $(F - V)$ have negative real parts if and only if $\rho(FV^{-1}) < 1$. It follows that the disease-free equilibrium is locally asymptotically if $\mathcal{R}_0 = \rho(FV^{-1}) < 1$.

Instability for $\mathcal{R}_0 > 1$ can be established by a continuity argument. If $\mathcal{R}_0 \leq 1$, then for any $\epsilon > 0$, $((1 + \epsilon)I - FV^{-1})$ is a non-singular M-matrix and, by Lemma 5.1, $((1 + \epsilon)I - FV^{-1})^{-1} \geq 0$. By the proof of Lemma 5.2, all eigenvalues of $((1 + \epsilon)V - F)$ have positive real parts. Since $\epsilon > 0$ is arbitrary, and eigenvalues are continuous functions of the entries of the matrix, it follows that all eigenvalues of $(V - F)$ have non-negative real parts. To reverse the argument, suppose all the eigenvalues of $(V - F)$ have non-negative real parts. For any positive ϵ , $(V + \epsilon I - F)$ is a non-singular M-matrix, and by the proof of Lemma 5.2, $\rho(F(V + \epsilon I)^{-1}) < 1$. Again, since $\epsilon > 0$ is arbitrary, it follows that $\rho(FV^{-1}) \leq 1$. Thus, $(F - V)$ has at least one eigenvalue with positive real part if and only if $\rho(FV^{-1}) > 1$, and the disease-free equilibrium is unstable whenever $\mathcal{R}_0 > 1$. \square

These results validate the extension of the definition of the reproduction number to more general situations. In the vaccination model (5.1) of Sect. 5.1 we calculated a pair of final size relations which contained the elements of a matrix K . This matrix is precisely the next generation matrix with large domain $K_L = FV^{-1}$ that has been introduced in this section.

Example 1 Consider the SEIR model with infectivity in the exposed stage,

$$\begin{aligned} S' &= -\frac{\alpha}{N}S(I + \varepsilon E) \\ E' &= \frac{\alpha}{N}S(I + \varepsilon E) - \kappa E \\ I' &= \kappa E - \alpha I \\ R' &= \alpha I. \end{aligned} \tag{5.10}$$

Here the disease states are E and I ,

$$\mathcal{F} = \begin{bmatrix} \varepsilon Ea + Ia \\ 0 \end{bmatrix}$$

and

$$F = \begin{bmatrix} \varepsilon a & a \\ 0 & 0 \end{bmatrix}, \quad V = \begin{bmatrix} \kappa & 0 \\ -\kappa & \alpha \end{bmatrix}, \quad V^{-1} = \begin{bmatrix} 1/\kappa & 0 \\ 1/\alpha & 1/\alpha \end{bmatrix}.$$

Then we may calculate

$$K_L = FV^{-1} = \begin{bmatrix} \frac{\varepsilon a}{\kappa} + \frac{a}{\alpha} & \frac{a}{\alpha} \\ 0 & 0 \end{bmatrix}.$$

It is clear that since \mathcal{R}_0 is equal to the trace of FV^{-1} ,

$$\mathcal{R}_0 = \frac{\varepsilon a}{\kappa} + \frac{a}{\alpha},$$

the element in the first row and first column of FV^{-1} . If all new infections are in a single compartment, as is the case here, the basic reproduction number is the trace of the matrix FV^{-1} .

In general, it is possible to reduce the size of the next generation matrix with large domain to the number of *state at infection* [18]. The states at infection are those disease states in which there can be new infections. Suppose that there are n disease states and k states at infection with $k < n$. Then we may define an auxiliary $n \times k$ matrix P in which each column corresponds to a state at infection and has 1 in the corresponding row and 0 elsewhere. Then the next generation matrix is the $k \times k$ matrix

$$K = P^T K_L P.$$

It is easy to show, using the fact that $PP^T K_L = K_L$, that the $n \times n$ matrix K_L and the $k \times k$ matrix K have the same non-zero eigenvalues and therefore the same spectral radius. Construction of the next generation matrix which has lower dimension than the next generation matrix with large domain may simplify the calculation of the basic reproduction number.

In Example 1 above, the only disease state at infection is E , the matrix P is

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

and the next generation matrix K is the 1×1 matrix

$$K = \left[\frac{\varepsilon a}{\kappa} + \frac{a}{\alpha} \right].$$

5.2.1 Some More Complicated Examples

The next generation approach is very general and can be applied to models with heterogeneous mixing and control measures applied differently in different groups.

Example 2 Consider the vaccination model (5.1) of Sect. 5.1. The disease states are I_U and I_V . Then

$$\mathcal{F} = \begin{bmatrix} a_U(I_U + \delta I_V) \\ \sigma a_V(I_U + \delta I_V) \end{bmatrix}$$

and

$$F = \begin{bmatrix} a_U & \delta a_U \\ \sigma a_V & \sigma \delta a_V \end{bmatrix}, \quad V = \begin{bmatrix} \alpha_U & 0 \\ 0 & \alpha_V \end{bmatrix}.$$

It is easy to see that the next generation matrix with large domain is the matrix K calculated in Sect. 5.1. Since each disease state is a disease state at infection, the next generation matrix is K , the same as the next generation matrix with large domain. As in Example 1, the determinant of K is zero and K has rank 1. Thus the control reproduction number is the trace of K ,

$$\mathcal{R}_c = \frac{w_U}{\alpha_U} + \frac{\delta \sigma N_V}{\alpha_V}.$$

5.3 Heterogeneous Mixing

In disease transmission models not all members of the population make contacts at the same rate. In sexually transmitted diseases there is often a “core” group of very active members who are responsible for most of the disease cases, and control measures aimed at this core group have been very effective in control [27]. In epidemics there are often “super-spreaders”, who make many contacts and are instrumental in spreading disease and in general some members of the population make more contacts than others. Recently there has been a move to complicated network models for simulating epidemics [23, 24, 32–34, 38]. These assume knowledge of the mixing patterns of groups of members of the population and

make predictions based on simulations of a stochastic model. A basic description of network models may be found in [44]. While network models can give very detailed predictions, they have some serious disadvantages. For a detailed network model, simulations take long enough to make it difficult to examine a significant range of parameter values, and it is difficult to estimate the sensitivity with respect to parameters of the model. The theoretical analysis of network models is a very active and rapidly developing field [38–40].

However, it is possible to consider models more realistic than simple compartmental models but simpler to analyze than detailed network models. To model heterogeneity in mixing we may assume that the population is divided into subgroups with different activity levels. We will analyze an *SIR* model in which there are two groups with different contact rates. The approach extends easily to models with more compartments, such as exposed periods or a sequence of infective stages and also to models with an arbitrary number of activity levels. In this way, we may hope to give models intermediate between the too simple compartmental models and the too complicated network models.

In this section, we describe the formulation of models for two groups with different activity levels and give the main results for the simplest compartmental epidemic models. The analysis of models of the same type with more complicated compartmental structure is given in [11] and the analysis of models with more groups is given in [12]. There is no problem, other than technical calculation difficulties, in extending everything in this section to an arbitrary number of sub-populations.

Consider two sub-populations of constant sizes N_1, N_2 , respectively, each divided into susceptibles, infectives, and removed members with subscripts to identify the sub-population. In this section, we will assume that the number of contacts per member in unit time is a constant. Suppose that each group i member makes a_i contacts sufficient to transmit infection in unit time, and that the fraction of contacts made by a member of group i that is with a member of group j is p_{ij} , ($i, j = 1, 2$). Then

$$p_{i1} + p_{i2} = 1, \quad i = 1, 2.$$

We assume that all contacts between a susceptible and an infective transmit infection to the susceptible. Suppose the mean infective period in group i is $1/\alpha_i$. We assume that there are no disease deaths, so that the population size of each group is constant.

A two-group *SIR* epidemic model is

$$\begin{aligned} S'_i &= -a_i S_i \left[p_{i1} \frac{I_1}{N_1} + p_{i2} \frac{I_2}{N_2} \right] \\ I'_i &= a_i S_i \left[p_{i1} \frac{I_1}{N_1} + p_{i2} \frac{I_2}{N_2} \right] - \alpha_i I_i, \quad i = 1, 2. \end{aligned} \tag{5.11}$$

The initial conditions are

$$S_i(0) + I_i(0) = N_i, \quad i = 1, 2.$$

The two-group model includes two possibilities. It may describe a population with groups differing by activity levels and possibly by vulnerability to infection. For an epidemic model, in which we assume the time scale is short enough that members do not age over the course of the epidemic, the grouping could be by age. However, for a longer term disease transmission model with age-dependent transmission it would be necessary to take into account the fact that the ages of members of the population change over the course of the disease and use a different type of model, to be studied in Chap. 13.

The two-group model (5.11) assumes the same infectivities and susceptibilities in each group. A more general model would be

$$\begin{aligned} S'_i &= -\sigma_i a_i S_i \left[\delta_1 p_{i1} \frac{I_1}{N_1} + \delta_2 p_{i2} \frac{I_2}{N_2} \right] \\ I'_i &= \sigma_i a_i S_i \left[\delta_1 p_{i1} \frac{I_1}{N_1} + \delta_2 p_{i2} \frac{I_2}{N_2} \right] - \alpha_i I_i, \quad i = 1, 2. \end{aligned} \tag{5.12}$$

This is just the model (5.11) with the addition of susceptibility factors σ_1, σ_2 for susceptibles in the two groups and infectivity factors δ_1, δ_2 for infectives in the two groups. As before a_1, a_2 are effective contact rates, and this model adds transmission probabilities to (5.11).

It is not possible to calculate the reproduction number for the two-group model (5.11) directly by counting secondary infections. It is necessary to use the next generation matrix approach of [45] described in Section 5.2 and calculate the reproduction number as the largest eigenvalue of the matrix FV^{-1} , where

$$F = \begin{bmatrix} p_{11}a_1 & p_{12}a_1 \frac{N_1}{N_2} \\ p_{21}a_2 \frac{N_2}{N_1} & p_{22}a_2 \end{bmatrix}, \quad V = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix}.$$

Then

$$FV^{-1} = \begin{bmatrix} \frac{p_{11}a_1}{\alpha_1} & \frac{p_{12}a_1}{\alpha_2} \frac{N_1}{N_2} \\ \frac{p_{21}a_2}{\alpha_1} \frac{N_2}{N_1} & \frac{p_{22}a_2}{\alpha_2} \end{bmatrix}.$$

The eigenvalues of the matrix FV^{-1} are the roots of the quadratic equation

$$\lambda^2 - \left(\frac{p_{11}a_1}{\alpha_1} + \frac{p_{22}a_2}{\alpha_2} \right) \lambda + (p_{11}p_{22} - p_{12}p_{21}) \frac{a_1 a_2}{\alpha_1 \alpha_2} = 0. \tag{5.13}$$

The basic reproduction number \mathcal{R}_0 is the larger of these two eigenvalues,

$$\mathcal{R}_0 = \frac{\frac{p_{11}a_1}{\alpha_1} + \frac{p_{22}a_2}{\alpha_2} + \sqrt{\left(\frac{p_{11}a_1}{\alpha_1} - \frac{p_{22}a_2}{\alpha_2}\right)^2 + 4\frac{p_{12}p_{21}a_1a_2}{\alpha_1\alpha_2}}}{2}.$$

In order to obtain a more useful expression for \mathcal{R}_0 , it is necessary to make some assumptions about the nature of the mixing between the two groups. The mixing is determined by the two quantities p_{12} , p_{21} since $p_{11} = 1 - p_{12}$ and $p_{22} = 1 - p_{21}$.

There has been much study of mixing patterns, see, for example, [8, 9, 13]. One possibility is proportionate mixing, that is, that the number of contacts between groups is proportional to the relative activity levels. In other words, mixing is random but constrained by the activity levels [42]. Under the assumption of proportionate mixing,

$$p_{ij} = \frac{a_j N_j}{a_1 N_1 + a_2 N_2},$$

and we may write

$$p_{11} = p_{21} = p_1, \quad p_{12} = p_{22} = p_2,$$

with $p_1 + p_2 = 1$. In particular,

$$p_{11}p_{22} - p_{12}p_{21} = 0,$$

and thus

$$\mathcal{R}_0 = a_1 \frac{p_1}{\alpha_1} + a_2 \frac{p_2}{\alpha_2}.$$

Another possibility is preferred mixing [42], in which a fraction π_i of each group mixes randomly with its own group and the remaining members mix proportionately. Thus, preferred mixing is given by

$$\begin{aligned} p_{11} &= \pi_1 + (1 - \pi_1)p_1, & p_{12} &= (1 - \pi_1)p_2 \\ p_{21} &= (1 - \pi_2)p_1, & p_{22} &= \pi_2 + (1 - \pi_2)p_2, \end{aligned} \tag{5.14}$$

with

$$p_i = \frac{(1 - \pi_i)a_i N_i}{(1 - \pi_1)a_1 N_1 + (1 - \pi_2)a_2 N_2}, \quad i = 1, 2.$$

Proportionate mixing is the special case of preferred mixing with $\pi_1 = \pi_2 = 0$.

It is also possible to have like-with-like mixing, in which members of each group mix only with members of the same group. This is the special case of preferred mixing with $\pi_1 = \pi_2 = 1$. For like-with-like mixing,

$$p_{11} = p_{22} = 1, \quad p_{12} = p_{21} = 0.$$

Then the roots of (5.13) are a_1/α_1 and a_2/α_2 , and the reproduction number is

$$\mathcal{R}_0 = \max \left\{ \frac{a_1}{\alpha_1}, \frac{a_2}{\alpha_2} \right\}.$$

By calculating the partial derivatives of p_{ij} ($i, j = 1, 2$) with respect to π_1 and π_2 , we may show that p_{11} and p_{22} increase when either π_1 or π_2 is increased, while p_{12} and p_{21} decrease when either π_1 or π_2 is increased. From this, we may see from the general expression for \mathcal{R}_0 that increasing either of the preferences π_1, π_2 increases the basic reproduction number.

We may follow the analysis of the *SIR* model (5.11) to obtain the basic reproduction number for the *SIR* model (5.12) with susceptibility and infectivity reduction factors

$$\mathcal{R}_0 = \frac{\sum_{i=1}^2 \sigma_i \delta_i \frac{p_{ii} a_i}{\alpha_i} + \sqrt{\left(\sigma_1 \delta_1 \frac{p_{11} a_1}{\alpha_1} - \sigma_2 \delta_2 \frac{p_{22} a_2}{\alpha_2} \right)^2 + 4 \sigma_1 \delta_2 \sigma_2 \delta_1 \frac{p_{12} p_{21} a_1 a_2}{\alpha_1 \alpha_2}}}{2}.$$

In the special case of proportionate mixing, where $p_{11}p_{22} - p_{12}p_{21} = 0$, this reduces to

$$\mathcal{R}_0 = \sum_{i=1}^2 \sigma_i \delta_i \frac{p_{ii} a_i}{\alpha_i}.$$

The vaccination model of Sect. 5.1 is an example of a two-group model of the form (5.12), with

$$\sigma_1 = \sigma_2 = \delta_1 = 1, \quad \delta_2 = \delta.$$

It is easy to show [11] that, just as for a one-group model [10],

$$S_1 \rightarrow S_1(\infty) > 0, \quad S_2 \rightarrow S_2(\infty) > 0,$$

as $t \rightarrow \infty$.

Example 1 Consider a two-group SEIR model,

$$\begin{aligned} S'_i &= -a_i S_i \left[p_{ii} \frac{I_i}{N_i} + p_{ij} \frac{I_j}{N_j} \right] \\ E'_i &= a_i S_i \left[p_{ii} \frac{I_i}{N_i} + p_{ij} \frac{I_j}{N_j} \right] - \kappa_i E_i \\ I'_i &= \kappa_i E_i - \alpha_i I_i \\ R'_i &= \alpha_i I_i, \quad i, j = 1, 2, i \neq j. \end{aligned} \tag{5.15}$$

The disease states are E_i and I_i ($i = 1, 2$).

Now

$$\mathcal{F} = \begin{bmatrix} a_1 p_{11} I_1 + a_1 p_{12} I_2 \frac{N_1}{N_2} \\ 0 \\ a_2 p_{21} I_1 \frac{N_2}{N_1} \\ 0 \end{bmatrix}, \quad F = \begin{bmatrix} 0 & a_1 p_{11} & 0 & a_1 p_{12} \frac{N_1}{N_2} \\ 0 & 0 & 0 & 0 \\ 0 & a_2 p_{21} \frac{N_2}{N_1} & 0 & a_2 p_{22} \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

and

$$V = \begin{bmatrix} \kappa_1 & 0 & 0 & 0 \\ -\kappa_1 & \alpha_1 & 0 & 0 \\ 0 & 0\kappa_2 & 0 & 0 \\ 0 & 0 & -\kappa_2 & \alpha_2 \end{bmatrix}, \quad V^{-1} = \begin{bmatrix} \frac{1}{\kappa_1} & 0 & 0 & 0 \\ \frac{1}{\alpha_1} & \frac{1}{\alpha_1} & 0 & 0 \\ 0 & 0 & \frac{1}{\kappa_2} & 0 \\ 0 & 0 & \frac{1}{\alpha_2} & \frac{1}{\alpha_2} \end{bmatrix}.$$

Then we may calculate

$$K_L = FV^{-1} = \begin{bmatrix} a_1 \frac{p_{11}}{\alpha_1} & a_1 \frac{p_{11}}{\alpha_1} & a_1 \frac{p_{12}}{\alpha_1} \frac{N_1}{N_2} & a_1 \frac{p_{12}}{\alpha_1} \frac{N_1}{N_2} \\ 0 & 0 & 0 & 0 \\ a_2 \frac{p_{21}}{\alpha_1} \frac{N_2}{N_1} & a_2 \frac{p_{21}}{\alpha_1} \frac{N_2}{N_1} & a_2 \frac{p_{22}}{\alpha_2} & a_2 \frac{p_{22}}{\alpha_2} \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

In this example, it is advantageous to construct the next generation matrix by reducing the next generation matrix with large domain K_L . In order to do this, we use the auxiliary matrix

$$E = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

to construct the next generation matrix

$$K = E^T K_L E = \begin{bmatrix} \frac{p_{11}a_1}{\alpha_1} & \frac{p_{12}a_1}{\alpha_2} N_1 \\ \frac{p_{21}a_2}{\alpha_1} N_2 & \frac{p_{22}a_2}{\alpha_2} \end{bmatrix}. \quad (5.16)$$

This is the same matrix as the next generation matrix obtained for the two-group *SIR* model (5.11) introduced in this section.

For a one-group epidemic model there is a final size relation that makes it possible to calculate the size of the epidemic from the reproduction number [10, 35]. There is a corresponding final size relation for the two-group model (5.1), established in much the same way, combining expressions for the integrals of $(S_1 + I_1)', (S_2 + I_2)', (S_1)'/S_1, (S_2)'/S_2$. This relation does not involve the reproduction number explicitly but still makes it possible to calculate the size of the epidemic from the model parameters.

The final size relation for the model (5.11) is the pair of equations

$$\log \frac{S_i(0)}{S_i(\infty)} = a_i \left[\frac{p_{ii}}{\alpha_i} \left(1 - \frac{S_i(\infty)}{N_i} \right) + \frac{p_{ij}}{\alpha_j} \left(1 - \frac{S_j(\infty)}{N_j} \right) \right], \quad i, j = 1, 2, \quad i \neq j. \quad (5.17)$$

Just as with the vaccination model (5.1), the final size relation may be written in terms of the matrix

$$\mathcal{R} = \begin{bmatrix} \mathcal{R}_{11} & \mathcal{R}_{12} \\ \mathcal{R}_{21} & \mathcal{R}_{22} \end{bmatrix} = \begin{bmatrix} \frac{a_1 p_{11}}{\alpha_1} & \frac{a_1 p_{12}}{\alpha_2} \\ \frac{a_2 p_{21}}{\alpha_1} & \frac{a_2 p_{22}}{\alpha_2} \end{bmatrix}.$$

The matrix \mathcal{R} is similar to the next generation matrix K (and therefore has the same eigenvalues), since

$$\mathcal{R} = T^{-1} K T,$$

where

$$T = \begin{bmatrix} N_1 & 0 \\ 0 & N_2 \end{bmatrix}.$$

The final size relation makes it possible to calculate $S_1(\infty)$ and $S_2(\infty)$ and thence the number of disease cases

$$[N_1 - S_1(\infty)] + [N_2 - S_2(\infty)].$$

The final size relation takes a simpler form if the mixing is proportionate. With proportionate mixing, since

$$p_{11} = p_{21} = p_1, \quad p_{12} = p_{22} = p_2,$$

(5.17) implies

$$a_2 \log \frac{S_1(0)}{S_1(\infty)} = a_1 \log \frac{S_2(0)}{S_2(\infty)},$$

and we may write the final size relations as

$$\begin{aligned} \log \frac{S_1(0)}{S_1(\infty)} &= \frac{a_1 p_1}{\alpha_1} \left[1 - \frac{S_1(\infty)}{N_1} \right] + \frac{a_1 p_2}{\alpha_2} \left[1 - \frac{S_2(\infty)}{N_2} \right], \\ \left[\frac{S_1(\infty)}{S_1(0)} \right]^{a_2} &= \left[\frac{S_2(\infty)}{S_2(0)} \right]^{a_1}. \end{aligned} \tag{5.18}$$

We recall that in the case of proportionate mixing

$$\mathcal{R}_0 = \frac{p_1 a_1}{\alpha_1} + \frac{p_2 a_2}{\alpha_2}.$$

The second equation of (5.18) implies that if $a_1 > a_2$, then

$$1 - \frac{S_1(\infty)}{S_1(0)} > 1 - \frac{S_2(\infty)}{S_2(0)},$$

that is, the attack ratio is greater in the more active group.

It is not difficult to show that the final size relations (5.18) give a unique set of final numbers of susceptibles in each group. The final size relation can also be obtained in a similar way for more complicated compartmental models [4–6].

The model can be extended easily to an arbitrary number of groups with different activity levels. It is also important to be able to describe models with more stages in the progression through compartments, and models in which there are differences between groups in susceptibility. For example, influenza has two characteristics not included in the model (5.11) that are of importance. There is a latent period between infection and the development of infectivity and influenza symptoms. Also, only a fraction of latent members develop symptoms, while the remainder go through an asymptomatic stage in which there is some infectivity. Another important aspect is treatment, which could be directed at either or both group, and could be used for making decisions on how to target groups for treatment. A natural way to proceed in this direction is to build an age of infection model for populations with multiple groups.

Example 2 Consider a two-group endemic *SIR* model with preferential mixing and group-targeted vaccination:

$$\begin{aligned}\frac{dS_i}{dt} &= \mu N_i(1 - \phi_i) - (\lambda_i(t) + \mu)S_i, \\ \frac{dI_i}{dt} &= \lambda_i(t)S_i - (\alpha + \mu)I_i, \\ \frac{dR_i}{dt} &= \mu N_i\phi_i + \alpha I_i - \mu R_i, \quad i = 1, 2,\end{aligned}\tag{5.19}$$

where $N_i = S_i + I_i + R_i$. Here, λ_i represents the force of infection for susceptibles in group i given by

$$\lambda_i = a_i \sigma \sum_{j=1}^n p_{ij} \frac{I_j}{N_j},\tag{5.20}$$

where a_i denotes the average number of contacts an individual in sub-population i has per unit of time (which represents the activity level of group i), and σ is the probability of infection per contact when the contact is with an infectious individual, ϕ_i denotes the proportion of susceptibles in group i vaccinated (and removed) when entering the population. The fraction I_j/N_j gives the probability that a contact is with an infectious individual in sub-population j . The contact matrix (p_{ij}) has the same form as the preferential mixing considered earlier with

$$p_{ij} = \pi_i \delta_{ij} + (1 - \pi_i)p_j, \quad i, j = 1, 2.\tag{5.21}$$

The parameter π_i is the fraction of contacts with individuals in the same sub-population, δ_{ij} is the Kronecker delta (i.e., 1 when $i = j$ and 0 otherwise), and

$$p_j = \frac{(1 - \pi_j)a_j N_j}{(1 - \pi_1)a_1 N_1 + (1 - \pi_2)a_2 N_2}, \quad j = 1, 2.$$

Clearly, unless all the sub-groups are isolated (i.e., no interactions between the groups), there must be some i with $\pi_i < 1$.

For each sub-population i , if all contacts are with people within the same group (i.e., $p_{ii} = 1$ and $p_{ij} = 0$ for $i \neq j$), then the basic and control reproduction numbers for group i are, respectively,

$$\mathcal{R}_{0i} = \frac{\sigma a_i}{\mu + \alpha}, \quad \mathcal{R}_{vi} = \mathcal{R}_{0i}(1 - \phi_i), \quad i = 1, 2.\tag{5.22}$$

When there are contacts between sub-populations, i.e., $p_{ii} < 1$ or $\pi_i < 1$ for some i , we can derive the basic and control reproduction numbers for the metapopulation.

These reproduction numbers will be functions of \mathcal{R}_{0i} or \mathcal{R}_{vi} . The next generation matrix K_v (v for vaccination) is

$$K_v = \begin{pmatrix} \mathcal{R}_{v1}p_{11} & \mathcal{R}_{v1}p_{12} \\ \mathcal{R}_{v2}p_{21} & \mathcal{R}_{v2}p_{22} \end{pmatrix}. \quad (5.23)$$

The control reproduction number \mathcal{R}_v for the metapopulation is given by

$$\mathcal{R}_v = \frac{1}{2} \left[A + D + \sqrt{(A - D)^2 + 4BC} \right], \quad (5.24)$$

where

$$\begin{aligned} A &= \mathcal{R}_{01}p_{11}(1 - \phi_1), \quad B = \mathcal{R}_{01}p_{12}(1 - \phi_1), \\ C &= \mathcal{R}_{02}p_{21}(1 - \phi_2), \quad D = \mathcal{R}_{02}p_{22}(1 - \phi_2), \end{aligned}$$

and \mathcal{R}_{0i} ($i = 1, 2$) are given in (5.22). If $\phi_1 = \phi_2 = 0$, then \mathcal{R}_v reduces to

$$\mathcal{R}_0 = \frac{1}{2} \left[\mathcal{R}_{01}p_{11} + \mathcal{R}_{02}p_{22} + \sqrt{(\mathcal{R}_{01}p_{11} - \mathcal{R}_{02}p_{22})^2 + 4\mathcal{R}_{01}p_{12}\mathcal{R}_{02}p_{21}} \right].$$

To study the effects of vaccination strategies, assume that $\mathcal{R}_0 > 1$ in the absence of vaccination and

$$\mathcal{R}_{01} > 1, \quad \mathcal{R}_{02} > 1. \quad (5.25)$$

Let

$$\Omega = \{(\phi_1, \phi_2) \mid 0 \leq \phi_1 < 1, 0 \leq \phi_2 < 1\}. \quad (5.26)$$

Then each point $(\phi_1, \phi_2) \in \Omega$ represents a vaccination strategy.

Because we are interested in the case when the two groups are not isolated, either $\pi_1 < 1$ or $\pi_2 < 1$. This will be assumed for the results below. It can be shown that, for each fixed $(\phi_1, \phi_2) \in \Omega$, \mathcal{R}_v increases with both π_1 and π_2 , i.e.,

$$\frac{\partial \mathcal{R}_v}{\partial \pi_1} > 0, \quad \frac{\partial \mathcal{R}_v}{\partial \pi_2} > 0 \quad \text{for all } (\pi_1, \pi_2) \in \Omega. \quad (5.27)$$

For each fixed (π_1, π_2) , there are different combinations of ϕ_1 and ϕ_2 that can reduce \mathcal{R}_v to be below 1. For ease of presentation, consider the simpler case in which

$$\pi_1 = \pi_2 = \pi,$$

and consider $\mathcal{R}_v = \mathcal{R}_v(\pi)$ as a function of π . Then, for each fixed $\pi \in [0, 1]$, the curve determined by $\mathcal{R}_v(\pi) = 1$ divides the region Ω into two parts: one is the region

$$\Omega_\pi = \{(\phi_1, \phi_2) \mid 0 \leq \mathcal{R}_v(\pi) < 1, (\phi_1, \phi_2) \in \Omega, 0 \leq \pi < 1\},$$

which includes all points above the curve (see Fig. 5.1), and another is the region

$$D_\pi = \{(\phi_1, \phi_2) \mid \mathcal{R}_v(\pi) > 1, (\phi_1, \phi_2) \in \Omega, 0 \leq \pi < 1\},$$

which includes all points below the curve. It can be shown that the region Ω_π decreases as π increases and reduces to the region Ω^* as $\pi \rightarrow 0$, while the region D_π decreases as π decreases and reduces to the region D^* as $\pi \rightarrow 1$ (see Fig. 5.1). All these curves intersect at a single point (ϕ_{1c}, ϕ_{2c}) with

$$\phi_{1c} = 1 - \frac{1}{\mathcal{R}_{01}}, \quad \phi_{2c} = 1 - \frac{1}{\mathcal{R}_{02}}. \quad (5.28)$$

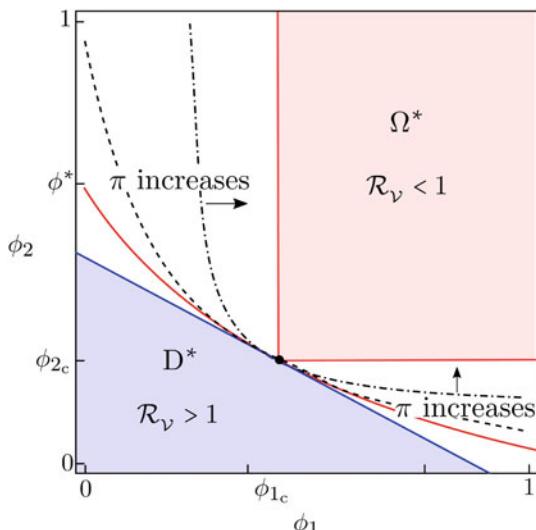
We observe from Fig. 5.1 that the region Ω^* (lighter-shaded) is determined by the two inequalities

$$\phi_{1c} < \phi_1 < 1, \quad \phi_{2c} < \phi_2 < 1, \quad (5.29)$$

where ϕ_{1c} and ϕ_{2c} are defined in (5.28). For region D^* (darker-shaded), the upper bound is determined by the line

$$\phi_2 = -\mathcal{A}\phi_1 + \mathcal{B}, \quad (5.30)$$

Fig. 5.1 Plot showing the regions Ω^* and D^* . Several curves of $\mathcal{R}_v(\pi) = 1$ for different π values are also shown, with the dashed curves corresponding to $0 < \pi < 1$, the thin solid lines (boundary of Ω^*) corresponding to $\pi = 1$, and the thick line corresponding to $\pi = 0$ (the upper bound of the region D^*). The arrows indicate the direction of change of the curve $\mathcal{R}_v(\pi) = 1$ as π increases from 0 to 1. All of the $\mathcal{R}_v(\pi) = 1$ curves intersect at the single point (ϕ_{1c}, ϕ_{2c}) .
Source: [15]



where

$$\begin{aligned}\mathcal{A} &= \frac{\mathcal{R}_{01}a_1N_1}{\mathcal{R}_{02}a_2N_2}, \\ \mathcal{B} &= \frac{(\mathcal{R}_{01}-1)a_1N_1 + (R_{02}-1)a_2N_2}{\mathcal{R}_{02}a_2N_2}.\end{aligned}\tag{5.31}$$

The two regions intersect at the point (ϕ_{1c}, ϕ_{2c}) .

This result suggests that there is a “lower bound” for vaccination efforts (ϕ_1, ϕ_2) , above which the infection can be eradicated regardless of mixing patterns. Similarly, it provides an “upper bound” for vaccination efforts (ϕ_1, ϕ_2) , below which the infection cannot be eradicated regardless of mixing patterns (see the definitions for ϕ_1^* and ϕ_2^* defined in (5.32) and see Fig. 5.2 for an illustration of the lower and upper bound). For an “intermediate level” vaccination strategy (ϕ_1, ϕ_2) , mixing parameters π_1 and π_2 can play an important role in influencing the effect of vaccination strategies on reducing \mathcal{R}_v . Thus, when designing vaccination strategies, one should take into consideration mixing patterns within and between sub-populations.

For given π_i ($i = 1, 2$), it can be shown that $\frac{\partial \mathcal{R}_v}{\partial \phi_i} < 0$. When the curve $\mathcal{R}_v = 1$ lies between regions D^* and Ω^* , the curve intersects the ϕ_1 -axis and ϕ_2 -axis at $(\phi_1^*, 0)$ and $(0, \phi_2^*)$, respectively, where

$$\begin{aligned}\phi_1^* &= 1 - \frac{1 - \mathcal{R}_{02}p_{22}}{\mathcal{R}_{01}p_{11}(1 - \mathcal{R}_{02}p_{22}) + \mathcal{R}_{01}\mathcal{R}_{02}p_{12}p_{21}}, \\ \phi_2^* &= 1 - \frac{1 - \mathcal{R}_{01}p_{11}}{\mathcal{R}_{02}p_{22}(1 - \mathcal{R}_{01}p_{11}) + \mathcal{R}_{22}\mathcal{R}_{01}p_{12}p_{21}}.\end{aligned}\tag{5.32}$$

Because $\mathcal{R}_{0i} > 1$ for $i = 1, 2$, it is possible that $\mathcal{R}_{01}p_{11} > 1$ and/or $\mathcal{R}_{02}p_{22} > 1$. Thus, it is possible that $\phi_1^* > 1$ and/or $\phi_2^* > 1$. When $\phi_1^* > 1$, we know from $\partial \mathcal{R}_v / \partial \phi_1 < 1$ that $\mathcal{R}_v > 1$ for any vaccination strategy $(\phi_1, 0)$. Thus, it is impossible to eradicate the infection if only sub-population 1 is vaccinated.

The results described above are based on the control reproduction number. Figure 5.2 shows some simulation results illustrating the effect of vaccination on the prevalence of infection. Different preference levels are used: $\pi_1 = 0.2$ and $\pi_2 = 0.4$, i.e., group 2 has a higher preference contacting people in its own group. Other parameter values used are $\sigma = 0.03$, $\alpha = 0.15$ (an infective period of about 6 days), and $a_1 = 12$, $a_2 = 8$, $\mu = 0.00016$ (a duration of 17 years in school). These values correspond to $\mathcal{R}_{01} = 2.4$ and $\mathcal{R}_{02} = 1.6$. The initial conditions are $x_1(0) = S_1(0)/N_1(0) = 0.4$, $y_1(0) = I_1(0)/N_1(0) = 0.00002$, $x_2(0) = S_2(0)/N_2(0) = 0.6$, $y_2(0) = I_2(0)/N_2(0) = 0.00002$. For this set of parameters, $\phi_1^* = 0.77$ and $\phi_2^* \gg 1$. Figure 5.2a is for a vaccination strategy $(\phi_1, 0)$ with $\phi_1 = 0.2 < \phi_1^*$, for which the infection persists ($\mathcal{R}_v = 1.8 > 1$), while Fig. 5.2b is for a vaccination strategy $(\phi_1, 0)$ with $\phi_1 = 0.8 > \phi_1^*$, in which case the infection dies out ($\mathcal{R}_v = 0.97 < 1$).

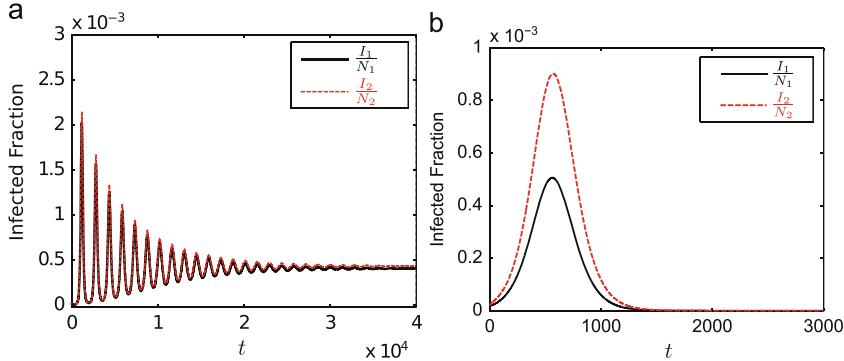


Fig. 5.2 When $\phi_1^* < 1$ and $\phi_2^* > 1$, the disease is eventually eradicated if the vaccination is applied to sub-population 1 alone at a level above ϕ_1^* . (a) $(\phi_1, \phi_2) = (0.2, 0)$ and $\phi_1 < \phi_1^* = 0.77$, the disease persists ($\mathcal{R}_v = 1.8$); (b) $(\phi_1, \phi_2) = (0.8, 0)$ and $\phi_1 > \phi_1^* = 0.77$, the disease eventually disappears ($\mathcal{R}_v = 0.97$). Source: [15]

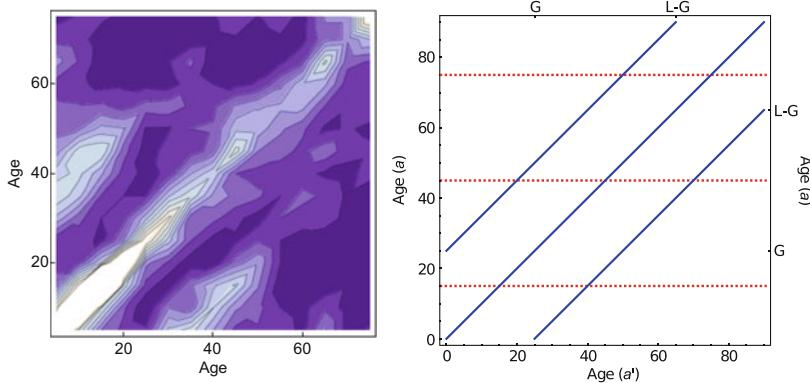


Fig. 5.3 The figure on the left shows observed patterns of contacts between age groups [41]. The lighter areas correspond to higher level of contacts. The figure on the right shows schematic contact matrices illustrating the main and off-diagonals representing contacts among contemporaries, between children and parents, and vice versa

Example 3 This example considers a more general preferential mixing function than the one given in (5.21). This is motivated by the observed mixing pattern shown in Fig. 5.3. Figure 5.3 illustrates the recently collected data reported in [41], which reveals the preferential mixing between parents and children in addition to that among contemporaries.

In the case of n groups, the similar function as (5.21) can be written as

$$p_{ij} = \pi_i \delta_{ij} + (1 - \pi_i) p_j, \quad i, j = 1, 2, \dots, n, \quad j \neq i \quad (5.33)$$

with

$$p_i = \frac{(1 - \pi_i)a_i N_i}{\sum_{k=1}^n (1 - \pi_k)a_k N_k}, \quad i = 1, 2, \dots, n.$$

Here, a_i 's are the contact rates, δ_{ij} 's are the Kronecker delta (i.e., $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$).

To capture the feature shown in Fig. 5.3 (left), Glasser et al. in [25] extended the mixing function (5.33) to include not only the preference along the diagonal but also along the sub- and super-diagonals as indicated in Fig. 5.3 (right). The function is described by

$$p_{ij} = \phi_{ij} + \left(1 - \sum_{l=1}^3 \pi_{li}\right)p_j, \quad p_j = \frac{\left(1 - \sum_{l=1}^3 \pi_{li}\right)a_j N_j}{\sum_{k=1}^n \left(1 - \sum_{l=1}^3 \pi_{lk}\right)a_k N_k} \quad (5.34)$$

with

$$\phi_{ij} = \begin{cases} \delta_{ij}\pi_{1i} + \delta_{i(j+G)}\pi_{2i}, & i \geq G, \\ \delta_{ij}\pi_{1i} + \delta_{i(j-G)}\pi_{3i}, & i \leq L - G. \end{cases} \quad (5.35)$$

G is the generation time (i.e., average age at which women bear children), L is longevity (i.e., average expectation of life at birth), and $L > G$. The parameters $\varepsilon_1 - \varepsilon_3$ represent the fractions of contacts reserved for contemporaries, children ($j - G$), and parents ($j + G$), respectively, and the corresponding delta function is defined as

$$\delta_{i(j \pm G)} = \begin{cases} 1 & \text{if } i = j \pm G, \\ 0 & \text{otherwise.} \end{cases} \quad (5.36)$$

Only people whose ages equal or exceed G can have children, and only those whose ages equal to or less than $L - G$ can have parents, but people aged at least G but not more than $L - G$ can have both children and parents.

Denote the preferential vectors by $\Pi_l = (\pi_{l1}, \pi_{l2}, \dots, \pi_{ln})$, $l = 1, 2, 3$. When $\Pi_2 = \Pi_3 = 0$, the expression (5.34) reduces to the formula (5.33). For ease of notation, we mix indices and real numbers, but if age classes are 0–4, 5–9, … and $G = 25$ years, for example, by $i > G$ we mean $i >$ class 5. Notice that the non-zero elements of Π_2 and Π_3 are related. If $G = 25$ years, for example, then

$$a_i N_i \pi_{2i} = a_j N_j \pi_{3j}, \quad i = 6, 7, \dots, \quad j = i - 5.$$

Notice also that $0 \leq \sum_{l=1}^3 \pi_{li} < 1$.

The mixing function was further extended in [25] to replace the delta functions in (5.36) by Gaussian kernels. Delta formulations are convenient mathematically, but do not allow the age range of one's contemporaries to vary as one ages (e.g., the range narrows perceptively among adolescents), much less differences between the age ranges of one's contemporaries and one's parents or children. By virtue of secular patterns in childbearing, moreover, the age ranges of parents and children may change with age. But the Gaussian formulation allows such variation, reproducing the essential features of these observations. Figure 5.4 demonstrates the comparison between observed mixing patterns (left panel) and the model outcomes (right panel) using the extended mixing functions.

5.3.1 *Optimal Vaccine Allocation in Heterogeneous Populations

One of the significant benefits of models with heterogeneous mixing is that it provides an approach to identifying optimal allocation of vaccines in a metapopulation, particularly when resources are limited. Consider a metapopulation with n sub-populations connected by heterogeneous mixing, which is described by an $n \times n$ matrix $P = (p_{ij})$ for $i, j = 1, 2, \dots, n$. The following model is studied in [21, 26, 43]:

$$\begin{aligned} \frac{dS_i}{dt} &= (1 - \phi_i)\theta N_i - (\lambda_i + \theta)S_i \\ \frac{dI_i}{dt} &= \lambda_i S_i - (\gamma + \theta)I_i \\ \frac{dR_i}{dt} &= \phi_i\theta N_i + \gamma I_i - \theta R_i \\ N_i &= S_i + I_i + R_i \\ \lambda_i &= \sigma a_i \sum_{j=1}^n p_{ij} I_j / N_j, \quad i = 1, 2, \dots, n, \end{aligned} \tag{5.37}$$

where ϕ_i are proportions immunized at entry into sub-population i , γ is the *per-capita* recovery rate, θ is the *per-capita* rate for entering and leaving sub-population i so that the population size N_i remains constant. The function λ_i is the force of infection, i.e., *per-capita* hazard rate of infection of susceptible individuals in sub-population i , in which σ is the probability of infection upon contacting an infectious person, a_i is the average contact rate (activity) in sub-population i , p_{ij} is the proportion of i th sub-population's contacts that are with members of j th sub-population, and I_j/N_j is the probability that a randomly encountered member of sub-population j is infectious. A similar model with two-level mixing is studied in [22].

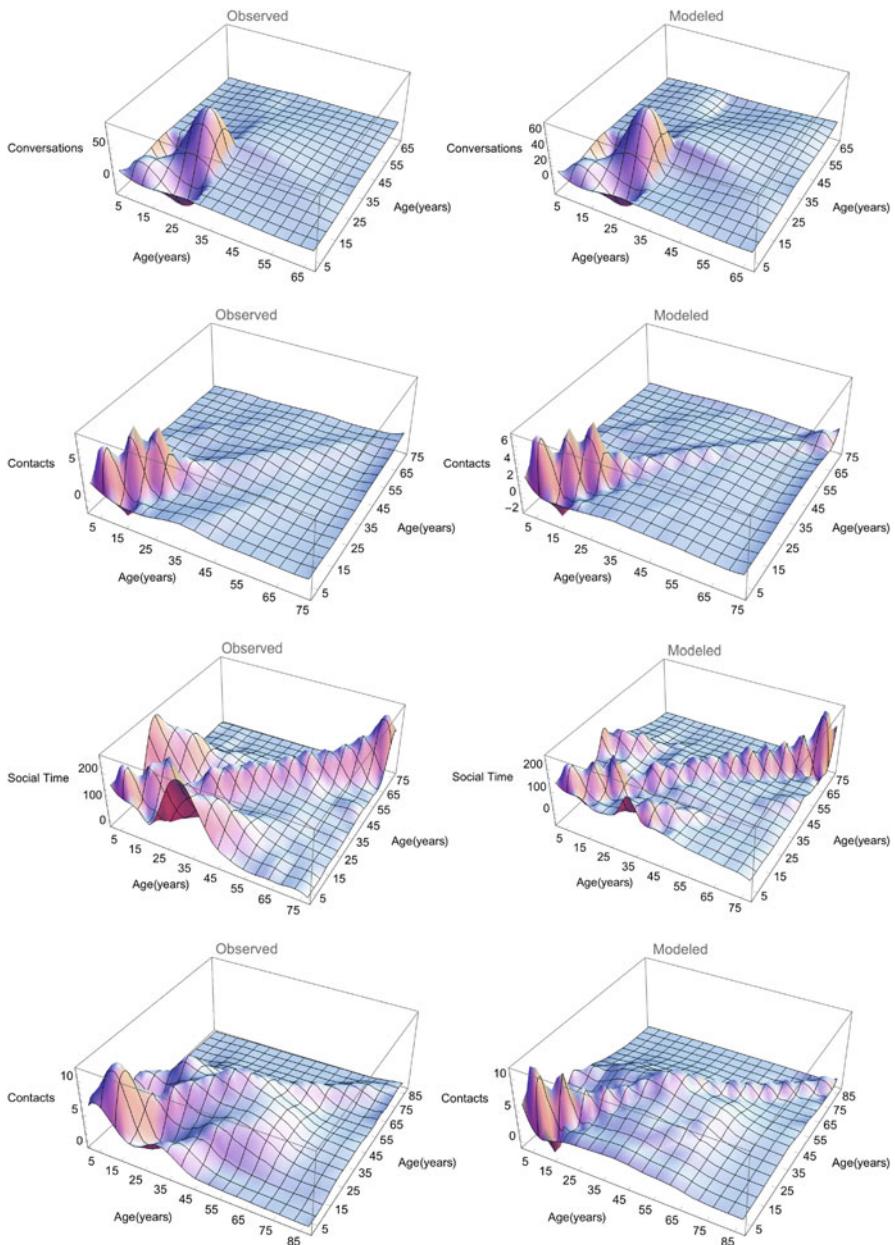


Fig. 5.4 Comparison of the mixing patterns generated by the four empirical studies (top to bottom): [17, 41, 47, 49] (left panel) and the fitted model (5.34) and (5.35) (right panel). Interpolating functions are fitted to geometric means of corresponding row- and column-elements of the mixing matrix *Source*: [25]

The mixing matrix P can incorporate heterogeneities in contact rate (a_i), population size (N_i), preference for mixing within the same sub-population (π_i), etc. Because of these heterogeneities, the optimal combination of vaccination coverages ϕ_i is unlikely to be homogeneous. Typically, the matrix P has to satisfy the following conditions of [14]:

$$\begin{aligned} p_{ij} &\geq 0, \quad i, j = 1, \dots, n, \\ \sum_{j=1}^n p_{ij} &= 1, \quad i = 1, \dots, n, \\ a_i N_i p_{ij} &= a_j N_j p_{ji}, \quad i, j = 1, \dots, n. \end{aligned} \tag{5.38}$$

A commonly used non-homogeneous mixing function that satisfies conditions in (5.38) is the preferred mixing function of [30] given by

$$p_{ij} = \pi_i \delta_{ij} + (1 - \pi_i) \frac{(1 - \pi_j) a_j N_j}{\sum_{k=1}^n (1 - \pi_k) a_k N_k}, \quad i, j = 1, \dots, n, \tag{5.39}$$

where $\pi_i \in [0, 1]$ is the fraction of contacts of group i that is reserved for itself (preferential mixing), whereas the complement $(1 - \pi_i)$ is distributed among all sub-populations in proportion to the unreserved contacts, including i (proportionate mixing). We refer to the mixing given by (5.39) as Jacquez-type preferred mixing. Two extreme cases of (5.39) are the *proportionate mixing* when $\pi_i = 0$ and the *isolated mixing* (i.e., no interactions between sub-populations) when $\pi_i = 1$.

For model (5.37), the basic and effective sub-population reproduction numbers, denoted, respectively, by \mathcal{R}_{0i} and \mathcal{R}_{vi} , for sub-population i ($i = 1, 2, \dots, n$) are given by

$$\mathcal{R}_{0i} = \rho a_i, \quad \mathcal{R}_{vi} = \mathcal{R}_{0i}(1 - \phi_i), \quad i = 1, 2, \dots, n, \tag{5.40}$$

where

$$\rho = \frac{\sigma}{\gamma + \theta}.$$

The next generation matrix (NGM) corresponding to this metapopulation model is

$$K_v = \begin{pmatrix} \mathcal{R}_{v1} p_{11} & \mathcal{R}_{v1} p_{12} & \cdots & \mathcal{R}_{v1} p_{1n} \\ \mathcal{R}_{v2} p_{21} & \mathcal{R}_{v2} p_{22} & \cdots & \mathcal{R}_{v2} p_{2n} \\ \vdots & \ddots & & \vdots \\ \mathcal{R}_{vn} p_{n1} & \mathcal{R}_{vn} p_{n2} & \cdots & \mathcal{R}_{vn} p_{nn} \end{pmatrix}. \tag{5.41}$$

Then the effective reproduction number for the metapopulation is given as

$$\mathcal{R}_v = r(K_v),$$

which is the spectral radius (and the dominant eigenvalue, by the Perron–Frobenius Theorem) of the non-negative matrix K_v . Let $\phi = (\phi_1, \phi_2, \dots, \phi_n)$. Naturally, $\mathcal{R}_v = \mathcal{R}_v(\phi)$ is a function of ϕ . The total number of vaccine doses, denoted by η , is $\eta = \sum_{i=1}^n \phi_i N_i$. For demonstration purposes, we will assume that vaccine efficacy is 100%. We focus on identifying the most efficient allocation of vaccine $\phi = (\phi_1, \phi_2, \dots, \phi_n) \in [0, 1]^n$ for reducing \mathcal{R}_v with limited vaccine doses η or using fewest doses to achieve $\mathcal{R}_v < 1$ (to prevent outbreaks). More specifically, we consider the following two constrained optimization problems:

- (I) Minimize $\mathcal{R}_v = \mathcal{R}_v(\phi)$, subject to $\ell(\phi) := \sum_{i=1}^n \phi_i N_i = \eta$, for $\phi \in [0, 1]^n$.
- (II) Minimize $\eta = \sum_{i=1}^n \phi_i N_i$, subject to $\mathcal{R}_v(\phi) \leq 1$, for $\phi \in [0, 1]^n$.

Consider the optimization problem only for the case of $\mathcal{R}_0 = \mathcal{R}_v(0) \geq 1$, as there will be no outbreak if $\mathcal{R}_0 < 1$. If a solution to Problem (I) exists for a given value of η , let $\Phi^* = \Phi^*(\eta)$ and $\mathcal{R}_{v\{\min\}}(\eta)$ denote the optimal vaccination allocation and corresponding minimum reproduction number, respectively. Let

$$\Omega_\phi^{(n)}(\eta) := \left\{ (\phi_1, \phi_2, \dots, \phi_n) : \ell(\phi) = \eta \right\}.$$

Then, for the solution Φ^* to be feasible, we need to have

$$\begin{aligned} \Phi^*(\eta) &= (\phi_1^*(\eta), \phi_2^*(\eta), \dots, \phi_n^*(\eta)) \in [0, 1]^n, \\ \mathcal{R}_{v\{\min\}}(\eta) &= \min_{\Omega_\phi^{(n)}(\eta) \cap [0, 1]^n} \mathcal{R}_v = \mathcal{R}_v|_{\Phi^*(\eta)}. \end{aligned} \quad (5.42)$$

An optimal solution $\Phi^*(\eta)$ to Problem (I) that lies in the interior of the unit hypercube must satisfy the following equations:

$$\begin{aligned} \nabla \mathcal{R}_v|_{\Phi^*(\eta)} &= \tilde{\lambda} \nabla \ell = \tilde{\lambda} (N_1, \dots, N_n), \quad \Phi^*(\eta) \in (0, 1)^n, \\ \ell|_{\Phi^*(\eta)} &= \sum_{i=1}^n \phi_i^*(\eta) N_i = \eta, \end{aligned} \quad (5.43)$$

where the constant $\tilde{\lambda}$ is the Lagrange multiplier.

Similarly, if an optimal solution to Problem (II) exists, it is useful practically to have an explicit expression or estimate of the bounds for the minimum vaccine doses needed, which we denote by η_* . The significance of η_* is that it is the smallest number of vaccination doses that can prevent outbreaks under an optimal vaccination policy.

To find η_* , notice that $\mathcal{R}_v(\phi)$ is a monotonically decreasing function of ϕ_i , and thus, a decreasing function of $\eta = \sum_{i=1}^n \phi_i N_i$. Therefore, the inequality constraint $\mathcal{R}_v(\phi) \leq 1$ can be replaced by an equality constraint $\mathcal{R}_v(\phi) = 1$, and thus,

$$\eta_* = \min_{\{\mathcal{R}_v(\phi)=1\} \cap [0,1]^n} \ell(\phi).$$

It follows that η_* is the minimum of $\eta \in [0, N]$ such that $\mathcal{R}_{v\{\min\}}(\eta) = 1$ and can be found by solving the equation:

$$\mathcal{R}_{v\{\min\}}(\eta_*) = \mathcal{R}_v|_{P^*(\eta_*)} = 1. \quad (5.44)$$

Before we present the results of optimal solutions for Problems (I) and (II), we state the following important results, which is proved in [43], regarding the bounds for the effective reproduction number \mathcal{R}_v for general mixing matrices (not just Jacquez-type preferred mixing given in (5.39)).

Theorem 5.2 (Bounds for $\mathcal{R}_v(\phi)$) *Let P be a non-negative, invertible, irreducible matrix such that $-P^{-1}$ is essentially non-negative and the conditions (5.38) are satisfied. Then*

(a) *The lower and upper bounds of $\mathcal{R}_v(\phi)$ are*

$$\sum_{i=1}^n \omega_i \mathcal{R}_{vi} \leq \mathcal{R}_v \leq \max\{\mathcal{R}_{v1}, \dots, \mathcal{R}_{vn}\}, \quad \text{where } \omega_i = \frac{a_i N_i}{\sum_{k=1}^n a_k N_k}. \quad (5.45)$$

(b) *The lower and upper bounds of $\mathcal{R}_v(\phi)$ correspond to the cases of proportionate mixing and isolated mixing, respectively.*

For the optimal solutions of Problems (I) and (II), it is shown in [43] that for the case of $n = 2$ explicit expressions for Φ^* and η_* are possible, and for the case of $n > 2$ their upper and lower bounds can be obtained.

For ease of presentation, introduce the following notation:

$$\begin{aligned} \kappa_1 &:= p_{22} \sqrt{N_1 N_2 \mathcal{R}_{02}} - N_2 \sqrt{p_{12} p_{21} \mathcal{R}_{01} \mathcal{R}_{02}}, \\ \kappa_2 &:= p_{11} \sqrt{N_1 N_2 \mathcal{R}_{01}} - N_1 \sqrt{p_{12} p_{21} \mathcal{R}_{01} \mathcal{R}_{02}}, \\ \eta_0 &:= N - \frac{\kappa_1 N_1 + \kappa_2 N_2}{\max\{\kappa_1, \kappa_2\}}. \end{aligned} \quad (5.46)$$

For the mixing given in (5.39), it is easy to verify the following fact:

$$|P| = \begin{vmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{vmatrix} = \pi_1 \pi_2 + \frac{\pi_1 (1 - \pi_2)^2 a_2 N_2 + \pi_2 (1 - \pi_1)^2 a_1 N_1}{(1 - \pi_1) a_1 N_1 + (1 - \pi_2) a_2 N_2} > 0, \quad (5.47)$$

provided that $\pi_i \in (0, 1)$, $a_i > 0$, and $N_i > 0$, $i = 1, 2$.

Theorem 5.3 (Optimal Solution to Problem (I) When $n = 2$) Consider $\mathcal{R}_v = \mathcal{R}_v(\phi_1, \phi_2)$ as a function of ϕ_1 and ϕ_2 , and let η_0 and κ_i be given in (5.46). Assume that condition (5.47) holds.

- (a) For a given value of η , the optimal point $\Phi^*(\eta)$ exists and lies in the interior of the unit square if and only if

$$\eta_0 < \eta < N \quad \text{and} \quad \kappa_i > 0 \text{ for } i = 1, 2. \quad (5.48)$$

- (b) For each $\eta_0 < \eta < N$, the explicit formulae for $P^*(\eta)$ and $\mathcal{R}_{v\{\min\}}(\eta)$ are given by

$$\Phi^*(\eta) = (1, 1) - \frac{N - \eta}{\kappa_1 N_1 + \kappa_2 N_2} (\kappa_1, \kappa_2) \quad (5.49)$$

$$\mathcal{R}_{v\{\min\}}(\eta) = |P| \mathcal{R}_{01} \mathcal{R}_{02} \sqrt{N_1 N_2} \frac{N - \eta}{\kappa_1 N_1 + \kappa_2 N_2}. \quad (5.50)$$

- (c) If $0 < \eta < \eta_0$, the minimum point $\Phi^*(\eta)$ is one of the boundary points $(\eta/N_1, 0)$ or $(0, \eta/N_2)$ and hence

$$\mathcal{R}_{v\{\min\}}(\eta) = \min\{\mathcal{R}_v(\eta/N_1, 0), \mathcal{R}_v(0, \eta/N_2)\}.$$

For the general case of $n > 2$, Theorem 5.2 can be used to derive the lower and upper bounds for the minimum reproduction number $\mathcal{R}_{v\{\min\}}(\eta)$. To facilitate biological interpretations, introduce the following notation:

$$\begin{aligned} f_i &:= N_i/N, \quad 1 \leq i \leq n \quad \text{Population fraction of sub-population } i; \\ \mathcal{U} &:= \sum_{i=1}^n (1 - \phi_i) f_i \quad \text{Population fraction unvaccinated;} \\ \widehat{\mathcal{R}}_0 &:= \sum_{i=1}^n \mathcal{R}_{0i} f_i \quad \text{Population weighted reproduction number;} \\ \mathcal{R}_0^\diamond &:= \left(\sum_{i=1}^n \frac{1}{\mathcal{R}_{0i}} f_i \right)^{-1} \quad \text{Harmonic mean of } \mathcal{R}_{0i} \text{ weighted by} \\ &\quad \text{sub-population fractions } f_i; \\ \tilde{\mathcal{R}}_0 &:= \min_i \mathcal{R}_{0i}^2 / \widehat{\mathcal{R}}_0 \quad \text{Analogous to a scaled reproduction number.} \end{aligned} \quad (5.51)$$

The following results provide the lower and upper bounds for the minimum $\mathcal{R}_{v\{\min\}}(\eta)$ in Problem (I):

Theorem 5.4 Assume that the conditions of Theorem 5.2 hold. Let $\eta < N$, and let \mathcal{U} , $\widehat{\mathcal{R}}_0$, \mathcal{R}_0^\diamond , and $\tilde{\mathcal{R}}_0$ be defined in (5.51).

(a) The bounds of $\mathcal{R}_{v\{\min\}}(\eta)$ for $\phi \in \Omega_p^{(n)}(\eta) \cap [0, 1]^n$ are

$$\tilde{\mathcal{R}}_0 \mathcal{U} \leq \mathcal{R}_{v\{\min\}}(\eta) \leq \mathcal{R}_0^\diamond \mathcal{U}. \quad (5.52)$$

(b) If $\mathcal{R}_{0i} > 1$ for all i , then

$$\frac{\eta_*}{N} \leq 1 - \frac{1}{\mathcal{R}_0^\diamond}. \quad (5.53)$$

Remarks The bounds for the optimal solutions have clear biological meanings based on the biological interpretations of the quantities in (5.51). (i) Note that \mathcal{R}_0^\diamond and $\tilde{\mathcal{R}}_0$ are weighted basic reproduction numbers, and the factor \mathcal{U} is the fraction of the overall population that remains susceptible. In light of this, we see that the lower and upper bounds for $\mathcal{R}_{v\{\min\}}(\eta)$ in (5.52) take the familiar form of an effective reproduction number. (ii) For the upper bound of η_* , if $a_i = a$ are all the same, we have $\mathcal{R}_0^\diamond = \mathcal{R}_0$, in which case the upper bound in (5.53) becomes $1 - 1/\mathcal{R}_0$. This is similar to the usual formula for the critical vaccination fraction $\phi_c = 1 - 1/\mathcal{R}_0$, for which the number of vaccinated is $\eta_c = \phi_c N = N(1 - 1/\mathcal{R}_0)$.

Although various observations about the effect of mixing on reproduction numbers have been made in previous studies, the result stated in Theorem 5.2 provided a definitive lower and upper bounds corresponding to the proportionate and the isolated mixing (a rigorous proof can be found in [43]) for a large class of mixing matrix P (not just Jacquez-type). Using a model metapopulation composed of a city and several villages, May and Anderson [36, 37] showed that heterogeneity in relevant sub-population characteristics also increased \mathcal{R}_v . Hethcote and van Ark [28] argued that person-to-person contact rates in densely populated urban areas should be no more than twice those in sparsely populated rural ones. This change in parameter values diminished the apparent effect of heterogeneity. The facts that population heterogeneities tend to increase \mathcal{R}_0 and that models assuming proportionate mixing generate lower values of \mathcal{R}_0 have been suggested by other researchers [1, 3, 20].

Example 4 Figure 5.5 illustrates an example from [21], which extends May and Anderson's [36] conclusion that "under a uniformly applied immunization program, the overall fraction that must be immunized is larger than would be estimated by (incorrectly) assuming the population to be homogeneously mixed." Consider $\mathcal{R}_v = \mathcal{R}_v(\phi_1, \phi_2)$ as a function of vaccination coverage (ϕ_1, ϕ_2) . The two contour plots of \mathcal{R}_v are for the cases of (a) homogeneous contacts ($a_1 = a_2 = 10$) and (b) heterogeneous contacts ($a_1 = 8, a_2 = 12$), while other parameters are the same for the two sub-populations ($N_1 = N_2, \pi_1 = \pi_2 = 0.6, \sigma = 0.05, \gamma = 1/7, \theta = 1/(365 \times 70)$). Note that the total number of vaccine doses is given by $\phi_1 N_1 + \phi_2 N_2$. Because $N_1 = N_2$, a vaccination pair (ϕ_1, ϕ_2) that minimizes the total doses if and only if it minimizes the quantity $\phi_1 + \phi_2$. The thicker curve is the contour for $\mathcal{R}_v = 1$. The thick dashed line with slope -1 is $\phi_1 + \phi_2 = c$ for a

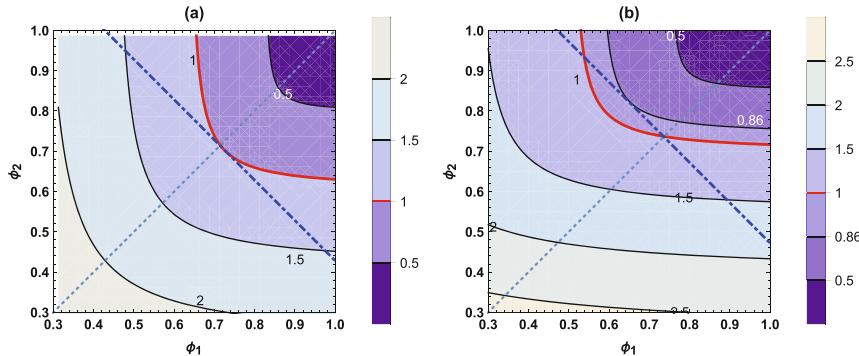


Fig. 5.5 Contour plots of \mathcal{R}_v as a function of ϕ_1 and ϕ_2 for (a) homogeneous population ($a_1 = a_2$) and (b) heterogeneous population ($a_1 \neq a_2$). For both plots, $\pi_1 = \pi_2$ and $N_1 = N_2$. The thick solid curve represents the contour $\mathcal{R}_v = 1$. The thin and thick dashed lines correspond to $\phi_1 = \phi_2$ and $\phi + \phi_2 = 2 \times 0.74$, respectively. All points (ϕ_1, ϕ_2) on the line $\phi_1 + \phi_2 = c$ for a constant $c > 0$ correspond to the same total vaccination doses. Source: [21]

constant $c > 0$. It shows in (a) that $\mathcal{R}_v(0.74, 0.74) = 1$ and that $\mathcal{R}_v(\phi_1, \phi_2) > 1$ for all other pairs (ϕ_1, ϕ_2) with $\phi_1 + \phi_2 = 2 \times 0.74$. This suggests that the optimal allocation is the homogeneous coverage $\phi_1 = \phi_2$. However, the plot in (b) shows a very different result. Particularly, among all pairs (ϕ_1, ϕ_2) on the line $\phi_1 + \phi_2 = 2 \times 0.74$, some can make $\mathcal{R}_v(\phi_1, \phi_2) < 1$. In fact, there is one point (ϕ_{1c}, ϕ_{2c}) at which the minimum $\mathcal{R}_v(\phi_{1c}, \phi_{2c}) = 0.86$ is achieved. This suggests that, in a non-homogeneous population, uniform coverage (equal ϕ_i for all i) may not be the most efficient.

5.4 A Heterogeneous Mixing Age of Infection Model

The basic age of infection model extends the simple *SIR* epidemic model by allowing an arbitrary number of stages in the model and arbitrary distributions of stay in each stage. However, it does not include the possibility of subgroups with different activity levels and heterogeneous mixing between subgroups. This possibility can be included in a heterogeneous mixing age of infection model. As in homogeneous mixing models, the age of infection approach is more general than simpler models in several respects. Age of infection models allow arbitrary distributions of stay in compartments and arbitrary sequences of compartments. In addition, they allow variable infectivity. This can be included in the kernel $A(s)$ which leads to the infectivity function $\varphi(t)$ describing infectivity rather than simply counting the number of infectives.

As in the previous section, we consider two sub-populations of sizes N_1, N_2 , respectively, each divided into susceptibles and infected members with subscripts to identify the sub-population. Suppose that $A_i(s)$ is the mean infectivity of individuals who have been infected s time units previously, and that a_1, a_2 are the contact rates of the two sub-populations. It is necessary to describe also the mixing between the two groups. Suppose that the fraction of contacts made by a member of group i that is with a member of group j is p_{ij} , $i, j = 1, 2$. Then

$$p_{11} + p_{12} = p_{21} + p_{22} = 1.$$

A two-group model may describe a population with groups differing by activity levels and possibly by vulnerability to infection, so that $a_1 \neq a_2$ but $A_1(s) = A_2(s)$. It may also describe a population with one group which has been vaccinated against infection, so that the two groups have the same activity level but different disease model parameters. In this case, $a_1 = a_2$ but $A_1(\tau) \neq A_2(\tau)$. In this model, any differences between groups in susceptibility or infectivity are included in the factors $A_1(s), A_2(s)$.

An age of infection model with two subgroups is

$$\begin{aligned} S'_i &= -a_i S_i \left[\frac{p_{ii}}{N_i} \varphi_i + \frac{p_{ij}}{N_j} \varphi_j \right] \\ \varphi_i(t) &= \int_0^\infty [-S'_i(t-\tau)] A_i(\tau) d\tau, \quad i, j = 1, 2, \quad i \neq j. \end{aligned}$$

Here, $\varphi_i(t)$ is the total infectivity of infected members of group i ($i = 1, 2$).

As for the homogeneous mixing model, we may write this model using only the equations for S_i ,

$$\begin{aligned} S'_i(t) &= -a_i S_i(t) \left[\frac{p_{ii}}{N_i} \int_0^\infty A_i(s) S'_i(t-s) ds + \frac{p_{ij}}{N_j} \int_0^\infty A_j(s) S'_j(t-s) ds \right], \\ &\quad i, j = 1, 2, \quad i \neq j. \end{aligned} \tag{5.54}$$

The next generation matrix is

$$P = \begin{bmatrix} a_1 p_{11} \int_0^\infty A_1(\tau) ds & a_1 p_{12} \frac{N_1}{N_2} \int_0^\infty A_2(s) ds \\ a_2 p_{21} \frac{N_2}{N_1} \int_0^\infty A_1(s) ds & a_2 p_{22} \int_0^\infty A_2(s) ds \end{bmatrix}.$$

The matrix P is similar to the matrix $Q = R^{-1} P R$, with

$$R = \begin{bmatrix} N_1 & 0 \\ 0 & N_2 \end{bmatrix}$$

and

$$Q = \begin{bmatrix} a_1 p_{11} \int_0^\infty A_1(s) ds & a_1 p_{12} \int_0^\infty A_2(s) ds \\ a_2 p_{21} \int_0^\infty A_1(s) ds & a_2 p_{22} \int_0^\infty A_2(s) ds \end{bmatrix}.$$

Thus \mathcal{R}_0 is the largest root of

$$\det \begin{bmatrix} a_1 p_{11} \int_0^\infty A_1(s) ds - \lambda & a_1 p_{12} \int_0^\infty A_2(s) ds \\ a_2 p_{21} \int_0^\infty A_1(s) ds & a_2 p_{22} \int_0^\infty A_2(s) ds - \lambda \end{bmatrix} = 0. \quad (5.55)$$

In order to obtain an invasion criterion, initially when $S_1(t)$ is close to $S_1(0) = N_1$ and $S_2(t)$ is close to $S_2(0) = N_2$, we replace $S_1(t)$ and $S_2(t)$ by N_1, N_2 , respectively, to give a linear system, and the condition that this linear system has a solution $S_i(t) = N_i e^{rt}$ ($i = 1, 2$) is

$$1 = a_i p_{i1} \int_0^\infty e^{-rs} A_1(s) ds + a_i p_{i2} \int_0^\infty e^{-rs} A_2(s) ds, \quad i = 1, 2. \quad (5.56)$$

The initial exponential growth rate is the solution r of the equation

$$\det \begin{bmatrix} a_1 p_{11} \int_0^\infty e^{-rs} A_1(s) ds - 1 & a_1 p_{12} \int_0^\infty e^{-rs} A_2(s) ds \\ a_2 p_{21} \int_0^\infty e^{-rs} A_1(s) ds & a_2 p_{22} \int_0^\infty e^{-rs} A_2(s) ds - 1 \end{bmatrix} = 0. \quad (5.57)$$

In the special case of proportionate mixing, in which $p_{11} = p_{21}, p_{12} = p_{22}$, so that $p_{12}p_{21} = p_{11}p_{22}$, the basic reproduction number is given by

$$\mathcal{R}_0 = a_1 p_{11} \int_0^\infty A_1(s) ds + a_2 p_{22} \int_0^\infty A_2(s) ds,$$

and Eq. (5.57) reduces to

$$a_1 p_{11} \int_0^\infty e^{-rs} A_1(s) ds + a_2 p_{22} \int_0^\infty e^{-rs} A_2(s) ds = 1. \quad (5.58)$$

There is an epidemic if and only if $\mathcal{R}_0 > 1$.

In the special case in which the two groups have the same infectivity distribution but may have different activity levels and possibly vulnerability to infection, so that $A_1(s) = A_2(s) = A(s)$, \mathcal{R}_0 is the largest root of

$$\det \begin{bmatrix} a_1 p_{11} \int_0^\infty A(s) ds - \lambda & a_1 p_{12} \int_0^\infty A(s) ds \\ a_2 p_{21} \int_0^\infty A(s) ds & a_2 p_{22} \int_0^\infty A(s) ds - \lambda \end{bmatrix}$$

and the initial exponential growth rate is the solution r of the equation

$$\det \begin{bmatrix} a_1 p_{11} \int_0^\infty e^{-rs} A(s) ds - 1 & a_1 p_{12} \int_0^\infty e^{-rs} A(s) ds \\ a_2 p_{21} \int_0^\infty e^{-rs} A(s) ds & a_2 p_{22} \int_0^\infty e^{-rs} A(s) ds - 1 \end{bmatrix} = 0. \quad (5.59)$$

Comparing Eqs. (5.55) and (5.59), we see that each of $\mathcal{R}_0 / \int_0^\infty A(\tau) d\tau$ and $1 / \int_0^\infty e^{-r\tau} A(\tau) d\tau$ is the largest eigenvalue of the matrix

$$\begin{bmatrix} a_1 p_{11} & a_1 p_{12} \\ a_2 p_{21} & a_2 p_{22} \end{bmatrix},$$

the largest root of the equation

$$x^2 - (a_1 p_{11} + a_2 p_{22})x + a_1 a_2 (p_{11} p_{22} - p_{12} p_{21}) = 0.$$

Thus

$$\frac{\mathcal{R}_0}{\int_0^\infty A(s) ds} = \frac{1}{\int_0^\infty e^{-rs} A(s) ds},$$

which implies the same relation as for the homogeneous mixing model. Thus, if we assume heterogeneous mixing, we obtain the same estimate of the reproduction number from observation of the initial exponential growth rate, and this conclusion remains valid for an arbitrary number of groups with different contact rates. The estimate of the basic reproduction number from the initial exponential growth rate does not depend on heterogeneity of the model. This result does not generalize to the case $A_1(s) \neq A_2(s)$, but it does remain valid for an arbitrary number of groups with different contact rates.

5.4.1 The Final Size of a Heterogeneous Mixing Epidemic

With homogeneous mixing, knowledge of the basic reproduction number translates into knowledge of the final size of the epidemic. However, with heterogeneous mixing, even in the simplest case of proportionate mixing, the size of the epidemic is not determined uniquely by the basic reproduction number.

For the heterogeneous mixing model (5.54) there is a pair of final size relations. We divide the equation for S_1 in (5.54) by $S_i(t)$ and integrate with respect to t from 0 to ∞ . Much as in the derivation of the final size relation for the homogeneous

mixing model we obtain a pair of final size relations which may be solved for $S_1(\infty)$ and $S_2(\infty)$:

$$\log \frac{S_i(0)}{S_i(\infty)} = \sum_{j=1}^2 \left[a_i \frac{p_{ij}}{N_j} (N_j - S_j(\infty)) \int_0^\infty A_j(s) ds \right], \quad i = 1, 2. \quad (5.60)$$

The system of equations (5.60) has a unique solution $(S_1(\infty), S_2(\infty))$. In order to prove this, we define

$$g_i(x_1, x_2) = \log \frac{S_i(0)}{x_i} - a_i \sum_{j=1}^2 p_{ij} \left[1 - \frac{x_j}{N_j} \right] \int_0^\infty A_j(s) ds.$$

A solution of (5.60) is a solution (x_1, x_2) of the system

$$g_i(x_1, x_2) = 0, \quad i = 1, 2.$$

For each x_2 , $g_1(0^+, x_2) > 0$, $g_1(S_1(0), x_2) < 0$. Also, as a function of x_1 , $g_1(x_1, x_2)$ either decreases or decreases initially and then increases to a negative value when $x_1 = S_1(0)$. Thus for each $x_2 < S_2(0)$, there is a unique $x_1(x_2)$ such that $g_1(x_1(x_2), x_2) = 0$. Also, since $g_1(x_1, x_2)$ is an increasing function of x_2 , the function $x_1(x_2)$ is increasing. Now, since $g_2(x_1, 0^+) > 0$, $g_2(x_1, S_2(0)) < 0$, there exists x_2 such that $g_2(x_1(x_2), x_2) = 0$. Also, $g_2(x_1(x_2), x_2)$ either decreases monotonically or decreases initially and then increases to a negative value when $x_2 = S_2(0)$. Therefore this solution is also unique. This implies that

$$(x_1(x_2), x_2)$$

is the unique solution of the final size relations.

Numerical simulations indicate that models with heterogeneous mixing may give very different epidemic sizes than models with the same basic reproduction number and homogeneous mixing. The reproduction number of an epidemic model is not sufficient to determine the size of the epidemic if there is heterogeneity in the model. We conjecture that for a given value of the basic reproduction number the maximum epidemic size for any mixing is obtained with homogeneous mixing.

Assume that the parameters $N_1, N_2, \int_0^\infty A(\tau) d\tau$ remain fixed and attempt to minimize $S_1(\infty) + S_2(\infty)$ as a function of a_1, a_2 (with a_1, a_2 constrained to keep $p_1 a_1 + p_2 a_2 = k$ fixed and p_1, p_2 as specified by proportionate mixing). Homogeneous mixing corresponds to $a_1 = a_2$.

The constraint relating a_1, a_2 implies that

$$\frac{da_2}{da_1} = \frac{2a_1 - k}{k - 2a_2} \cdot \frac{N_1}{N_2},$$

when $a_1 = a_2 = k$ we have

$$\frac{da_2}{da_1} = -\frac{N_1}{N_2}.$$

Also, when $a_1 = a_2$,

$$p_1 = \frac{N_1}{N_1 + N_2}, \quad p_2 = \frac{N_2}{N_1 + N_2}, \quad \frac{S_1}{N_1} = \frac{S_2}{N_2}, \quad \frac{dp_1}{da_1} = \frac{N_1}{kN}.$$

If we differentiate with respect to a_1 , we can calculate that

$$\frac{d[S_1(\infty) + S_2(\infty)]}{da_1} = 0$$

when $a_1 = a_2$. We believe that $a_1 = a_2$ is the only critical point of $S_1(\infty) + S_2(\infty)$, although we have not been able to verify this analytically. If $a_1 = a_2$ is the only critical point of $S_1(\infty) + S_2(\infty)$, this critical point must be a minimum. We conjecture that this result is also valid if we allow arbitrary mixing, that is, we conjecture that for a given value of the basic reproduction number the maximum epidemic size for any mixing is obtained with homogeneous mixing.

While we have confined the description of the heterogeneous mixing situation to a two-group model, the extension to an arbitrary number of groups is straightforward. We suggest that in advance planning for a pandemic, the number of groups to be considered for different treatment rates should determine the number of groups to be used in the model. On the other hand, the number of groups to be considered should also depend on the amount and reliability of data, and these two criteria may be contradictory. A model with fewer groups and parameters chosen as weighted averages of the parameters for a model with more groups may give predictions that are quite similar to those of the more detailed models. We suggest also that use of the final size relations for a model with total population size assumed constant is a good time-saving procedure for making predictions if the disease death rate is small.

We have seen that in the case of homogeneous mixing, knowledge of the initial exponential growth rate and the infective period distribution is sufficient to determine the basic reproduction number and thence the final size of an epidemic. In the case of heterogeneous mixing, knowledge of the initial exponential growth rate and the infective period distribution is sufficient to determine the basic reproduction number, but not to determine the final size of the epidemic.

This raises the question of what additional information that may be measured at the start of a disease outbreak would suffice to determine the epidemic final size if the mixing is heterogeneous.

We assume that $A_1(s)$, $A_2(s)$, and the mixing matrix

$$M = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$$

are known. The next generation matrix is

$$K = \begin{bmatrix} a_1 \frac{p_{11}}{N_1} \int_0^\infty A_1(s)ds & a_1 \frac{p_{12}}{N_2} \int_0^\infty A_2(s)ds \\ a_2 \frac{p_{21}}{N_1} \int_0^\infty A_1(s)ds & a_2 \frac{p_{22}}{N_2} \int_0^\infty A_2(s)ds \end{bmatrix},$$

and \mathcal{R}_0 is the largest (positive) eigenvalue of this matrix. There is a corresponding eigenvector with positive components

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}.$$

Since the components of this eigenvector give the proportions of infective cases in the two groups initially, it is reasonable to hope to be able to determine this eigenvector from early outbreak data.

The general final size relation is

$$\log \frac{S_i(0)}{S_i(\infty)} = \sum_{j=1}^2 \left[a_i p_{ij} \left(1 - \frac{S_j(\infty)}{N_j} \right) \int_0^\infty A_j(s)ds \right], \quad i = 1, 2.$$

These equations may be solved for $S_1(\infty)$, $S_2(\infty)$ if the contact rates a_1 , a_2 can be determined from the available information.

The condition that the vector \mathbf{u} with components (u_1, u_2) is an eigenvector of the next generation matrix corresponding to the eigenvalue \mathcal{R}_0 is

$$a_i(p_{i1}u_1 + p_{i2}u_2) \int_0^\infty A_i(s)ds = \mathcal{R}_0 u_i, \quad i = 1, 2,$$

and since it is assumed that the function $A(\tau)$, the vector \mathbf{u} , and the mixing matrix (p_{ij}) are known these equations determine a_1 and a_2 .

In vector notation, if we define the column vector

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

and the row vectors

$$\mathbf{M}_j = [p_{j1} \ p_{j2}],$$

we have

$$a_j = \frac{\mathcal{R}_0}{\mathbf{M}_j \mathbf{u} \int_0^\infty A_j(s)ds} u_j.$$

When these values are substituted into the final size system, $S_1(\infty)$ and $S_2(\infty)$ may be determined. This argument extends easily to models with an arbitrary number of activity groups.

In real-life applications, there are usually many groups, and the final size of an epidemic is obtained most efficiently by numerical simulations. The results obtained here are more likely to be useful in theoretical applications, such as comparisons of different control strategies.

One may think of the case $a_1 \neq a_2$, $A_1(s) = A_2(s)$ as a model for a disease with heterogeneous mixing but not treatment and the case $a_1 = a_2$, $A_1(s) \neq A_2(s)$ as a model for a disease in which the mixing is homogeneous but treatment that changes the infective period distribution has been applied to a part of the population. Of course, if the treatment also includes quarantine that also changes the contact rate, the case $a_1 \neq a_2$, $A_1(s) \neq A_2(s)$ would be appropriate.

We suggest that in advance planning for a pandemic, the number of groups to be considered for different treatment rates should determine the number of groups to be used in the model. On the other hand, the number of groups to be considered should also depend on the amount and reliability of data, and these two criteria may be contradictory. A model with fewer groups and parameters chosen as weighted averages of the parameters for a model with more groups may give predictions that are quite similar to those of the more detailed models. We suggest also that use of the final size relations for a model with total population size assumed constant is a good time-saving procedure for making predictions if the disease death rate is small.

5.5 Some Warnings

An actual epidemic differs considerably from the idealized models such as (5.1) as well as the extensions considered later. Some notable differences are:

1. When it is realized that an epidemic has begun, individuals are likely to modify their behavior by avoiding crowds to reduce their contacts and by being more careful about hygiene to reduce the risk that a contact will produce infection.
2. If a vaccine is available for the disease which has broken out, public health measures will include vaccination of part of the population. Various vaccination strategies are possible, including vaccination of health care workers and other first line responders to the epidemic, vaccination of members of the population who have been in contact with diagnosed infectives, or vaccination of members of the population who live in close proximity to diagnosed infectives.
3. Diagnosed infectives may be hospitalized, both for treatment and to isolate them from the rest of the population. Isolation may be imperfect; in-hospital transmission of infection was a major problem in the SARS epidemic.
4. Contact tracing of diagnosed infectives may identify people at risk of becoming infective, who may be quarantined (instructed to remain at home and avoid

contacts) and monitored so that they may be isolated immediately if and when they become infective.

5. In some diseases, exposed members who have not yet developed symptoms may already be infective, and this would require inclusion in the model of new infections caused by contacts between susceptibles and asymptomatic infectives from the exposed class.
6. In the SARS epidemic of 2002–2003 in-hospital transmission of disease from patients to health care workers or visitors because of imperfect isolation accounted for many of the cases. This points to an essential heterogeneity in disease transmission which must be included whenever there is any risk of such transmission.

5.6 *Projects: Reproduction Numbers for Discrete Models

This project concerns the computation of the reproduction number for discrete models using the approach of the next generation matrix. A formula for the reproduction number \mathcal{R} (either \mathcal{R}_0 or \mathcal{R}_C) is derived by adopting the method used in [2] based on the next generation matrix approach. That is, in the discrete-time case

$$\mathcal{R} = \varrho(F(I - T)^{-1}), \quad (5.61)$$

where ϱ represents the spectral radius, F is the matrix associated with new infections, and T is the matrix of transitions with $\varrho(T) < 1$ (see [2, 16, 31, 48]). Here F and T are calculated on the infected variables only evaluated at the disease-free equilibrium, and the Jacobian on these variables is $F + T$, which is assumed to be irreducible.

Consider the simple discrete SEIR model with geometric distributions for the latent and infectious period with parameters α and δ ($\alpha < 1, \delta < 1$), respectively. This is equivalent to assuming constant transition probabilities $1 - \alpha$ and $1 - \delta$ per unit time from E to I and from I to R , respectively. The model reads

$$\begin{aligned} S_{n+1} &= S_n e^{-\beta \frac{I_n}{N}}, \\ E_{n+1} &= S_n (1 - e^{-\beta \frac{I_n}{N}}) + \alpha E_n \\ I_{n+1} &= (1 - \alpha) E_n + \delta I_n, \quad n = 1, 2, \dots . \end{aligned} \quad (5.62)$$

For system (5.62), the matrices associated with new infections and transitions are

$$F = \begin{bmatrix} 0 & \beta \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad T = \begin{bmatrix} \alpha & 0 \\ 1 - \alpha & \delta \end{bmatrix},$$

respectively. Then $\varrho(T) = \max\{\alpha, \delta\} < 1$; thus,

$$\mathcal{R} = \varrho(F(I - T)^{-1}) = \frac{\beta}{1 - \delta}. \quad (5.63)$$

Question 1 Extend the model (5.62) by incorporating isolation or hospitalization of infectious individuals. For the extended model, compute the reproduction number using the formula (5.61).

Question 2 Extend the model (5.62) by considering different transmission rates β_i for individuals in different infectious stages I_i ($i = 1, 2, \dots$). Use the formula (5.61) to compute the reproduction number for the extended model.

Project 2 Consider next the case when the infective period follows an arbitrary discrete (bounded) distribution, which is denoted by Y . Let $q_i = \mathbb{P}(Y > i)$ and $\mathbb{P}(Y = i) = q_{i-1} - q_i$. It is easy to see that q_i is a decreasing function, i.e., $q_i \geq q_{i+1}$. In fact, $q_0 = 1$ and $q_m = 0$ for all $m \geq M$, where M is the maximum number of units of time that an individual takes to recover.

Because the geometric is the only memoryless discrete distribution, when other distributions are considered it is necessary to keep track of the past in order to know the values at the present. In fact, it is impossible to use the next generation matrix approach directly because the disease stages (S , E , and I) at time $n + 1$ cannot be written in the form

$$[E_{n+1}, I_{n+1}, S_{n+1}]^T = \mathcal{M}([E_n, I_n, S_n]^T),$$

where $\mathcal{M} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$. To overcome this difficulty we can consider multiple I stages, similar to the approach known as the “linear chain trick” used in continuous models to convert a gamma distribution to a sequence of exponential distributions. Thus, we introduce the subclasses $I^{(1)}, I^{(2)}, \dots, I^{(M)}$ (see Fig. 5.6). The superscript i corresponds to the time since becoming infectious. Notice that these subclasses $I^{(i)}$ are different from those in the negative binomial model because

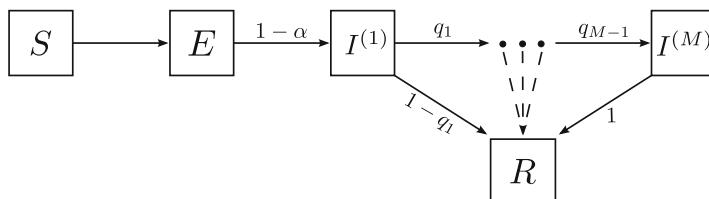


Fig. 5.6 A transition diagram for the case when the stage duration of the infective period has an arbitrary bounded distribution with upper bound M . The superscript i is the stage age in the infectious period and individuals in $I^{(i)}$ (for all i) can enter the recovered class R with a certain probability

here an individual can only stay in $I^{(i)}$ for one unit of time, and must go to either the $I^{(i+1)}$ class with probability q_i or the recovered class R with probability $1 - q_i$.

From Fig. 5.6 the model equations can be written as

$$\begin{aligned} S_{n+1} &= S_n e^{-\sum_{i=1}^M \beta_i \frac{I_n^{(i)}}{N}}, \\ E_{n+1} &= S_n \left[1 - e^{-\sum_{i=1}^M \beta_i \frac{I_n^{(i)}}{N}} \right] + \alpha E_n, \\ I_{n+1}^{(1)} &= (1 - \alpha)E_n, \quad I_{n+1}^{(2)} = q_1 I_n^{(1)}, \quad I_{n+1}^{(j)} = \frac{q_{j-1}}{q_{j-2}} I_n^{(j-1)}, \quad 3 \leq j \leq M, \end{aligned} \quad (5.64)$$

where β_i denote the transmission rates at the infective stage i , $1 \leq i \leq M$. As q_i is the probability that an infective individual remains infective i time units after becoming infective, the transition probability from $I_n^{(2)}$ to $I_{n+1}^{(3)}$ is given by the probability that an infective individual is still infective two time units after becoming infectious given that the person remained infective one time unit ago, i.e., q_2/q_1 . This explains the $I_{n+1}^{(3)}$ equation and similarly $I_{n+1}^{(j)}$ equations for $3 \leq j \leq M$.

Question 1 For the case when transmission rates β_i are stage-dependent, show that

$$\mathcal{R}_0 = \varrho(F(I - T)^{-1}) = \sum_{i=1}^M \beta_i q_{i-1}. \quad (5.65)$$

Question 2 Derive \mathcal{R}_0 from its biological definition. **Hint:** Using the fact that the distribution Y has an upper bound M and that for a given function f

$$\sum_{m=1}^M \mathbb{P}(Y = m) f(m) = \mathbb{E}(f(Y)).$$

The reproduction number from the biological definition (with $f(m) = \sum_{i=1}^m \beta_i$) is

$$\mathcal{R} = \sum_{i=0}^{M-1} \beta_{i+1} q_i.$$

The formula (5.63) can also be applied to models with various heterogeneities. Consider a model that includes two sub-populations, female and male populations, with heterogeneous mixing (i.e., no sexual contacts between individuals of the same sex). Assume that the infective periods for female and male populations follow arbitrary discrete (bounded) distributions denoted by Y_f and Y_m , respectively. Here the subscripts f and m stand for female and male, respectively. Let $q_{f,i} = \mathbb{P}(Y_f > i)$ and $q_{m,i} = \mathbb{P}(Y_m > i)$ with $q_{f,0} = q_{m,0} = 1$, and the upper bounds for the two distributions (i.e., the maximum numbers of units of time that an individual takes to recover) be M_w for $w = f, m$.

The model equations are

$$\begin{aligned} S_{w,n+1} &= S_{w,n} e^{-\sum_{i=1}^{M_{\tilde{w}}} \beta_{\tilde{w},i} I_{\tilde{w},n}^{(i)}/N}, \\ E_{w,n+1} &= S_{w,n} [1 - e^{-\sum_{i=1}^{M_{\tilde{w}}} \beta_{\tilde{w},i} I_{\tilde{w},n}^{(i)}/N}] + \alpha_w E_{w,n}, \\ I_{w,n+1}^{(1)} &= (1 - \alpha_w) E_{w,n}, \quad I_{w,n+1}^{(2)} = q_{w,1} I_{w,n}^{(1)}, \\ I_{w,n+1}^{(j)} &= \frac{q_{w,j-1}}{q_{w,j-2}} I_{w,n}^{(j-1)}, \quad 3 \leq j \leq M_w, \quad \text{for } w = f, m. \end{aligned}$$

Here \tilde{w} represents the opposite sex of w , i.e., $\tilde{f} = m$, $\tilde{m} = f$. The constant $\beta_{\tilde{f},i}$ ($\beta_{\tilde{m},i}$) represents the infection rate to a female (male) transmitted by infectious male (female) individuals with stage age i .

Question 3 Show that the reproduction number is given by

$$\mathcal{R} = \varrho(F(I - T)^{-1}) = \sqrt{\left(\sum_{i=1}^{M_f} \beta_{f,i} q_{f,i-1} \right) \left(\sum_{i=1}^{M_m} \beta_{m,i} q_{m,i-1} \right)}. \quad (5.66)$$

The square root in (5.66) is a consequence of the fact that the secondary infections need to be computed from one female (male) to other females (males) through the male (female) population.

Hint: Consider the order of variables:

$$(E_{f,n}, I_{f,n}^{(1)}, I_{f,n}^{(2)}, \dots, I_{f,n}^{(M_f)}, E_{m,n}, I_{m,n}^{(1)}, I_{m,n}^{(2)}, \dots, I_{m,n}^{(M_m)}).$$

First show that

$$F(I - T)^{-1} = \begin{bmatrix} 0 & F_m(I - T_m)^{-1} \\ F_f(I - T_f)^{-1} & 0 \end{bmatrix},$$

where

$$F_w(I - T_w)^{-1} = \begin{bmatrix} \sum_{i=1}^{M_w} \beta_{w,i} q_{w,i-1} \sum_{i=1}^{M_w} \beta_{w,i} q_{w,i-1} \cdots \beta_{w,M_w} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \quad w = f, m.$$

References: [2, 16, 31, 48].

5.7 *Project: Modeling the Synergy Between HIV and HSV-2

Consider the following model for HSV-2, which includes a male population (specified by subscript m) and a female population with two sub-groups representing low-risk and high-risk groups (specified by subscripts f_1 and f_2 , respectively):

$$\begin{aligned}\frac{dS_i}{dt} &= \mu_i N_i - \lambda_i(t) S_i - \mu_i S_i, \\ \frac{dA_i}{dt} &= \lambda_i(t) S_i + \gamma_i(\theta_i) L_i - (\omega_i + \theta_i + \mu_i) A_i, \\ \frac{dL_i}{dt} &= (\omega_i + \theta_i) A_i - (\gamma_i(\theta_i) + \mu_i) L_i, \quad i = m, f_1, f_2,\end{aligned}\tag{5.67}$$

where $\lambda_i(t)$ ($i = m, f_1, f_2$) are the force of infection functions given by

$$\begin{aligned}\lambda_m(t) &= \sum_{i=1}^2 b_m c_i \beta_{f_i m} \frac{A_{f_i}}{N_{f_i}}, \\ \lambda_{f_j}(t) &= b_{f_j} \beta_{m f_j} \frac{A_m}{N_m}, \quad j = 1, 2,\end{aligned}\tag{5.68}$$

and $N_i = S_i + A_i + L_i$, $i = m, f_1, f_2$. Each group i ($i = m, f_1, f_2$) is divided into three subgroups: susceptible (S_i), infected with acute HSV-2 only (A_i), infected with latent HSV-2 only (L_i). The population within each group is assumed to be homogeneous in the sense that individuals have the same infective period, duration of immunity, contact rate, and so on. A transition diagram between these epidemiological classes within group i is depicted in Fig. 5.7.

For each sub-population i ($i = f_1, f_2, m$) there is a per-capita recruitment rate μ_i into the susceptible group. For all classes there is a constant per-capita rate μ_i of exiting the sexually active population. Thus, the total population N_i in group i remains constant for all time. Susceptible people in group i acquire infection with HSV-2 at the rate $\lambda_i(t)$. Upon being infected with HSV-2, people in group i enter the class A_i . These individuals become latent L_i at the constant rate ω_i (an average duration in A_i is $1/\omega_i$). Following an appropriate stimulus in individuals with latent HSV-2, reactivation may occur at the rate γ_i . Finally, the antiviral treatment rate for the A_i individuals is denoted by θ_i . Because antiviral medications will also suppress

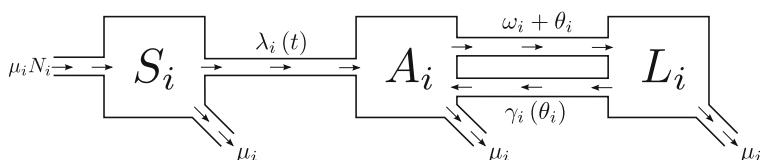


Fig. 5.7 A transition diagram for HSV-2 for subgroup i ($i = m, f_1, f_2$)

reactivation of latent HSV-2, we assume that the reactivation rate of people with latent HSV-2 γ_i is a decreasing function of θ_i , denoted by $\gamma_i(\theta_i)$.

For the forces of infection $\lambda_i(t)$ ($i = m, f_1, f_2$), b_i is the rate at which individuals in group i acquire new sexual partners (also referred to as contact rates), and c_j denotes the probability that a male chooses a female partner in group j ($j = f_1, f_2$). Then $c_1 + c_2 = 1$. For ease of notation, let

$$c_1 = c, \quad c_2 = 1 - c.$$

Overall, the number of female partners in groups j ($j = 1, 2$) that males acquire should be equal to the number of male partners that females in groups j acquire. These observations lead to the following balance conditions:

$$b_m c N_m = b_{f_1} N_{f_1}, \quad b_m (1 - c) N_m = b_{f_2} N_{f_2}. \quad (5.69)$$

To ensure that constraints in (5.69) are satisfied, we assume in numerical simulations that b_m and c are fixed constants with b_{f_1} and b_{f_2} being varied according to N_m , N_{f_1} , and N_{f_2} . The parameters $\beta_{im}(\beta_{mi})$, $i = f_1, f_2$ are the HSV-2 transmission probabilities per partner between females infected with acute HSV-2 in group i and susceptible males (between males infected with acute HSV-2 and susceptible females in group i).

Question 1 Let \mathcal{R}_{mf_jm} denote the average number of secondary male infections generated by one male individual through females in group f_j ($j = 1, 2$). Show that

$$\mathcal{R}_{mf_jm} = \sqrt{\frac{b_{f_j} \beta_{mf_j}}{\omega_m + \theta_m + \mu_m} \cdot P_m \cdot \frac{b_m c_j \beta_{f_jm}}{\omega_{f_j} + \theta_{f_j} + \mu_{f_j}} \cdot P_{f_j}}, \quad j = 1, 2$$

with P_i ($i = m, f_1, f_2$) representing the probability that an individual of group i is in the acute stage (A), which is given by

$$P_i = \frac{(\omega_i + \theta_i + \mu_i)(\gamma_i(\theta_i) + \mu_i)}{[\gamma_i^L(\theta_i) + \omega_i + \theta_i + \mu_i]\mu_i}, \quad i = m, f_1, f_2. \quad (5.70)$$

Question 2 Let \mathcal{R} denote the overall reproduction number for the entire population.

(a) Show that

$$\mathcal{R} = \sqrt{\left(\mathcal{R}_{mf_1m}\right)^2 + \left(\mathcal{R}_{mf_2m}\right)^2}, \quad (5.71)$$

where \mathcal{R}_{mf_jm} ($j = 1, 2$) are given in Question 1.

- (b) Provide a biological interpretation of the expression on the right-hand side of Eq.(5.71).

Question 3 Let E_0 denote the disease-free equilibrium of the system (5.67), and let E^* denote an endemic equilibrium.

- (a) Show that E_0 is locally asymptotically stable when $\mathcal{R} < 1$ and unstable when $\mathcal{R} > 1$.
- (b) Choose the function for $\gamma(\theta)$ to be in the following form: $\gamma_i(\theta_i) = \gamma_i(0)\alpha_i/(\alpha_i + \theta_i)$. Show via numerical simulations that E^* exists and is locally asymptotically stable when $\mathcal{R} > 1$. One case to consider is when $c > 0.5$, e.g., $c = 0.9$ (90% of male contacts are with the low-risk female group). Because of the constraint (5.69), b_i and N_i are not independent. Choose $b_m = 0.1$, $b_{f_1} = 0.0901$, $b_{f_2} = 9.01$ (so that $b_{f_2}/b_{f_1} = 100$), $N_m = N_{f_1} + N_{f_2} = 10^7$ (e.g., $N_{f_1} = 9.9889 \times 10^6$, $N_{f_2} = 1.1099 \times 10^4$). Consider the case when treatment is absent, i.e., $\theta = 0$. Other parameter values are $\omega = 2.5$, $\gamma_m(0) = 0.436$, $\gamma_{f_1} = \gamma_{f_2} = 0.339$, $\alpha_i = 2$. The time unit is month.
- (c) Explore numerically the effect of treatment θ . Consider various scenarios such as treatment in only one subgroup (male, low-risk female or high-risk female group). Summarize the observed outcomes in terms of effect of treatment on the prevalence of HSV-2.

5.8 Project: Effect of Heterogeneities on Reproduction Numbers

Consider the metapopulation model (5.37), which includes vaccination coverage $\phi = (\phi_1, \phi_2, \dots, \phi_n)$ in the n sub-populations. As pointed out in Sect. 5.3.1 that several types of heterogeneities including the activity (a_i), sub-population size (N_i), and preference for mixing within the sub-population (π_i) may affect the optimal vaccination strategy. In this project, we examine in more details how these heterogeneities may affect \mathcal{R}_v for the case of $n = 2$, and how to choose (ϕ_1, ϕ_2) to reduce \mathcal{R}_v below a certain level. For example, Fig. 5.8 illustrates the different parameter regions in the (ϕ_1, ϕ_2) plane in which $\mathcal{R}_v < 1$ in the cases of (a) proportionate mixing ($\pi_1 = \pi_2 = 0$) and (b) preferential mixing ($\pi_i > 0$).

For Questions 1–3 below, let $\sigma = 0.05$, $\gamma = 1/7$.

Question 1 For the cases in (a) and (b) below, determine the values of the basic reproduction number \mathcal{R}_0 for each case. Describe how preferential mixing and heterogeneous activity may influence the effect of heterogeneity in activity on \mathcal{R}_0 . Let $N_1 = N_2 = 500$. Determine the values of \mathcal{R}_0 for each case.

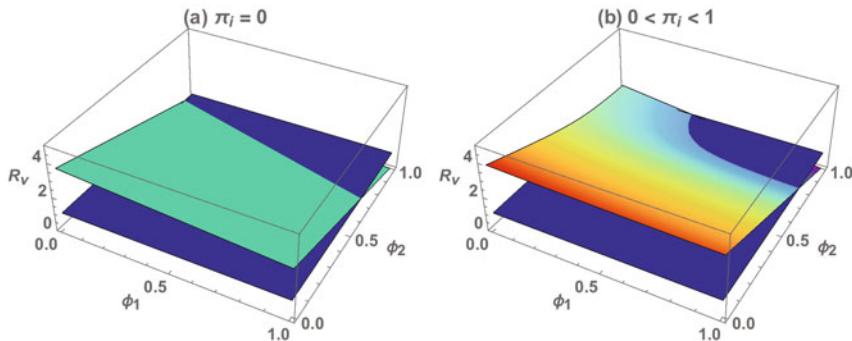


Fig. 5.8 Plots of \mathcal{R}_v as a function of ϕ_1 and ϕ_2 for (a) proportionate mixing ($\pi_i = 0$, no preference) and (b) preferential mixing ($\pi_i > 0$). Values at or below their intersection with the dark plane, $\mathcal{R}_v = 1$, are combinations of ϕ_i ($i = 1, 2$) at which population-immunity attains or exceeds this threshold

- (a) Homogeneous activity: $a_1 = a_2 = 10$ ($a_1 + a_2 = 20$).
 - (i) No preference: $\pi_1 = \pi_2 = 0$.
 - (ii) Homogeneous preference: $\pi_1 = \pi_2 = 0.5$.
 - (iii) Heterogeneous preference: $\pi_1 = 0.25, \pi_2 = 0.75$ and $\pi_1 = 0.75, \pi_2 = 0.25$.
- (b) Heterogeneous activity: $a_1 = 8$ and $a_2 = 12$ ($a_1 + a_2 = 20$).
 - Repeat (i)–(iii) in (a).

Question 2 Same as in *Question 1* but consider heterogeneities in both activity a_i and population size N_i . Given that $N_1 + N_2 = N = 1000$ and $a_1 = 5, a_2 = 10$. Repeat (i)–(iii) in *Question 1*(a) for the following two cases:

- (a) Homogeneous population size: $N_1 = N_2 = 0.5N$.
- (b) Heterogeneous population size: $N_1 = 0.1N$ and $N_2 = 0.9N$.

Question 3 Consider the effective reproduction number $\mathcal{R}_v(\phi_1, \phi_2)$ as a function of vaccination coverage (ϕ_1, ϕ_2) .

- (a) Find the optimal solution $\Phi^* = (\phi_1^*, \phi_2^*)$ for the following set of parameter values: $\pi_1 = \pi_2 = 0.3, a_1 = 15, a_2 = 12, N_1 = 1100, N_2 = 900$. The given number of vaccine doses is $\phi_1 N_1 + \phi_2 N_2 = 990$.
- (b) What is the minimum value $\mathcal{R}_{v\{\min\}} = \mathcal{R}_v(\Phi^*)$?

References

1. Adler, F.R. (1992) The effects of averaging on the basic reproduction ratio, *Math Biosci.* **111**(1): 89–98.
2. Allen,L.J. and P. van den Driessche (2008) The basic reproduction number in some discrete-time epidemic models, *J. Diff. Equ. App.* **14**: 1127–1147.
3. Andersson, H. & Britton, T. (1998) Heterogeneity in epidemic models and its effect on the spread of infection, *J. Appl. Prob.* **35**: 651–661.
4. Arino, J., F. Brauer, P. van den Driessche, J. Watmough & J. Wu (2006) Simple models for containment of a pandemic, *J. Roy. Soc. Interface*, **3**: 453–457.
5. Arino, J., F. Brauer, P. van den Driessche, J. Watmough & J. Wu (2007) A final size relation for epidemic models, *Math. Biosc. & Eng.* **4**: 159–176.
6. Arino, J., F. Brauer, P. van den Driessche, J. Watmough & J. Wu (2008) A model for influenza with vaccination and antiviral treatment, *Theor. Pop. Biol.* **253**: 118–130.
7. Berman, A. & R.J. Plemmons (1994) Nonnegative Matrices in the Mathematical Sciences, SIAM, Vol. **9**, 1994.
8. Blythe, S.P., S. Busenberg & C. Castillo-Chavez (1995) Affinity and paired-event probability, *Math. Biosc.* **128**: 265–84 .
9. Blythe, S.P. , C. Castillo-Chavez, J. Palmer & M. Cheng (1991) Towards a unified theory of mixing and pair formation, *Math. Biosc.* **107**: 379–405.
10. Brauer, F. (2005) The Kermack–McKendrick epidemic model revisited, *Math. Biosc.* **198**: 119–131.
11. Brauer, F. (2008) Epidemic models with treatment and heterogeneous mixing, *Bull. Math. Biol.* **70**: 1869–1885.
12. Brauer, F. & J. Watmough (2009) Age of infection epidemic models with heterogeneous mixing, *J. Biol. Dynamics* **3**: 324–330.
13. Busenberg, S. & C. Castillo-Chavez (1989) Interaction, pair formation and force of infection terms in sexually transmitted diseases, In Mathematical and Statistical Approaches to AIDS Epidemiology, Lect. Notes Biomath. **83**, C. Castillo-Chavez (ed.), Springer-Verlag, Berlin-Heidelberg-New York, 289–300.
14. Busenberg, S. & C. Castillo-Chavez (1991) A general solution of the problem of mixing of sub-populations and its application to risk- and age-structured epidemic models for the spread of AIDS, *IMA J. Math. Appl. Med. Biol.*, **8**: 1–29.
15. Chow, L., M. Fan, and Z. Feng (2011) Dynamics of a multi-group epidemiological model with group-targeted vaccination strategies, *J. Theor. Biol.* **29**: 56–64.
16. De Jong, M., O. Diekmann and J.A.P. Heesterbeek (1994) The computation of R_0 for discrete-time epidemic models with dynamic heterogeneity, *Math. Biosci.* **119**(1): 97–114.
17. Del Valle, S. Y., J.M. Hyman, H.W. Hethcote & S.G. Eubank (2007) Mixing patterns between age groups in social networks, *Social Networks*, **29**(4): 539–554.
18. Diekmann, O. & J.A.P. Heesterbeek (2000) Mathematical Epidemiology of Infectious Diseases. Wiley, Chichester (2000)
19. Diekmann, O., J.A.P. Heesterbeek & J.A.J. Metz (1990) On the definition and the computation of the basic reproductive ratio \mathcal{R}_0 in models for infectious diseases in heterogeneous populations, *J. Math. Biol.* **28**: 365–382.
20. Diekmann, O., J.A.P. Heesterbeek, & T. Britton (2012) Mathematical tools for understanding infectious disease dynamics, 2012, Princeton University Press.
21. Feng, Z., A.N. Hill, P.J. Smith, J.W. Glasser (2015) An elaboration of theory about preventing outbreaks in homogeneous populations to include heterogeneity or preferential mixing, *J. Theor. Biol.* **386**: 177–187.
22. Feng, Z., A.N. Hill, A.T. Curns, J.W. Glasser (2007) Evaluating targeted interventions via meta-population models with multi-level mixing, *Math. Biosci.* **287**: 93–104.
23. Ferguson, N.M., D.A.T. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, S. Iam-sirithaworn, & D.S. Burke (2005) Strategies for containing an emerging influenza pandemic in Southeast Asia, *Nature*, **437**: 209–214.

24. Gani, R., H. Hughes, T. Griffin, J. Medlock, & S. Leach (2005) Potential impact of antiviral use on hospitalizations during influenza pandemic, *Emerg. Infect. Dis.* **11**: 1355–1362.
25. Glasser, J., Z. Feng, A. Moylan, S. Del Valle & C. Castillo-Chavez (2012) Mixing in age-structured population models of infectious diseases, *Math. Biosci.*, **235**(1): 1–7.
26. Glasser, J.W., Z. Feng, S.B. Omer, P.J. Smith, L.E. Rodewald (2016) The effect of heterogeneity in uptake of the measles, mumps, and rubella vaccine on the potential for outbreaks of measles: a modelling study, *J. Math. Biol.* **16**(5): 599–605.
27. Hethcote, H.W. & J.A. Yorke (1984) *Gonorrhea Transmission Dynamics and Control*, Lect. Notes in Biomath. **56**, Springer-Verlag, Berlin-Heidelberg-New York (1984).
28. Hethcote, H.W. & J.W. van Ark (1987) Epidemiological models for heterogeneous populations: proportionate mixing, parameter estimation and immunization programs. *Math. Biosci.*, **84**(1): 85–118.
29. Hirsch, M.W. & S. Smale, Differential Equations (1974) Dynamical Systems, and Linear Algebra, Academic Press, Orlando, FL (1974).
30. Jacquez, J.A., C.P. Simon, J. Koopman, L. Sattenspiel, & T. Perry (1988) Modeling and analyzing HIV transmission: the effect of contact patterns, *Math. Biosci.*, **92**: 119–199.
31. Lewis, M.A., J. Renclawowicz, P. van Den Driessche, and M. Wonham (2006) A comparison of continuous and discrete-time West Nile Virus models, *Bull. Math. Biol.* **68**(3): 491–509.
32. Longini, I.M., M.E. Halloran, A. Nizam, & Y. Yang (2004) Containing pandemic influenza with antiviral agents, *Am. J. Epidemiol.* **159**: 623–633.
33. Longini, I.M., A. Nizam, S. Xu, K. Ungchusak, W. Hanshaoworakul, D.A.T. Cummings, & M.E. Halloran (2005) Containing pandemic influenza at the source, *Science* **309**, 1083–1087.
34. Longini, I. M. & M. E. Halloran (2005) Strategy for distribution of influenza vaccine to high - risk groups and children, *Am. J. Epidemiol.* **161**: 303–306.
35. Ma, J. & D.J.D. Earn (2006) Generality of the final size formula for an epidemic of a newly invading infectious disease, *Bull. Math. Biol.* **68**: 679–702.
36. May, R.M. & R.M. Anderson (1984) Spatial heterogeneity and the design of immunization programs. *Math Biosci.* **72**(1): 83–111.
37. May, R.M. & R.M. Anderson (1984) Spatial, temporal and genetic heterogeneity in host populations and the design of immunization programmes. *IMA J Math Appl Math Biol.* **1**(3): 233–66.
38. Meyers, L.A. (2007) Contact network epidemiology: Bond percolation applied to infectious disease prediction and control, *Bull. Am. Math. Soc.* **44**, 63–86.
39. Meyers, L.A. , M.E.J. Newman & B. Pourbohloul (2006) Predicting epidemics on directed contact networks, *J. Theor. Biol.* **240**, 400–418.
40. Meyers, L.A., B. Pourbohloul, M.E.J. Newman, D.M. Skowronski, & R.C. Brunham (2005) Network theory and SARS: predicting outbreak diversity. *J. Theor. Biol.*, **232**: 71–81.
41. Mossong, J., N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, ... & W. J. Edmunds (2008) Social contacts and mixing patterns relevant to the spread of infectious diseases, *PLoS Medicine*, **5**(3): e74.
42. Nold, A. (1980) Heterogeneity in disease transmission modeling, *Math. Biosci.* **52**: 227–240.
43. Poghotanyan, G., Z. Feng, J.W. Glasser, A.N. Hill (2018) Constrained minimization problems for the reproduction number in meta-population models, *J. Math. Biol.* <https://doi.org/10.1007/s00285-018-1216-z>.
44. Pourbohloul, B. & J. Miller (2008) Network theory and the spread of communicable diseases, Center for Disease Modeling Preprint 2008-03, www.cdm.yorku.ca/cdmprint03.pdf
45. van den Driessche, P. & J. Watmough (2002) Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math. Biosc.* **180**: 29–48.
46. van den Driessche, P. & J. Watmough (2002) Further notes on the basic reproduction number, in Mathematical Epidemiology, F. Brauer, P. van den Driessche, & J. Wu (eds.) Springer Lecture Notes, Vol. **1945**.

47. Wallinga, J., P. Teunis, & M. Kretzschmar (2006) Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents, *American Journal of Epi.*, **164**(10): 936–944.
48. Wesley, C. L., L.J. Allen, C.B. Jonsson, Y.-K. Chu, R.D. Owen (2009) A discrete-time rodent-hantavirus model structured by infection and developmental stages, *Adv. Stu. Pure Math.* **53**: 387–398.
49. Zagheni, E., F.C. Billari, P. Manfredi, A. Melegaro, J. Mossong & W.J. Edmunds (2008) Using time-use data to parameterize models for the spread of close-contact infectious diseases, *Am. J. Epi.*, **168**(9), 1082–1090.

Chapter 6

Models for Diseases Transmitted by Vectors



6.1 Introduction

Many diseases are transmitted from human to human indirectly, through a *vector*. Vectors are living organisms that can transmit infectious diseases between humans. Many vectors are bloodsucking insects that ingest disease-producing microorganisms during blood meals from an infected (human) host, and then inject it into a new host during a subsequent blood meal. The best known vectors are mosquitoes for diseases including malaria, dengue fever, chikungunya, Zika virus, Rift Valley fever, yellow fever, Japanese encephalitis, lymphatic filariasis, and West Nile fever, but ticks (for Lyme disease and tularemia), bugs (for Chagas' disease), flies (for onchocerciasis), sandflies (for leishmaniasis), fleas (for plague, transmitted by fleas from rats to humans), and some freshwater snails (for schistosomiasis) are vectors for some diseases.

Every year there are more than a billion cases of vector-borne diseases and more than a million deaths. Vector-borne diseases account for over 17% of all infectious diseases worldwide. Malaria is the most deadly vector-borne diseases and caused an estimated 627,000 deaths in 2012. The most rapidly growing vector-borne disease is dengue, for which the number of cases has multiplied by 30 in the last 50 years. These diseases are found more commonly in tropical and sub-tropical regions where mosquitoes flourish and in places where access to safe drinking water and sanitation systems is uncertain.

Some vector-borne diseases, such as dengue, chikungunya, and West Nile virus, are emerging in countries where they were unknown previously because of globalization of travel and trade and environmental challenges such as climate change.

Many of the important underlying ideas of mathematical epidemiology arose in the study of malaria begun by Sir. R.A. Ross [16]. Malaria is one example of a disease with vector transmission, the infection being transmitted back and forth between vectors (mosquitoes) and hosts (humans). It kills hundreds of thousands of

people annually, mostly children and mostly in poor countries in Africa. Among communicable diseases, only tuberculosis causes more deaths. We will analyze some models for malaria in a later chapter. Other vector diseases include West Nile virus and HIV with heterosexual transmission, yellow fever, and dengue fever (which will also be studied in a later chapter).

Vector-transmitted diseases require models that include both vectors and hosts. For most diseases transmitted by vectors, the vectors are insects, with a much shorter life span than the hosts, who may be humans as for malaria or animals as for West Nile virus. For heterosexually transmitted human diseases, transmission goes back and forth between males and females rather than between two different species, but a model still requires two separate groups.

The compartmental structure of the disease may be different in host and vector species; for many diseases with insects as vectors an infected vector remains infected for life so that the disease may have an *SI* or *SEI* structure in the vectors and an *SIR* or *SEIR* structure in the hosts. We will describe vector models with *SEIR* structure in the host species and *SEI* structure in the vector species, but the analysis of other types of vector-transmitted diseases is similar. The models are of the same type as those studied in Chaps. 2–4, but have a more complicated structure because they involve two different species.

6.2 A Basic Vector Transmission Models

Ross received the second Nobel Prize in Medicine for demonstrating the vector transmission nature of malaria, and he then constructed a model in 1909 that predicted the possibility of controlling malaria by decreasing the mosquito population size below a threshold value. This prediction was borne out in practice, but controlling the mosquito population size is very difficult because of the ability of mosquitoes to adapt to pesticides. Malaria remains a very dangerous disease.

We describe a basic model for a vector-transmitted disease, in which we assume throughout that vectors satisfy a simple *SEI* model, with no recovery from infection, and the hosts satisfy a simple *SEIR* model. This model is a template for vector disease transmission models for many specific diseases. We are thinking of mosquitoes as vectors, and because a mosquito lifetime is much shorter than that of the human hosts we must include demographics in the vector population. The model includes both epidemic and endemic situations.

We consider a constant total population size N_h of hosts (humans), divided into S_h susceptibles, E_h exposed members, I_h infectives, and R_h removed members. There is a birth rate Λ_h in the susceptible class and a proportional natural death rate μ_h in each class. Exposed hosts proceed to the infective class at rate η_h and infected hosts recover at rate γ .

There is a constant birth rate μN_v of vectors in unit time and a proportional vector death rate μ in each class, so that the total vector population size N_v is constant. The vector population is divided into S_v susceptibles, E_v exposed members, and I_v

infectives. Exposed vectors move to the infected class at rate η_v and do not recover from infection. For simplicity, we assume that there are no disease deaths of either hosts or vectors.

We assume that the mosquito biting rate is proportional to the size N_h of the host population. Thus an average mosquito makes bN_h bites in unit time. The total number of mosquito bites in unit time is $bN_h N_v$ and the number of bites received by an average host in unit time is bN_v . We assume that f_{vh} is the probability that a bite transmits infection from vector to host and f_{hv} is the probability that a bite transmits infection from host to vector. We define

$$\beta_h = b f_{vh} N_v, \quad \beta_v = b f_{hv} N_h.$$

If we eliminate b from these two equations, we obtain a balance relation

$$f_{vh} \beta_h N_v = f_{hv} \beta_v N_h. \quad (6.1)$$

In many models for vector-transmitted diseases, the vector population size is much larger than the host population size but the two contact rates β_h and β_v are of the same order of magnitude, suggesting that $f_{hv} >> f_{vh}$.

Then the number of new infective hosts in unit time is

$$b f_{vh} N_v S_h \frac{I_v}{N_v} = \beta_h S_h \frac{I_v}{N_v}.$$

A similar argument shows that the number of new mosquito infections is

$$\beta_v S_v \frac{I_h}{N_h}.$$

The model is

$$\begin{aligned} S'_h &= \Lambda(N_h) - \beta_h S_h \frac{I_v}{N_v} - \mu_h S_h \\ E'_h &= \beta_h S_h \frac{I_v}{N_v} - (\eta_h + \mu_h) E_h \\ I'_h &= \eta_h E_h I_v - (\gamma + \mu_h) I_h \\ S'_v &= \mu_v N_v - \beta_v S_v \frac{I_h}{N_h} - \mu_v S_v \\ E'_v &= \beta_v S_v \frac{I_h}{N_h} - (\eta_v + \mu_v) E_v \\ I'_v &= \eta_v E_v - \mu_v I_v. \end{aligned} \quad (6.2)$$

The corresponding epidemic model is (6.2) with $\Lambda(N_h) = \mu_h = 0$.

6.2.1 The Basic Reproduction Number

The basic reproduction number is defined as the number of secondary disease cases caused by introducing a single infective host into a wholly susceptible population of both hosts (humans) and vectors (mosquitoes). For the model (6.2) this may be calculated directly. There are two stages. First, the infective human infects mosquitoes, at a rate β_v for a time $1/(\gamma + \mu_h)$. This produces $\beta_v/(\gamma + \mu_h)$ infected mosquitoes, of whom a fraction $\eta_v/(\eta_v + \mu_v)$ proceeds to become infective.

The second stage is that these infective mosquitoes infect humans at a rate β_h for a time $1/\mu_v$, producing β_h/μ_v infected humans per mosquito. The net result of these two stages is

$$\frac{\beta_h}{\gamma + \mu_h} \frac{\eta_v}{\eta_v + \mu_v} \frac{\beta_v}{\mu_v} = \beta_v \beta_h \frac{\eta_v}{(\eta_v + \mu_v)(\gamma + \mu_h)\mu_v}$$

infected humans, and this is the basic reproduction number \mathcal{R}_0 .

We could also calculate the basic reproduction number by using the next generation matrix approach [21]. This would give the next generation matrix

$$K \begin{bmatrix} 0 & \beta_v \frac{\eta_v}{\mu_v(\mu_v + \eta_v)} \\ \beta_h \frac{1}{\gamma + \mu_h} & 0 \end{bmatrix}.$$

The basic reproduction number is the positive eigenvalue of this matrix,

$$\mathcal{R}_0 = \sqrt{\beta_h \beta_v \frac{\eta_v}{\mu_v(\gamma + \mu_h)(\mu_v + \eta_v)}}.$$

In this calculation, the transition from host to vector to host is considered as two generations. In studying vector-transmitted diseases it is common to consider this as one generation and use the value that we obtained by our direct approach

$$\mathcal{R}_0 = \beta_h \beta_v \frac{\eta_v}{\mu_v(\gamma + \mu_h)(\mu_v + \eta_v)}. \quad (6.3)$$

This choice is made in [4, 10] and is the choice that we make because it conforms to the result obtained by direct calculation, without using the next generation approach. However, other references, including [13], use the square root form, and it is important to be aware of which form is being used in any study. The two choices have the same threshold value.

In fact, different expressions are possible for the next generation matrix. This is shown in [6]. Using the next generation matrix approach but considering only host infections as new infections and vector infections as transitions, we would obtain the form (6.3).

6.2.2 The Initial Exponential Growth Rate

In order to determine the initial exponential growth rate from the model, a quantity that can be compared with experimental data, we linearize the model (6.2) about the disease-free equilibrium $S_h = N_h, E_h = I_h = 0, S_v = N_v, E_v = I_v = 0$. If we let $y = N_h - S_h, z = N_v - S_v$, we obtain the linearization

$$\begin{aligned} y' &= \beta_v I_v - \mu_h y \\ E'_h &= \beta_v I_v - (\eta + \mu_h) E_h \\ I'_h &= \eta E_h - (\gamma + \mu_h) I_h \\ z' &= -\mu_v z + \beta_h I_h \\ E_v &= \beta_h I_h - (\mu_v + \eta_v) E_v \\ I'_v &= \eta_v E_v - \mu_v I_v. \end{aligned} \tag{6.4}$$

The corresponding characteristic equation is

$$\det \begin{bmatrix} -\lambda & 0 & 0 & 0 & 0 & \beta_v \\ 0 & -(\lambda + \eta_h + \mu_h) & 0 & 0 & 0 & \beta_v \\ 0 & \eta_h & -(\lambda + \gamma + \mu_h) & 0 & 0 & 0 \\ 0 & 0 & -\beta_h & -(\lambda + \mu_v) & 0 & 0 \\ 0 & 0 & -\beta_h & 0 & -(\lambda + \mu_v + \eta_v) & 0 \\ 0 & 0 & 0 & 0 & \eta_v & -(\lambda + \mu_v) \end{bmatrix} = 0.$$

Solutions of the linearization (6.4) are linear combinations of exponential functions whose exponents are the roots of the characteristic equation (the eigenvalues of the coefficient matrix of (6.4)).

We can reduce this equation to a product of two factors and a fourth degree polynomial equation

$$\lambda(\lambda + \mu_v) \det \begin{bmatrix} -(\lambda + \eta_h + \mu_h) & 0 & 0 & \beta_v \\ \eta_h & -(\lambda + \gamma + \mu_h) & 0 & 0 \\ 0 & \beta_h & -(\lambda + \mu_v + \eta_v) & 0 \\ 0 & 0 & \eta_v & -(\lambda + \mu_v) \end{bmatrix} = 0.$$

The initial exponential growth rate is the largest root of this fourth degree equation, which reduces to

$$g(\lambda) = (\lambda + \eta_h)(\lambda + \gamma + \mu_h)(\lambda + \mu_v + \eta_v)(\lambda + \mu_v) - \beta_h \beta_v \eta_h \eta_v = 0. \tag{6.5}$$

Since $g(0) < 0$ if $\mathcal{R}_0 > 1$, and since $g(\lambda)$ is positive for large positive λ and $g'(\lambda) > 0$ for positive λ , there is a unique positive root of the equation $g(\lambda) = 0$, and this is the initial exponential growth rate. The initial exponential growth rate may be measured experimentally, and if the measured value is ρ , then from (6.5) we

obtain

$$(\rho + \eta_h)(\rho + \gamma)(\rho + \mu_v + \eta_v)(\rho + \mu_v) = \beta_h \beta_v \eta_h \eta_v = \mathcal{R}_0 \eta_h \mu_v \gamma (\mu_v + \eta_v).$$

From this, we obtain

$$\mathcal{R}_0 = \frac{(\rho + \eta_h)(\rho + \gamma)(\rho + \mu_v + \eta_v)(\rho + \mu_v)}{\eta_h \mu_v \gamma (\mu_v + \eta_v)}. \quad (6.6)$$

This gives a way to estimate the basic reproduction number from measurable quantities. In addition, the balance relation (6.1) allows us to calculate the values of β_h and β_v separately, which makes it possible to simulate the model.

The same reasoning may be used for a short term epidemic model in which we ignore demographics in the host population. The results would be the same except that $\Lambda(N_h)$ and μ_h would be replaced by zero.

By linearizing the system (6.2) at an equilibrium it is possible to show that the disease-free equilibrium is asymptotically stable if and only if $\mathcal{R}_0 < 1$ and that the endemic equilibrium exists only if $\mathcal{R}_0 > 1$ and is asymptotically stable.

6.3 Fast and Slow Dynamics

In vector-borne disease transmission models in which the vector is an insect, the vector time scale is often much faster than the host time scale. In such models it is possible to consider two separate time scales [3]. We can make a *quasi-steady-state* hypothesis assuming that the vector population sizes remain almost constant [17]. We treat the vector population sizes as constants which depend on the host population sizes and approximate the system by a host model, which has fewer equations than the full system but may be more complicated in form.

We describe the process in an *SIR/SI* epidemic model which is somewhat simpler than (6.2). Consider a host population with total population size N_h , and no demographics, and a vector population with birth and death rates μ_v and constant total population size N_v . We assume mass action contact rate with contact rates β_h , β_v , recovery rate γ for the host species and no recovery for the vector species.

The resulting model is

$$\begin{aligned} \frac{dS_h}{dt} &= -\beta_h S_h \frac{I_v}{N_v} \\ \frac{dI_h}{dt} &= \beta_h S_h \frac{I_v}{N_v} - \gamma I_h \end{aligned}$$

$$\begin{aligned}\frac{dS_v}{dt} &= \mu_v N_v - \beta_v S_v \frac{I_h}{N_h} - \mu_v S_v \\ \frac{dI_v}{dt} &= \beta_v S_v \frac{I_h}{N_h} - \mu_v I_v.\end{aligned}$$

Because $S_v + I_v$ is a constant N_v , we can reduce this model to a three-dimensional problem,

$$\begin{aligned}\frac{dS_h}{dt} &= -\beta_h S_h \frac{I_v}{N_v} \\ \frac{dI_h}{dt} &= \beta_h S_h \frac{I_v}{N_v} - \gamma I_h \\ \frac{dI_v}{dt} &= \beta_v (N_v - I_v) \frac{I_h}{N_h} - \mu_v I_v.\end{aligned}\tag{6.7}$$

In order to obtain the basic reproduction number, we observe that a single infective host infects

$$\frac{\beta_h}{\gamma}$$

vectors in a wholly susceptible population of vectors, and each of these infects

$$\frac{\beta_v}{\mu_v}$$

hosts in a wholly susceptible host population. Thus the number of secondary infections caused by an infective host (and also the number of secondary vector infections caused by an infective vector) is given by

$$\mathcal{R}_0 = \frac{\beta_h \beta_v}{\gamma \mu_v}.$$

However, one could also argue that the transition from host to vector to host should be viewed as two generations, and it may be more reasonable to say that

$$\mathcal{R}_0 = \sqrt{\frac{\beta_h \beta_v}{\gamma \mu_v}}.$$

With either choice, the transition is at $\mathcal{R}_0 = 1$.

It is possible to prove that there is an epidemic if and only if $\mathcal{R}_0 > 1$.

A model for dengue fever of the form (6.7) has been studied in [15]. In terms of the above model, with time measured in days, this model has parameter values with μ_h much smaller than μ_v and β_h much smaller than β_v , expressing the fact that the vector dynamics are much faster than the host dynamics, and this suggests a model

$$\begin{aligned}\frac{dS_h}{dt} &= -\beta_h S_h \frac{I_v}{N_v} \\ \frac{dI_h}{dt} &= \beta_h S_h \frac{I_v}{N_v} - \alpha I_h \\ \epsilon \frac{dI_v}{dt} &= \beta_v (N_v - I_v) \frac{I_h}{N_h} - \mu_v I_v,\end{aligned}\tag{6.8}$$

with ϵ a small positive constant.

Other models describing two time scales may be found in [18, 19]. In fact, such a form is valid for all vector disease transmission models in which the vector population size is much greater than the host population size. We assume that $N_h \ll N_v$, and we let

$$\epsilon = \frac{N_h}{N_v} \ll 1.\tag{6.9}$$

Then, if we replace N_h by ϵN_v (6.7) becomes

$$\begin{aligned}\frac{dS_h}{dt} &= -\beta_h S_h I_v \\ \frac{dI_h}{dt} &= \beta_h S_h I_v - \alpha I_h \\ \epsilon \frac{dI_v}{dt} &= \beta_v (N_v - I_v) \frac{I_h}{N_v} - \frac{N_h}{N_v} \mu_v I_v.\end{aligned}\tag{6.10}$$

For differential equations or systems of differential equations which depend on a parameter there is a general theorem to the effect that solutions are continuous functions of the parameter on any finite interval. However, if a derivative is multiplied by a parameter which may be allowed to tend to zero, this is not necessarily true. Such a situation is called a *singular perturbation*. Many problems in the biological sciences involve actions on very different time scales, and these may lead to a rapid change in some of the variables on a very short initial time interval, while other variables act more slowly.

6.3.1 Singular Perturbations

Singular perturbation problems arise in models (systems of differential equations) containing a small parameter ϵ , of the form

$$\begin{aligned}\epsilon \frac{dy}{d\tau} &= f(y, z, \epsilon), \quad y(0) = y_0 \\ \frac{dz}{d\tau} &= g(y, z, \epsilon), \quad z(0) = z_0\end{aligned}\tag{6.11}$$

with solution $(y(\tau, \epsilon), z(\tau, \epsilon))$. There is a corresponding reduced system obtained by setting $\epsilon = 0$,

$$\begin{aligned}f(y, z, 0) &= 0 \\ \frac{dz}{d\tau} &= g(y, z, 0), \quad z(0) = z_0\end{aligned}\tag{6.12}$$

with solution $(y_0(\tau), z_0(\tau))$.

Since ϵ is assumed to be small, the form (6.11) suggests that the y reaction time is much faster than the z reaction time. Thus y goes to its equilibrium value rapidly, and at its equilibrium $f(y, z, 0) = 0$. Then we might expect that the reduced problem (6.12) is a good approximation to the full problem (6.11) *after* a short initial time interval near $\tau = 0$ during which y moves to its equilibrium value. In applications one often makes a quasi-steady-state hypothesis that y remains almost constant, so that $\frac{dy}{d\tau} \approx 0$. This hypothesis is expressed as $f(y, z, 0) = 0$; in singular perturbation language the hypothesis is just that the full problem is approximated by the reduced problem.

Because (6.12) is a first-order differential equation (which requires one initial condition to identify a unique solution) and (6.11) is a two-dimensional system (requiring two initial conditions for a unique solution) we must expect to lose an initial condition in the reduction, and this suggests that the solutions of (6.12) and (6.11) (each derived on a different time scale) may not agree close to $\tau = 0$. Because $y(\tau, \epsilon)$ (the solution to the full problem) and $y_0(\tau)$ (the solution to the reduced problem) do not match at $\tau = 0$, we should expect that $y(\tau, \epsilon)$ changes rapidly for t close to 0. It is possible to analyze this by making a change of independent variable to change from the slow time scale to the fast time scale. However, since our interest is mainly in the long-term behavior of the system we do not explore this further here.

If the partial derivative $f_y(y, z, 0) \neq 0$ we may solve the equation $f(y, z, 0) = 0$ for y as a function of z , $y = \phi(z)$. Thus the reduced system (6.12) is equivalent to the first-order initial value problem

$$\frac{dz}{d\tau} = g(\phi(z), z, 0), \quad z(0) = z_0.\tag{6.13}$$

Then we have the solution of (6.12) with $z_0(\tau)$ the solution of (6.13) and $y_0(\tau) = \phi(z_0(\tau))$, and $y_0(0) = \phi(z_0)$. If $y_0 \neq \phi(z_0)$ it is not possible for the solution of the reduced problem (6.12) to satisfy the two initial conditions of the full problem (6.11). The solution $(y_0(\tau), z_0(\tau))$ of the reduced problem is called the *outer solution*. In order to use the solution of the reduced problem (6.12) as an approximation to the solution of the full problem (6.11), we would need a result to the effect that for each t away from $t = 0$ the solution of the reduced problem (6.12) is the limit as $\epsilon \rightarrow 0$ of the solution of the full problem (6.11).

A change of time scale $\tau = \epsilon t$ transforms the system (6.11) to the system

$$\begin{aligned}\frac{dy}{dt} &= f(y, z, \epsilon), \quad y(0) = y_0 \\ \frac{dz}{dt} &= \epsilon g(y, z, \epsilon), \quad z(0) = z_0.\end{aligned}\tag{6.14}$$

The second equation of (6.14) says that the second variable z remains almost constant on a large t -interval corresponding to a small τ -interval. This suggests considering the initial value problem

$$\frac{dy}{dt} = f(y, z_0, 0), \quad y(0) = y_0,\tag{6.15}$$

called the boundary layer system, as an approximation valid on a small t -interval called the boundary layer.

The mathematical treatment of singular perturbations began in the 1940s from the perspective of asymptotic expansions. A few years later the qualitative result which justifies the use of the reduced system as an approximation to the full system was obtained independently in the USA and the Soviet Union [12, 20].

Theorem 6.1 (Levinson–Tihonov) *Suppose that*

1. f, g are smooth functions,
2. the equation $f(y, z, 0) = 0$ can be solved for y as a smooth function of z , $y = \phi(z)$,
3. the reduced system (6.12) has a solution on an interval $0 \leq \tau \leq T$,
4. the boundary layer system (6.15) has an asymptotically stable equilibrium.

Then $y(\tau, \epsilon) \rightarrow y_0(\tau)$, $z(t, \epsilon) \rightarrow z_0(\tau)$ as $\epsilon \rightarrow 0+$ for $0 < \tau \leq T$. The convergence of y is non-uniform at $\tau = 0$.

There is an extension of this result to infinite time intervals [9]

Theorem 6.2 *Suppose, in addition to the hypotheses of the Levinson–Tihonov theorem that the reduced system (6.12) has a solution which is asymptotically stable and that the boundary layer system (6.15) has a solution which is asymptotically stable uniformly in z_0 . Then the convergence is uniform on closed subsets of $0 < t < \infty$.*

The essential content of these results is that if ϵ is sufficiently small the solution of the reduced system is a good approximation to the solution of the singularly perturbed system except very close to $t = 0$. The relation $f(y, z, 0) = 0$ is called the *quasi-steady-state*. Close to $\tau = 0$ the solution of the boundary layer system (6.15) describes the behavior of solutions. Thus the analysis of a singular perturbation problem can be decomposed into the analysis of two simpler problems, namely the boundary layer system and the reduced problem. Curiously, in fluid dynamic applications the focus of attention has been on the boundary layer system, whereas in most biological applications the primary interest has been in the long-term behavior, that is, the reduced problem.

The underlying idea in a singular perturbation problem is that there are two different time scales inherent in the problem, and this makes it possible to analyze the problem separately on each time scale. The reduction in dimension because of this separation simplifies the analysis. The singular perturbation idea will be applied to the malaria model of Sect. 11.3 which includes epidemiological and genetic processes that occur on very different time scales.

6.4 A Vector-Borne Epidemic Model

We begin with the epidemic model (6.10), which is in singular perturbation form. The quasi-steady-state, which is the case $\epsilon = 0$ of the equation for I_v in (6.10), is given by the equation

$$\beta_v(N_v - I_v) \frac{I_h}{N_v} = \frac{N_h}{N_v} \mu_v I_v.$$

This expresses I_v as a function of the other variables as

$$I_v = \frac{\beta_v N_v I_h}{\beta_v I_h + \mu_v N_h}. \quad (6.16)$$

We substitute this form for I_v into the equations for S_h and I_h to give the reduced equation

$$\begin{aligned} S'_h &= -\beta_h S_h \frac{\beta_v I_h}{\beta_v I_h + \mu_v N_h} \\ I'_h &= \beta_h S_h \frac{\beta_v I_h}{\beta_v I_h + \mu_v N_h} - \alpha I_h. \end{aligned} \quad (6.17)$$

This approach gives a way to formulate models for vector-transmitted diseases which consist of a system involving only the host variables. While this system is two-dimensional and the original system is three-dimensional, it does have a more complicated form.

6.4.1 A Final Size Relation

Addition of the first two equations of (6.17) (or (6.10)) gives

$$(S_h + I_h)' = -\alpha_h I_h.$$

Thus $S_h + I_h$ is a decreasing non-negative function and tends to a limit as $t \rightarrow \infty$. Also, its derivative tends to zero, so that $I_h(t) \rightarrow 0$ as $t \rightarrow \infty$. Integration of this equation with respect to t from 0 to ∞ gives

$$N_h - S_h(\infty) = \alpha \int_0^\infty I_h(t) dt. \quad (6.18)$$

Division of the first equation of (6.17) by S_h and integration gives

$$\log \frac{S_h(0)}{S_h(\infty)} = \beta_h \beta_v \int_0^\infty \frac{I_h(t)}{\mu_v N_h + \beta_v I_h(t)} dt.$$

Since

$$\mu_v N_h \leq \mu_v N_h + \beta_v I_h \leq (\mu_v + \beta_v) N_h,$$

we obtain

$$\begin{aligned} \log \frac{S_h(0)}{S_h(\infty)} &\leq \frac{\beta_h \beta_v}{\mu_v N_h} \int_0^\infty I_h(t) dt \\ &= \frac{\beta_h \beta_v}{\mu_v \alpha} [N_h - S_h(\infty)] \\ &= \mathcal{R}_0 \left[1 - \frac{S_h(\infty)}{N_h} \right], \end{aligned} \quad (6.19)$$

and

$$\begin{aligned} \log \frac{S_h(0)}{S_h(\infty)} &\geq \frac{\beta_h \beta_v}{(\mu_v + \beta_v) N_h} \int_0^\infty I_h(t) dt \\ &= \frac{\mu_v}{\mu_v + \beta_v} \mathcal{R}_0 \left[1 - \frac{S_h(\infty)}{N_h} \right]. \end{aligned} \quad (6.20)$$

Combining (6.19) and (6.20) we obtain the two-sided estimate final size relation

$$\frac{\mu_v}{\mu_v + \beta_v} \mathcal{R}_0 \left[1 - \frac{S_h(\infty)}{N_h} \right] \leq \log \frac{S_h(0)}{S_h(\infty)} \leq \mathcal{R}_0 \left[1 - \frac{S_h(\infty)}{N_h} \right]. \quad (6.21)$$

This final size estimate gives upper and lower bounds for $S_h(\infty)$. Further information about this topic may be found in [1, 2].

6.5 *Project: An SEIR/SEI Model

Consider the *SEIR/SEI* epidemic model

$$\begin{aligned} S'_h &= -\beta_v S_h I_v - \mu_h S_h \\ E'_h &= \beta_v S_h I_v - (\eta_h + \mu_h) E_h \\ I'_h &= \eta_h E_h - (\alpha + \mu_h) I_h \\ S'_v &= \mu_v N_v - \beta_h S_v I_h - \mu_v S_v \\ E'_v &= \beta_h S_v I_h - (\eta_v + \mu_v) E_v \\ I'_v &= \eta_v E_v - \mu_v I_v. \end{aligned} \tag{6.22}$$

Assume that the vector dynamics are much faster than the host dynamics.

Question 1 Determine the basic reproduction number of the system (6.22).

Question 2 Consider the model (6.22) as a two time scale system and find the quasi-steady-state.

Question 3 Obtain a final size relation for the model (6.22).

6.6 *Project: Models for Onchocerciasis

Onchocerciasis, also known as “River Blindness” is a vector-borne disease that affects the skin and eyes of humans. It is endemic in parts of Africa, Yemen, and Central America, and is especially prevalent in sub-Saharan Africa. It is transmitted by *Onchocerca volvulus*, a parasitic worm whose life cycle includes five larval stages, including a stage that requires a human host and another stage that requires a black fly host.

The peak biting time for black flies is during the daylight hours, and the black flies remain near their breeding sites on well-oxygenated water. Thus communities on the river’s edge are most at risk. The vector stage is very complicated, and it would be difficult to include all of its stages in a model. We will assume a constant total vector population size with an *SI* model for vectors.

The standard medication for treatment of onchocerciasis is ivermectin [8]; oral administration kills larvae rapidly but does not kill the adult worms. However, it does reduce their reproductive rate for several months [5, 14]. Ivermectin treatment

is available to a fraction p of the population, limited by restrictions on who can receive the medication, limited health care, and willingness to participate.

For humans we assume a model of *SEIR* type but with infectives divided into those who do not participate in treatment (I), participants who are not yet being treated (P), and infectives who have received ivermectin treatment (M). We let $H = P + M$, the population of infected hosts who are currently or will eventually be treated.

Because the total human population size $N = S + E + I + H$ and the total vector population size $F = U + V$ are constants, we do not need differential equations for S and U . The effective human population size is $W = I + P + (1 - \nu)M = I + H - \nu M$, where ν is the relative decrease in infectivity of a medicated host compared to an untreated host. In reality, ν is 1 shortly after treatment but falls gradually to about 0.35. We will take ν to have a constant value of 0.6. We assume proportional death rates μ for hosts and d for vectors a rate σ of progression from exposed to infective stage for hosts. The rates of progression from E to P and I are $p\sigma$ and $q\sigma$, respectively, where $q = 1 - p$. The progression rate from the premedication class P to the medicated class M is φ . Since infected humans recover only through the death of all adult worms, there is a rate of recovery of all infected host classes γ . With time measured in years, approximate parameter values are

$$\mu = 0.02, d = 12, \sigma = 1, p = 0.65, \varphi = 2, \nu = 0.6, \gamma = 0.8.$$

The values of the contact rate β and the infection rate α are difficult to measure. Since N is constant, we can omit $S = N - E - I - R$ from the model.

The model is

$$\begin{aligned} \frac{dE}{dt} &= \beta SV - (\sigma + \mu)E \\ \frac{dI}{dt} &= q\sigma E - (\gamma + \mu)I \\ \frac{dH}{dt} &= p\sigma E - (\gamma + \mu)H \\ \frac{dM}{dt} &= \varphi H - (\varphi + \gamma + \mu)M \\ \frac{dV}{dt} &= \alpha(F - V)W - dV. \end{aligned} \tag{6.23}$$

The parameters α and β can be estimated from known endemic fractions of infected humans and vectors in the absence of ivermectin treatment. Field estimates from Cameroon suggest

$$\alpha = \frac{1.08}{N}, \quad \beta = \frac{0.3}{F}.$$

Question 1 Use the next generation matrix to obtain an expression for the basic reproduction number \mathcal{R}_0 and estimate a numerical value for \mathcal{R}_0 using the given parameter values.

Question 2 We rescale the system (6.23) with respect to the constant total host population size N and the constant total vector population size F . We make the substitutions

$$S = Nx, E = Ny, I = Ni, H = Nh, M = Nm, V = Fv, W = Fw.$$

Obtain the resulting system, with $a = \alpha/d$,

$$\begin{aligned} \frac{dy}{dt} &= \beta xv - (\sigma + \mu)y \\ \frac{di}{dt} &= q\sigma y - (\gamma + \mu)i \\ \frac{dh}{dt} &= p\sigma y - (\gamma + \mu)h \\ \frac{dm}{dt} &= \varphi h - (\varphi + \gamma + \mu)m \\ \frac{1}{d} \frac{dv}{dt} &= aw(1 - v) - v. \end{aligned} \tag{6.24}$$

Question 3 Rescale the system (6.24) further by making a change of independent variable $\tau = (\gamma + \mu)t$.

Question 4 The model (6.24) has different time scales for vectors and hosts. We may view it as a singular perturbation and approximate it by letting $1/d \rightarrow 0$. There is a quasi-steady state given by $aw(1 - v) - v = 0$ or

$$v = \frac{aw}{1 + aw}, \tag{6.25}$$

and the model is given by the first four equations of (6.24) together with (6.25). The error in this approximation to (6.24) is in the *boundary layer* close to $t = 0$. Show that the reduced model has a disease-free equilibrium $y = 1, i = h = m = 0$. Determine the basic reproduction number \mathcal{R}_0 and show that the disease-free equilibrium is asymptotically stable if $\mathcal{R}_0 < 1$. Show that there is an endemic equilibrium which is asymptotically stable if $\mathcal{R}_0 > 1$.

In practice, delivery of ivermectin is administered at fixed intervals. This suggests that a pulse vaccination model would be appropriate, and this would admit the possibility of stable periodic orbits. However, we do not explore this generalization here.

References: [7, 11].

6.7 Exercises

1. Consider the model

$$\begin{aligned} S'_h &= \Lambda_h(N_h) - \beta_h S_h I_v - \mu_h S_h \\ I'_h &= \beta_h S_h I_v - (\mu_h + \alpha_h) I_h \\ S'_v &= \Lambda_v(N_v) - \beta_h S_v I_h - \mu_v S_v \\ I'_v &= \beta_v S_v I_h - \mu_v I_v, \end{aligned}$$

a simple model for malaria with a host species (h) and a vector species (v) under the assumptions $\Lambda_i(N_h) = \mu_i N_i$, $i = h, v$, and $\alpha_h = 0$.

- (a) Derive the expression of \mathcal{R}_0 .
- (b) Find all possible equilibrium points.
- (c) Find the conditions (in terms of \mathcal{R}_0) under which an endemic equilibrium exists.
- (d) Determine the critical value of β_{vc} such that $\mathcal{R}_0 < 1$ for all $\beta_v < \beta_{vc}$.

2. Consider the model

$$\begin{aligned} S'_h &= -\beta_h S_h I_v \\ I'_h &= \beta_h S_h I_v - \alpha_h I_h \\ S'_v &= \Lambda_v(N_v) - \beta_v S_v I_h - \mu_v S_v - \sigma S_v \\ I'_v &= \beta_v S_v I_h - \mu_v I_v - \sigma S_v, \end{aligned}$$

describing a vector-borne *SIR/SI* epidemic model that includes removal of vectors at rate σ . Calculate the basic reproduction number.

3. Formulate a model for a vector-borne epidemic with a rate of treatment of infective hosts, and calculate its reproduction number.

References

1. Brauer, F. (2017) A final size relation for epidemic models of vector - transmitted diseases, *Infectious Disease Modelling*, **2**: 12–20.
2. Brauer, F. (2019) A singular perturbation approach to epidemics of vector-transmitted diseases, to appear.
3. Brauer, F. and C. Kribs (2016) *Dynamical Systems for Biological Modeling: An Introduction*, CRC Press.
4. Chowell, G., P. Diaz-Duenas, J.C. Miller, A. Alcazar-Velasco, J.M. Hyman, P.W. Fenimore, and C. Castillo-Chavez (2007) Estimation of the reproduction number of dengue fever from spatial epidemic data, *Math. Biosc.* **208**: 571–589.

5. Coffeng, L.E., W.A. Stolk, HGM Zouré, J.L. Vetterman, & K.B. Agblewonu (2013) African programme for onchocerciasis control 1995–2015:Model-estimated health impact and cost, PLoS Neglected tropical diseases, **7**: e2032. <https://doi.org/10.1371/journal.ptnd.0002032>.
6. Cushing, J.M. & O. Diekmann (2016) The many guises of \mathcal{R}_0 (a didactic note), J. Theor. Biol. **404**: 295–302.
7. Dietz, K. (1982) The population dynamics of onchocerciasis, in Population Dynamics of Infectious Diseases (R.m. Anderson, ed.), Chapman and Hall, London, pp.209–241.
8. Hopkins, A. and B.A. Boatman (2011) Onchocerciasis, water and sanitation-related diseases and the environment: Challenges, interventions, and preventive measures (J.M.H. Selendy, ed.), John Wiley and Sons: pp.133–149.
9. Hoppensteadt, F.C. (1966) Singular perturbations on the infinite interval, Trans. Amer. Math. Soc. **123**: 521–535.
10. Kucharski, A.J., S. Funk, R.M. Egge, H-P. Mallet, W.J. Edmunds and E.J. Nilles (2016) Transmission dynamics of Zika virus in island populations: a modelling analysis of the 2013–14 French Polynesia outbreak, PLOS Neglected tropical Diseases DOI 101371
11. Ledder, G., D. Sylvester, R.R. Bouchat, and J.A. Thiel (2017) Continued and pulsed epidemiological models for onchocerciasis with implications for eradication strategy, Math. Biosc. & Eng., to appear.
12. Levinson, N. (1950) Perturbations of discontinuous solutions of nonlinear systems of differential equations, Acta Math. **82**: 71–106.
13. Pinho, S.T.R., C.P. Ferreira, L. Esteva, F.R. Barreto, V.C. Morato e Silva and M.G.L Teixeira (2010) Modelling the dynamics of dengue real epidemics, Phil. Trans. Roy. Soc. A **368**: 5679–5693.
14. Plaisier,A.P., E.S. Alley, B.A. Boutin, G.J. van Oortmarsen, & H. Remme (1995) Irreversible effects of ivermectin on adult parasites in onchocerciasis patients in the onchocerciasis control programme in West Africa, J. Inf. Diseases **172**; 204–210.
15. Rocha, F., M. Aguiar, M. Souza, & N. Stollenwerk (2013) Time-scale separation and centre manifold analysis describing vector-borne disease dynamics, Int. J. Comp. Math **90**: 2105–2125.
16. Ross, R. (1911) *The Prevention of Malaria*, 2nd ed., (with Addendum), John Murray, London.
17. Segel, L.A. & M. Slemrod (1989) The quasi-steady-state assumption: A case study in perturbation, SIAM Review **31**: 446–477.
18. Souza, M.O. (2014) Multiscale analysis for a vector-borne epidemic model (2014) J. Math. Biol. **68**: 1269–1291.
19. Takeuchi, Y., W. Ma, & E. Beretta (2000) Global asymptotic properties of a delay SIR epidemic model with finite incubation times, Nonlin. Analysis **42**; 931–947.
20. Tikhonov, A.N. (1948) On the dependence of the solutions of differential equations on a small parameter, Mat. Sbornik NS **22**; 193–204.
21. van den Driessche, P. and J. Watmough (2002) Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission, Math. Biosc. **180**:29–48.

Part II

Models for Specific Diseases

Chapter 7

Models for Tuberculosis



In this chapter we describe several models for tuberculosis (TB). The disease is endemic in many areas of the world. The models in this chapter will be extensions of the standard SIR or SEIR type of endemic models presented in Chap. 3. Depending on the typical characteristics of a specific disease, various modifications of the standard models will be considered.

According to the recent WHO report [25], there were 8.6 million new TB cases in 2012 and 1.3 million TB deaths. TB remains a major global health problem and is the leading cause of death by an infectious disease, after the human immunodeficiency virus (HIV). It is reported that about three million people who developed TB in 2012 were missed by national notification systems. Key actions needed to detect people with the illness and ensure that they get the right treatment and care include: expanded services (including rapid tests) throughout health systems bolstered by the support of nongovernmental organizations, community workers, and volunteers to diagnose and report cases.

A typical epidemiological feature associated with TB is its long period of latency. As pointed out by G.W. Comstock, “tuberculosis is an infectious disease with an incubation period from weeks to a lifetime.” Figure 7.1 illustrates that TB has a long and variable period of latency. Treating a patient with an active TB is more difficult and requires a much longer time to complete the treatment than treating a latent TB infection (LTBI). This makes it important to identify and treat latent people before they develop the disease. One of the approaches to achieve this is through screening. However, such screening programs require resources. An optimal control problem can be formulated using mathematical models for TB.

Figure 7.2 shows the data from observation in adolescents who had developed clinical tuberculosis following primary infection [24]. It suggests that among the 10% of latent individuals who eventually develop active TB, around 60% will do so during the first year post-infection. The rest will develop active TB in either 2 years (20%), 5 years (15%), 20 years (5%), or even longer.

Fig. 7.1 A diagram showing the progression from latent to active TB during the period of infection (adopted from www.biomerieux-Rdiagnos6cs.com). It shows that only about 10% of latently infected will develop active TB and 5% of those will stay in the latent stage for long time

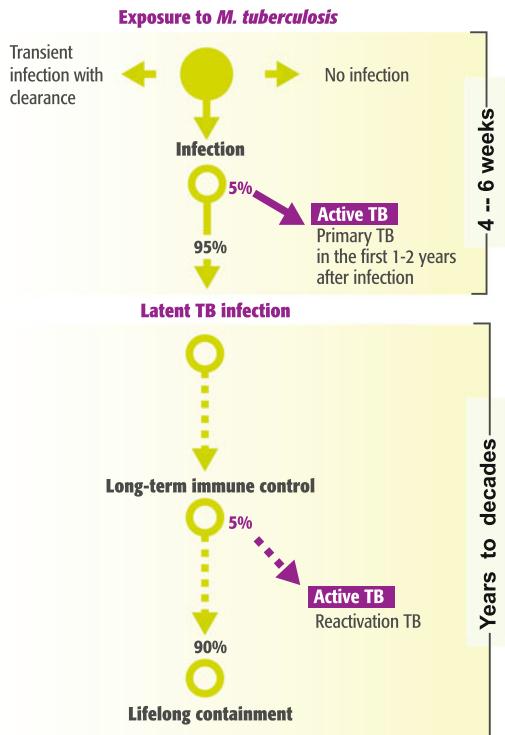
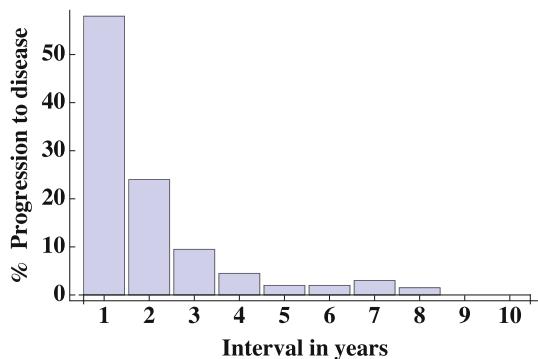


Fig. 7.2 An example of distribution of progression from latent to active TB [24]



The good news is that latent and active TB can be treated with antibiotics. The bad news is that its treatment has side effects (sometimes quite serious) and takes a long time. Carriers of the tubercle bacillus who have not developed TB disease can be treated with a single drug *INH*; unfortunately, it must be taken religiously for 6 months [6]. Treatment for those with active TB requires the simultaneous use of three drugs for a period of about 9 months. Lack of compliance with these drug treatments (a very serious problem) may lead to not only a relapse but also to the development of multidrug-resistant TB (MDR-TB)—one of the most serious

public health problems facing society today. According to the WHO report [25], globally in 2012, approximately 450,000 people developed MDR-TB and there were approximately 170,000 deaths from MDR-TB. An individual can become infected with the resistant strain of TB in two ways, one is the so-called primary resistance which is obtained by direct transmission from someone with resistant TB, and the other is the acquired resistance which is developed from the sensitive TB due to incomplete or inappropriate treatment. This also creates a challenge for designing treatment policy, and should be incorporated in the modeling of optimal control for TB.

In this chapter we present several TB models that can be used to study the above mentioned problems. We begin with a relatively simple TB model with a single strain, and then extend it to include both drug-sensitive and drug-resistant strains. The two-strain model will be further extended to include two control measures representing “case-finding” (i.e., identifying people with LTBI) and “case-holding” (i.e., making sure the treatment of active TB infections is complete) and study the optimal control strategies.

7.1 A One-Strain Model with Treatment

Because there is no permanent immunity and an individual after treatment for TB can still become infected with possibly reduced susceptibility, we divide the population into four epidemiological classes: susceptible (S), latently infected (L), infectious (I), and treated (T). Assume that latent and infectious individuals are treated at rates r_1 and r_2 , respectively, and latent individuals develop active TB at rate κ . The model reads

$$\begin{aligned} S' &= \mu N - cS \frac{I}{N} - \mu S + r_1 L + r_2 I, \\ L' &= cS \frac{I}{N} + c^* T \frac{I}{N} - (\kappa + r_1 + \mu)L, \\ I' &= \kappa L - (r_2 + \mu)I, \\ T' &= r_1 L + r_2 I - c^* T \frac{I}{N} - \mu T, \end{aligned} \tag{7.1}$$

where $N = S + L + I + T$ is the total population, which will remain constant for all time due to the balanced birth and natural death rate μ . The parameters c and c^* denote the average numbers of susceptible and treated individuals, respectively, infected by one infective individual per unit of time. If the treated individuals have a reduced susceptibility to infection, then $c^* < c$.

The dynamics of system (7.1) is standard in the sense that the disease will either go extinct or persist depending on whether the reproduction number

$$\mathcal{R}_0 = \frac{c\kappa}{(\kappa + r_1 + \mu)(r_2 + \mu)} \quad (7.2)$$

is less or greater than 1. It is clear from (7.2) that \mathcal{R}_0 is a decreasing function of treatment rates r_1 and r_2 . The effect of treatment on the disease prevalence can be examined by considering the fraction of infectives I/N at the endemic equilibrium in the case $\mathcal{R}_0 > 1$. In the simpler case when $c = c^*$, the equilibrium value of I/N is given by

$$\frac{I^*}{N^*} = \frac{\kappa}{\kappa + r_2 + \mu} \left(1 - \frac{1}{\mathcal{R}_0}\right), \quad (7.3)$$

which is a decreasing function of r_1 and r_2 as well. This shows that in the absence of the resistant strain, treatment is beneficial in reducing the disease burden. This is not the case when a resistant strain is considered, as shown in the two-strain model presented next.

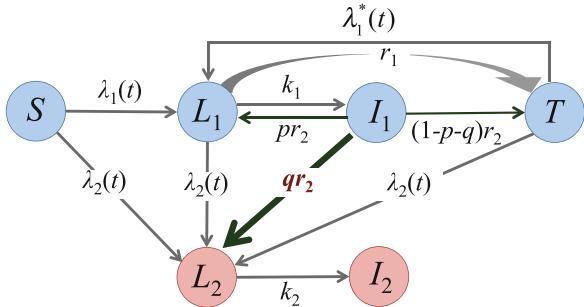
7.2 A Two-Strain TB Model

As mentioned earlier, treatment of active TB may take as long as 12 months, and lack of compliance with these treatments may lead to the development of antibiotic resistant TB. The one-strain model (7.1) can be extended to include both drug-sensitive (DS) and drug-resistant (DR) strains of TB, with possible development of resistance due to treatment failure. A transition diagram between the epidemiological classes is shown in Fig. 7.3. The latent and infectious individuals with sensitive TB are denoted by L_1 and I_1 , respectively. Two additional classes are included for the resistant strain, i.e., the latent and infectious classes with resistant strain denoted by L_2 and I_2 , respectively. The sensitive and resistant strains will be referred to as strain 1 and 2, respectively. Because it is very hard to cure a patient with resistant TB, we ignore the treatment of the resistant strain. Furthermore, assume that I_2 individuals can infect S , L_1 , and T individuals. The λ functions represent the forces of infection, which are given by

$$\lambda_i(t) = c_i \frac{I_i}{N}, \quad \lambda_1^*(t) = c_1^* \frac{I_1}{N}, \quad i = 1, 2,$$

where c_1 and c_1^* have the meaning as c and c^* in the one-strain model (7.1); c_2 is similar to c_1 but for the resistant strain; r_i ($i = 1, 2$) and μ are the same as in (7.1); κ_i denotes the progression rate from latent to infectious stage of strain i .

Fig. 7.3 A diagram for the two-strain TB model showing transitions between the epidemiological classes. The thicker arrow and the rate qr_2 represent the development of resistant TB due to the failure of treatment for infections with the sensitive strain. The birth and death rates are omitted



The additional parameters are related to treatment failure and the development of resistant TB. For example, $p + q$ denotes the proportion of those treated infectious individuals who did not complete their treatment, where the proportion p modifies the rate that departs from the sensitive TB latent class; $qr_2 I_1$ gives the rate at which individuals develop resistant TB because they did not complete the treatment of active TB. Therefore, $p \geq 0$, $q \geq 0$ and $p + q \leq 1$. From the disease transmission diagram (see Fig. 7.3) we can write the following system of ordinary differential equations:

$$\begin{aligned} S' &= \mu N - c_1 S \frac{I_1}{N} - c_2 S \frac{I_2}{N} - \mu S, \\ L'_1 &= c_1 S \frac{I_1}{N} - (\mu + \kappa_1) L_1 - r_1 L_1 + pr_2 I_1 + c_1^* T \frac{I_1}{N} - c_2 L_1 \frac{I_2}{N}, \\ I'_1 &= \kappa_1 L_1 - \mu I_1 - r_2 I_1, \\ L'_2 &= qr_2 I_1 - (\mu + \kappa_2) L_2 + c_2 (S + L_1 + T) \frac{I_2}{N}, \\ I'_2 &= \kappa_2 L_2 - \mu I_2, \\ T' &= r_1 L_1 + (1 - p - q) r_2 I_1 - c_1^* T \frac{I_1}{N} - c_2 T \frac{I_2}{N} - \mu T, \end{aligned} \tag{7.4}$$

where $N = S + L_1 + I_1 + T + L_2 + I_2$ is the total population, which remains constant.

The detailed analysis of the model (7.4) is presented in [4]. System (7.4) has up to four possible equilibria denoted by E_0 (infection-free), E_1 (only the sensitive strain is present), E_2 (only the resistant strain is present), and E^* (coexistence of both strains). The existence of these equilibria depends on the reproduction numbers for the sensitive and resistant strains, which are given by

$$\mathcal{R}_S = \left(\frac{c_1 + pr_2}{\mu + r_2} \right) \left(\frac{\kappa_1}{\mu + \kappa_1 + r_1} \right) \tag{7.5}$$

and

$$\mathcal{R}_R = \left(\frac{c_2}{\mu} \right) \left(\frac{\kappa_2}{\mu + \kappa_2} \right), \quad (7.6)$$

respectively.

The dynamics of system (7.4) are dramatically different for the cases $q = 0$ and $q > 0$ (development of resistant TB due to treatment failure), particularly in terms of the number of equilibria and their stability, as well as the likelihood for coexistence of both strains. In addition to the reproduction numbers, there are two functions, $\mathcal{R}_R = f(\mathcal{R}_S)$ and $\mathcal{R}_R = g(\mathcal{R}_S)$, which divide the parameter region in the $(\mathcal{R}_S, \mathcal{R}_R)$ plane into sub-regions for the stability of equilibria:

$$\begin{aligned} f(\mathcal{R}_S) &= \frac{1}{1 + \frac{1-\mathcal{R}_S}{(\mathcal{R}_S-AB)(1+1/B)}} \\ g(\mathcal{R}_S) &= \frac{1}{C} \left(AB + C - 1 \pm \sqrt{(AB + C - 1)^2 + 4(\mathcal{R}_S - AB)C} \right), \end{aligned} \quad (7.7)$$

for $\mathcal{R}_S \geq 1$ where

$$A = \frac{pr_2}{\mu + \kappa_1 + r_1}, \quad B = \frac{\kappa_1}{\mu + d_1 + r_2}, \quad C = \frac{\mu}{\mu + \kappa_1 + r_1}.$$

The properties of f and g include

$$f(1) = g(1) = 1, \quad f(\mathcal{R}_S) < g(\mathcal{R}_S) \quad \text{for } \mathcal{R}_S > 1$$

(see Fig. 7.4). The two curves of f and g and the lines $\mathcal{R}_i = 1$ ($i = S, R$) divide the first quadrant of the $(\mathcal{R}_S, \mathcal{R}_R)$ plane into either four regions in the case $q = 0$ (see Fig. 7.4a) or three regions in the case $q > 0$ (see Fig. 7.4b). The stability results for system (7.4) in the case of $q = 0$ and $q > 0$ are summarized in Theorems 7.1 and 7.2, respectively.

Theorem 7.1 Assume $q = 0$. Let Regions I–IV be as in Fig. 7.4a.

- (a) The disease-free equilibrium E_0 is g.a.s. if $(\mathcal{R}_S, \mathcal{R}_R)$ is in Region I.
- (b) For $\mathcal{R}_S > 1$, the boundary equilibrium E_1 is locally asymptotically stable if $(\mathcal{R}_S, \mathcal{R}_R)$ is in Region II and unstable in Regions III and IV.
- (c) For $\mathcal{R}_R > 1$, the boundary equilibrium E_2 is locally asymptotically stable if $(\mathcal{R}_S, \mathcal{R}_R)$ is in Region IV and unstable if in Regions II and III.
- (d) The coexistence equilibrium E^* exists and is locally asymptotically stable if $(\mathcal{R}_S, \mathcal{R}_R)$ is in Region III.

When $q > 0$, the equilibrium E_1 (sensitive strain only) is never stable, and the coexistence region III is much larger than that in the case of $q = 0$, as stated in the following theorem and illustrated in Fig. 7.4b.

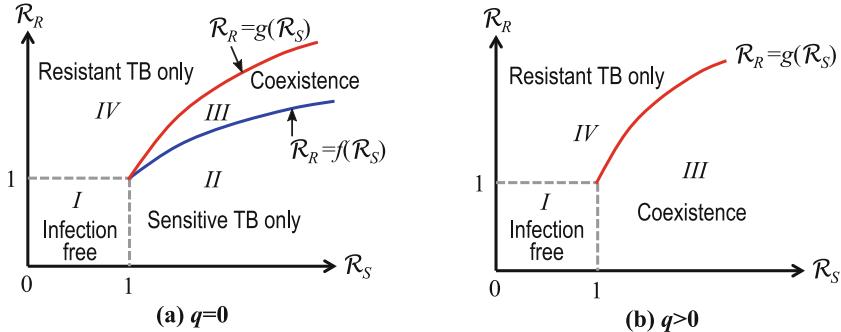


Fig. 7.4 (a) A bifurcation diagram for the system in the case $q = 0$. There are four Regions I , II , III , and IV in the parameter space $(\mathcal{R}_S, \mathcal{R}_R)$. In Region I , E_0 is a global attractor and other equilibria are unstable when they exist. In Regions II and IV , E^* does not exist, while E_1 and E_2 are locally asymptotically stable, respectively. In Region III , E^* exists and is locally asymptotically stable. (b) A bifurcation diagram for the system in the case $q > 0$. There are three Regions I , III , and IV in the parameter space $(\mathcal{R}_S, \mathcal{R}_R)$ (E_1 does not exist), in which E_0 , E_2 , and E^* are stable, respectively

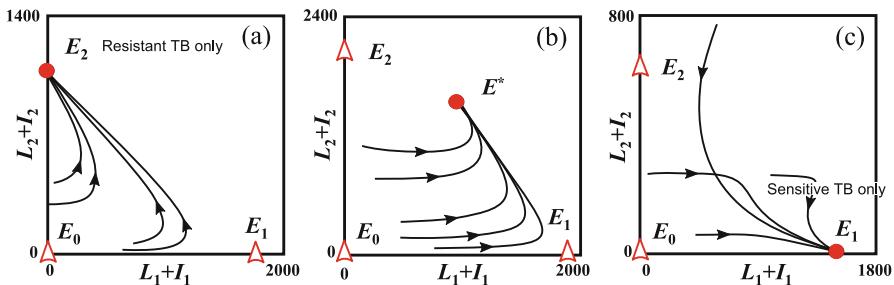


Fig. 7.5 Phase portraits of solutions to (7.4) in the case of $q = 0$. The choice of parameter values gives a fixed value $\mathcal{R}_S = 3.45$. In (a) $\mathcal{R}_R = 2$ and $(\mathcal{R}_S, \mathcal{R}_R) \in IV$. In (b) $\mathcal{R}_R = 2.4$ and $(\mathcal{R}_S, \mathcal{R}_R) \in III$. In (c) $\mathcal{R}_R = 1.2$ and $(\mathcal{R}_S, \mathcal{R}_R) \in II$. A circle indicates a stable equilibrium, and a triangle indicates an unstable equilibrium

Theorem 7.2 Assume that $q > 0$. Let Regions I – III be as in Fig. 7.4b.

- (a) The disease-free equilibrium E_1 is g.a.s. if $\mathcal{R}_S < 1$ and $\mathcal{R}_R < 1$ (Region I).
- (b) For $\mathcal{R}_R > 1$, the boundary equilibrium E_2 is locally asymptotically stable if $\mathcal{R}_S < 1$ or if $\mathcal{R}_S > 1$ and $\mathcal{R}_R > g(\mathcal{R}_S)$ (Region IV). E_2 is unstable if $\mathcal{R}_S > 1$ and $\mathcal{R}_R < g(\mathcal{R}_S)$ (Region III).
- (c) The equilibrium E_3 exists and is locally asymptotically stable iff $\mathcal{R}_S > 1$ and $\mathcal{R}_R < g(\mathcal{R}_S)$ (Region III).

Figure 7.5 shows some simulation results of the model in the case of $q = 0$, illustrating the disease outcomes for $(\mathcal{R}_S, \mathcal{R}_R)$ in different regions as shown in Fig. 7.4a. In this figure, the parameter values used are $\mu = 0.143$, $c_1 = 13$, $\kappa_1 = 1$, $q = 0$, $p = 0.5$, $r_1 = 1$, $r_2 = 2$, $\kappa_2 = 1$. For this set of values, $\mathcal{R}_S = 3.45$.

Figure 7.5a–c corresponds to different values of \mathcal{R}_R (or equivalently c_2), for which $(\mathcal{R}_S, \mathcal{R}_R)$ is in Regions IV, III, and II, respectively.

These results demonstrate that lack of drug treatment compliance by TB patients may have an important implication for the maintenance of antibiotic resistant strains. To make the role of antibiotic resistance transparent, we first studied a special version of our two-strain model with two competing strains of TB: the typical strain plus a resistant strain that was not the result of antibiotic resistance ($q = 0$). In this last situation, we found that coexistence is possible but rare while later we noticed that coexistence is almost certain when the second strain is the result of antibiotic resistance. In our two-strain model there is a superinfection-like term $c_2 L_1 I_2 / N$. Is this necessary to obtain the coexistence result because it is well known that superinfection can cause coexistence (see [16, 18])? The answer is no. In fact, it can be shown that in the absence of the superinfection-like term coexistence is still almost the rule when the second strain is the result of antibiotic resistance (see Fig. 7.4).

7.3 Optimal Treatment Strategies

Analysis of the two-strain model in Sect. 7.2 demonstrated that treatment may facilitate the spread of resistant TB and increase the level of TB prevalence. Thus, the effort levels devoted to treating latent and infectious TB individuals may lead to different outcomes.

Let $u_1(t)$ and $u_2(t)$ denote the time-dependent control efforts, which represent the fractions of individuals in L_1 and I_1 classes receiving prophylaxis and drug treatment, respectively, at time t . The state system with controls $u_1(t)$ and $u_2(t)$ reads:

$$\begin{aligned} S' &= \mu N - c_1 S \frac{I_1}{N} - c_2 S \frac{I_2}{N} - \mu S, \\ L'_1 &= c_1 S \frac{I_1}{N} - (\mu + \kappa_1) L_1 - u_1(t) r_1 L_1 \\ &\quad + (1 - u_2(t)) p r_2 I_1 + c_1^* T \frac{I_1}{N} - c_2 L_1 \frac{I_2}{N}, \\ I'_1 &= \kappa_1 L_1 - \mu I_1 - r_2 I_1, \\ L'_2 &= (1 - u_2(t)) q r_2 I_1 - (\mu + \kappa_2) L_2 + c_2 (S + L_1 + T) \frac{I_2}{N}, \\ I'_2 &= \kappa_2 L_2 - \mu I_2, \\ T' &= u_1(t) r_1 L_1 + [1 - (1 - u_2(t))(p + q)] r_2 I_1 - c_1^* T \frac{I_1}{N} - c_2 T \frac{I_2}{N} - \mu T, \end{aligned} \tag{7.8}$$

with initial values $S(0), L_1(0), I_1(0), L_2(0), I_2(0), T(0)$.

The control functions, $u_1(t)$ and $u_2(t)$, are bounded, *Lebesgue* integrable functions. The “case-finding” control, $u_1(t)$, represents the fraction of typical TB latent individuals that is identified and will be put under treatment (to reduce the number of individuals that may be infectious). The coefficient, $1 - u_2(t)$, represents the effort that prevents the failure of the treatment in the typical TB infectious individuals (to reduce the number of individuals developing resistant TB). When the “case-holding” control $u_2(t)$ is near 1, there is low treatment failure and high implementation costs.

Our objective function to be minimized is

$$J(u_1, u_2) = \int_0^{t_f} \left[L_2(t) + I_2(t) + \frac{B_1}{2} u_1^2(t) + \frac{B_2}{2} u_2^2(t) \right] dt, \quad (7.9)$$

where we want to minimize the latent and infectious groups with resistant-strain TB while also keeping the cost of the treatments low. We assume that the costs of the treatments are nonlinear and take quadratic form here. The coefficients, B_1 and B_2 , are balancing cost factors due to size and importance of the three parts of the objective functional. We seek to find an optimal control pair, u_1^* and u_2^* , such that

$$J(u_1^*, u_2^*) = \min_{\mathcal{Q}} J(u_1, u_2), \quad (7.10)$$

where $\mathcal{Q} = \{(u_1, u_2) \in L^1(0, t_f) \mid a_i \leq u_i \leq b_i, i = 1, 2\}$ and a_i and b_i , $i = 1, 2$, are fixed positive constants.

The necessary conditions that an optimal pair must satisfy come from Pontryagin’s maximum principle [19]. This principle converts (7.8)–(7.10) into a problem of minimizing pointwise a Hamiltonian, H , with respect to u_1 and u_2 :

$$H = L_2 + I_2 + \frac{B_1}{2} u_1^2 + \frac{B_2}{2} u_2^2 + \sum_{i=1}^6 \lambda_i g_i, \quad (7.11)$$

where g_i is the right-hand side of the differential equation of the i th state variable. By applying Pontryagin’s maximum principle [19] and the existence result for the optimal control pairs from [13], we know that there exists an optimal control pair u_1^* , u_2^* and corresponding solution, S^* , L_1^* , I_1^* , L_2^* , I_2^* , and T^* , that minimizes $J(u_1, u_2)$ over \mathcal{Q} . Furthermore, there exist adjoint functions, $\lambda_1(t), \dots, \lambda_6(t)$, such that

$$\lambda'_1 = \lambda_1(c_1 \frac{I_1^*}{N} + c_1^* \frac{I_2^*}{N} + \mu) + \lambda_2(-c_1 \frac{I_1^*}{N}) + \lambda_4(-c_1^* \frac{I_2^*}{N}),$$

$$\lambda'_2 = \lambda_2(\mu + \kappa_1 + u_1(t)r_1 + c_1^* \frac{I_2^*}{N}) + \lambda_3(-\kappa_1) + \lambda_4(-c_1^* \frac{I_2^*}{N}) + \lambda_6(-u_1^*(t)r_1),$$

$$\lambda'_3 = \lambda_1(c_1 \frac{S^*}{N}) + \lambda_2(-c_1 \frac{S^*}{N}) - (1 - u_2^*(t))pr_2 - c_2 \frac{T^*}{N} + \lambda_3(\mu + r_2)$$

$$\begin{aligned}
& + \lambda_4(-(1 - u_2^*(t))qr_2) + \lambda_6(-(1 - (1 - u_2^*(t))(p + q))r_2 + c_2 \frac{T^*}{N}), \\
\lambda'_4 &= -1 + \lambda_4(\mu + \kappa_2) + \lambda_5(-\kappa_2), \\
\lambda'_5 &= -1 + \lambda_1(c_1^* \frac{S^*}{N}) + \lambda_2(c_1^* \frac{L_1^*}{N}) - \lambda_4(\beta^* \frac{S^* + L_1^* + T^*}{N}) + \lambda_5\mu + \lambda_6(c_1^* \frac{T^*}{N}), \\
\lambda'_6 &= \lambda_2(-c_2 \frac{I_1^*}{N}) + \lambda_4(-c_1^* \frac{I_2^*}{N}) + \lambda_6(c_2 \frac{I_1^*}{N} + c_1^* \frac{I_2^*}{N} + \mu),
\end{aligned} \tag{7.12}$$

with transversality conditions

$$\lambda_i(t_f) = 0, \quad i = 1, \dots, 6 \tag{7.13}$$

and $N = S^* + L_1^* + I_1^* + L_2^* + I_2^* + T^*$. Moreover, the characterization holds:

$$\begin{aligned}
u_1^*(t) &= \min \left(\max(a_1, \frac{1}{B_1}(\lambda_2 - \lambda_6)r_1 L_1^*), b_1 \right) \\
u_2^*(t) &= \min \left(\max(a_2, \frac{1}{B_2}(\lambda_2 p + \lambda_4 q - \lambda_6(p + q)r_2 I_1^*)), b_2 \right).
\end{aligned} \tag{7.14}$$

Due to the a priori boundedness of the state and adjoint functions and the resulting Lipschitz structure of the ODEs, we obtain the uniqueness of the optimal control for small t_f . The uniqueness of the optimal control follows from the uniqueness of the optimality system, which consists of (7.8) and (7.12)–(7.14). There is a restriction on the time interval in order to guarantee the uniqueness of the optimality system. This smallness restriction on the length on the time interval is due to the opposite time orientations of (7.8), (7.12), and (7.13); the state problem has initial values and the adjoint problem has final values. This restriction is very common in control problems (see [12, 15]).

The optimal treatment is obtained by solving the optimality system, consisting of the state and adjoint equations. An iterative method is used for solving the optimality system. We start to solve the state equations with a guess for the controls over the simulated time using a forward fourth-order Runge–Kutta scheme. Because of the transversality conditions (7.13), the adjoint equations are solved by a backward fourth-order Runge–Kutta scheme using the current iteration solution of the state equations. Then, the controls are updated by using a convex combination of the previous controls and the value from the characterizations (7.14). This process is repeated and iteration is stopped if the values of unknowns at the previous iteration are very close to the ones at the present iteration.

For the figures presented here, we assume that the weight factor B_2 associated with control u_2 is greater than or equal to B_1 which is associated with control u_1 . This assumption is based on the following facts: The cost associated with u_1 will include the cost of screening and treatment programs, and the cost associated with u_2 will include the cost of holding the patients in the hospital or sending people to watch the patients to finish their treatment. Treating an infectious TB individual takes longer (by several months) than treating a latent TB individual. In these three

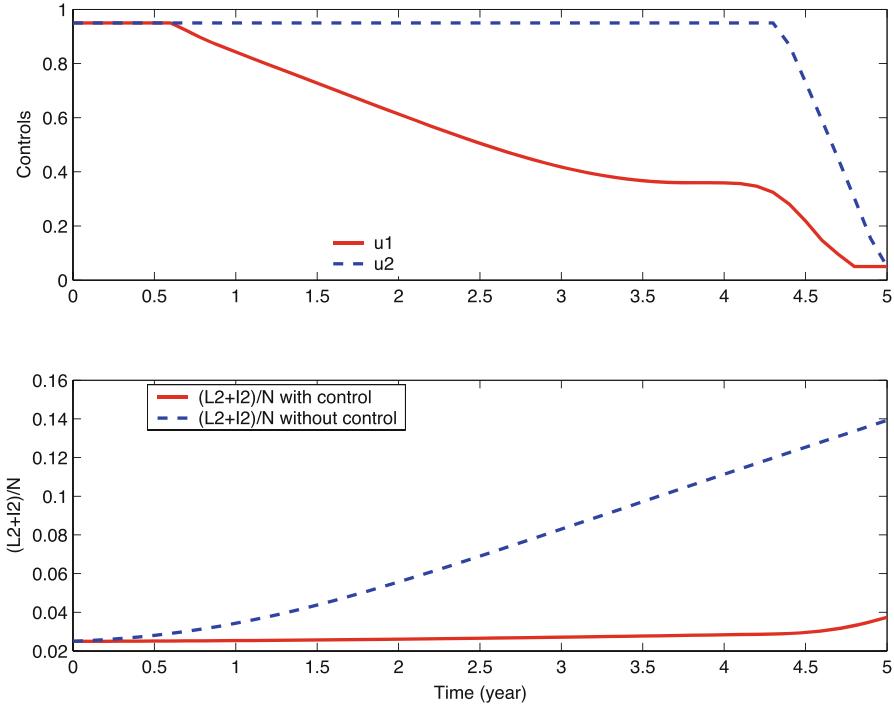


Fig. 7.6 The optimal control strategy for the case of $B_1 = 50$ and $B_2 = 500$

figures, the set of the weight factors, $B_1 = 50$ and $B_2 = 500$, is chosen to illustrate the optimal treatment strategy. Other epidemiological and numerical parameters are presented in [14].

In the top frame of Fig. 7.6, the controls, u_1 (solid curve) and u_2 (dashdot curve), are plotted as a function of time. In the bottom frame, the fractions of individuals infected with resistant TB, $(L_2 + I_2)/N$, with control (solid curve) and without control (dashed curve) are plotted. Parameters $N = 30,000$ and $\beta^* = 0.029$ are chosen. Results for other parameters are presented in [14]. To minimize the total number of the latent and infectious individuals with resistant TB, $L_2 + I_2$, the optimal control u_2 is at the upper bound during almost 4.3 years and then u_2 is decreasing to the lower bound, while the steadily decreasing value for u_1 is applied over the most of the simulated time, 5 years. The total number of individuals $L_2 + I_2$ infected with resistant TB at the final time $t_f = 5$ (years) is 1123 in the case with control and 4176 without control, and the total number of cases of resistant TB prevented at the end of the control program is 3053 ($= 4176 - 1123$).

In Fig. 7.7, the controls, u_1 and u_2 , are plotted as a function of time for $N = 6000, 12,000$, and $30,000$ in the top and bottom frame, respectively. Other parameters except the total number of individuals and $c_1^* = 0.029$ are fixed for

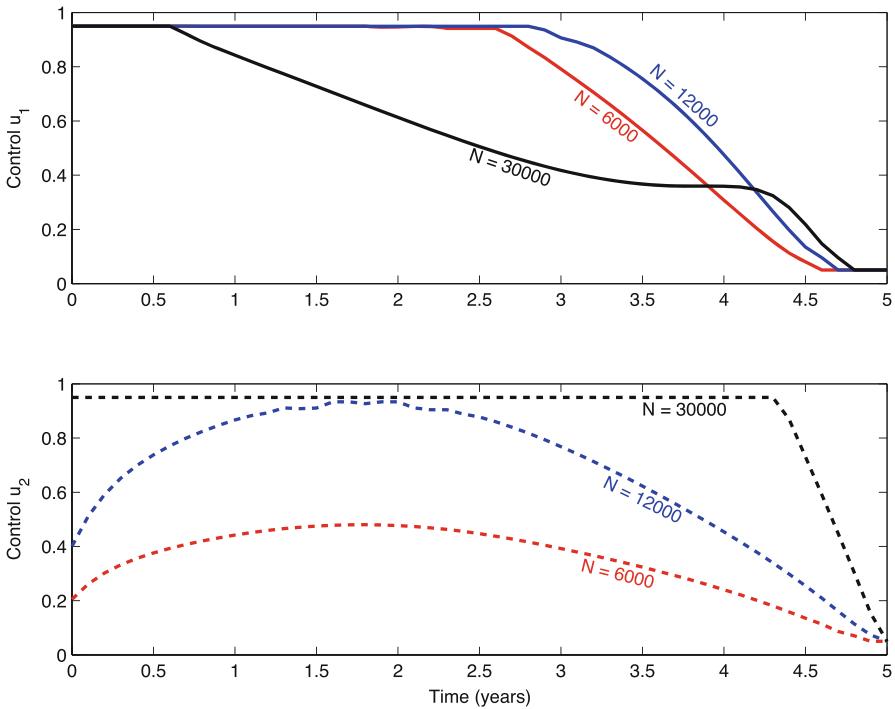


Fig. 7.7 The controls, μ_1 and μ_2 , are plotted as a function of time for $N = 6000, 12,000$, and $30,000$ in the top and bottom frame, respectively

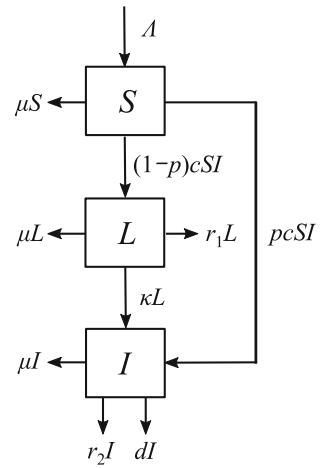
these three cases. These results show that more effort should be devoted to “case-finding” control u_1 if the population size is small, but “case-holding” control u_2 will play a more significant role if the population size is big. Note that, in general, with B_1 fixed, as B_2 increases, the amount of u_2 decreases. A similar result holds if B_2 is fixed and B_1 increases.

In conclusion, our optimal control results show how a cost-effective combination of treatment efforts (case holding and case finding) may depend on the population size, cost of implementing treatments controls, and the parameters of the model. We have identified optimal control strategies for several scenarios. Control programs that follow these strategies can effectively reduce the number of latent and infectious resistant-strain TB cases.

7.4 Modeling of the Long and Variable Latency of TB

As indicated in Figs. 7.1 and 7.2, the latent period of TB can range from a couple of years to lifetime. One of the approaches to incorporate this feature is to divide latent individuals into two classes based on the rates of progression to the disease stage,

Fig. 7.8 A transition diagram of TB with fast and slow progressions [3]



with one class having a faster progression than the other. For example, based on the transition diagram shown in Fig. 7.8, Blower et al. considered the following model in [3]:

$$\begin{aligned} S' &= \Lambda - cSI - \mu S, \\ L' &= (1 - p)cSI - (\kappa + r_1 + \mu)L, \\ I' &= pcSI + \kappa L - (r_2 + \mu + d)I, \\ C' &= r_1 L - \mu C, \\ T' &= r_2 I - \mu T, \end{aligned} \quad (7.15)$$

which includes five epidemiological classes: susceptible (S), latently infected (L), infectious (I), effectively chemoprophylaxed (C), and effectively treated (T). The model assumes that a fraction p of the newly infected individuals will become infectious within 1 year (fast progression, $p \approx 0.05$), while the remaining $1 - p$ fraction of newly infected individuals will enter the latent stage first and develop active TB at a rate κ (slow progression, approximately 1/20 (years)). The model suggests that for fast progression, an infected individual enters the infectious class I immediately. The parameters r_1 and r_2 denote the rates of prophylaxis and treatment, respectively. The per-capita natural and disease mortalities are μ and d , respectively.

The control reproduction number of the model (7.15) is

$$\mathcal{R}_C = \mathcal{R}_C^{\text{fast}} + \mathcal{R}_C^{\text{slow}}, \quad (7.16)$$

where

$$\mathcal{R}_C^{\text{fast}} = \left(\frac{cp\Lambda}{\mu} \right) \left(\frac{1}{r_2 + \mu + d} \right)$$

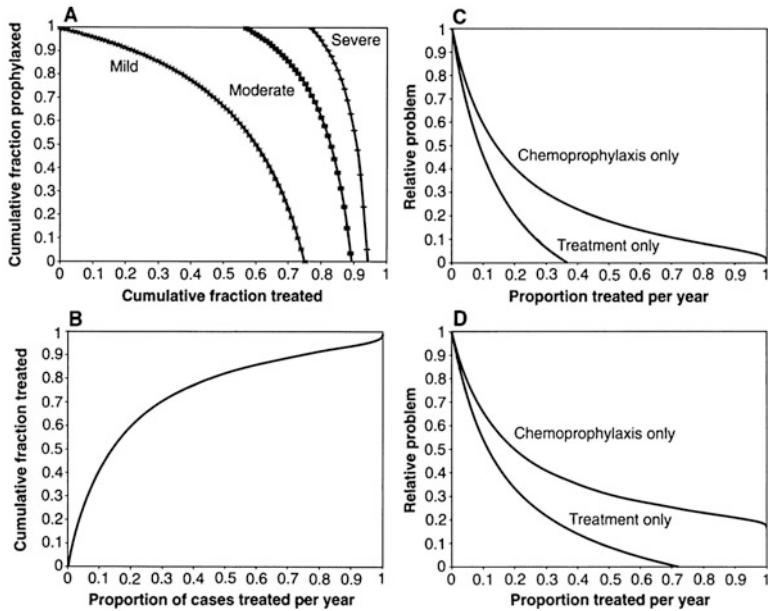


Fig. 7.9 Plots of the curve $\mathcal{R}_C = 1$ for severity levels of mild, moderate, and severe, which correspond to the basic reproduction number \mathcal{R}_0 values of 4, 9, and 17, respectively [3]

and

$$\mathcal{R}_C^{\text{slow}} = \left(\frac{c(1-p)\Lambda}{\mu} \right) \left(\frac{\kappa}{r_1 + \kappa + \mu} \right) \left(\frac{1}{r_2 + \mu + d} \right)$$

represent the reproduction numbers associated with the fast and slow paths, respectively. The formula (7.16) can be used to evaluate the possibility to eradicate tuberculosis either by treatment alone or by a combination of treatment and chemoprophylaxis, as demonstrated in Fig. 7.9 (adopted from [3]).

To investigate the role of treatment failure for drug-sensitive (DS) cases in the prevalence of drug-resistant (DR) TB, Blower et al. [3] extended the one-strain TB model (7.15) to include a drug-resistant strain as follows:

$$\begin{aligned}
 S' &= \Lambda - (c_S I_S + c_R I_R)S - \mu S, \\
 L'_S &= (1-p)c_S I_S S - (r_1 + \kappa + \mu)L_S, \\
 I'_S &= p c_S I_S S + \kappa L_S - (r_2 + \mu + d)I_S, \\
 C'_S &= r_1 L_S - \mu C_S, \\
 T'_S &= r_2(1-q)I_S - \mu T_S, \\
 L'_R &= (1-p)c_R I_R S - (\kappa + \mu)L_R, \\
 I'_R &= p c_R I_R S + q r_2 I_S + \kappa L_R - (r_2 \delta + \mu + d)I_R, \\
 T'_R &= \delta r_2 T_R - \mu T_R,
 \end{aligned} \tag{7.17}$$

where the epidemiological classes associated with the DS and DR tuberculosis are indicated by the subscripts S and R , respectively. The parameter q represents the fraction of treatment failure for DS tuberculosis that leads to the development of DR tuberculosis, and the parameter δ denotes the relative effectiveness of treatment for a resistant case.

Blower et al. [3] showed that the model (7.17) can help gain insights into the impact of treatment failure (q) on the development of DR tuberculosis.

Another approach to incorporate the long and variable latency is to consider an arbitrarily distributed latency, in which case the model consists of a system of integro-differential equations. Examples can be found in [10, 11]. It is shown in [10] that the model with a distributed latency has the same dynamical behavior as the ODE model, although it can provide a more detailed description for the reproduction number as it involves a more realistic distribution of the latent period. The model considered in [11] includes both DS and DR strains with infection age-dependent progression (e.g., the progression distribution shown in Fig. 7.2).

7.5 Backward Bifurcation in a TB Model with Reinfection

As mentioned earlier, the latent period of TB can be as long as many years and even lifetime. Re-exposure to TB bacilli through repeated contacts with individuals with active TB may accelerate the progression of LTBI towards active TB, and exogenous reinfection (i.e., acquiring a new infection from another infected individual) may occur. To investigate the impact of exogenous reinfection in the spread and control of TB, we can extend the one-strain model (7.1) by incorporating the exogenous reinfection. It is demonstrated in [9] that exogenous reinfection may play a fundamental role in the transmission dynamics and the epidemiology of TB at the population level. Particularly, the model is capable of exhibiting a backward bifurcation, i.e., a stable endemic equilibrium can exist even when the reproduction number is less than 1. Although some studies (e.g., see [23]) find that, for parameter values in a certain range, the onset of the backward bifurcation is unlikely to occur, other scenarios are possible in which the conditions for backward bifurcation can be satisfied. In either case, it is helpful to know that exogenous reinfection may play an important role in TB dynamics, which can be critical for the design of control programs for TB.

An extension of the one-strain model (7.1) when reinfection is included takes the form:

$$\begin{aligned} S' &= \mu N - cS \frac{I}{N} - \mu S, \\ L' &= cS \frac{I}{N} + c^*T \frac{I}{N} - pcL \frac{I}{N} - (\kappa + \mu)L, \\ I' &= pcL \frac{I}{N} + \kappa L - (r + \mu)I, \\ T' &= rI - c^*T \frac{I}{N} - \mu T. \end{aligned} \tag{7.18}$$

The model (7.18) ignores the treatment of latent individuals and only infectious individuals may receive treatment at the rate r . The parameter p is a factor reflecting the difference between primary infections (infection from susceptibles) and exogenous reinfections (infection from latent individuals). A value of $p \in (0, 1)$ implies that a primary infection provides some degree of cross-immunity to exogenous reinfections. Other parameters have the same meaning as in (7.1).

The reproduction number for (7.18) is

$$\mathcal{R}_0 = \frac{c\kappa}{(\kappa + \mu)(r + \mu)}. \quad (7.19)$$

The usual result that the disease-free equilibrium is locally asymptotically stable when $\mathcal{R}_0 < 1$ still holds. However, the usual result that there is no endemic equilibrium when $\mathcal{R}_0 < 1$ no longer holds. To show that an endemic equilibrium may exist when $\mathcal{R}_0 < 1$, consider the simpler case when $c^* = c$. Let $U^* = (S^*, L^*, I^*, T^*)$ denote an endemic equilibrium, i.e., $I^* > 0$. Let $x = I^*/N$, then

$$\frac{S^*}{N} = \frac{\mu}{\mu + cx^*}, \quad \frac{L^*}{N} = \frac{(\mu + r)x^*}{\kappa + p\mu x^*}, \quad \frac{T^*}{N} = \frac{rx}{\mu + cx},$$

and x^* is a solution of the quadratic equation

$$Ax^2 + Bx + C = 0, \quad (7.20)$$

where

$$A = p\mathcal{R}_0, \quad B = (1 + p + Q)D_E - p\mathcal{R}_0, \quad C = D_E Q \left(\frac{1}{\mathcal{R}_0} - 1 \right),$$

and $D_E = \kappa/(\mu + \kappa)$ and $Q = \kappa/(\mu + r)$. Note that $D_E < 1$ denotes the probability that a latent individual survives and becomes infective. Let

$$\mathcal{R}_p = \frac{1}{p} \left(D_E (1 + p - Q) + s\sqrt{D_E Q(p - pD_E - D_E)} \right) \quad (7.21)$$

and

$$p_0 = \frac{(1 + Q)D_E}{1 - D_E}. \quad (7.22)$$

Then $\mathcal{R}_p < 1$ if $p > p_0$, and $B^2 - 4AC > (=, <) 0$ if $\mathcal{R}_0 > (=, <) \mathcal{R}_p$, in which case Eq. (7.20) has two (one, none) positive solutions x_{\pm} (see [9] for more detailed proofs). This establishes the following result:

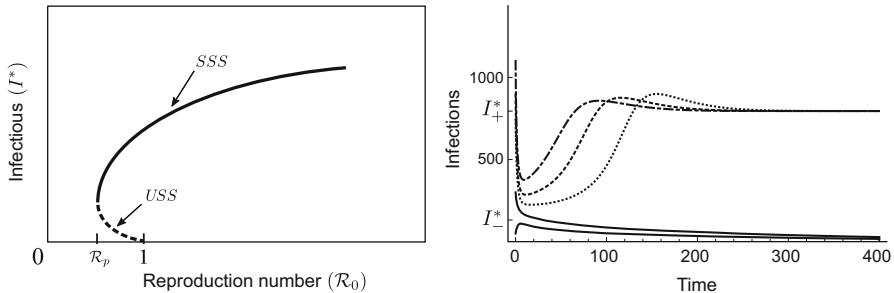


Fig. 7.10 The left plot is a depiction of the backward bifurcation for the model with reinfection (7.18). The I^* component of an equilibria is plotted as a function of \mathcal{R}_0 . The curve demonstrates that there are two positive steady states for $\mathcal{R}_p < \mathcal{R}_0 < 1$. The solid part of the curve corresponds to the stable steady state (SSS) and the dashed part corresponds to the unstable steady state (USS). The right figure shows the numerical solutions of system (7.18) for parameter values in the region where the backward bifurcation occurs ($\mathcal{R}_p < \mathcal{R}_0 < 1$ and $p > p_0$). The number of infectious individuals $I(t)$ is plotted. It illustrates that, depending on the initial conditions, the solution will converge to either the infection-free equilibrium (the solid curves) or the stable positive equilibrium (dashed curves)

Theorem 7.3 Let \mathcal{R}_0 , \mathcal{R}_p , and p_0 be as defined in (7.19), (7.21), and (7.22).

- (a) If $\mathcal{R}_0 > 1$, then system (7.18) has exactly one endemic equilibrium and it is locally asymptotically stable
- (b) If $\mathcal{R}_0 < 1$, then the disease-free equilibrium is locally asymptotically stable. Moreover,
 - (i) for each $p > p_0$ there exists a positive constant $\mathcal{R}_p < 1$ such that system (7.18) has exactly two (one, none) positive equilibria when $\mathcal{R}_0 > (=, <) \mathcal{R}_p$. In the case of two positive equilibria, the one with large (smaller) I^* component is stable (unstable), as depicted in Fig. 7.10 (left).
 - (ii) for $p = (<) p_0$, system (7.18) has exactly one (no) positive equilibrium.

Numerical simulations of the model (7.18) confirm the backward bifurcation. Figure 7.10 (right) shows the I component of the solutions with different initial values in the case $p > p_0$ and $\mathcal{R}_p < \mathcal{R}_0 < 1$. We observe that solutions with initial I values near 0 converge to the disease-free steady state (see the solid curves), whereas other solutions converge to the endemic steady state with a larger I value (see the dashed curves).

7.6 Other TB Models with More Complexities

The models considered above assume homogeneity in several aspects including mixing and age structure of the population. In many cases, the problems under investigation require the consideration of some of the heterogeneities. For example,

in [1] TB models involving non-homogeneous mixing that incorporates “household” contacts and age-dependent mixing are considered to assess the possible causes for the observed historical decline of tuberculosis notifications. In [5], an age-structured TB model is used to study optimal age-dependent vaccination strategies $\psi(a)$, where a denotes the chronological age of an individual. One of the optimal control problems is to minimize the reproduction number corresponding to $\psi(a)$, \mathcal{R}_ψ , under the constraint on the cost $C(\psi) < C^*$, where C^* is a fixed constant. Results suggest that the optimal vaccination strategy is either a one-age strategy (vaccinate every one at a single age A determined by the parameter and parameter functions in the model) or two-age strategy (vaccinate the population at two fixed ages A_1 and A_2 determined by model parameters).

Models with multiple resistant strains have been used to answer various questions associated with the establishment and spread of drug-resistant TB. Blower and Chou [2] use a model with three strains of resistant TB to study how to effectively control MDR-TB in “hot zones” (i.e., areas that have $> 5\%$ prevalence of MDR-TB). The findings of the model suggest that the levels of MDR are driven by case-finding rates, cure rates, and amplification probabilities (the probability that a case will develop further resistance during treatment). In [7], heterogeneity in the relative fitness of MDR strains is incorporated in a TB model. The model includes two resistant strains (one is more fit than the other), as well as a drug-sensitive strain, and is used to study the impact of initial fitness estimates on the emergence of MDR-TB. The model results show that “even when the average relative fitness of MDR strains is low and a well-functioning control program is in place, a small sub-population of a relatively fit MDR strain may eventually outcompete both the drug-sensitive strains and the less fit MDR strains.”

A two-strain TB model with reinfection is considered in [21] to study the role of reinfection in the transmission dynamics of drug-resistant TB and the coexistence of sensitive and resistant strains of TB. In [17] multiple sub-populations are considered with one sub-population being genetically susceptible to TB. Different strategies involving treatment of latent TB infections and active TB disease are examined and in different populations are examined. Results from the model analysis suggest that the presence of a genetically susceptible sub-population dramatically alters effects of treatment. To study the impact of treatment failure and its influence in the criteria for the control of drug-resistant TB, a model with multiple stages of treatment failure with different probabilities of leading to resistant TB is considered in [8]. Model results indicate that case detection and treatment can be a critical factor in the control of MDR-TB.

To explore the influence of HIV on TB prevalence, models including the interaction between TB and HIV can be used, in which case the model analysis can be very challenging due to the possible coinfections of TB and HIV. For example, a model incorporating both TB and HIV with multiple stages of HIV is studied in [20]. The model results suggest that “an HIV epidemic can significantly increase the

frequency and severity of tuberculosis outbreaks, but that this amplification effect of HIV on tuberculosis outbreaks is very sensitive to the tuberculosis treatment rate.” Another model that includes both TB and HIV is studied in [22]. The model results suggest that the accelerated progression from LTBI to active TB in individuals co-infected with HIV can have a significant influence in TB prevalence.

7.7 Project: Some Calculations for the Two-Strain Model

Exercise 1 Derive the formulas of the reproduction numbers \mathcal{R}_S and \mathcal{R}_R given in (7.5) and (7.6), respectively, for the two-strain TB model (7.4) by considering the stability of the disease-free equilibrium

$$E_0 = (S^{(0)}, L_1^{(0)}, I_1^{(0)}, L_2^{(0)}, I_2^{(0)}, T^{(0)}) = (N, 0, 0, 0, 0, 0).$$

The Jacobian at E_0 is

$$J(E_0) = \begin{bmatrix} -\mu & 0 & -c_1 & 0 & -c_2 & 0 \\ 0 & -(\mu + \kappa_1 + r_1) & c_1 + pr_2 & 0 & 0 & 0 \\ 0 & \kappa_1 & -(\mu + r_2) & 0 & 0 & 0 \\ 0 & 0 & qr_2 & -(\mu + \kappa_2) & c_2 & 0 \\ 0 & 0 & 0 & \kappa_2 & -\mu & 0 \\ 0 & r_1 & (p + q)r_2 & 0 & 0 & -\mu \end{bmatrix}.$$

Clearly, $J(E_0)$ has the negative eigenvalues μ with multiplicity 2, and other eigenvalues are given by the eigenvalues of J_1 and J_2 , where

$$J_1 = \begin{bmatrix} -(\mu + \kappa_1 + r_1) & c_1 + pr_2 \\ \kappa_1 & -(\mu + r_2) \end{bmatrix}, \quad J_2 = \begin{bmatrix} -(\mu + \kappa_2) & c_2 \\ \kappa_2 & -\mu \end{bmatrix}.$$

Show that J_1 has negative eigenvalues if and only if $\mathcal{R}_S < 1$ and J_2 has negative eigenvalues if and only if $\mathcal{R}_R < 1$.

Exercise 2 Consider the two-strain TB model (7.4) and conduct numerical simulations of the system for parameter values specified below. Given the parameter values $\mu = 0.0143$ (or $1/\mu = 70$), $c_1 = 13$, $\kappa_1 = \kappa_2 = 1$, $r_1 = 1$, $r_2 = 2$, $q = 0$, $p = 0.5$. Choose the values of c_2 such that

- (a) E_1 is stable;
- (b) E_2 is stable;
- (c) E^* is stable.

Exercise 3 Similar to Exercise 2 but for the case when $q > 0$. Given the same values for other parameters as in Exercise 2, and choose the values of c_2 such that

- (a) E_2 is stable;
- (b) E^* is stable.

7.8 Project: Refinements of the One-Strain Model

The model (7.1) is a one-strain model for TB, which assumes that infected people stay in the latent stage L before entering the infective stage I . The model (7.15) includes fast and slow progression (represented by p and $1 - p$, respectively), which assumes that with fractions $1 - p$ and p infected individuals will enter the latent L and infectious I stages, respectively. In both models, the stage duration is assumed to be exponentially distributed with parameter k . That is, the probability that an infected individual has not become infective s units of time after infection is e^{-ks} . This assumption may not be appropriate for TB due to the long and variable latency, as shown in Figs. 7.1 and 7.2. To examine whether or not a model with a more realistic assumption on the distribution of latent period, we can consider a model in which the exponential survival function e^{-ks} is replaced by a general distribution $p(s)$, as described below.

Let $p(s)$ be a function representing the proportion of those individuals latent at time t and who, if alive, are still infected (but not infectious) at time $t + s$. Then $-\dot{p}(\tau)$ is the rate of removal of individuals from E class into I class τ units of time after becoming latent. Assume that

$$p(s) \geq 0, \quad \dot{p}(s) \leq 0, \quad p(0) = 1, \quad \int_0^\infty p(s)ds < \infty.$$

Let $S(t)$, $E(t)$, $I(t)$, and $T(t)$ denote the number of individuals in the susceptible, latent, infectious, and treated classes, respectively. Consider the following model with $p(s)$ being the arbitrary distribution of the latent stage:

$$\begin{aligned} S' &= \Lambda - cS\frac{I}{N} - \mu S, \\ E(t) &= E_0(t) + \int_0^t [cS(s) + c^*T(s)] \frac{I(s)}{N(s)} p(t-s) e^{-(\mu+r_1)(t-s)} ds, \\ I(t) &= \int_0^t \int_0^\tau [cS(s) + c^*T(s)] \frac{I(s)}{N(s)} e^{-(\mu+r_1)(\tau-s)} \\ &\quad \times [-\dot{p}(\tau-s) e^{-(\mu+r_2)(t-\tau)}] ds d\tau + I_0 e^{-(\mu+r_2)t} + I_0(t), \\ T' &= r_1 E + r_2 I - c^*T \frac{I}{N} - \mu T, \\ N &= S + E + I, \end{aligned} \tag{7.23}$$

where $E_0(t)$ denotes those individuals in E class at time $t = 0$ and still in the latent class, $I_0(t)$ denotes those initially in class E who have moved into class I and are still alive at time t , and $I_0 e^{-(\mu+r_2)t}$ with $I_0 = I(0)$ represents those who are infective at time 0 and are still alive and in I class. $E_0(t)$ and $I_0(t)$ are assumed to have compact support (that is, they vanish for large enough t). All other parameters have the same meanings as in model (7.1).

Question 1 Derive the formula below for the reproduction number \mathcal{R}_0

$$\mathcal{R}_0 = c \int_0^\infty a(\tau) d\tau, \quad (7.24)$$

where $a(u)$ is defined by

$$a(t-s) = \int_s^t e^{-(\mu+r_1)(\tau-s)} [-\dot{p}(\tau-s)e^{-(\mu+r_2)(t-\tau)}] d\tau.$$

Provide biological interpretations for the $a(u)$ and the factors in the expression of $a(u)$.

Question 2 Show that the disease will die out if $\mathcal{R}_0 < 1$ and persist if $\mathcal{R}_0 > 1$.

Question 3 Find all biologically feasible steady-state solutions of the model (7.23). Determine the condition for the existence of a coexistence steady state (i.e., $I > 0$ and $J > 0$).

Question 4 For each steady state, identify the conditions under which the steady state is locally asymptotically stable.

Question 5 Does the model with a more realistic distribution for the latent stage provide new qualitative behavior of the disease dynamics in comparison with the ODE model (7.1)? Identify the additional insights that the model (7.23) can provide.

7.9 Project: Refinements of the Two-Strain Model

Model (7.4) is a two-strain model for TB with drug-sensitive and drug-resistant strains. In this model, the latent stages for both strain are assumed to be exponentially distributed with parameters κ_1 and κ_2 . The long and variable latency for the sensitive strain make this assumption unrealistic (the latent period for the resistant strain is much shorter). The model below is an alternative two-strain model with distributed delay in latency, in which $p(\theta)$ is used to describe the progression from latent stage to infectious stage and θ is the age of infection (time since becoming infected):

$$\begin{aligned} \frac{dS}{dt} &= \mu N - (\mu + \lambda_1(t) + \lambda_2(t)) S + (1 - r)\chi \int_0^\infty p(\theta)i(\theta, t)d\theta, \\ \frac{\partial i}{\partial t} + \frac{\partial i}{\partial \theta} + \mu i(\theta, t) + (1 - r + qr)\chi p(\theta)i(\theta, t) &= 0, \\ \frac{dJ}{dt} &= \lambda_2(t)S - \mu J + qr\chi \int_0^\infty p(\theta)i(\theta, t)d\theta, \\ i(0, t) &= \lambda_1(t)S(t), \quad S(0) = S_0 > 0, \quad i(\theta, 0) = i_0(\theta) \geq 0, \quad J(0) = J_0 > 0. \end{aligned} \quad (7.25)$$

$S(t)$ is the number of susceptibles at time t ; $i(\theta, t)$ denotes the density of infection age θ individuals with the drug-sensitive strain at time t ; $J(t)$ is the number of infected individuals with a drug-resistant strain at time t ; and $N = S + I + J$, where

$$I(t) = \int_0^\infty i(\theta, t)d\theta$$

is the total number of individuals infected with the sensitive strain. The function $p(\theta)$ ($0 \leq p(\theta) \leq 1$) is assumed constant in time and is based on experimental data (e.g., the distribution in Fig. 7.2). Note that $p(\theta)i(\theta, t)$ represents the age density of infectious individuals. The transmission rates of sensitive and resistant strains are

$$\lambda_1(t) = \frac{c_1}{N(t)} \int_0^\infty p(\theta)i(\theta, t)d\theta \quad \text{and} \quad \lambda_2(t) = c_2 \frac{J(t)}{N(t)}, \quad (7.26)$$

respectively, with c_1 and c_2 being the per-capita transmission rates of the sensitive and resistant strains. The rate at which sensitive-strain-infected individuals leave the i class due to treatment is $(1 - r + qr)\chi p(\theta)$, where χ denotes the treatment rate for individuals with drug-sensitive TB. The factor $(1 - r + qr)$ introduces the effect of incomplete treatment: a fraction r of the treated individuals with sensitive TB do not recover due to incomplete treatment, and the remaining fraction $1 - r$ is actually cured and becomes susceptible again. Moreover, among the individuals who do not finish their treatment, a fraction q of them will develop drug-resistant TB and the remaining fraction will remain latent. The per-capita birth and natural death rates are assumed to be the same and equal to μ . In this model, for simplicity, treated individuals are assumed to have the same susceptibility.

Question 1 Let $v(t) = i(0, t) = \lambda_1(t)S(t)$. Show that system (7.25) is equivalent to the following system, which is easier to analyze:

$$\begin{aligned}
v(t) &= \frac{N(t) - J(t) - \int_0^t K_0(\theta)v(t-\theta)d\theta}{N(t)} \int_0^t K_1(\theta)v(t-\theta)d\theta + \tilde{F}_1(t), \\
\frac{dJ}{dt} &= \beta_2 \left(N(t) - J(t) - \int_0^t K_0(\theta)v(t-\theta)d\theta \right) \frac{J(t)}{N(t)} \\
&\quad - mJ(t) + \int_0^t K_2(\theta)v(t-\theta)d\theta + \tilde{F}_2(t), \\
\frac{dN}{dt} &= b(N)N(t) - \mu N(t) - \delta J(t),
\end{aligned} \tag{7.27}$$

where $\tilde{F}_i(t)$ involve parameters and initial condition with $\lim_{t \rightarrow \infty} \tilde{F}_i(t) = 0$, $i = 1, 2$, and

$$\begin{aligned}
K_0(\theta) &= e^{-\mu\theta - \int_0^\theta (1-r+qr)\chi p(s)ds}, \\
K_1(\theta) &= \beta_1 p(\theta)K_0(\theta) = -\frac{\beta_1}{\chi(1-r+qr)} \left(\frac{d}{d\theta} K_0(\theta) + \mu K_0(\theta) \right), \\
K_2(\theta) &= qr\chi p(\theta)K_0(\theta) = -\frac{qr}{(1-r+qr)} \left(\frac{d}{d\theta} K_0(\theta) + \mu K_0(\theta) \right).
\end{aligned} \tag{7.28}$$

Question 2 Let \mathcal{R}_1 and \mathcal{R}_2 denote the reproduction numbers for the sensitive and resistant strains. Derive a formula for \mathcal{R}_1 and \mathcal{R}_2 .

Question 3 Explore how the existence and stability of steady states of the model (7.25) depend on \mathcal{R}_1 and \mathcal{R}_2 .

Question 4 Does the model with a more realistic distribution for the latent stage for the sensitive strain provide new qualitative behavior of the disease dynamics in comparison with the ODEs model (7.4)? Identify the additional insights that the model (7.25) may provide.

References

1. Aparicio, J. P., and C. Castillo-Chavez (2009) Mathematical modelling of tuberculosis epidemics, *Math. Biosc. Eng.* **6**: 209–237.
2. Blower, S. M., and T. Chou (2004) Modeling the emergence of the ‘hot zones’: tuberculosis and the amplification dynamics of drug resistance, *Nature Medicine* **10**: 1111–1116.
3. Blower, S. M., P.M. Small, and P.C. Hopewell (1996) Control strategies for tuberculosis epidemics: new models for old problems, *Science* **273**: 497–500.
4. Castillo-Chavez, C., and Z. Feng (1997) To treat or not to treat: the case of tuberculosis. *J. Math. Biol.* **35**: 629–656.
5. Castillo-Chavez, C., and Z. Feng (1998) Global stability of an age-structure model for TB and its applications to optimal vaccination strategies, *Math.Biosc.* **151**: 135–154.
6. CDC, Tuberculosis treatment (2014) <http://www.cdc.gov/tb/topic/treatment/>.

7. Cohen, T., and M. Murray (2004) Modeling epidemics of multidrug-resistant *M. tuberculosis* of heterogeneous fitness *Nature Medicine* **10**: 1117–1121.
8. Dye, C., and B.G. Williams (2000) Criteria for the control of drug-resistant tuberculosis, *Proc. Nat. Acad. Sci.* **97**: 8180–8185.
9. Feng, Z., C. Castillo-Chavez, and A.F. Capurro (2000) A model for tuberculosis with exogenous reinfection, *Theor. Pop. Biol.* **57**: 235–247.
10. Feng, Z., W. Huang, and C. Castillo-Chavez (2001) On the role of variable latent periods in mathematical models for tuberculosis, *J. Dyn. Diff. Eq.* **13**: 435–452.
11. Feng, Z., M. Iannelli, and F.A. Milner (2002) A two-strain tuberculosis model with age of infection, *SIAM J. Appl. Math.* **62**: 1634–1656.
12. Fister, K. R., S. Lenhart, and J.S. McNally (1998) Optimizing chemotherapy in an HIV model, *Electronic J. Diff. Eq.* **32**: 1–12.
13. Fleming, W., and R. Rishel (1975) *Deterministic and Stochastic Optimal Control*, Springer.
14. Jung, E., S. Lenhart, and Z. Feng (2002) Optimal control of treatments in a two-strain tuberculosis model, *Discrete and Continuous Dynamical Systems Series B* **2**: 473–482.
15. Kirschner, D., S. Lenhart, and S. Serbin (1997) Optimal control of the chemotherapy of HIV *J. Math. Biol.* **35**: 775–792.
16. Levin, S. and D. Pimentel (1981) Selection of intermediate rates of increase in parasite-host systems, *Am. Naturalist* **117**: 308–315.
17. Murphy, B. M., B.H. Singer, and D. Kirschner (2003) On treatment of tuberculosis in heterogeneous populations, *J. Theor. Biol.* **223**: 391–404.
18. Nowak, M.A., and R.M. May (1994) Superinfection and the evolution of parasite virulence, *Proc. Roy. Soc. London, Series B: Biological Sciences* **255**: 81–89.
19. Pontryagin, L.S. (1987) *Mathematical Theory of Optimal Processes*, CRC Press.
20. Porco, T.C., P.M. Small, and S.M. Blower (2001) Amplification dynamics: predicting the effect of HIV on tuberculosis outbreaks, *J. Acquired Immune Deficiency Syndromes* **28**: 437–444.
21. Rodrigues, P., M.G.M. Gomes, and C. Rebelo (2007) Drug resistance in tuberculosis, a reinfection model, *Theor. Pop. Biol.* **71**: 196–212.
22. Roeger, L-I.W., Z. Feng, and C. Castillo-Chavez (2009) Modeling TB and HIV co-infections *Math. Biosc. Eng.* **6**: 815–837.
23. Singer, B.H., and D.E. Kirschner (2004) Influence of backward bifurcation on interpretation of \mathcal{R}_0 in a model of epidemic tuberculosis with reinfection, *Math. Biosc. Eng.* **1**: 81–93.
24. Styblo, K. (1991) Selected papers. vol. 24, *Epidemiology of tuberculosis*, Hague, The Netherlands: Royal Netherlands Tuberculosis Association.
25. WHO. Global tuberculosis report 2013 http://www.who.int/tb/publications/global_report/.

Chapter 8

Models for HIV/AIDS



8.1 Introduction

Acquired immunodeficiency syndrome (AIDS) was first identified as a new disease in the homosexual community in San Francisco in 1981. The human immunodeficiency virus (HIV) was identified as the causative agent for AIDS in 1983. The disease has several very unusual aspects. After the initial infection, there are symptoms, including headaches and fever for 2 or 3 weeks. Transmissibility is high for about 2 months, and then there is a very long latent period during which transmissibility is low. At the end of this latent period, which may last 10 years, transmissibility rises, signaling the development of full-blown AIDS. In the absence of treatment, AIDS is invariably fatal. Now, HIV can be treated with a combination of highly active antiretroviral therapy (HAART) drugs, which both reduce the symptoms and prolong the period of low infectivity. While there is still no cure for AIDS, treatment has made it no longer a necessarily fatal disease. To describe the variation of infectivity for HIV, one possibility would be to use a staged progression model, with multiple infective stages having different infectivity. Another possibility would be to use an age of infection model.

HIV is transmitted in many ways, the most common of which are sexual contact, either heterosexual or homosexual, shared drug injection needles, and contaminated blood transfusions. Vertical transmission from mother to child is also possible. In the past, transfusions of contaminated blood were another source of disease transmission, but in developed countries screening of blood since 1985 has eliminated blood transfusions as a transmission mode.

A full model for HIV/AIDS should include a variety of transmission modes, and might take into account of many factors including the level of sexual activity, drug use, condom use, and the sexual contact network, resulting in large scale systems with many parameters that need to be estimated from data. Models were developed first for homosexual transmission. In this chapter, we will consider not only models for disease transmission in a homosexual community (the current terminology is

men having sex with men, or MSM), but also models that include heterosexual transmission through female sex workers. We also consider modes that include the joint disease dynamics of HIV and TB and the synergy between HIV and HSV-2.

The identification of the human immunodeficiency virus [11, 55, 56, 58, 107] captured the attention of theoreticians and modelers as AIDS became one of the most feared diseases nearly three decades ago. Most of the initial modeling contributions focused on the study of the transmission dynamics of HIV at the population level since little was known about the epidemiology of HIV and, as expected, modeling was carried out first under simple settings and crude assumptions [3–7, 10, 19, 26, 32–35, 46, 48, 49, 57, 59, 65, 67–71, 75, 76, 79, 81, 88, 95–97, 102, 103]. An overview of the “state of the art” on the transmission dynamics of HIV modeling in the 1980’s is found in [30], the review papers [97, 99], or in the books [8, 30, 63].

The modeling studies in [32–35, 67, 102, 103] focused on the impact that changes in the pool of susceptibles, disease-induced mortality, heterogeneous mixing, vertical transmission, asymptomatic carriers, variable infectivity, and incubation (or latent) and infective periods may have on the dynamics of sexually transmitted HIV. Efforts to model the risk of infection from sexual partner selection or from within and between group mixing became central to the research of various groups studying HIV dynamics. Other studies focused on the role of gender, core populations, and heterogenous mixing contact rates on HIV dynamics. These were naturally involved in the development of sexual-behavior surveys and data collection on sexual and “dating” activity, as well as on the mathematical modeling and analysis of heterogenous “mixing” frameworks (see [20, 21, 23, 24, 27, 28, 36–39, 41, 47, 78, 93]). The overview in [83] highlights the potential role of sexual activity and drinking on the dynamics of STDs [47, 65, 66, 93] and while the adaptive dynamics generated by changing behaviors in response, to a multitude of factors, were rarely explored, some earlier attempts were also carried out as a result of the HIV pandemic [22, 60].

As described in these historical papers [3–5], knowledge of these periods was quickly identified as critical to the initial efforts to predict the dynamics of HIV. In [35] it is observed that: “The duration of the latent period is thought to be a few days to a few weeks [3–5], and while the duration of the infectious period is not yet known, those individuals that develop full-blown AIDS have an average incubation period estimated variously at 35–47 months [88], 66 months [3], and as high as 96 months [81].” This estimate is continually being revised as information and experience accumulate. However, even the most conservative estimate suggests that it may be reasonable to approximate the infectious period by the incubation period; that is, to assume a negligible latent period. Pickering et al. [88] stress that the ability to transmit HIV is not constant, as individuals are most infectious 3–16 months following exposure, and recent studies [58, 77, 94] report the existence of two peaks of infectiousness, one taking place a few weeks after exposure and the other before the onset of “full-blown” AIDS.

In the context of the dynamics of a homosexually-active homogeneously-mixing population, the reproduction number is given by $\mathcal{R}_0 = \lambda C(T)D$, where λ denotes the probability of transmission per partner, $C(T)$ the mean number of sexual

partners an average individual has per unit time when the population density is T , and D the death-adjusted mean infective period (see [35]). Since HIV is a slow disease, if $\mathcal{R}_0 \leq 1$ it will die out while if $\mathcal{R}_0 \geq 1$ it will persist in the presence of a small number of infected/infective individuals. The mathematical analysis and numerical simulations in [35] suggest that whenever the incubation period distribution is exponential the reproduction number \mathcal{R}_0 is a global bifurcation parameter (transcritical bifurcation), that is, as \mathcal{R}_0 crosses 1 a global transfer of stability from the disease-free state to the endemic equilibrium takes place, and vice versa. Local results do not depend on the distribution of times spent in the infective categories (the survivorship functions). Keeping a suite of parameters fixed [35] allowed for the comparison of the exponential incubation period distribution versus a piecewise constant survivorship (individuals remain infective for a fixed length of time). It was found that for “...some realistic parameters we can see (at least in these cases) that the reproduction numbers corresponding to these two extreme cases do not differ by more than 18% whenever the two distributions have the same mean [35].”

The inclusion of heterogeneity via the introduction of a large number of subgroups limited the forecasting capability of these models due to factors that included increased levels of uncertainty (more parameters). The use of multi-group models raised the expected modeling and parameter estimation challenges [20, 21, 23, 24, 27, 28, 36–38, 41, 65, 66, 93]. In addition, the analyses of some of these models generated novel dynamic behavior, questioning, possibly for the first time in epidemiology, the centrality of the role of the basic reproduction number in the identification and development of control, or education, or intervention measures. For example, the natural asymmetry present in disease transmission as a result of prevalent alternative modes of sexual engagement proved to be capable of giving rise to the existence of multiple equilibria [33, 34, 67]; an unexpected outcome at that time.

8.2 A Model with Exponential Waiting Times

A single homosexually-active population is divided into three classes. S denoting the number of susceptible individuals, I infective individuals, and A former I -individuals who have developed full-blown AIDS (see Fig. 8.9). We assume that all HIV-infected individuals will eventually develop full-blown AIDS (unless they die first from other causes). This, unfortunately, may be the most realistic as evidence accumulates that AIDS is a progressive disease. Later, we will suggest a project to develop a model under the assumption that some fraction of infected individuals will escape progression to full-blown AIDS. Originally, a latent class (i.e., those exposed individuals that are not yet infectious) was not included because it was believed then that the time spent in that class is short. It is further assumed that individuals who develop full-blown AIDS are no longer actively infective, that is, that they have no sexual contacts; it is also assumed that infected individuals become

infective immediately. Finally, it is assumed that infective individuals acquire AIDS at the constant rate α_I per unit time and become sexually inactive at the constant rate α per unit time. Therefore, $1/(\mu + \alpha_I)$ gives the mean incubation period and $1/(\mu + \alpha)$ gives the mean sexual life expectancy.

The introduction of the model requires additional definitions. Λ will denote the constant recruitment rate into the susceptible class (individuals who are sexually active); μ the constant per-capita natural mortality rate; d the per-capita constant disease-induced mortality due to AIDS. The function $C(T)$ models the mean number of sexual partners an average individual has per unit time when the population density is T ; λ (a constant) denotes the average sexual risk per infected partner; λ is often thought as the product $i\phi$ [68], where ϕ is the average number of contacts per sexual partner and i the conditional probability of infection from a sexual contact when the latter is infected. Kingsley et al. [72] had presented (not surprising) evidence that the probability of seroconversion (infection) increases with the number of infected sexual partners. Hence, $\lambda C(T)$ models the transmission rate per unit time per infected partner when the size of the sexually active population is T . Using the modeling framework published in [3, 4] with the help of Fig. 8.1, we arrive at the following epidemiological model [35] for sexually transmitted HIV under the assumption of exponential waiting times in the infection classes.

$$\begin{aligned}\frac{dS(t)}{dt} &= \Lambda - \lambda C(T(t)) \frac{S(t)I(t)}{T(t)} - \mu S(t) \\ \frac{dI(t)}{dt} &= \lambda C(T(t)) \frac{S(t)I(t)}{T(t)} - (\alpha_I + \mu)I(t) \\ \frac{dA(t)}{dt} &= \alpha_I I(t) - (\alpha + \mu)A(t)\end{aligned}\quad (8.1)$$

where

$$T = I + S. \quad (8.2)$$

The fraction I/T can be thought of as representing the fraction of contacts that a susceptible individual has with a randomly selected infective individual. Here $\lambda C(T)SI/T$ denotes the number of newly-infected individuals per unit time since

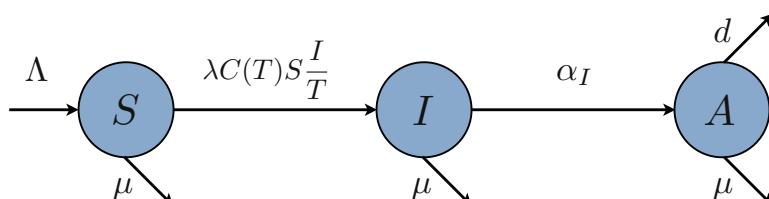


Fig. 8.1 Flow diagram: single group model in the case when all infected people will progress to AIDS

individuals in classes A are sexually inactive. A plausible assumption for modeling $C(T)$ would be to assume that it is approximately linear for small T approaching a saturation level for large values of T [62]. Here, it is assumed that $C(T)$ is a differentiable and increasing function of T (except when noted). Anderson et. al. [4] observe that $C(T)$, the mean number of sexual partners per unit time, underestimates the importance of highly active individuals and that consequently, modifications should be made to this framework in order to properly account for their role.

The analysis of the system (8.1) found in [35] makes the following assumptions concerning $C(T)$:

$$C(T) > 0, \quad C'(T) \geq 0, \quad (8.3)$$

with prime denoting the derivative with respect to T . The dynamics of S and I are independent of A (by construction). The system is well-posed, that is, if $S(0) \geq 0$, $I(0) \geq 0$, $A(0) \geq 0$ then a unique solution exists with $S(t) \geq 0$, $I(t) \geq 0$, $A(t) \geq 0$ for $t \geq 0$.

As it is the case with most of the epidemiological systems addressed in this book, system (8.1) always has the disease-free equilibrium given by

$$(S, I, A) = \left(\frac{\Lambda}{\mu}, 0, 0 \right), \quad (8.4)$$

and under certain assumptions it also supports a unique endemic equilibrium.

The stability of the disease-free equilibrium (8.4) is determined by the non-dimensional ratio

$$\mathcal{R}_0 = \lambda \left(\frac{1}{\sigma_I} \right) C \left(\frac{\Lambda}{\mu} \right), \quad (8.5)$$

that is, by the *basic reproduction number*. In the definition of \mathcal{R}_0 , $\sigma_I = \alpha_I + \mu$, and \mathcal{R}_0 denotes the number of secondary infections generated by a single infective individual in a population of susceptibles at a demographic steady state. \mathcal{R}_0 is given by the product of the three factors (epidemiological parameters): λ (the probability of transmission per partner), $C(\Lambda/\mu)$ (the mean number of sexual partners that an average susceptible individual has per unit time when everybody in the population is susceptible), and

$$D = \left(\frac{1}{\sigma_I} \right). \quad (8.6)$$

The death-adjusted mean infective period is $D = D_I$ with D_I denoting the death-adjusted mean infectious period $1/\sigma_I$ of the I class. The use of the dimensionless ratio, $\mathcal{R}_0 = \lambda C(\Lambda/\mu)D$ leads to the following result [35]:

Theorem 8.1 *If $\mathcal{R}_0 < 1$ then the equilibrium $(\Lambda/\mu, 0, 0)$ of the system (8.1) is globally asymptotically stable.*

That is, every solution of (8.1) $(S(t), I(t), A(t))$ with $S(0) \geq 0$, $I(0) \geq 0$, $A(0) \geq 0$ tends to $(\Lambda/\mu, 0, 0)$ as $t \rightarrow +\infty$. That is, the condition $\mathcal{R}_0 \leq 1$ is sufficient to guarantee that the disease will eventually die out in this population.

An endemic equilibrium (S^*, I^*, A^*) of (8.1) satisfies

$$\Lambda = \left[\frac{\Lambda - \mu S^*}{\sigma_I} - \mu \right] S^*, \quad I^* = \frac{\Lambda - \mu S^*}{\alpha_I + \mu}, \quad A^* = \frac{\alpha_I}{\alpha_I + \mu} I^*.$$

In [35] it has also been established (following some of the same arguments used in other chapters) that:

Theorem 8.2 *If $\mathcal{R}_0 > 1$, there is a unique endemic equilibrium (S^*, I^*, A^*) , which is locally asymptotically stable, and the disease-free state $(\Lambda/\mu, 0, 0)$ is unstable.*

We can summarize the situation (full details of all proofs are in [35]) as follows: The disease-free state of system (8.1) is globally asymptotically stable when $\mathcal{R}_0 > 1$ and unstable if $\mathcal{R}_0 < 1$. When $\mathcal{R}_0 > 1$, this system has a unique locally asymptotically stable endemic equilibrium. There is a transfer of stability to the endemic state as \mathcal{R}_0 crosses one. Further, when $\mathcal{R}_0 > 1$ it was shown, as one would expect, that the endemic equilibrium is also globally asymptotically stable.

The reproduction number (\mathcal{R}_0) increases proportionately with the transmission probability and the average number of sexual partners; it may also increase in proportion to the rate of recruitment of individuals to the susceptible class through $C(T)$. \mathcal{R}_0 is an increasing function of the mean infective period D and may be a decreasing function of the mortality rate (depending on the functional expression for $C(T)$).

8.3 An HIV Model with Arbitrary Incubation Period Distributions

The use of exponential waiting distributions in the I class corresponds to the requirement that the per-capita removal rate from the I class (by the development of full-blown AIDS symptoms) into the A class is constant. It would be clearly an improvement in the model of Sect. 8.2 if we were to move from constant to variable removal rates and this is what we do in this section (the ideas follow those in [19, 35]). Hence, it is still assumed that individuals become immediately infective (that is, we continue to neglect the latent period) and continue to divide the at risk population into the three classes: S , I , and A . The parameters $\lambda = i\phi$, Λ , μ , d , and p have the same meaning as in Sect. 8.2; however, the removal rates are modified through the introduction of the function $P_I(s)$ representing the proportion of individuals who become I -infective at time t and that, if alive, are still infective

at time $t + s$ (survive as infective). The survivorship function P_I is non-negative and non-increasing, and $P_I(0) = 1$. It is further assumed that

$$\int_0^\infty P_I(s)ds < \infty,$$

and thus, $-\dot{P}_I(x)$ is the rate of removal of individuals from the class I into the class A , x time units after infection.

The number of new infections occurring at time x is $\lambda C(T(x))S(x)I(x)/T(x)$ where we have kept the meaning of $C(T)$, I , and T as in Sect. 8.2. The rate of change in the susceptible class is given now by the expression:

$$\frac{dS(t)}{dt} = A - \lambda C(T(t))S(t)\frac{I(t)}{T(t)} - \mu S(t), \quad (8.7)$$

with

$$\int_0^t \lambda C(T(x))S(x)\frac{I(x)}{T(x)}e^{-\mu(t-x)}P_I(t-x)dx$$

representing the number of individuals who have been infected from time 0 to t and are still in class I . The discount factor $\exp(-\mu(t-x))$ takes into account removals due to deaths by natural causes (not HIV). Hence, if $I_0(t)$ denotes individuals in class I at time $t = 0$ that are still infective at time t then the total number of infectives at time t is given by

$$I(t) = I_0(t) + \int_0^t \lambda C(T(x))S(x)\frac{I(x)}{T(x)}e^{-\mu(t-x)}P_I(t-x)dx, \quad (8.8)$$

where $I_0(t)$ is assumed (for biological and mathematical reasons) to have compact support (vanishing for large enough t).

The expression for $A(t)$ turns out to be the sum of three terms: $A_0e^{-(\mu+d)t}$, where $A_0 = A(0)$, representing individuals who had full-blown AIDS at time zero and are still alive; $A_0(t)$ representing individuals initially in class I who moved into class A and are still alive at time t ; and those who joined the I class after time $t = 0$ (see below). We assume that $A_0(t)$ approaches zero as t approaches infinity. The term representing infected individuals “born” after time $t = 0$ is given by

$$\int_0^\tau \left\{ \int_0^t \lambda C(T(x))S(x)\frac{I(x)}{T(x)}e^{-\mu(\tau-x)}[-\dot{P}_I(\tau-x)e^{-(\mu+d)(t-\tau)}]dx \right\} d\tau,$$

where $-\dot{P}_I(\tau-x)$, denotes the rate of removal from the class I at time τ or $(\tau-x)$ units after infection. Therefore

$$\begin{aligned} A(t) = & p \int_0^\tau \left\{ \int_0^t \lambda C(T(x))S(x)\frac{I(x)}{T(x)}e^{-\mu(t-x)}[-\dot{P}_A(\tau-x)e^{-(t-\tau)}]dx \right\} d\tau \\ & + A_0e^{-(\mu+d)t} + A_0(t). \end{aligned} \quad (8.9)$$

The model given by equations (8.7) is a system of nonlinear integral equations. The standard results on well-posedness for these systems, as found in [82] guarantee the existence and uniqueness of solutions and their continuous dependence on parameters. The proof of positivity is given in [32].

The basic reproduction number \mathcal{R}_0 of the system (8.7) is given by

$$\mathcal{R}_0 = \lambda C \left(\frac{\Lambda}{\mu} \right) \int_0^\infty P_I(x) e^{-\mu x} dx, \quad (8.10)$$

where

$$\int_0^\infty P_I(x) e^{-\mu x} dx,$$

is the death-adjusted mean infective period, D . If $P_I(x) = e^{-\alpha_I x}$ then (8.10) reduces to (8.5). We also observe that as before

$$D_I = \int_0^\infty P_I(s) e^{-\mu s} ds$$

denotes the mean infective periods of the class I .

System (8.7) with $I_0(t) = 0$ always has the equilibrium

$$(S, I) = \left(\frac{\Lambda}{\mu}, 0 \right), \quad (8.11)$$

but no other constant solutions. However, since $I_0(t)$ must be zero for large t , one would expect, under appropriate assumptions, that $(\Lambda/\mu, 0)$ will be an attractor or “asymptotic equilibrium” as $t \rightarrow +\infty$. The following results have been shown in [32, 35].

Theorem 8.3 *The infection-free state $(\Lambda/\mu, 0)$ of the limiting system (8.7) is a global attractor; that is, $\lim_{t \rightarrow +\infty} (S(t), I(t)) = (\Lambda/\mu, 0)$ for any positive solution of system (8.7) as long as $\mathcal{R}_0 \leq 1$.*

Theorem 8.4 *The infection-free state of system (8.7) is unstable when $\mathcal{R}_0 > 1$ and there exists a constant $W^* > 0$, such that, any positive solution $(S(t), I(t))$ of (8.7) satisfies $\limsup_{t \rightarrow +\infty} I(t) \geq W^*$.*

In other words, if $\mathcal{R}_0 > 1$ the disease-free state (8.4) cannot be an attractor of any positive solution. That is, every solution has at least approximately W^* infectives (this W^* is the same as that in the statement of Theorem 8.5 below) for a sequence of times t tending to $+\infty$ and if $S(t), I(t)$ approach nonzero constants as $t \rightarrow +\infty$, when $\mathcal{R}_0 > 1$ then the results in [82] guarantee that these constants must satisfy

the limiting system associated with (8.7), which is given by the following set of equations:

$$\begin{aligned}\frac{dS}{dt} &= \Lambda - \lambda C(T(t))S(t)\frac{I(t)}{T(t)} - \mu S(t) \\ I(t) &= \int_0^t \lambda C(T(x))S(x)\frac{I(x)}{T(x)}e^{-\mu(t-x)}P_I(t-x)dx.\end{aligned}\quad (8.12)$$

The limiting system (8.12) is an autonomous system for which we have established the following result:

Theorem 8.5 *If $\mathcal{R}_0 > 1$ the limiting system (8.12) has a unique positive equilibrium S^*, I^* . If in addition $(d/dT)(C(T)/T) \leq 0$, then this endemic equilibrium is locally asymptotic.*

Theorem 8.5 indicates that there is a switch of stability from $(\Lambda/\mu, 0)$ to (S^*, I^*) as \mathcal{R}_0 crosses 1. We also conjecture but have not proved that the asymptotic dynamics of system (8.7) and the limiting system (8.12) agree. An alternative approach can be found in [61]. The proofs of these results can be found in [35].

8.4 An Age of Infection Model

The model presented here is developed in [103]. We consider a homogeneously-mixing male homosexual population with infected members stratified by infection age (time since having been infected). We divide the population into three compartments: S (uninfected, but susceptible), I (HIV infected but with minimal or no symptoms), and A (fully developed AIDS). We assume members of the class A are no longer sexually active, and we let $T = S + I$ be the size of the sexually active population.

We let t denote time and τ denote age of infection, and we stratify the infected population by writing

$$I(t) = \int_0^\infty i(t, \tau)d\tau,$$

where $i(t, \tau)$ denotes the infection age density at time t . We assume:

- the mean number of sexual contacts per individual in unit time is α ,
- there is a mean transmission rate $\lambda(\tau)$ at which a typical susceptible individual contracts the infection by contact with an infected individual of infection age τ ,
- there is a rate $\alpha(\tau)$ of leaving the sexually active population (because of progression to AIDS) that depends on the age of infection,
- there is a constant rate of recruitment Λ into the sexually active population,

- there is a constant rate μ of departure of uninfected members from the sexually active population,
- there is a constant death rate v from full-blown AIDS.

Under these assumptions, the fraction of members remaining in the class I τ time units after having been infected is

$$P(\tau) = e^{-\mu\tau - \int_0^\tau \alpha(\sigma)d\sigma}.$$

Then

$$i(t, \tau) = i(t - \tau, 0)P(\tau).$$

We define the total infectivity at time t ,

$$W(t) = W_0(t) + \int_0^t \lambda(\tau)i(t, \tau)d\tau = W_0(t) + \int_0^t \lambda(\tau)i(t - \tau, 0)P(\tau)d\tau,$$

where $W_0(t)$ is the infectivity at time t of those individuals who were infected at time $t = 0$. Then the rate of new infections in unit time is

$$B(t) = i(t, 0) = a \frac{S(t)}{T(t)} W(t),$$

and

$$W(t) = W_0(t) + \int_0^t \lambda(\tau)P(\tau)B(t - \tau)d\tau.$$

We will take a to be constant, but one could assume more generally that a is a function of the total sexually active population size T .

These assumptions lead to the model

$$\begin{aligned} S'(t) &= \Lambda - B(t) - \mu S(t) \\ W(t) &= W_0(t) + \int_0^t \lambda(\tau)P(\tau)B(t - \tau)d\tau \\ B(t) &= a \frac{S(t)}{T(t)} W(t) \\ I(t) &= I_0(t) + \int_0^t B(t - \tau)P(\tau)d\tau. \end{aligned} \tag{8.13}$$

Since we wish to study equilibria and their stability, we consider the limit system of (8.13), namely

$$\begin{aligned} S'(t) &= \Lambda - B(t) - \mu S(t) \\ W(t) &= \int_0^\infty \lambda(\tau) P(\tau) B(t-\tau) d\tau \\ B(t) &= a \frac{S(t)}{T(t)} W(t) \\ I(t) &= \int_0^\infty B(t-\tau) P(\tau) d\tau. \end{aligned} \tag{8.14}$$

In order to obtain an expression for the number of active AIDS cases, not part of the model since individuals in the class A are assumed not to have any further sexual contacts, but included because it provides a relation that may be compared with data, we differentiate the equation

$$I'(t) = \int_0^t B(s) P(t-s) ds$$

of (8.14), using

$$P'(u) = -[\mu + \alpha(u)]P(u).$$

We obtain

$$I'(t) = B(t) - \mu I(t) - \int_0^t B(s) \alpha(t-s) P(t-s) ds.$$

The input to the AIDS class A is

$$\int_0^t B(s) \alpha(t-s) P(t-s) ds.$$

Thus the number of active AIDS cases is given by

$$A'(t) = \int_0^\infty \alpha(t-s) P(t-s) B(s) ds - v A(t).$$

Analysis of the model (8.14) would be considerably simpler if we had assumed mass action incidence rather than standard incidence, because use of standard incidence brings $T(t) = S(t) + I(t)$ into the model. However, mass action incidence is much less plausible for sexual transmission models than standard incidence. For the model it is not difficult to show that the basic reproduction number is given by

$$\mathcal{R}_0 = a \int_0^\infty \lambda(\tau) P(\tau) d\tau,$$

and that there is a disease-free equilibrium $S = \Lambda/\mu$, $I + B = W = 0$ which is asymptotically stable if $\mathcal{R}_0 < 1$. Calculation of the endemic equilibrium is more difficult, but it is possible to show that there is an endemic equilibrium that is asymptotically stable at least for values of \mathcal{R}_0 larger than 1 but close to 1. For larger values of \mathcal{R}_0 the endemic equilibrium may be unstable, and there may be a Hopf bifurcation [64] and sustained oscillatory solutions of the model.

8.5 *HIV and Tuberculosis: Dynamics of Coinfections

HIV diminishes the ability of the immune system to respond to invasions by infectious agents such as *M. tuberculosis*. Furthermore, as HIV infection progresses, immunity often declines with patients becoming more susceptible to typical or rare infections. In wealthier societies HIV and TB treatments are common; these drugs have altered significantly the joint dynamics of TB and HIV.

The modeling literature on the independent dynamics of HIV or TB is quite extensive. TB efforts include, for example, [9, 18, 40, 42, 43, 50, 51, 89] while HIV/AIDS include [31, 63, 80, 103] to name a few more. TB/HIV coinfection modeling efforts have also been published. Kirschner [73] developed an immunological model describing HIV-1 and TB coinfections within a host. Naresh et al. [86] introduced a model involving a population sub-divided into four epidemiological classes: susceptible, TB infective, HIV infective, and those with AIDS; a model focusing on the transmission dynamics of HIV and treatable TB in variable size populations. Schulzer et al. [101] looked at HIV/TB joint dynamics using actuarial methods. West and Thompson [105] introduced models for the joint dynamics of HIV and TB that were explored via numerical simulations; their main goal was to estimate parameters and use their estimates to forecast the future transmission of TB in the United States. Porco et al. [90] looked, using a discrete event simulation model, at the impact of HIV on the probability and expected severity of TB outbreaks. Additional efforts include those in [91, 98].

A system of differential equations is used in [92] to model the joint dynamics of TB and HIV. The total population is divided into the following epidemiological subgroups: S , susceptible; L , latent with TB; I , infectious with TB; T , successfully treated with TB; J_1 , HIV infectious; J_2 , HIV infectious and TB latent; J_3 , infectious with both TB and HIV; and A , “full-blown” AIDS. The compartmental diagram in Fig. 8.2 illustrates the flow of individuals as they face the possibility of acquiring specific-disease infections or even coinfections.

The TB/HIV model is given by the following systems of eight ordinary differential equations:

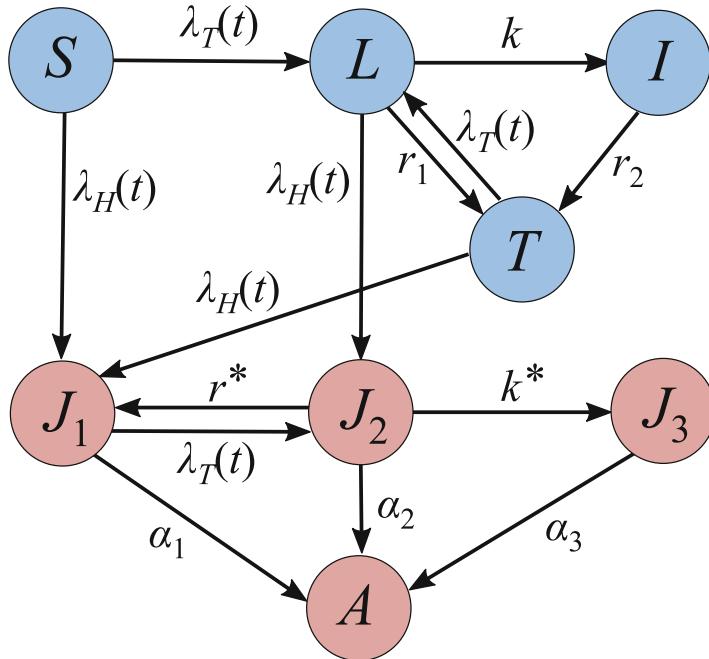


Fig. 8.2 Transition diagram between classes for the dynamics of TB and HIV coinfections. The force of infection for TB is $\lambda_T = c(I + J_3)/N$, and the force of infection for HIV is $\lambda_H = \sigma J^*/R$, where $J^* = J_1 + J_2 + J_3$

$$\begin{aligned}
 \frac{dS}{dt} &= \Lambda - cS \frac{I + J_3}{N} - \sigma S \frac{J^*}{R} - \mu S \\
 \frac{dL}{dt} &= c(S + T) \frac{I + J_3}{N} - \sigma L \frac{J^*}{R} - (\mu + k + r_1)L \\
 \text{TB : } \frac{dI}{dt} &= kL - (\mu + d + r_2)I \\
 \frac{dT}{dt} &= r_1 L + r_2 I - cT \frac{I + J_3}{N} - \sigma T \frac{J^*}{R} - \mu T,
 \end{aligned} \tag{8.15a}$$

$$\begin{aligned}
 \frac{dJ_1}{dt} &= \sigma(S + T) \frac{J^*}{R} - cJ_1 \frac{I + J_3}{N} - (\alpha_1 + \mu)J_1 + r^* J_2 \\
 \frac{dJ_2}{dt} &= \sigma L \frac{J^*}{R} + cJ_1 \frac{I + J_3}{N} - (\alpha_2 + \mu + k^* + r^*)J_2 \\
 \text{HIV : } \frac{dJ_3}{dt} &= k^* J_2 - (\alpha_3 + \mu + d^*)J_3 \\
 \frac{dA}{dt} &= \alpha_1 J_1 + \alpha_2 J_2 + \alpha_3 J_3 - (\mu + f)A,
 \end{aligned} \tag{8.15b}$$

Table 8.1 Definition of parameters and state variables used in the TB/HIV model (8.15)

Symbol	Definition
N	Total population
R	Total active population ($= N - I - J_3 - A = S + L + T + J_1 + J_2$)
J^*	Individuals with HIV who have not developed AIDS ($= J_1 + J_2 + J_3$)
A	Constant recruitment rate
c	Transmission rate of TB
σ	Transmission rate of HIV
μ	Per-capita natural death rate
k	Per-capita TB progression rate for individuals not infected with HIV
k^*	Per-capita TB progression rate for individuals infected also with HIV
d	Per-capita TB-induced death rate
d^*	Per-capita HIV-induced death rate
f	Per-capita AIDS-induced death rate
r_1	Per-capita latent TB treatment rate for individuals with no HIV
r_2	Per-capita active TB treatment rate for individuals with no HIV
r^*	Per-capita latent TB treatment rate for individuals with also HIV
α_i	Per-capita AIDS progression rate for individuals in the J_i ($i = 1, 2, 3$)

where

$$\begin{aligned} N &= S + L + I + T + J_1 + J_2 + J_3 + A, \\ R &= N - I - J_3 - A = S + L + T + J_1 + J_2, \\ J^* &= J_1 + J_2 + J_3. \end{aligned} \quad (8.16)$$

The variable R here collects non-infectious “circulating” individuals, that is, those who do not have active TB or AIDS. Definitions of model parameters are collected in Table 8.1.

The modeling assumptions include: homogenous mixing; HIV positive and TB infective (J_3) showing severe HIV symptoms cannot be effectively treated for active TB; TB infections are only acquired through contacts with TB infectious individuals (I and J_3); and individuals may acquire HIV infections only through contacts with HIV infectious individuals (J^* group). Further, the “probability” of infection per contact is assumed to be the same for T and S classes (β and λ). Furthermore, I (TB infectious), J_3 (TB and HIV infectious), and A (AIDS) individuals are too ill to remain sexually active and, consequently, they do not transmit HIV through sexual activity. Hence, $R \equiv N - I - J_3 - A$ and the HIV incidence is modeled by $\sigma J^*/R$ (see [29, 74, 108]).

The probability of having a contact with HIV infectious individuals is modeled as J^*/R and the number of new HIV infections in a unit time is therefore $\sigma S J^*/R$ [IV drug injections, vertically-transmitted HIV (children of birth), or HIV transmission via breast feeding, forms of HIV transmission are ignored]. The most drastic in this model comes from the incorporation of sexual transmission as an indirect risk

factor, a function of HIV prevalence. Further, demographic changes are ignored or alternatively, it is assumed that the time scale under consideration is such that changes in population size are not too significant.

The TB control reproduction number (under treatment) is given by the expression

$$\mathcal{R}_1 = \frac{ck}{(\mu + k + r_1)(\mu + d + r_2)} \quad (8.17)$$

while the HIV reproduction number is

$$\mathcal{R}_2 = \frac{\sigma}{\alpha_1 + \mu}. \quad (8.18)$$

\mathcal{R}_1 is the product of the average number of susceptible infected by one TB infective individual over its effective infective period, $c/(\mu+d+r_2)$, and the fraction of the population that survives the TB latent period, $k/(\mu+k+r_1)$. \mathcal{R}_1 denotes the number of secondary TB infectious cases generated by a typical TB infective individual during its effective infective period if introduced in a population of mostly TB-susceptible individuals, in a population where TB treatment is accessible. \mathcal{R}_2 is the HIV reproduction number in a TB-free society, the number of secondary HIV infections produced by an HIV infectious (but not TB-infected) individual during its infectious period if introduced in a population of HIV-susceptible individuals (in a TB-free world). The reproduction numbers do not involve the parameters tied in to the dynamics of TB-HIV coinfection, that is, k^* and α_3 .

Consequently, the reproduction number for system (8.15) under TB treatment is given by

$$\mathcal{R} = \max\{\mathcal{R}_1, \mathcal{R}_2\}.$$

We have shown in [92] that TB and HIV will die out if $\mathcal{R} < 1$ while either or both diseases may become endemic if $\mathcal{R} > 1$.

In [92], it was shown that system (8.15) is well-posed, that is, solutions that start in this octant where all the variables are non-negative stay there. It was also shown that system (8.15) has three possible non-negative boundary equilibria: the disease-free equilibrium (DFE) or E_0 , the TB-only (HIV-free) equilibrium or E_T , and the HIV-only (TB-free) equilibrium or E_H . The components of E_0 are

$$S_0 = \frac{\Lambda}{\mu}, \quad L_0 = I_0 = T_0 = J_{01} = J_{02} = J_{03} = A_0 = 0.$$

The E_T components are

$$S_T = \frac{\Lambda}{\mu + cI_T/N_T}, \quad L_T = \frac{I_T}{R_{1b}}, \quad I_T = \frac{N_T(\mathcal{R}_1 - 1)}{\mathcal{R}_1 + \mathcal{R}_{1a}}, \quad T_T = \frac{(r_1L + r_2I_T)S_T}{\Lambda}, \\ J_{1T} = J_{2T} = J_{3T} = A_T = 0,$$

where

$$N_T = \frac{\Lambda}{\mu + d(\mathcal{R}_1 - 1)/(\mathcal{R}_1 + \mathcal{R}_{1a})},$$

with

$$\mathcal{R}_{1a} = \frac{c}{\mu + k + r_1}, \quad \mathcal{R}_{1b} = \frac{k}{\mu + d + r_2}. \quad (8.19)$$

The E_H components are

$$\begin{aligned} S_H &= \frac{\Lambda}{\mu\mathcal{R}_2 + \alpha_1(\mathcal{R}_2 - 1)}, & L_H = I_H = T_H = 0, \\ J_{1H} &= (\mathcal{R}_2 - 1)S_H, & J_{2H} = J_{3H} = 0, & A_H = \frac{\alpha_1 J_{1H}}{\mu + f}. \end{aligned}$$

The following results were established in [92]:

Theorem 8.6 *The disease-free equilibrium E_0 is locally asymptotically stable if $\mathcal{R} < 1$, and it is unstable if $\mathcal{R} > 1$.*

Theorem 8.7 *The HIV-free equilibrium E_T is locally asymptotically stable if $\mathcal{R}_1 > 1$ and $\mathcal{R}_2 < 1$.*

We observe that E_H may not be locally asymptotically stable under the conditions $\mathcal{R}_1 < 1$ and $\mathcal{R}_2 > 1$. Our numerical studies show that when $\mathcal{R}_1 < 1$ and $\mathcal{R}_2 > 1$ it is possible that the equilibrium E_H is unstable and TB can coexist with HIV [92]. Further, whenever both reproduction numbers are greater than 1, that is, $\mathcal{R}_1 > 1$ and $\mathcal{R}_2 > 1$, E_T and E_H both exist and E_0 is unstable. Our numerical studies show that all three boundary equilibria are unstable and solutions converge to an interior equilibrium point. Furthermore, partial analytical results and numerical simulations support the existence of an interior equilibrium \hat{E} when both reproduction numbers, \mathcal{R}_1 and \mathcal{R}_2 , are greater than 1. The numerical simulations of the system further suggest that the interior equilibrium is LAS in most cases although the possibility of stable periodic solutions seems likely [92].

When both reproduction numbers are greater than 1, i.e., $\mathcal{R}_1 > 1$ and $\mathcal{R}_2 > 1$, E_T and E_H both exist and E_0 is unstable. In this case, the numerical simulations of the model show that it is possible that all three boundary equilibria are unstable and solutions converge to an interior equilibrium point. Although explicit expressions for an interior equilibrium are very difficult to compute analytically, we have managed to obtain some relationships that can be used to determine the existence of an interior equilibrium.

Let $\hat{E} = (\hat{S}, \hat{L}, \hat{I}, \hat{J}_1, \hat{J}_2, \hat{J}_3, \hat{A})$ denote an interior equilibrium with all components positive, and let x and y denote the fractions:

$$x = \frac{\hat{I} + \hat{J}_3}{\hat{N}} > 0 \quad \text{and} \quad y = \frac{\hat{J}^*}{\hat{R}} > 0. \quad (8.20)$$

Note that x and y correspond to the levels of disease prevalence for TB and HIV, respectively.

By setting the right-hand-side of the system (8.15) equal to zero we can obtain the following two equations for x and y :

$$\begin{aligned} x &= xF(x, y), \\ y &= yG(x, y), \end{aligned} \quad (8.21)$$

where

$$\begin{aligned} F(x, y) &= \frac{c}{\hat{N}} \left[\frac{k\hat{S}}{(\mu + d + r_2)B_1} + \frac{k^*}{\Delta_2 \Delta_3} \left(\frac{\hat{S}\sigma y}{B_1} + \hat{J}_1 \right) \right], \\ G(x, y) &= \frac{\sigma}{\hat{R}} \left\{ \frac{1}{B_2} \left(\hat{S} + \hat{T} + \frac{r^*\hat{L}}{\Delta_2} \right) \left(1 + \frac{cx}{\Delta_2} \left[1 + \frac{k^*}{\Delta_3} \right] \right) + \frac{\hat{L}}{\Delta_2} \left[1 + \frac{k^*}{\Delta_3} \right] \right\}, \end{aligned} \quad (8.22)$$

in which

$$\begin{aligned} \hat{S} &= \frac{\Lambda}{\mu + cx + \sigma y}, \quad \hat{L} = \frac{c\Lambda}{B_1(\mu + cx + \sigma y)}x, \quad \hat{I} = \frac{k}{\mu + d + r_2} \hat{L}, \\ \hat{T} &= \frac{r_1 + \frac{r_2 k}{\mu + d + r_2}}{cx + \sigma y + \mu} \hat{L}, \quad \hat{J}_1 = \frac{(\hat{S} + \hat{T} + \frac{r^*\hat{L}}{\Delta_2})\sigma y}{B_2}, \quad \hat{J}_2 = \frac{\hat{L}\sigma y + \hat{J}_1 cx}{\Delta_2}, \\ \hat{J}_3 &= \frac{k^*(\hat{L}\sigma y + \hat{J}_1 cx)}{\Delta_2 \Delta_3}, \quad \hat{A} = \frac{1}{\mu + f} (\alpha_1 \hat{J}_1 + \alpha_2 \hat{J}_2 + \alpha_3 \hat{J}_3), \end{aligned} \quad (8.23)$$

and

$$\begin{aligned} \Delta_2 &= \alpha_2 + \mu + k^* + r^*, \\ \Delta_3 &= \alpha_3 + \mu + d, \\ B_1 &= \sigma y + \mu + k + r_1 - \frac{cx(r_1 + \frac{r_2 k}{\mu + d + r_2})}{cx + \sigma y + \mu} \\ &\geq \sigma y + \mu + k + r_1 - (r_1 + k) \\ &> 0, \\ B_2 &= \frac{cx(\alpha_1 + \mu + k^*)}{\Delta_2} + \alpha_1 + \mu. \end{aligned} \quad (8.24)$$

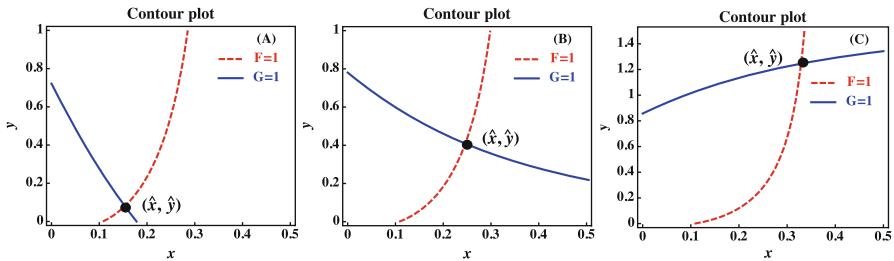


Fig. 8.3 Contour plots showing the intersection points of the curves $F(x, y) = 1$ (dashed curve) and $G(x, y) = 1$ (solid curve) for various values of \mathcal{R}_2 with \mathcal{R}_1 being fixed at 1.5 ($c = 12$). The values of \mathcal{R}_2 in (A)–(C) are 3.6, 4.6, and 7, respectively (corresponding to $\lambda\sigma = 0.41, 0.52$, and 0.8). The axes are $x = (I + J_3)/N$ and $y = J^*/R$, representing the factors in the incidence functions for TB and HIV, respectively. The intersection $(\hat{x}, \hat{y}) = (\frac{\hat{I}+\hat{J}_3}{\hat{N}}, \frac{\hat{J}^*}{\hat{R}})$ determines components of the interior equilibrium \hat{E} if $0 < \hat{x} < 1$ and $\hat{y} > 0$

Note that $x > 0$ and $y > 0$, Eq. (8.21) reduces to

$$F(x, y) = 1, \quad G(x, y) = 1, \quad (8.25)$$

and an intersection of the two curves determined by Eq. (8.25), denoted by \hat{x} and \hat{y} , corresponds to a coexistence equilibrium of TB and HIV. We can consider \hat{x} as a measure for the TB prevalence. The intersection property of the two curves given by $F(x, y) = 1$ and $G(x, y) = 1$ are illustrated in Fig. 8.3.

Figure 8.3 plots the intersection point (\hat{x}, \hat{y}) of the contour plots of $F(x, y) = 1$ (dashed curve) and $G(x, y) = 1$ (solid curve) for several values of \mathcal{R}_2 with \mathcal{R}_1 being fixed ($\mathcal{R}_1 = 1.5$ corresponding to $c = 12$). Again, an interior equilibrium \hat{E} can be determined by \hat{x} and \hat{y} if $0 < \hat{x} < 1$ and $\hat{y} > 0$. This figure illustrates how \hat{x} changes with increasing \mathcal{R}_2 . We have chosen $k^* = 5k$ (i.e., the progression rate to active TB in individuals with both latent TB and HIV is five times higher than that in individuals with latent TB only), $\alpha_3 = 5\alpha_1$ (i.e., the progression to AIDS in individuals with active TB is five times higher than that in individuals without TB). For this set of parameter values, the values of \mathcal{R}_2 in Fig. 8.3A–C are 3.6, 4.6, and 7, respectively. It shows that when \mathcal{R}_2 increases from 3.8 to 4.6, the $F(x, y) = 1$ curve does not change much while the right-end of the $G(x, y) = 1$ curve moves to the right of the $F = 1$ curve. This leads to an intersection point of the two curves (see (A) and (B)), which corresponds to an interior equilibrium \hat{E} . When \mathcal{R}_2 is further increased to 7, the $G(x, y) = 1$ curve changes from decreasing to increasing (see (C)). Although there is still a unique intersection point, the $y = \hat{J}^*/\hat{R}$ component may become greater than 1. This is still biologically feasible as J/R can exceed 1 (see (C)). The intersection points in (A)–(C) are $(\hat{x}, \hat{y}) = (\frac{\hat{I}+\hat{J}_3}{\hat{N}}, \frac{\hat{J}^*}{\hat{R}}) = (0.15, 0.07), (0.25, 0.4), (0.33, 1.25)$, respectively. We observe that \hat{x} increases with \mathcal{R}_2 from 0.15 to 0.33. This implies that the prevalence of HIV may have significant impact on the infection level of TB.

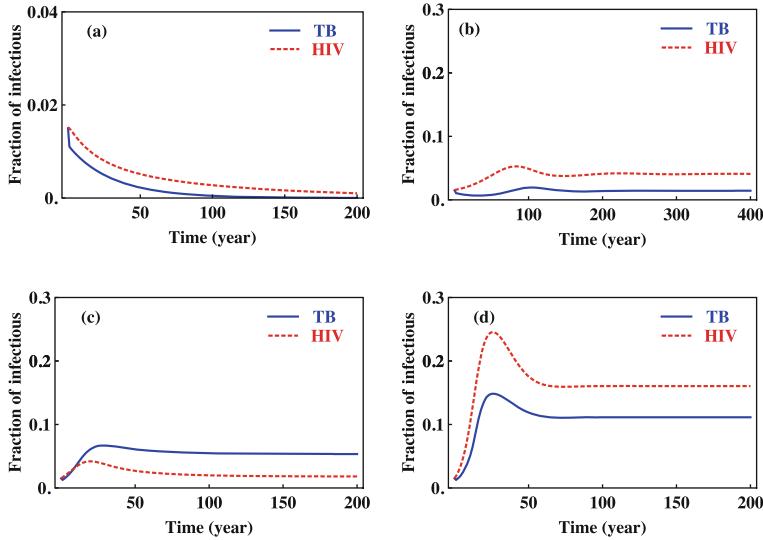


Fig. 8.4 Time plots of prevalence of TB and HIV. The TB curves (solid) represent the fraction of active TB ($(I + J_3)/N$), and the HIV curve (dashed) represents the activity-adjusted fraction of HIV (J^*/R)

Figure 8.4 examines changes in infection levels over time. It plots the time series of $[I(t) + J_3(t)]/N(t)$ (fraction of active TB) and $J^*(t)/R(t)$ (activity-adjusted fraction of HIV infectious) for fixed \mathcal{R}_1 and various \mathcal{R}_2 . The top two figures are for the case when the reproduction number for TB is less than 1 ($\mathcal{R}_1 = 0.96 < 1$ or $c = 7.5$), and the reproduction number for HIV is $\mathcal{R}_2 = 0.9 < 1$ (or $\sigma = 0.105$) in (a) and $\mathcal{R}_2 = 1.3 > 1$ (or $\sigma = 0.15$) in (a). It illustrates in Fig. 8.4a that TB cannot persist if $\mathcal{R}_2 < 1$. However, if $\mathcal{R}_2 > 1$ then it is possible that TB can become prevalent even if $\mathcal{R}_1 < 1$ (see Fig. 8.4b). The bottom two figures are for the case when the reproduction number of TB is greater than 1 ($\mathcal{R}_1 = 1.2$, or $c = 9.1$), and $\mathcal{R}_2 = 2$ (or $\sigma = 0.23$) in (c) and $\mathcal{R}_2 = 3$ (or $\sigma = 0.34$) in (d). It demonstrates that an increase in \mathcal{R}_2 will lead to an increase in the level of TB prevalence as well. All other parameters are the same as in Fig. 8.3 except that $k^* = 3k$.

Another way to look at the role of HIV on TB dynamics is to compare the outcomes between the cases where HIV is absent or present (instead of varying the value of \mathcal{R}_2). This result is presented in Fig. 8.5. The reproduction numbers are identical in Fig. 8.5A, B: $\mathcal{R}_1 = 0.98 < 1$ ($c = 7.7$) and $\mathcal{R}_2 = 1.2 > 1$ ($\sigma = 0.137$). Other parameter values are the same as in Fig. 8.4 except that $k^* = k$. The variables plotted are $(I + J_2)/N$ and J^*/N . Figure 8.5A is for the case when HIV is absent by letting $J^*(0) = 0$. It shows that TB cannot persist. In Fig. 8.5B, the initial value of HIV is positive (i.e., $J^*(0) > 0$). It shows that both TB and HIV coexist.

Examples of other mathematical models on dynamics of TB/HIV coinfections include [73, 86, 90, 91, 101].

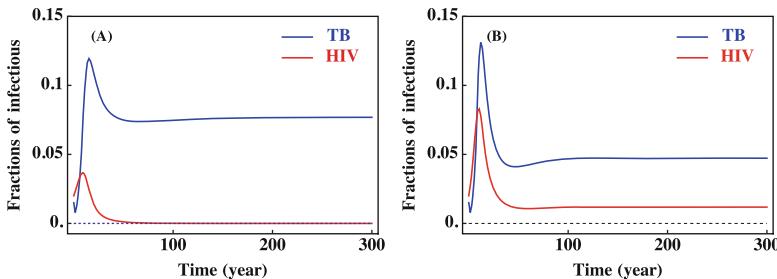


Fig. 8.5 For the plot in (A), HIV is absent by letting $J^*(0) = 0$. It shows that TB cannot persist. In (B), the initial value of HIV is positive (i.e., $J^*(0) > 0$). It shows that both TB and HIV will coexist

8.6 *Modeling the Synergy Between HIV and HSV-2

The example presented in this section considers the synergy between HIV and HSV-2. One of the questions that is of interest for public health officials is how treatment of HSV-2 may influence the prevalence and control of HIV.

Several mathematical models have been developed to investigate the transmission dynamics of HSV-2 (e.g., [17, 53, 87, 100] and references therein) and HIV (e.g., [12–14, 44, 84, 85] and references therein). To our knowledge, however, there have only been a few modeling studies of the epidemiological synergy between HSV-2 and HIV. Using the individual-based model STDSIM, White et al. [106] studied the population-level effect of HSV-2 therapy on the incidence of HIV in sub-Saharan Africa. Foss et al. [54] developed a dynamic HSV/HIV model to estimate the contribution of HSV-2 to HIV transmission from clients to female sex workers in southern India and the maximum potential impact of “perfect” HSV-2 suppressive therapy on HIV incidence. Blower and Ma [16] used a transmission model that specifies the dynamics of HIV and HSV-2 to predict the effect of a high-prevalence HSV-2 epidemic on HIV incidence. Abu-Raddad et al. [1] constructed a deterministic compartmental model to describe HIV and HSV-2 transmission dynamics and their interaction. However, the model studied in [16] does not include heterogeneity in sexual activity and assumes that individuals mix randomly, whereupon each infective individual is equally likely to spread the disease to all others. Also, gender is not incorporated into the models studied by either [16] or [1]. The models in [54, 106] incorporate various heterogeneities, including gender and/or age, but not sexual activity, and only numerical simulations are conducted.

Gender may be an important factor in modeling the epidemiological synergy between HSV-2 and HIV as shown in the meta-analysis of several studies that male parameters differ from the corresponding female parameters. For example, the male-to-female HSV-2 transmission probability is greater than the female-to-male transmission probability [45, 104], and thus the risk of female-to-male transmission per sex act is less than the risk of male-to-female transmission [84, 85]. Thus, to fully understand the epidemiological synergy between HSV-2 and HIV and to investigate

measures for controlling these sexually transmitted diseases, it is important to analyze models that consider heterogeneities in sexual activity, mixing within and between different activity groups and genders.

In [2, 52], a model incorporating both HIV and HSV-2 infections was analyzed. The model considers one male population and multiple female populations based on their activity levels with variable male preferences to different female groups. Results from the model demonstrate that the heterogeneity in activity levels and male preference in mixing may play an important role in model outcomes. More details of the model analysis are presented below.

Consider a population consisting of sexually active female and male individuals. Consider the case in which the female population is divided into subgroups based on levels of sexual activity (e.g., number of partners) with a low-risk group (e.g., members of the general population) and a high-risk group (e.g., sex workers), while all individuals in the male population have the same activity level. These subpopulations are labeled by the subscripts f_1, f_2, m , which denote low- and high-risk females and males, respectively. Let N_i denote the population sizes of groups i , where $i = m, f_1, f_2$. The population in each group is assumed to be homogeneous in the sense that individuals have the same infectious period, duration of immunity, contact rate, and so on. We divide the progression of HIV into two stages, acute infection and AIDS. Similarly, HSV-2 is represented by acute and latent infection stages. Because individuals infected with HIV alone or HSV-2 alone can become coinfective with both HIV and HSV-2, each group i ($i = m, f_1, f_2$) is further divided into seven epidemiological classes or subgroups: susceptible, infected with acute HSV-2 only (A_i), infected with latent HSV-2 only (L_i), infected with HIV only (H_i), infected with HIV and acute HSV-2 (P_i), infected with HIV and latent HSV-2 (Q_i) and AIDS ($AIDS$). A transition diagram between these epidemiological classes within group i is depicted in Fig. 8.6.

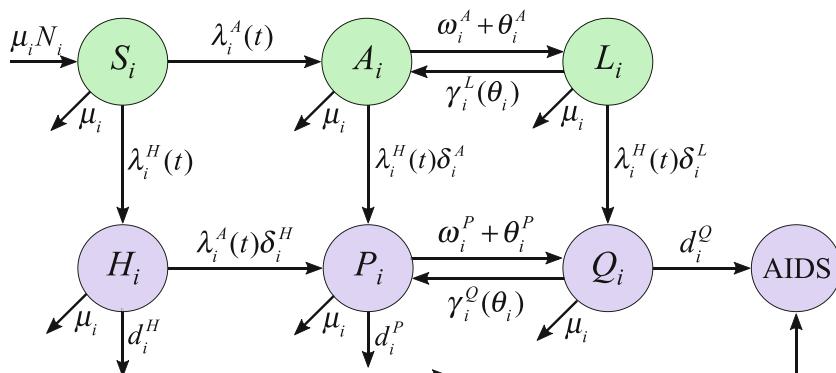


Fig. 8.6 Transition diagram of the coupled dynamics between HIV and HSV-2. The top row includes classes infected with HSV-2 only, and the bottom row includes classes infected with either HIV only or coinfective with HIV and HSV-2

For each sub-population i ($i = f_1, f_2, m$) there is a per-capita recruitment rate μ_i into the susceptible group. For all classes there is a constant per-capita rate μ_i of exiting the sexually active population. Thus, the total population N_i in group i remains constant for all time. Susceptible people in group i acquire infection with HSV-2 or HIV at the rate $\lambda_i^A(t)$ or $\lambda_i^H(t)$, respectively. Upon being infected with HSV-2, people in group i enter the class A_i (infected with acute HSV-2 only). These individuals become latent L_i at the constant rate ω_i^A (an average duration in A_i is $1/\omega_i^A$). Following an appropriate stimulus in individuals with latent HSV-2, reactivation may occur [17]. We assume that people with latent HSV-2 only reactivate at the rate γ_i^L . Individuals with HIV are assumed to develop AIDS at the rate d_i^H . Let δ_i^A and δ_i^L denote the enhanced susceptibility to HIV infection for individuals in group i with acute or latent HSV-2 infection. Classes P_i and Q_i are similar to A_i and L_i , respectively, except that A_i and L_i denote individuals with HSV-2 only whereas P_i and Q_i denote individuals with coinfections. The difference in stage durations is indicated by the superscripts (e.g., $1/\gamma_i^L$ for the L class and $1/\gamma_i^Q$ for the Q class). Finally, the antiviral treatment rates for the A_i and P_i individuals are denoted by θ_i^A and θ_i^Q , respectively. Because antiviral medications will also suppress reactivation of latent HSV-2, we assume that the reactivation rate of people with latent HSV-2 γ_i^L (or γ_i^Q) is a decreasing function of θ_i^A (or θ_i^P), denoted by $\gamma_i^L(\theta_i^A)$ (or $\gamma_i^Q(\theta_i^P)$). The sources for most of the parameter values are from [1, 53] (see [52] for more details).

Based on Fig. 8.6, Alvey et al. [2] studied the following model:

$$\begin{aligned} \frac{dS_i}{dt} &= \mu_i N_i - (\lambda_i^A(t) + \lambda_i^H(t)) S_i - \mu_i S_i, \\ \frac{dA_i}{dt} &= \lambda_i^A(t) S_i + \gamma_i^L(\theta_i^A) L_i - \delta_i^A \lambda_i^H(t) A_i - (\omega_i^A + \theta_i^A + \mu_i) A_i, \\ \frac{dL_i}{dt} &= (\omega_i^A + \theta_i^A) A_i - \delta_i^L \lambda_i^H(t) L_i - (\gamma_i^L(\theta_i^A) + \mu_i) L_i, \\ \frac{dH_i}{dt} &= \lambda_i^H(t) S_i - \delta_i^H \lambda_i^A(t) H_i - (\mu_i + d_i^H) H_i, \\ \frac{dP_i}{dt} &= \delta_i^A \lambda_i^H(t) A_i + \delta_i^H \lambda_i^A(t) H_i + \gamma_i^Q(\theta_i^P) Q_i - (\omega_i^P + \theta_i^P + \mu_i + d_i^P) P_i, \\ \frac{dQ_i}{dt} &= \delta_i^L \lambda_i^H(t) L_i + (\omega_i^P + \theta_i^P) P_i - (\gamma_i^Q(\theta_i^P) + \mu_i + d_i^Q) Q_i, \quad i = m, f_1, f_2, \end{aligned} \tag{8.26}$$

where the functions $\lambda_i^j(t)$ represent the forces of infection given below. Let b_i ($i = m, f_1, f_2$) be the rate at which individuals in group i acquire new sexual partners (also referred to as contact rates), and let c_j denote the probability that a male chooses a female partner in group j ($j = f_1, f_2$). Then $c_1 + c_2 = 1$. For ease of notation, let

$$c_1 = c, \quad c_2 = 1 - c.$$

Overall, the number of female partners in groups j ($j = f_1, f_2$) that males acquire should be equal to the number of male partners that females in groups j acquire. These observations lead to the following balance conditions:

$$b_m c N_m = b_{f_1} N_{f_1}, \quad b_m (1 - c) N_m = b_{f_2} N_{f_2}. \quad (8.27)$$

To ensure that constraints in (8.27) are satisfied, we assume in numerical simulations that b_m and c are fixed constants with b_{f_1} and b_{f_2} being varied according to N_m , N_{f_1} , and N_{f_2} .

The force of infection functions can be expressed as

$$\begin{aligned} \lambda_m^H(t) &= \sum_{i=1}^2 b_m c_i \beta_{f_i m}^H \frac{H_{f_i} + \delta_{f_i}^P P_{f_i} + \delta_{f_i}^Q Q_{f_i}}{N_{f_i}}, \\ \lambda_{f_j}^H(t) &= b_{f_j} \beta_{m f_j}^H \frac{H_m + \delta_m^P P_m + \delta_m^Q Q_m}{N_m}, \quad j = 1, 2, \\ \lambda_m^A(t) &= \sum_{i=1}^2 b_m c_i \beta_{f_i m}^A \frac{A_{f_i} + \sigma_{f_i}^P P_{f_i}}{N_{f_i}}, \\ \lambda_{f_j}^A(t) &= b_{f_j} \beta_{m f_j}^A \frac{A_m + \sigma_m^P P_m}{N_m}, \quad j = 1, 2, \end{aligned} \quad (8.28)$$

where

$$N_i = S_i + A_i + L_i + H_i + P_i + Q_i, \quad i = m, f_1, f_2$$

denotes the total population size of group i . In (8.28), β_{im}^H (β_{mi}^H), $i = f_1, f_2$ are the HIV transmission probabilities per partner between females infected with HIV in group i and susceptible males (between males infected with HIV and susceptible females in group i); β_{im}^A (β_{mi}^A), $i = f_1, f_2$ are the HSV-2 transmission probabilities per partner between females infected with acute HSV-2 in group i and susceptible males (between males infected with acute HSV-2 and susceptible females in group i); δ_i^P and δ_i^Q ($i = m, f_1, f_2$) are the enhanced HIV infectiousness of coinfected individuals, and σ_i^P ($i = m, f_1, f_2$) are the enhanced HSV-2 infectiousness of coinfected individuals.

8.6.1 Reproduction Numbers for Individual Diseases

For each of the two diseases, we can compute the reproduction number in the absence of the other disease. Let \mathcal{R}_0^A and \mathcal{R}_0^H denote these reproduction numbers for HSV-2 and HIV, respectively. Due to the loop between the symptomatic and

asymptomatic stages of HSV-2, the derivation of analytical expression for \mathcal{R}_0^A for model (8.26) is not straightforward. A detailed derivation of the following formula for \mathcal{R}_0^A can be found in [2, 52]:

$$\mathcal{R}_0^A = \sqrt{\left(\mathcal{R}_{mf_1m}^A\right)^2 + \left(\mathcal{R}_{mf_2m}^A\right)^2}, \quad (8.29)$$

where

$$\mathcal{R}_{mf_jm}^A = \sqrt{\frac{b_{f_j}\beta_{mf_j}^A}{\omega_m^A + \theta_m^A + \mu_m} \cdot P_m^A \cdot \frac{b_m c_j \beta_{f_jm}^A}{\omega_{f_j}^A + \theta_{f_j}^A + \mu_{f_j}} \cdot P_{f_j}^A}, \quad j = 1, 2$$

with P_i^A ($i = m, f_1, f_2$) representing the probability that an individual of group i is in the acute stage (A), which is given by

$$P_i^A = \frac{(\omega_i^A + \theta_i^A + \mu_i)(\gamma_i^L(\theta_i^A) + \mu_i)}{[\gamma_i^L(\theta_i^A) + \omega_i^A + \theta_i^A + \mu_i]\mu_i}, \quad i = m, f_1, f_2. \quad (8.30)$$

The formulas for P_i^A in (8.30) can be explained as follows. Let

$$p = \frac{\omega_i^A + \theta_i^A}{\omega_i^A + \theta_i^A + \mu_i}, \quad q = \frac{\gamma_i^L}{\gamma_i^L + \mu_i},$$

where p represents the probability that an individual moves from the acute stage (A) to the latent stage (L), and q represents the probability that an individual moves from L to A . Thus, the probability that an individual is in the acute stage within the $A \rightleftharpoons L$ loop is

$$\sum_{k=1}^{\infty} (pq)^k = \frac{(\omega_i^A + \theta_i^A + \mu_i)(\gamma_i^L + \mu_i)}{(\gamma_i^L + \omega_i^A + \theta_i^A + \mu_i)\mu_i} = P_i^A.$$

Notice that in the formula for \mathcal{R}_0^A the balance conditions in (8.27) have been used. Other factors in $\mathcal{R}_{mf_jm}^A$ ($i = 1, 2$) also have clear biological interpretations:

- $b_{f_j}\beta_{mf_j}^A$ is the number of new infections that a male will cause in females of group j ($j = 1, 2$) per unit of time;
- $b_m c_j \beta_{f_jm}^A$ is the number of new infections that a female in group j ($j = 1, 2$) will cause in males per unit of time;
- $\frac{1}{\omega_i^A + \theta_i^A + \mu_i}$ ($i = m, f_1, f_2$) represents the mean time that an individual in group i remains infected (i.e., in either A or L).

Thus, $\sqrt{\mathcal{R}_{mf_jm}^A}$ represents the average secondary HSV-2 male infections by one male individual through females in group j ($j = 1, 2$) while in the infectious stage (A) in a completely susceptible population. The square root is associated with the fact that we need to consider both the male-to-female and female-to-male processes to obtain the number of secondary infections. The overall reproduction number \mathcal{R}_0^A is an average of $\mathcal{R}_{mf_im}^A$ ($i = 1, 2$).

Let \mathcal{R}_0^H denote the basic reproduction number for HIV in the absence of HSV-2. Then

$$\mathcal{R}_0^H = \sqrt{\left(\mathcal{R}_{mf_1m}^H\right)^2 + \left(\mathcal{R}_{mf_2m}^H\right)^2},$$

where

$$\mathcal{R}_{mf_jm}^H = \sqrt{\frac{b_{fj}\beta_{mf_j}^H}{d_m^H + \mu_m} \cdot \frac{b_m c_j \beta_{fjm}^H}{d_{fj}^H + \mu_{fj}}}, \quad j = 1, 2.$$

The biological meanings of $\mathcal{R}_{mf_1m}^H$ and $\mathcal{R}_{mf_2m}^H$ can be explained in the similar way as those of $\mathcal{R}_{mf_1m}^A$ and $\mathcal{R}_{mf_2m}^A$. It is clear that \mathcal{R}_0^H represents the average secondary HIV male infections by one male individual (through both female groups) during the whole HIV infective period in a completely susceptible population.

8.6.2 Invasion Reproduction Numbers

Let \mathcal{R}_A^H denote the invasion reproduction number for HIV in a population where the HSV-2 infection is already established at the endemic equilibrium, which is denoted by E_ϑ^A . The nonzero components of E_ϑ^A are S_i^0 , A_i^0 , and L_i^0 , representing the density of susceptible, acute HSV-2, and HSV-2 latent, respectively, in group i . Let $N_i^0 = S_i^0 + A_i^0 + L_i^0$. For ease of notation, let

$$\lambda_m^{A0} = b_m \sum_{i=1}^2 c_i \beta_{fjm}^A \frac{A_{fj}^0}{N_{fj}^0}, \quad \lambda_{fj}^{A0} = b_{fj} \beta_{mfj}^A \frac{A_m^0}{N_m^0}, \quad j = 1, 2$$

and

$$\mathbf{d}_i = \left(1, \delta_i^P, \delta_i^Q\right), \quad \mathbf{x}_i^0 = \left(S_i^0, \delta_i^A A_i^0, \delta_i^L L_i^0\right)^T, \quad i = m, f_1, f_2.$$

Note that the system (8.26) has 9 infected variables with HIV ($H_i, P_i, Q_i, i = m, f_1, f_2$). Consider the HIV-free equilibrium E_ϑ^A of system (8.26). The matrices

\mathcal{F}^H and \mathcal{V}^H (corresponding to the new infection and remaining transfer terms, respectively) are given by

$$\mathcal{F}^H = \begin{pmatrix} 0 & F_{f_1 m}^H & F_{f_2 m}^H \\ F_{m f_1}^H & 0 & 0 \\ F_{m f_2}^H & 0 & 0 \end{pmatrix}, \quad \mathcal{V}^H = \begin{pmatrix} V_m^H & 0 & 0 \\ 0 & V_{f_1}^H & 0 \\ 0 & 0 & V_{f_2}^H \end{pmatrix}, \quad (8.31)$$

where

$$F_{f_j m}^H = b_m c_j \beta_{f_j m}^H \frac{\mathbf{x}_m^0}{N_m^0} \mathbf{d}_{f_j}, \quad F_{m f_j}^H = b_{f_j} \beta_{m f_j}^H \frac{\mathbf{x}_{f_j}^0}{N_{f_j}^0} \mathbf{d}_m, \quad j = 1, 2$$

and

$$V_i^H = \begin{pmatrix} (\mu_i + d_i^H + \delta_i^H \lambda_i^{A0}) & 0 & 0 \\ -\delta_i^H \lambda_i^{A0} & \omega_i^P + \theta_i^P + \mu_i + d_i^P & -\gamma_i^Q(\theta_i^P) \\ 0 & -(\omega_i^P + \theta_i^P) & \gamma_i^Q(\theta_i^P) + \mu_i + d_i^Q \end{pmatrix}, \quad (8.32)$$

for $i = m, f_1, f_2$. Then, the next generation matrix for HIV, denoted by K_H , can be expressed by

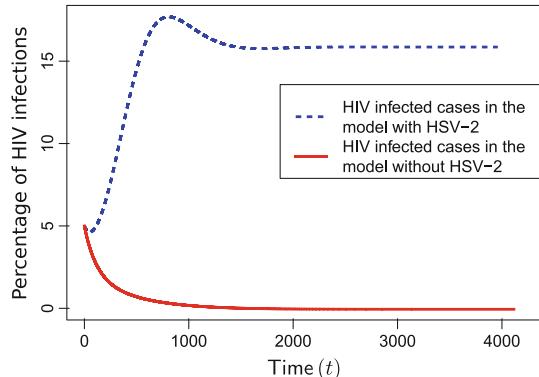
$$K_H = \mathcal{F}^H (\mathcal{V}^H)^{-1} = \begin{pmatrix} 0 & F_{f_1 m}^H (V_{f_1}^H)^{-1} & F_{f_2 m}^H (V_{f_2}^H)^{-1} \\ F_{m f_1}^H (V_m^H)^{-1} & 0 & 0 \\ F_{m f_2}^H (V_m^H)^{-1} & 0 & 0 \end{pmatrix} := (k_{ij})_{9 \times 9}, \quad (8.33)$$

where the entries k_{ij} of the matrix K_H can be found in the Appendix A of [52].

Noting that $\text{Rank}(K_H) = 2$ and that the sum of the diagonal elements in matrix K_H is zero, it follows from Vieta's formulas that if the numbers of susceptible people and those with acute and latent HSV-2 in group i are S_i^0, A_i^0, L_i^0 , respectively, the reproduction number for HIV infection is given by

$$\begin{aligned} \mathcal{R}_A^H &= R_A^H(S_i^0, A_i^0, L_i^0, 0, 0, 0) := \rho(K_H) = \sqrt{-E_2(K_H)} \\ &= \sqrt{\sum_{i=1}^3 \sum_{j=4}^9 k_{ij} k_{ji}}, \end{aligned} \quad (8.34)$$

Fig. 8.7 Numerical solutions of the system (8.26) for the case when $\mathcal{R}_0^A > 1$, $\mathcal{R}_0^H < 1$, and $\mathcal{R}_A^H > 1$. The dashed and solid curves represent levels of HIV infections with and without HSV-2 present, respectively. It shows that even when the basic reproduction number for HIV \mathcal{R}_0^H is less than 1, HIV can still persist if the invasion reproduction number \mathcal{R}_A^H is greater than 1



where $\rho(K_H)$ represents the spectral radius of the matrix K_H and $E_2(K_H)$ is the sum of all the 2×2 principal minors of matrix K_H . It is shown in [52] that invasion is possible if and only if $\mathcal{R}_A^H > 1$.

Similarly, an invasion reproduction number \mathcal{R}_H^A for HSV-2 to invade a population in which HIV is present (see [52]). Detailed results on the existence and local stability of the boundary equilibria can also be found in [52].

8.6.3 Influence of HSV-2 on the Dynamics of HIV

Figure 8.7 illustrates the result of numerical simulations showing how the joint disease dynamics of HIV and HSV-2 may depend on the basic and invasion reproduction numbers. It is for the case when enhancement of HIV by HSV-2 is relatively strong with $\mathcal{R}_0^A > 1$, $\mathcal{R}_0^H < 1$, and $\mathcal{R}_A^H > 1$. It shows that while HIV can invade and persist in the presence of HSV-2 (the dashed curve), it dies out in the absence of HSV-2 (the solid curve), suggesting that HSV-2 infection can favor the invasion of HIV.

8.7 An HIV Model with Vaccination

Blower et al. [15] studied model for HIV with live attenuated HIV vaccines (LAHVs). Consider two viral strains, one wild strain and one vaccine strain. Divide the total population into the following epidemiological classes: susceptible individuals (S), unvaccinated individuals infected with the wild-type HIV (I_w) or the vaccine strain (either by vaccination or by transmission) (I_v) or dually infected

with both strains (I_{vw}), and individuals with AIDS (A). The model consists of the following ordinary differential equations:

$$\begin{aligned} S' &= (1 - p)\pi - (c\lambda_v + c\lambda_w + \mu)S, \\ I'_v &= p\pi + c\lambda_v S - (1 - \psi)c\lambda_w I_v - (\nu_v + \mu)I_v, \\ I'_w &= c\lambda_w S - (\nu_w + \mu)I_w, \\ I'_{vw} &= (1 - \psi)c\lambda_w I_v - (\nu_{vw} + \mu)I_{vw}, \\ A' &= \nu_w I_w + \nu_v I_v + \nu_{vw} I_{vw} - (\mu_A + \mu)A, \end{aligned} \tag{8.35}$$

where λ_v and λ_w are per-capita risks of infection with the vaccine and wild-type strains, respectively, given by

$$\lambda_v = \beta_v \frac{I_v}{N_{SA}}, \quad \lambda_w = \beta_w \frac{I_w + gI_{vw}}{N_{SA}},$$

and $N_{SA} = X + I_v + I_w + I_{vw}$ denotes the number of sexually active population. Other parameters include: β_v and β_w are infection rates for vaccine and wild-type strains, respectively, p is the fraction of new susceptibles vaccinated, π is the number of new susceptibles that join the sexually active population per unit time, c is the average rate of acquiring new sex partners, $1/\mu$ is the average period of acquisition of new sex partners, $1/\mu_A$ is the average survival time with AIDS, ψ denotes the degree of protection that the vaccine provides against infection with the wild-type strain, ν is the progression rate to AIDS in individuals infected with the LAHV strain (ν_v), the wild-type strain (ν_w), or both strains (ν_{vw}), $1/\mu_A$ is the average survival time from AIDS to death. The disease progression rates are related by the expression $\nu_{vw} = \delta\nu_w$, where δ specifies the vaccine-induced degree of reduction in the wild-type disease progression rate.

A time-dependent uncertainty analysis of model (8.35) can be used [15] to predict the potential impact of LAHVs on the annual AIDS death rate, as illustrated in Fig. 8.8. It shows the result of infection with the wild-type strain of HIV for Zimbabwe (A) and Thailand (B), and the result of the LAHV strain for Zimbabwe (C) and Thailand (D). Parameter values used include the following probability density functions (pdfs): $1/\mu_A$ (pdf: 9 months to 1 year to 18 months), β_w (pdf: 0.05 to 0.1 to 0.2), $\beta_v = \alpha\beta_w$ where α (range of pdf: (0.001, 0.1)), ν_w (pdf: range from 50% progression to AIDS in 7.5 years to 50% progression in 10 years). Consider a mass vaccination campaign (with follow-up programs) that would vaccinate anywhere from 80% to 95% of susceptibles with p (range of pdf: (0.8, 0.95)), ψ (range of pdf: (0.5, 0.95)), δ (range of pdf: (0.1, 1)). The population size of sexually active adults are chosen to be 5,560,000 (Zimbabwe), 34,433,00 (Thailand).

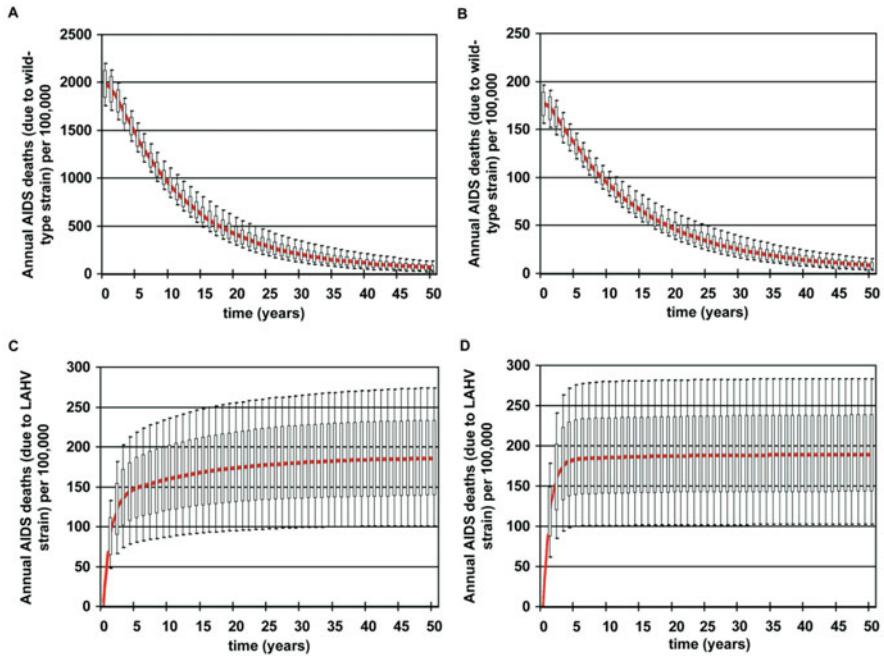


Fig. 8.8 Simulation results of model (8.35). It plots annual AIDS deaths (per 100,000 individuals) for Zimbabwe (A), Thailand (B), Zimbabwe (C), and Thailand (D). Source: [15]

8.8 A Model with Antiretroviral Therapy (ART)

A mathematical model with antiretroviral therapy (ART) is considered in [25] to study the effect of ART on risk behaviors and sexually transmitted infections (STI). Individuals are divided into two risk groups $i = 1, 2$, with $s = 1$ for STI+ and $s = 0$ for STI-. The population size for fix i and s is divided into the following epidemiological classes: susceptible to HIV (S_{is}), untreated HIV+ (I_{is}^u), untreated with AIDS (A_{is}^u), treated HIV+ (I_{is}^τ), treated with AIDS (A_{is}^τ). The group sizes are

$$N_{is} = S_{is} + I_{is}^u + I_{is}^\tau + A_{is}^\tau, \quad i = 1, 2, s = 0, 1.$$

The per-capita rates of STI and HIV infection of a susceptible individual in risk group i are denoted by $\xi_i(t)$ and $\lambda_{is}(t)$, respectively, and are given by

$$\xi_i(t) = \theta_i \sum_j \rho_{ij} \frac{N_{j1}}{\sum_s N_{js}}, \quad i = 1, 2, \quad (8.36)$$

$$\lambda_{i0}(t) = \beta_i \sum_{j=1}^2 \rho_{ij} \frac{\sum_s (I_{js}^u + (1 - \eta)(I_{js}^\tau + A_{js}^\tau))}{\sum_s N_{js}}, \quad \lambda_{i1} = 3\lambda_{i0}, \quad i = 1, 2, \quad (8.37)$$

where θ_i and β_i are transmission rates of STI and HIV, respectively, for susceptibles of risk level i , ρ_{ij} represents the sexual mixing between types i and j individual (e.g., proportionate mixing), η_1 represents the reduction in HIV infectiousness due to treatment with ART.

The following model is a simplified version of the model considered in [25]:

$$\begin{aligned} \frac{dS_{is}}{dt} &= \Lambda_i - (\lambda_{is}(t) + \mu)S_{is} + (1-s)[-\xi_i(t)S_{i0} + \delta S_{i1}] + s[\xi_i(t)S_{i0} - \delta S_{i1}], \\ \frac{dI_{is}^u}{dt} &= \lambda_{is}(t)S_{is} - (\gamma^u + \mu + r^h)I_{is}^u \\ &\quad + \omega I_{is}^\tau + (1-s)[-\xi_i(t)I_{i0}^u + \delta I_{i1}^u] + s[\xi_i(t)I_{i0}^u - \delta I_{i1}^u] \\ \frac{dI_{is}^\tau}{dt} &= r^h I_{is}^u - (\gamma^\tau + \mu + \omega)I_{is}^\tau \\ &\quad + (1-s)[-\xi_i(t)I_{i0}^\tau + \delta I_{i1}^\tau] + s[\xi_i(t)I_{i0}^\tau - \delta I_{i1}^\tau] \quad (8.38) \\ \frac{dA_{is}^u}{dt} &= \gamma^u I_{is}^u - (\alpha^u + \mu + r^a)A_{is}^u + \omega A_{is}^\tau + (1-s)\delta A_{i1}^u - s\delta A_{i1}^u \\ \frac{dA_{is}^\tau}{dt} &= \gamma^\tau I_{is}^\tau - (\alpha^\tau + \mu + \omega)A_{is}^\tau \\ &\quad + r^a A_{is}^u + (1-s)[-\xi_i(t)A_{i0}^\tau + \delta A_{i1}^\tau] + s[\xi_i(t)A_{i0}^\tau - \delta A_{i1}^\tau], \end{aligned}$$

where $i = 1, 2$, $s = 0, 1$, Λ_i is the recruitment rate to group i ($i = 1, 2$), γ^u and γ^τ are the rates of progression to AIDS for untreated and treated individuals, respectively, α^u and α^τ are the rates of AIDS mortality for untreated and treated individuals, respectively, δ is the recovery rate from the STI infection, η represents the reduction in HIV infectiousness as a result of ART, w is the withdraw rate from treatment, r^a and r^h are treatment coverage rates of AIDS and HIV-positive individuals, respectively, $1/\mu$ represents the average duration of sexually active life.

A more general model is studied in [25], in which a detailed model analysis is presented to demonstrate the impact of the wide-scale use of ART on HIV transmission.

8.9 Project: What If Not All Infectives Progress to AIDS?

In the model (8.1) it is assumed that all HIV-infected individuals eventually progress to full-blown AIDS, as this appears to be the case. Suppose, however, that only a fraction p , $0 < p < 1$ progress to AIDS while the remaining infectives remain in this class until they are no longer sexually active. In addition to the classes S , I , and A , a model must now also include a class Y of infective individuals that will not develop full-blown AIDS and a class Z of former Y -individuals who are no longer sexually active. The corresponding model is

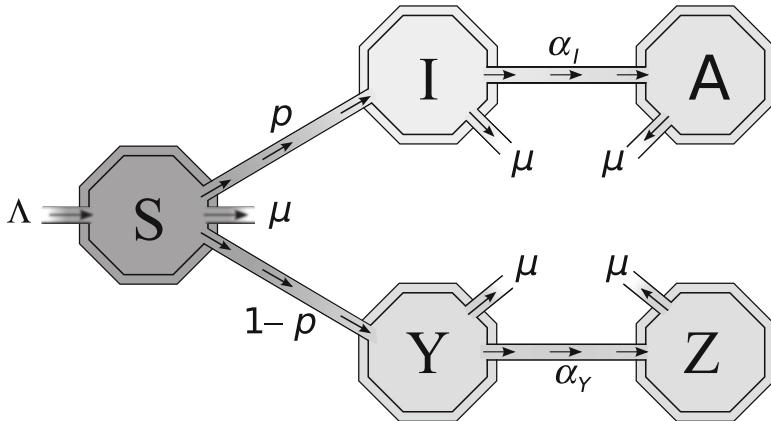


Fig. 8.9 Flow diagram; single group model for the case when only a fraction of infected people will progress to AIDS

$$\begin{aligned}
 \frac{dS(t)}{dt} &= \Lambda - \lambda C(T(t)) \frac{S(t)W(t)}{T(t)} - \mu S(t) \\
 \frac{dI(t)}{dt} &= \lambda p C(T(t)) \frac{S(t)W(t)}{T(t)} - (\alpha_I + \mu)I(t) \\
 \frac{dY(t)}{dt} &= \lambda(1-p)C(T(t)) \frac{S(t)W(t)}{T(t)} - (\alpha_Y + \mu)Y(t) \\
 \frac{dA(t)}{dt} &= \alpha_I I(t) - (d + \mu)A(t) \\
 \frac{dZ(t)}{dt} &= \alpha_Y Y(t) - \mu Z(t)
 \end{aligned} \tag{8.39}$$

where

$$W = I + Y \quad \text{and} \quad T = W + S. \tag{8.40}$$

A flow diagram is shown in Fig. 8.9.

It is further assumed that individuals who develop full-blown AIDS are no longer actively infective, that is, that they have no sexual contacts; it is also assumed that infected individuals become immediately infective. Finally, it is assumed that individuals in this population become sexually inactive or acquire AIDS at the constant rates α_Y and α_I (respectively) per unit time. Therefore, $1/(\mu + \alpha_Y)$ gives the average or mean incubation period with the fraction $1/(\mu + \alpha_Y)$ denoting the average or mean sexual life expectancy. For simplicity, we assume $\alpha_I = \alpha_Y$, but it is possible to extend the model to the case $\alpha_I \neq \alpha_Y$.

As before, the function $C(T)$ models the mean number of sexual partners an average individual has per unit time when the population density is T ; λ (a constant) denotes the average sexual risk per infected partner; λ is often thought as the product $i\phi$ [68], where ϕ is the average number of contacts per sexual partner and i the conditional probability of infection from a sexual contact when the latter is infected. Kingsley et al. [72] had presented (not surprising) evidence that the probability of seroconversion (infection) increases with the number of infected sexual partners. Hence, $\lambda C(T)$ models the transmission rate per unit time per infected partner when the size of the sexually active population is T . We continue to assume

$$C(T) > 0, \quad C'(T) \geq 0, \quad (8.41)$$

Question 4 Show that for the model (8.39) the basic reproduction number is given by

$$\mathcal{R}_0 = \lambda \left(\frac{p}{\sigma_I} + \frac{1-p}{\sigma_Y} \right) C \left(\frac{\Lambda}{\mu} \right), \quad (8.42)$$

where $\sigma_I = \alpha_I + \mu$, $\sigma_Y = \alpha_Y + \mu$.

\mathcal{R}_0 is given by the product of the three factors (epidemiological parameters): λ (the probability of transmission per partner), $C(\Lambda/\mu)$ (the mean number of sexual partners that an average susceptible individual has per unit time when everybody in the population is susceptible), and

$$D = \left(\frac{p}{\sigma_I} + \frac{1-p}{\sigma_Y} \right). \quad (8.43)$$

The death-adjusted mean infective period is $D = pD_I + (1-p)D_Y$ with D_I and D_Y denoting the death-adjusted mean infective periods, $1/\sigma_I$ and $1/\sigma_Y$ of the I and Y classes, respectively.

Question 5 Show that if $\mathcal{R}_0 < 1$, the disease-free equilibrium $(\Lambda/\mu, 0, 0)$ of the system (8.39) is asymptotically stable, and if $\mathcal{R}_0 > 1$ there is a unique endemic equilibrium (S^*, I^*, Y^*) , which is locally asymptotically stable, and the disease-free state $(\Lambda/\mu, 0, 0)$ is unstable.

Next, we allow arbitrary incubation period distributions by introducing two functions $P_I(s)$ and $P_Y(s)$ representing the proportion of individuals who become I - or Y -infective at time t and that, if alive, are still infective at time $t+s$ (survive as infectious). P_I and P_Y , survivorship functions, are non-negative and non-increasing, and $P_I(0) = P_Y(0) = 1$. It is further assumed that

$$\int_0^\infty P_I(s)ds < \infty, \quad \int_0^\infty P_Y(s)ds < \infty,$$

and thus, $-\dot{P}(x)$ and $-\dot{P}_Y(x)$ are the rates of removal of individuals from classes I and Y into classes A and Z , x time units after infection.

Question 6 Derive the corresponding model and determine its basic reproduction number.

References

1. Abu-Raddad, L.J., A.S. Magaret, C., Celum, A. Wald, I.M. Longini Jr, S.G. Self, and L. Corey (2008) Genital herpes has played a more important role than any other sexually transmitted infection in driving HIV prevalence in Africa. *PLoS one*, **3**(5): e2230.
2. Alvey, C., Z. Feng, and J.W. Glasser (2015) A model for the coupled disease dynamics of HIV and HSV-2 with mixing among and between genders. *Math. Biosci.* **265**: 82–100.
3. Anderson, R.M., R.M. May, G.F. Medley, and A. Johnson (1986) A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS, *IMA J. Math. Med. Bio.* **3**: 229–263.
4. Anderson, R.M. and R.M. May (1987) Transmission dynamics of HIV infection, *Nature* **326**: 137–142.
5. Anderson, R.M. (1988) The epidemiology of HIV infection: variable incubation plus infectious periods and heterogeneity in sexual activity, *J. Roy. Statistical Society A*. **151**: 66–93.
6. Anderson, R.M., D.R. Cox, and H.C. Hillier (1989) Epidemiological and statistical aspects of the AIDS epidemic: introduction, *Phil. Trans. Roy. Soc. Lond. B* **325**: 39–44.
7. Anderson, R. M., S.P. Blythe, S. Gupta, and E. Konings (1989) The transmission dynamics of the human immunodeficiency virus type 1 in the male homosexual community in the United Kingdom: the influence of changes in sexual behavior, *Phil. Trans. R. Soc. Lond. B* **325**: 145–198.
8. Anderson, R.M. and R.M. May (1991) *Infectious Diseases of Humans*, Oxford Science Publications, Oxford.
9. Aparicio J.P., A.F. Capurro, and C. Castillo-Chávez (2002) Markers of disease evolution: the case of tuberculosis, *J. Theor. Biol.* **215**: 227–237.
10. Bailey, N.T.J. (1988) Statistical problems in the modeling and prediction of HIV/AIDS, *Aust. J. Stat.* **30A**: 41–55.
11. Barré-Sinoussi, F., J.C. Chermann, F. Rey, M.T. Nugeyre, S. Chamaret, J. Gruest, C. Dauguet, C. Axler-Blin, F. Vézinet-Brun, C. Rouzioux, et al (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS), *Science* **220**: 868–870.
12. Blower, S.M., A.N. Aschenbach, H.B. Gershengorn, and J.O. Kahn (2001) Predicting the unpredictable: transmission of drug-resistant HIV. *Nature medicine*, **7**(9): 1016.
13. Blower, S.M. and H. Dowlatabadi (1994) Sensitivity and uncertainty analysis of complex models of disease transmission: an HIV model, as an example. *International Statistical Review/Revue Internationale de Statistique*, 229–243.
14. Blower, S.M., H.B. Gershengorn, and R.M. Grant (2000) A tale of two futures: HIV and antiretroviral therapy in San Francisco. *Science*, **287**(5453): 650–654.
15. Blower, S. M., K. Koelle, D.E. Kirschner, and J. Mills (2001) Live attenuated HIV vaccines: predicting the tradeoff between efficacy and safety. *Proc. Natl. Acad. Sci.* **98**(6): 3618–3623.
16. Blower, S., and L. Ma (2004) Calculating the contribution of herpes simplex virus type 2 epidemics to increasing HIV incidence: treatment implications. *Clinical Infectious Diseases*, **39**(Supplement 5), S240–S247.

17. Blower, S.M., T.C. Porco, and G. Darby (1998) Predicting and preventing the emergence of antiviral drug resistance in HSV-2. *Nature medicine*, **4**(6): 673.
18. Blower S., P. Small, and P. Hopewell (1996) Control strategies for tuberculosis epidemics: new models for old problems, *Science*, **273**: 497–500.
19. Blythe, S.P. and R.M. Anderson (1988) Distributed incubation and infectious periods in models of the transmission dynamics of the human immunodeficiency virus (HIV), *IMA J. Math. Med. Bio.* **5**: 1–19.
20. Blythe, S.P. and C. Castillo-Chavez (1989) Like-with-like preference and sexual mixing models, *Math. Biosci.* **96**: 221–238.
21. Blythe, S.P., C. Castillo-Chavez, J. Palmer, and M. Cheng (1991) Towards a unified theory of mixing and pair formation, *Math. Biosc.* **107**: 379–405.
22. Blythe S.P., K. Cooke, C. Castillo-Chavez (1991) Autonomous risk-behavior change, and non-linear incidence rate, in models of sexually transmitted diseases, *Biometrics Unit Technical Report B-1048-M*.
23. Blythe, S.P., C. Castillo-Chavez and G. Casella (1992) Empirical methods for the estimation of the mixing probabilities for socially structured populations from a single survey sample, *Math. Pop. Studies.* **3**: 199–225.
24. Blythe, S.P., S. Busenberg and C. Castillo-Chavez (1995) Affinity and paired-event probability, *Math. Biosc.* **128**: 265–284.
25. Boily, M.C., F.I. Bastos, K. Desai, and B. Masse (2004) Changes in the transmission dynamics of the HIV epidemic after the wide-scale use of antiretroviral therapy could explain increases in sexually transmitted infections: results from mathematical models, *Sexually transmitted diseases*, **31**(2): 100–113.
26. Brookmeyer, R. and M. H. Gail (1988) A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic, *J. Am. Stat. Assoc.*, **83**:301–308.
27. Busenberg, S., and C. Castillo-Chavez (1989) Interaction, Pair Formation and Force of Infection Terms in Sexually Transmitted Diseases, *Lect. Notes Biomath.* **83**, Springer-Verlag, New York.
28. Busenberg, S., and C. Castillo-Chavez (1991) A general solution of the problem of mixing subpopulations, and its application to risk- and age-structured epidemic models for the spread of AIDS. *IMA J. Math. Applied in Med. and Biol.* **8**: 1–29.
29. Castillo-Chavez, C., H.W. Hethcote, V. Andreasen, S.A. Levin, S.A. and W-M, Liu (1988) Cross-immunity in the dynamics of homogeneous and heterogeneous populations, *Mathematical Ecology*, T. G. Hallam, L. G. Gross, and S. A. Levin (eds.), World Scientific Publishing Co., Singapore, pp. 303–316.
30. Castillo-Chavez, C., ed. (1989) Mathematical and Statistical Approaches to AIDS Epidemiology, *Lect. Notes Biomath.* **83**, Springer-Verlag, Berlin-Heidelberg-New York.
31. Castillo-Chavez, C. (1989) Review of recent models of HIV/AIDS transmission, in *Applied Mathematical Ecology* (ed. S. Levin), *Biomathematics Texts*, Springer-Verlag, **18**: 253–262.
32. Castillo-Chavez, C., K. Cooke, W. Huang, S.A. Levin (1989) The role of long periods of infectiousness in the dynamics of acquired immunodeficiency syndrome. In: Castillo-Chavez, C., S.A. Levin, C. Shoemaker (eds.) *Mathematical Approaches to Resource Management and Epidemiology*, (*Lecture Notes Biomathematics*, **81**, Springer-Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, pp. 177–189.
33. Castillo-Chavez, C., K.L. Cooke, W. Huang, and S.A. Levin (1989) Results on the dynamics for models for the sexual transmission of the human immunodeficiency virus, *Applied Math. Letters*, **2**: 327–331.
34. Castillo-Chavez, C., K. Cooke, W. Huang, and S.A. Levin (1989) On the role of long incubation periods in the dynamics of HIV/AIDS. Part 2: Multiple group models, *Mathematical and Statistical Approaches to AIDS Epidemiology*, C. Castillo-Chávez, (ed.), *Lecture notes in Biomathematics* **83**, Springer-Verlag, Berlin-Heidelberg-New York, pp. 200–217.
35. Castillo-Chavez, C., K. Cooke, W. Huang, and S.A. Levin (1989) The role of long incubation periods in the dynamics of HIV/AIDS. Part 1: Single populations models, *J. Math. Biol.* **27**: 373–398.

36. Castillo-Chavez, C. and S. Busenberg (1990) On the solution of the two-Sex mixing problem, Proceedings of the International Conference on Differential Equations and Applications to Biology and Population Dynamics, S. Busenberg and M. Martelli (eds.), Lecture Notes in Biomathematics Springer-Verlag, Berlin-Heidelberg-New York **92**: 80–98.
37. Castillo-Chavez, C., S. Busenberg and K. Gerow (1990) Pair formation in structured populations, Differential Equations with Applications in Biology, Physics and Engineering, J. Goldstein, F. Kappel, W. Schappacher (eds.), Marcel Dekker, New York. pp. 4765.
38. Castillo-Chavez, C., J.X. Velasco-Hernandez, and S. Fridman (1993) Modeling contact structures in biology,(Lect. Notes Biomath. **100**, Springer-Varlag.
39. Castillo-Chavez, C., W. Huang and J. Li (1996) On the existence of stable pair distributions, *J. Math. Biol.* **34**: 413–441.
40. Castillo-Chavez, C. and Z. Feng (1998) Mathematical models for the disease dynamics of tuberculosis, in Advances in mathematical population dynamics-molecules, cells and man (eds. M.A. Horn, G. Simonett, and G. Webb), Vanderbilt University Press, 117–128.
41. Castillo-Chavez, C. and S-F Hsu Schmitz (1997) The evolution of age-structured marriage functions: It takes two to tango, In, Structured-Population Models Marine, Terrestrial, and Freshwater Systems. S. Tuljapurkar and H. Caswell, (eds.), Chapman & Hall, New York, pages 533–550.
42. Castillo-Chavez, C. and Z. Feng (1997) To treat or not to treat: The case of tuberculosis, *J. Math. Biol.*, **35**: 629–656.
43. Castillo-Chavez, C. and Z. Feng (1998) Global stability of an age-structure model for TB and its applications to optimal vaccination, *Math. Biosc.* **151**: 135–154.
44. Cohen, M.S., N. Hellmann, J.A. Levy, K. DeCock, and J. Lange (2008) The spread, treatment, and prevention of HIV-1: evolution of a global pandemic. *The Journal of clinical investigation*, **118**(4): 1244–1254.
45. Corey, L., A. Wald, R. Patel, S.L. Sacks, S.K. Tyring, T. Warren, T., ... and L.S. Stratchounsky (2004) Once-daily valacyclovir to reduce the risk of transmission of genital herpes. *New England Journal of Medicine*, **350**(1): 11–20.
46. Cox, D.R. and G.F. Medley (1989) A process of events with notification delay and the forecasting of AIDS, *Phil. Trans. Roy. Soc. Lond. B* **325**: 135–145.
47. Crawford, C.M., S.J. Schwager, and C. Castillo-Chavez (1990) A methodology for asking sensitive questions among college undergraduates, Technical Report #BU-1105-M in the Biometrics Unit series, Cornell University, Ithaca, NY.
48. Dietz, K. (1988) On the transmission dynamics of HIV, *Math. Biosc.* **90**: 397–414.
49. Dietz, K. and K.P. Hadeler (1988) Epidemiological models for sexually transmitted diseases, *J. Math. Biol.* **26**: 1–25.
50. Feng, Z. and C. Castillo-Chavez (2000) A model for Tuberculosis with exogenous reinfection, *Theor. Pop. Biol.*, **57**: 235–247.
51. Feng, Z., W. Huang, and C. Castillo-Chavez (2001) On the role of variable latent periods in mathematical models for tuberculosis, *J. Dyn. Differential Equations*, **13**: 425–452.
52. Feng, Z., Z. Qiu, Z. Sang, C. Lorenzo, and J.W. Glasser (2013) Modeling the synergy between HSV-2 and HIV and potential impact of HSV-2 therapy. *Math. Biosci.* **245**(2): 171–187.
53. Foss, A.M., P.T. Vickerman, Z. Chalabi, P. Mayaud, M. Alary, and C.H. Watts (2009) Dynamic modeling of herpes simplex virus type-2 (HSV-2) transmission: issues in structural uncertainty. *Bull Math. Biol.* **71**(3): 720–749.
54. Foss, A.M., P.T. Vickerman, P. Mayaud, H.A. Weiss, B.M. Ramesh, S. Reza-Paul, S., ... and M. Alary (2011) Modelling the interactions between herpes simplex virus type 2 and HIV: implications for the HIV epidemic in southern India. *Sexually transmitted infections*, **87**(1): 22–27.
55. Gallo, R.C., S.Z. Salahuddin, M. Popovic, G.M. Shearer, M. Kaplan, B.F. Haynes, T. Palker, R. Redfield, J. Oleske, B. Safai, G. White, P. Foster, P.D., Markhamet (1984) Frequent detection and isolation of sytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS, *Science* **224**: 500–503.
56. Gallo, R.C. (1986) The first human retrovirus, *Scientific American* **255**: 88–98.

57. Gupta S., R.M. Anderson, and R.M. May (1989) Networks of sexual contacts: implications for the pattern of spread of HIV, *AIDS* **3**: 1–11.
58. Francis, D.P., P.M. Fenton, J.R. Broder, H.M. McClure, J.P. Getchell, C.R. McGrath, B. Swenson, J.S. McDougal, E.L. Palmer, and A.K. Harrison (1984) Infection of chimpanzees with lymphadenopathy-associated virus, *Lancet* **2**: 1276–1277.
59. Hadeler, K.P. (1989) Modeling AIDS in structured populations, 47th Session of the International Statistical Institute, Paris, August/September. Conf. Proc., C1-2: 83–99.
60. Hadeler, K.P. and C. Castillo-Chavez (1995) A core group model for disease transmission, *Math. Biosc.* **128**: 41–55.
61. Hethcote, H.W., H.W. Stech, P. van den Driessche (1981) Nonlinear oscillations in epidemic models, *SIAM J. Appl. Math.* **40**: 1–9.
62. Hethcote, H.W., J.W. van Ark (1987) Epidemiological methods for heterogeneous populations: proportional mixing, parameter estimation, and immunization programs. *Math. Biosc.* **84**: 85–118.
63. Hethcote, H.W., and J.W. Van Ark (1992) Modeling HIV Transmission and AIDS in the United States, *Lecture Notes in Biomathematics* **95**, Springer-Verlag, Berlin-Heidelberg-New York.
64. Hopf, E. (1942) Abzweigung einer periodischen Lösungen von einer stationären Lösung eines Differentialsystems, *Berlin Math-Phys. Sachsische Akademie der Wissenschaften*, Leipzig, **94**: 1–22.
65. Hsu Schmitz, S.F. (1993) Some theories, estimation methods and applications of marriage functions and two-sex mixing functions in demography and epidemiology. Unpublished doctoral dissertation, Cornell University, Ithaca, New York, U.S.A.
66. Hsu Schmitz S.F. and C. Castillo-Chavez (1994) Parameter estimation. *Brit. Med. J.* **293**: 1459–1462.
67. Huang, W., K.L. Cooke, and C. Castillo-Chavez, (1992) Stability and bifurcation for a multiple-group model for the dynamics of HIV/AIDS transmission, *SIAM J. Appl. Math.* ltextbf{52}: 835–854.
68. Hyman, J.M., E.A. Stanley (1988) *A risk base model for the spread of the AIDS virus*. *Math. Biosciences* **90** 415–473.
69. Hyman, J.M. and E.A. Stanley (1989) *The Effects of Social Mixing Patterns on the Spread of AIDS*, Mathematical Approaches to Problems in Resource Management and Epidemiology, (Ithaca, NY, 1987), 190–219, *Lecture Notes in Biomathematics*, **81**, C. Castillo-Chavez, S. A. Levin, and C. A. Shoemaker (Eds.), Springer, Berlin.
70. Isham, V. (1989) Estimation of the incidence of HIV infection, *Phil. Trans. Roy. Soc. Lond. B*, **325**: 113–121.
71. Kaplan, E.H. *What Are the Risks of Risky Sex?*, Operations Research, 1989.
72. Kingsley, R. A., R. Kaslow, C.R. Jr Rinaldo, K. Detre, N. Odaka, M. VanRaden, R. Detels, B.F. Polk, J. Chimel, S.F. Kersey, D. Ostrow, B. Visscher (1987) *Risk factors for seroconversion to human immunodeficiency virus among male homosexuals*, *Lancet* **1**, 345–348.
73. Kirschner, D. (1999) Dynamics of co-infection with *M. tuberculosis* and HIV-1, *Theor. Pop. Biol.*, **55**: 94–109.
74. Koelle, K., S. Cobey, B. Grenfell, M. Pascual (2006) Epochal evolution shapes the phylogenetics of interpandemic influenza A (H3N2) in Humans Science **314**: 1898–1903.
75. Koopman, J., C.P. Simon, J.A. Jacquez, J. Joseph, L. Sattenspiel and T Park (1988) Sexual partner selectiveness effects on homosexual HIV transmission dynamics. *Journal of AIDS* **1**: 486–504.
76. Lagakos, S.W., L. M. Barraj, and V. de Gruttola (1988) Nonparametric analysis of truncated survival data, with applications to AIDS, *Biometrika*, **75**: 515–523.
77. Lange, J. M. A., Paul, D. A., Huisman, H. G., De Wolf, F., Van den Berg, H., Roe!, C. A., Danner, S. A., Van der Noordaa, J., Goudsmit, J. *Persistent HIV antigenemia and decline of HIV core antibodies associated with transition to AIDS*. *Brit. Med. J.* **293**, 1459–1462 (1986).

78. Luo, X., and C. Castillo-Chavez. (1991) Limit behavior of pair formation for a large dissolution rate. *J. Mathematical Systems, Estimation, and Control*, **3**: 247–264.
79. May, R.M. and R.M. Anderson (1989) Possible demographic consequence of HIV/AIDS epidemics: II, assuming HIV infection does not necessarily lead to AIDS, in: *Mathematical Approaches to Problems in Resource Management and Epidemiology*, C. Castillo-Chavez, S.A. Levin, and C.A. Shoemaker (Eds.) *Lecture Notes in Biomathematics* **81**, Springer-Verlag, Berlin-Heidelberg, New York, London, Paris, Tokyo, Hong Kong, pp. 220–248.
80. May, R.M. and R.M. Anderson (1989) The transmission dynamics of human immunodeficiency virus (HIV), in *Applied Mathematical Ecology*, (ed. S. Levin), *Biomathematics Texts*, **18**, Springer-Verlag, New York.
81. Medley, G.F., R.M. Anderson, D.R. Cox, and L. Billiard (1987) Incubation period of AIDS in patients infected via blood transfusions, *Nature* **328**: 719–721.
82. Miller, R.K. (1971) The implications and necessity of affinity, *J. Biol. Dyn.* **4**: 456–477.
83. Morin, B., Castillo-Chavez, C. Hsu Schmitz, S-F. Mubayi, A., and X. Wang. Notes From the Heterogeneous: A Few Observations on the Implications and Necessity of Affinity. *Journal of Biological Dynamics*, Vol. **4**, No. 5, 2010, 456–477.
84. Mukandavire, Z., and W. Garira (2007) Age and sex structured model for assessing the demographic impact of mother-to-child transmission of HIV/AIDS. *Bull. Math. Biol.* **69**: 2061–2092.
85. Mukandavire, Z., and W. Garira (2007) Sex-structured HIV/AIDS model to analyse the effects of condom use with application to Zimbabwe. *J. Math. Biol.* **54**(5): 669–699.
86. Naresh, R. and A. Tripathi (2005) Modelling and analysis of HIV-TB Co-infection in a variable size population, *Mathematical Modelling and Analysis*, **10**: 275–286.
87. Newton, E. A., and J.M. Kuder (2000) A model of the transmission and control of genital herpes. Sexually transmitted diseases, **27**: 363–370.
88. Pickering, J., J.A. Wiley, N.S. Padian, et al. (1986) Modeling the incidence of acquired immunodeficiency syndrome (AIDS) in San Francisco, Los Angeles, and New York, *Math. Modelling* **7**: 661–688.
89. Porco T. and S. Blower (1998) Quantifying the intrinsic transmission dynamics of tuberculosis, *Theor. Pop. Biol.*, **54**: 117–132.
90. Porco, T., P. Small, and S. Blower (2001) Amplification dynamics: predicting the effect of HIV on tuberculosis outbreaks, *Journal of AIDS*, **28**: 437–444.
91. Raimundo, S.M., A.B. Engel, H.M. Yang, and R.C. Bassanezi (2003) An approach to estimating the transmission coefficients for AIDS and for tuberculosis using mathematical models, *Systems Analysis Modelling Simulation*, **43**: 423–442.
92. Roeger, L.-I.W., Z. Feng and C. Castillo-Chavez (2009) The impact of HIV infection on tuberculosis, *Math. Biosc. Eng.* **6**: 815–837.
93. Rubin, G., D. Umbauch, D., S.-F. Shyu and C. Castillo-Chavez (1992) Application of capture-recapture methodology to estimation of size of population at risk of AIDS and/or Other sexually-transmitted diseases, *Statistics in Medicine* **11**: 1533–49.
94. Salahuddin, S.Z., J.E. Groopman, P.D. Markham, M.G. Sarngaharan, R.R. Redfield, M.F. McLane, M. Essex, A. Sliski, R.C. Gallo (1984) HTLV-III in symptom-free seronegative persons, *Lancet* **2**: 1418–1420.
95. Sattenspiel, L. (1989) The structure and context of social interactions and the spread of HIV. In *Mathematical and Statistical Approaches to AIDS Epidemiology*, Castillo-Chavez, C. (ed.) *Lecture Notes in Biomathematics* **83**. Berlin: Springer-Verlag, pp. 242–259.
96. Sattenspiel, L., J. Koopman, C.P. Simon, and J.A. Jacquez (1990) The effects of population subdivision on the spread of the HIV infection, *Am. J. Physical Anthropology* **82**: 421–429.
97. Sattenspiel, L. and C. Castillo-Chavez (1990) Environmental context, social interactions, and the spread of HIV, *Am. J. Human Biology* **2**: 397–417.
98. Schinazi, R.B. (2003) Can HIV invade a population which is already sick? *Bull. Braz. Math. Soc. (N.S.)*, **34**: 479–488.

99. Schwager, S., C. Castillo-Chavez, and H.W. Hethcote (1989) Statistical and mathematical approaches to AIDS epidemiology: A review, In: C. Castillo-Chávez (ed.), Mathematical and Statistical Approaches to AIDS Epidemiology, pp. 2–35. Lecture Notes in Biomathematics, Vol. **83**, Springer-Verlag: Berlin.
100. Schinazi, R. B. (1999) Strategies to control the genital herpes epidemic. *Math. Biosci.* **159**(2): 113–121.
101. Schulzer, M., M.P. Radhamani, S. Grybowski, E. Mak, and J.M. Fitzgerald (1994) A mathematical model for the prediction of the impact of HIV infection on tuberculosis, *Int. J. Epidemiol.*, **23**: 400–407.
102. Thieme, H. and C. Castillo-Chavez (1989) On the role of variable infectivity in the dynamics of the human immunodeficiency virus epidemic, Mathematical and statistical approaches to AIDS epidemiology, C. Castillo-Chavez, (ed.), pp. 157–176. Lecture Notes in Biomathematics **83**, Springer-Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong.
103. Thieme, H.R. and C. Castillo-Chavez (1993) How may infection-age dependent infectivity affect the dynamics of HIV/AIDS?, *SIAM J. Appl. Math.*, **53**: 1447–1479.
104. Wald, A., A.G. Langenberg, K. Link, A.E. Izu, R. Ashley, T. Warren, ... and L. Corey (2001) Effect of condoms on reducing the transmission of herpes simplex virus type 2 from men to women. *JAMA*, **285**(24): 3100–3106.
105. West R. and J. Thompson (1996) Modeling the impact of HIV on the spread of tuberculosis in the United States, *Math. Biosci.*, **143**: 35–60.
106. White, R.G., E.E. Freeman, K.K. Orroth, R. Bakker, H.A. Weiss, N. O'farrell, ... and J.R. Glynn (2008) Population-level effect of HSV-2 therapy on the incidence of HIV in sub-Saharan Africa. Sexually transmitted infections, **84**(Suppl 2): ii12–ii18.
107. Wong-Staal, F., R.C. Gallo (1985) Human T-lymphotropic retroviruses. *Nature* **317**: 395–403.
108. Wu L.-I., and Z. Feng (2000) Homoclinic bifurcation in an SIQR model for childhood diseases, *J. Diff. Equ.* **168**: 150–167.

Chapter 9

Models for Influenza



9.1 Introduction to Influenza Models

Influenza causes more morbidity and more mortality than all other respiratory diseases together. There are annual seasonal epidemics that cause about 500,000 deaths worldwide each year. During the twentieth century there were three influenza pandemics. The World Health Organization estimates that there were 40,000,000–50,000,000 deaths worldwide in the 1918 pandemic, 2,000,000 deaths worldwide in the 1957 pandemic, and 1,000,000 deaths worldwide in the 1968 pandemic. There has been concern since 2005 that the H5N1 strain of avian influenza could develop into a strain that can be transmitted readily from human to human and develop into another pandemic, together with a widely held belief that even if this does not occur there is likely to be an influenza pandemic in the near future. More recently, the H1N1 strain of influenza did develop into a pandemic in 2009, but fortunately its case mortality rate was low and this pandemic turned out to be much less serious than had been feared. There were 18,500 confirmed deaths, but the actual number of deaths caused by the H1N1 influenza may have been as many as 200,000. This history has aroused considerable interest in modeling both the spread of influenza and comparison of the results of possible management strategies.

Vaccines are available for annual seasonal epidemics. Influenza strains mutate rapidly, and each year a judgment is made of which strains of influenza are most likely to invade. A vaccine is distributed that protects against the three strains considered most dangerous. However, if a strain radically different from previously known strains arrives, vaccine provides little or no protection and there is danger of a pandemic. As it would take at least 6 months to develop a vaccine to protect against such a new strain, it would not be possible to have a vaccine ready to protect against the initial onslaught of a new pandemic strain. Attempts are underway to try to develop a more universal vaccine.

Antiviral drugs are available to treat pandemic influenza, and they may have some preventive benefits as well, but such benefits are present only while antiviral treatment is continued.

Various kinds of models have been used to describe influenza outbreaks. Many public health policy decisions on coping with a possible influenza pandemic are based on construction of a contact network for a population and analysis of disease spread through this network. This analysis consists of multiple stochastic simulations requiring a substantial amount of computer time. In advance of an epidemic it is not possible to know its severity, and it would be necessary to make estimates for a range of reproduction numbers. Also, model parameters for the H1N1 influenza pandemic of 2009, especially the susceptibility to infection for different age groups, were significantly different from those for seasonal epidemics. In advance of an anticipated pandemic it may be more appropriate to use simpler models until enough data are acquired to facilitate parameter estimation. Early estimation of model parameters is extremely important for coping with a serious epidemic, and one of the outcomes of the H1N1 influenza pandemic of 2009 was development of new methods, mainly based on network models, to achieve this.

Our approach is to begin with simple models and to add more structure later as more information is obtained. When an epidemic does begin, plans for management strategies need to be very detailed, and use of the simple models we describe here should be restricted to advance planning and broad understanding.

We begin this chapter by developing a simple compartmental influenza transmission model and then augmenting it to include both pre-epidemic vaccination and treatment during an epidemic. We will also describe some ways in which the model can be modified to be more realistic, though more complicated. The development follows the treatment in [6, 7]. We will describe the models and the results of their analyses, but omit proofs in order to focus attention on the applications of the models. Many of the results may be found in earlier chapters.

9.2 A Basic Influenza Model

Since influenza epidemics usually come and go in a time period of several months, we do not include demographic effects (births and natural deaths) in our model. Our starting point is the simple *SIR* model. Two aspects of influenza that are easily added are that there is an incubation period between infection and the appearance of symptoms, and that a significant fraction of people who are infected never develop symptoms but go through an asymptomatic period, during which they have some infectivity, and then recover and go to the removed compartment [35]. Thus a model should contain the compartments *S* (susceptible), *L* (latent), *I* (infective), *A* (asymptomatic), and *R* (removed). Specifically, we make the following assumptions:

1. There is a small number I_0 of initial infectives in a population of constant total size N .

2. The number of contacts in unit time per individual is a constant multiple β of total population size N .
3. Latent members (L) are not infective.
4. A fraction p of latent members proceed to the infective compartment at rate κ , while the remainder goes directly to an asymptomatic infective compartment (A), also at rate κ .
5. There are no disease deaths; infectives (I) recover and leave the infective compartment at rate α , and go to the removed compartment (R).
6. Asymptomatics have infectivity reduced by a factor δ , and go to the removed compartment at rate η .

These assumptions lead to the model

$$\begin{aligned} S' &= -S\beta(I + \delta A) \\ L' &= S\beta(I + \delta A) - \kappa L \\ I' &= p\kappa L - \alpha I \\ A' &= (1-p)\kappa L - \eta A \\ R' &= \alpha I + \eta A \end{aligned} \tag{9.1}$$

with initial conditions

$$S(0) = S_0, \quad L(0) = 0, \quad I(0) = I_0, \quad A(0) = 0, \quad R(0) = 0, \quad N = S_0 + I_0.$$

In analyzing this model we may remove one variable since $N = S + L + I + A + R$ is constant. It is usually convenient to remove the variable R . It is possible to show that the model (9.1) is properly posed in the sense that all variables remain non-negative for $0 \leq t < \infty$. A flow diagram for the model (9.1) is shown in Fig. 9.1. The model (9.1) is the simplest possible description for influenza having the property that there are asymptomatic infections. The question that should be in the back of our minds is whether it is a sufficiently accurate description for its predictions to be useful.

The model (9.1), like the other models that we will introduce later, consists of a system of ordinary differential equations. The number of susceptibles in the population tends to a limit S_∞ as $t \rightarrow \infty$ and the final size relation may be used to find this limit without the need to solve the system of differential equations. It is more realistic to assume saturation of contacts and that β is a function of the total population size N . In general, the final size relation is an inequality. If there are no disease deaths, N is constant and β is constant even with saturation of contacts. If the disease death rate is small, it appears that the final size relation is very close to an equality and it is reasonable to assume that β is constant and use the final size relation as an equation to solve for S_∞ .

We may use the next generation matrix approach of [48] to calculate the basic reproduction number

$$\mathcal{R}_0 = \beta N \left[\frac{p}{\alpha} + \frac{\delta(1-p)}{\eta} \right]. \tag{9.2}$$

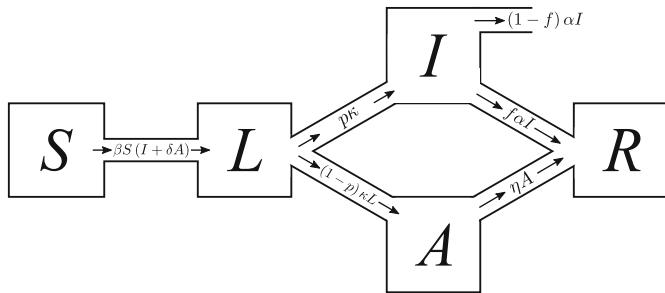


Fig. 9.1 Diagram for the basic influenza model (9.1)

A biological interpretation of this basic reproduction number is that a latent member introduced into a population of N susceptibles becomes infective with probability p , in which case he or she causes $\beta N/\alpha$ infections during an infective period of length $1/\alpha$, or becomes asymptomatic with probability $1 - p$, in which case he or she causes $\delta\beta N/\eta$ infections during an asymptomatic period of length $1/\eta$.

The *final size relation* is given by

$$\log \frac{S_0}{S_\infty} = \mathcal{R}_0 \left[1 - \frac{S_0}{N} \right]. \quad (9.3)$$

A very general form of the final size relation that is applicable to each of the models in this chapter is derived in Sect. 4.5. The final size relation shows that $S_\infty > 0$. This means that some members of the population are not infected during the epidemic. The size of the epidemic, the number of (clinical) cases of influenza during the epidemic, is

$$I_0 + S_0 - S_\infty = N - S_\infty,$$

and the number of symptomatic cases is

$$I_0 + p(S_0 - S_\infty).$$

If there are disease deaths, with a disease survival rate f and a disease death rate $(1 - f)$ among infectives (assuming no disease deaths of asymptomatics), the number of disease deaths is

$$I_0 + (1 - f)p(S_0 - S_\infty).$$

While mathematicians view the basic reproduction number as central in studying epidemiological models, epidemiologists may be more concerned with the *attack rate*, as this may be measured directly. For influenza, where there are asymptomatic

cases, there are two attack rates. One is the clinical attack rate, which is the fraction of the population that becomes infected, defined as

$$1 - \frac{S_\infty}{N}.$$

There is also the symptomatic attack rate, defined as the fraction of the population that develops disease symptoms, defined as

$$p \left[1 - \frac{S_\infty}{N} \right].$$

The attack rates and the basic reproduction number are connected through the final size relation (9.3). If we know the parameters of the model we can calculate \mathcal{R}_0 from (9.2) and then solve for S_∞ from (9.3).

We apply the model (9.1) using parameters appropriate for the 1957 influenza pandemic as suggested by [35]. The latent period is approximately 1.9 days and the infective period is approximately 4.1 days, so that

$$\kappa = \frac{1}{1.9} = 0.526, \quad \alpha = \eta = \frac{1}{4.1} = 0.244.$$

We also take

$$p = 2/3, \quad \delta = 0.5, \quad f = 0.98.$$

As in [35] we consider a population of 2000 members, of whom 12 are infective initially. In [35] a symptomatic attack rate was assumed for each of the four age groups, and the average symptomatic attack rate for the entire population was 0.326. This implies $S_\infty = 1022$. Then we obtain

$$\mathcal{R}_0 = 1.37$$

from (9.3). Now, we use this in (9.2) to calculate

$$\beta N = 0.402.$$

We will use these data as baseline values to estimate the effect that control measures might have had. The number of clinical cases is 978 (including the initial 12), the number of symptomatic cases is 664, again including the original 12, and the number of disease deaths is approximately 13.

The model (9.1) can be adapted to describe management strategies for both annual seasonal epidemics and pandemics.

9.2.1 Vaccination

To cope with annual seasonal influenza epidemics, there is a program of vaccination before the “flu” season begins. Each year a vaccine is produced aimed at protecting against the three influenza strains considered most dangerous for the coming season. We formulate a model to add vaccination to the model described by (9.1) under the assumption that vaccination reduces susceptibility (the probability of infection if a contact with an infected member of the population is made). In addition, if we assume that vaccinated members who develop infection are less likely to transmit infection, more likely not to develop symptoms, and are likely to recover more rapidly than unvaccinated members.

These assumptions require us to introduce additional compartments into the model to follow treated members of the population through the stages of infection. We use the classes S, L, I, A, R as before and introduce S_T , the class of treated susceptibles, L_T , the class of treated latent members, I_T , the class of treated infectives, and A_T , the class of treated asymptomatics. In addition to the assumptions made in formulating the model (9.1) we also assume

- A fraction γ of the population is vaccinated before a disease outbreak and vaccinated members have susceptibility to infection reduced by a factor σ_S .
- There are decreases σ_I and σ_A , respectively, in infectivity in I_T , and A_T ; it is reasonable to assume

$$\sigma_I < 1, \quad \sigma_A < 1.$$

- The rates of departure from L_T, I_T , and A_T are κ_T, α_T , and η_T , respectively. It is reasonable to assume

$$\kappa \leq \kappa_T, \quad \alpha \leq \alpha_T, \quad \eta \leq \eta_T.$$

- The fractions of members recovering from disease when they leave I and I_T are f and f_T , respectively. It is reasonable to assume $f \leq f_T$. In our analysis we will take $f = f_T = 1$ (no disease deaths).
- Vaccination decreases the fraction of latent members who will develop symptoms by a factor τ , with $0 \leq \tau \leq 1$.

For convenience we introduce the notation

$$Q = I + \delta A + \sigma_I I_T + \delta \sigma_A A_T. \tag{9.4}$$

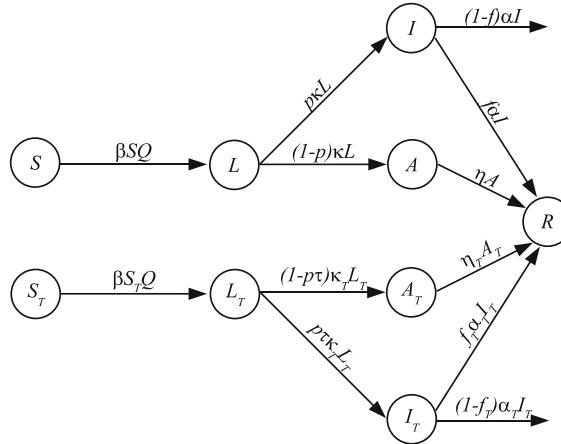
The resulting model is

$$S' = -S\beta Q$$

$$S'_T = -\sigma_S S_T \beta Q$$

$$L' = S\beta Q - \kappa L$$

Fig. 9.2 Transition diagram corresponding to the vaccination model (9.5)



$$\begin{aligned}
 L'_T &= \sigma_S S_T \beta Q - \kappa_T L_T \\
 I' &= p\kappa L - \alpha I \\
 I'_T &= \tau p \kappa_T L_T - \alpha_T I_T \\
 A' &= (1-p)\kappa L - \eta A \\
 A'_T &= (1-p\tau)\kappa_T L_T - \eta_T A_T \\
 R' &= \alpha I + \alpha_T I_T + \eta A + \eta_T A_T.
 \end{aligned} \tag{9.5}$$

The initial conditions are

$$\begin{aligned}
 S(0) &= (1-\gamma)S_0, & S_T(0) &= \gamma S_0, & I(0) &= I_0, & N &= S_0 + I_0, \\
 L(0) &= L_T(0) = I_T(0) = A(0) = A_T(0) = 0
 \end{aligned}$$

corresponding to pre-epidemic treatment of a fraction γ of the population. A flow diagram for the model (9.5) is shown in Fig. 9.2.

Since the infection now is beginning in a population which is not fully susceptible, we speak of the control reproduction number \mathcal{R}_c rather than the basic reproduction number. A computation using the next generation matrix leads to the control reproduction number

$$\mathcal{R}_c = (1-\gamma)\mathcal{R}_u + \gamma\mathcal{R}_v,$$

with

$$\begin{aligned}
 \mathcal{R}_u &= N\beta\left[\frac{p}{\alpha} + \frac{\delta(1-p)}{\eta}\right] = \mathcal{R}_0, \\
 \mathcal{R}_v &= \sigma_S N\beta\left[\frac{p\tau\sigma_I}{\alpha_T} + \frac{\delta(1-p\tau)\sigma_A}{\eta_T}\right].
 \end{aligned} \tag{9.6}$$

Then \mathcal{R}_u is a reproduction number for unvaccinated people and \mathcal{R}_v is a reproduction number for vaccinated people. There is a pair of final size relations for the two final sizes $S(\infty)$ and $S_T(\infty)$ in terms of the group sizes $N_u = (1 - \gamma)N$, $N_v = \gamma N$, namely

$$\begin{aligned} \log \frac{(1 - \gamma)S_0}{S_\infty} &= \mathcal{R}_u \left[1 - \frac{S_\infty}{N_u} \right] + \mathcal{R}_v \left[1 - \frac{S_T(\infty)}{N_v} \right], \\ \log \frac{\gamma S_0}{S_T(\infty)} &= \sigma_S \mathcal{R}_u \left[1 - \frac{S_\infty}{N_u} \right] + \mathcal{R}_v \left[1 - \frac{S_T(\infty)}{N_v} \right]. \end{aligned} \quad (9.7)$$

The number of symptomatic disease cases is

$$I_0 + p[(1 - \gamma)S_0 - S(\infty)] + p\tau[\gamma S_0 - S_T(\infty)],$$

and the number of disease deaths is

$$(1 - f)[I_0 + p(1 - \gamma)S_0 - S(\infty)] + (1 - f_T)p\tau[\gamma S_0 - S_T(\infty)].$$

These may be calculated with the aid of (9.7). By control of the epidemic we mean vaccinating enough people (i.e., taking γ large enough) to make $\mathcal{R}_c < 1$. We use the following parameters, suggested in [35] and [22],

$$\sigma_S = 0.3, \quad \sigma_I = \sigma_A = 0.2, \quad \kappa_T = 0.526, \quad \alpha_T = \eta_T = 0.323, \quad \tau = 0.4.$$

With these parameter values,

$$\mathcal{R}_u = 1.373, \quad \mathcal{R}_v = 0.047.$$

In order to make $\mathcal{R}_c = 1$, we need to take $\gamma = 0.28$. This is the fraction of the population that needs to be vaccinated to head off an epidemic.

We may solve the pair of final size equations with $S(0) = (1 - \gamma)S_0$, $S_T(0) = \gamma S_0$ for $S(\infty)$, $S_T(\infty)$ for different values of γ . We do this for the parameter values suggested above and we obtain the results shown in Table 9.1, giving the treatment fraction γ , the number of untreated susceptibles $S(\infty)$ at the end of the epidemic, the number of treated susceptibles $S_T(\infty)$ at the end of the epidemic, the number of treated cases of influenza $I_0 + p[(1 - \gamma)S_0 - S(\infty)]$, and the number of untreated cases $(\gamma S_0 - S_T(\infty))$. The results indicate the benefits of pre-epidemic vaccination of even a small fraction of the population in reducing the number of influenza cases. They also demonstrate the advantage of vaccination to an individual. The attack rate in the vaccinated portion of the population is much less than the attack rate in the unvaccinated portion of the population.

Table 9.1 Effect of vaccination

Fraction treated	S_∞	$S_T(\infty)$	Untreated cases	Treated cases
0	1015	0	660	0
0.05	1079	84	552	4
0.1	1149	174	439	7
0.15	1224	271	323	7
0.2	1305	375	201	6
0.25	1395	487	76	3
0.3	1391	596	13	0

9.3 Antiviral Treatment

If no vaccine is available for a strain of influenza it would be possible to use an antiviral treatment. However, antiviral treatment affords protection only while the treatment is continued and is more expensive. In addition, antivirals are in short supply and expensive, and treatment of enough of the population to control an anticipated epidemic may not be feasible. A policy of treatment aimed particularly at people who have been infected or who have been in contact with infectives once a disease outbreak has begun may be a more appropriate approach. This requires a model with treatment rates for latent, infective, and asymptotically infected members of the population that we construct building on the structure used for vaccination in (9.5).

Antiviral drugs have effects similar to vaccines in decreasing susceptibility to infection and decreasing infectivity, likelihood of developing symptoms, and length of infective period in case of infection. However, they are likely to be less effective than a well-matched vaccine, especially in the reduction of susceptibility.

Treatment may be given to diagnosed infectives. In addition, one may treat contacts of infectives who are thought to have been infected. This is modeled by treating latent members. In practice, some of those identified by contact tracing and treated would actually be susceptibles, but we neglect this in the model. Although we have allowed treatment of asymptomatics in the model, this is unlikely to be done, and we will describe the results of the model under the assumption $\varphi_A = \theta_A = 0$. However, for generality we retain the possibility of antiviral treatment of asymptomatics in the model. If treatment is given only to infectives, the compartments L_T , A_T are empty and may be omitted from the model.

We add to the model (9.5) antiviral treatment of latent, infective, and asymptotically infected members of the population, but we do not assume an initial treated class. In addition to the assumptions made earlier we also assume

- There is a treatment rate φ_L in L and a rate θ_L of relapse from L_T to L , a treatment rate φ_I in I and a rate θ_I of relapse from I_T to I , and a treatment rate φ_A in A and a rate θ_A of relapse from A_T to A .

The resulting model is

$$\begin{aligned}
 S' &= -\beta S Q \\
 L' &= \beta S Q - \kappa L - \varphi_L L + \theta_L L_T \\
 L'_T &= -\kappa_T L_T + \varphi_L L - \theta_L L_T \\
 I' &= p\kappa L - \alpha I - \varphi_I I + \theta_I I_T \\
 I'_T &= p\tau\kappa_T L_T - \alpha_T I_T + \varphi_I I - \theta_I I_T \\
 A' &= (1-p)\kappa L - \eta A - \varphi_A A + \theta_A A_T \\
 A'_T &= (1-p\tau)\kappa_T L_T - \eta_T A_T + \varphi_A A - \theta_A A_T \\
 N' &= -(1-f)\alpha I - (1-f_T)\alpha_T I_T,
 \end{aligned} \tag{9.8}$$

with Q as in (9.4). The initial conditions are

$$S(0) = S_0, \quad I(0) = I_0, \quad L(0) = L_T(0) = I_T(0) = A(0) = A_T(0) = 0, \quad N = S_0 + I_0.$$

A flow diagram for the model (9.8) is shown in Fig. 9.3.

The calculation of \mathcal{R}_c for the antiviral treatment model (9.8) is more complicated than for models considered previously, but it is possible to show that $\mathcal{R}_c = \mathcal{R}_I + \mathcal{R}_A$ with

$$\begin{aligned}
 \mathcal{R}_I &= \frac{N\beta}{\Delta_I \Delta_L} \left[(\alpha_T + \theta_I + \sigma_I \varphi_I) p\kappa(\kappa_T + \theta_L) + (\theta_I + \sigma_I(\alpha + \varphi_I)) p\tau\kappa_T \phi_L \right] \\
 \mathcal{R}_A &= \frac{\delta N\beta}{\Delta_L} \left[\frac{(1-p)\kappa(\kappa_T + \theta_L)}{\eta} + \frac{\sigma_A(1-p\tau)\kappa_T \varphi_L}{\eta_T} \right], \tag{9.9}
 \end{aligned}$$

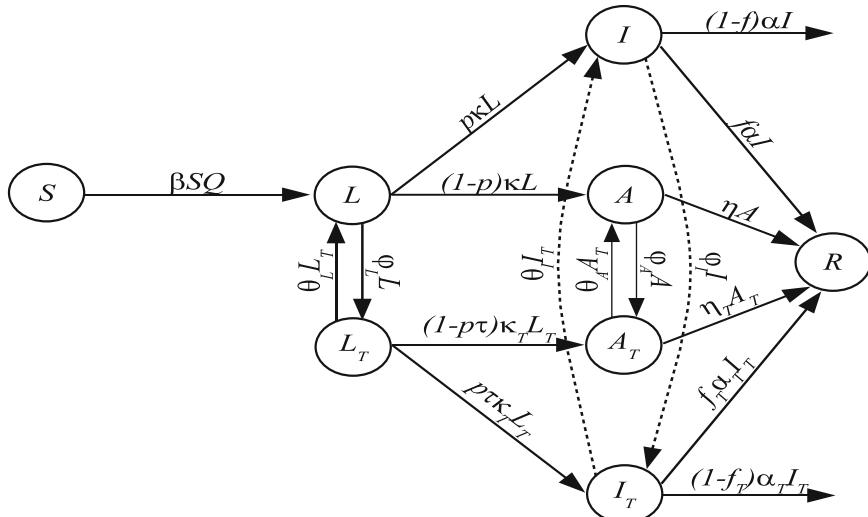


Fig. 9.3 Treatment model (9.8)

where

$$\begin{aligned}\Delta_I &= (\alpha + \varphi_I)(\alpha_T + \theta_I) - \theta_I \varphi_I \\ \Delta_L &= (\kappa + \varphi_L)(\kappa_T + \theta_L) - \theta_L \varphi_L.\end{aligned}$$

The final size equation is

$$\log \frac{S_0}{S_\infty} = \mathcal{R}_c \left[1 - \frac{S_\infty}{S_0} \right] + \frac{\beta I_0(\alpha_T + \theta_I + \sigma_I \varphi_I)}{\Delta_I}. \quad (9.10)$$

The number of people treated is

$$\int_0^\infty [\varphi_L L(t) + \varphi_I I(t)] dt$$

and the number of disease cases is

$$I_0 + \int_0^\infty [p\kappa L(t) + p\tau\kappa_T L_T] dt$$

which can be evaluated in terms of the parameters of the model. The number of people treated and the number of disease cases are constant multiples of $S_0 - S_\infty$ plus constant multiples of the (presumably small) number of initial infectives. Since the expressions in terms of $S_0 - S_\infty$ and I_0 are more complicated than in the cases of no treatment or vaccination, we have chosen to give these numbers in terms of integrals.

There is an important consequence of the calculation of the number of disease cases and treatments that is not at all obvious. If \mathcal{R}_c is close to 1 or ≤ 1 , $S_0 - S_\infty$ depends very sensitively on changes in I_0 . For example, in a population of 2000 with $\mathcal{R}_0 = 1.5$, a change in I_0 from 1 to 2 multiplies $S_0 - S_\infty$ and therefore treatments and cases by 1.4 and a change in I_0 from 1 to 5 multiplies $S_0 - S_\infty$ and therefore treatments and cases by 3. Thus, numerical predictions in themselves are of little value. However, comparison of different strategies is valid, and the model indicates the importance of early action while the number of infectives is small.

In the special case that treatment is applied only to infectives, \mathcal{R}_c is given by the simpler expression

$$\mathcal{R}_c = N\beta \left[\frac{p(\alpha_T + \theta_I + \sigma_I \varphi_I)}{\Delta_I} + \frac{\delta(1-p)}{\eta} \right].$$

The number of disease cases is $I_0 + p(S_0 - S_\infty)$, and the number of people treated is

$$\frac{\varphi_I(\alpha_T + \theta_I)}{\Delta_I} [I_0 + p(S_0 - S_\infty)].$$

Since the infective period is short, antiviral treatment would normally be applied as long as the patient remains infective. Thus we assume $\theta_I = 0$, and these relations become even simpler. The control reproduction number is

$$\mathcal{R}_c = N\beta \left[\frac{p(\alpha_T + \sigma_I \varphi_I)}{\alpha_T(\alpha + \varphi_I)} + \frac{\delta(1-p)}{\eta} \right], \quad (9.11)$$

and the number of people treated is

$$\frac{\varphi_I}{\alpha + \varphi_I} [I_0 + p(S_0 - S_\infty)]. \quad (9.12)$$

Since the basic reproduction number of a future pandemic cannot be known in advance, it is necessary to take a range of contact rates in order to make predictions. In particular, we could compare the effectiveness in controlling the number of infections or the number of disease deaths of different strategies such as treating only infectives, treating only latent members, or treating a combination of both infective and latent members. In making such comparisons, it is important to take into account that treatment of latent members must be supplied for a longer period than for infectives. In case of a pandemic, there are also questions of whether the supply of antiviral drugs will be sufficient to carry out a given strategy. For this reason, we must calculate the number of treatments corresponding to a given treatment rate. It is possible to do this from the model; the result is a constant multiple of I_0 plus a multiple of $S_0 - S_\infty$.

The results of such calculations appear to indicate that treatment of diagnosed infectives is the most effective strategy [6, 7]. However, there are other considerations that would go into any policy decision. For example, a pandemic would threaten to disrupt essential services and it could be decided to use antiviral drugs prophylactically in an attempt to protect health care workers and public safety personnel. A study including this aspect based on the antiviral treatment model given here is reported in [26].

For simulations, we use the initial values

$$S_0 = 1988, \quad I_0 = 12,$$

the parameters of Sect. 9.1, and for antiviral efficacy we assume

$$\sigma_S = 0.7, \quad \sigma_I = \sigma_A = 0.2,$$

based on data reported in [50].

We simulate the model (9.8) with $\theta_I = 0$, assuming that treatment continues for the duration of the infection. We assume that 80% of diagnosed infectives are treated within 1 day. Since the assumption of treatment at a constant rate φ_I implies treatment of a fraction $1 - \exp(-\varphi_I t)$ after a time t , we take φ_I to satisfy

$$1 - e^{-\varphi_I} = 0.8,$$

Table 9.2 Control by antiviral treatment

βN	\mathcal{R}_0	\mathcal{R}_c	Disease cases	Treatments
0.402	1.37	0.92	64	56
0.435	1.49	1.00	130	113
0.5	1.71	1.15	373	314
0.7	2.39	1.61	865	751

or $\varphi_I = 1.61$. We use different values of βN , corresponding to different values of \mathcal{R}_0 , and use (9.10) and (9.11), obtaining the results shown in Table 9.2.

We calculate from (9.11) that $\mathcal{R}_c = 1$ for an attack rate of 39%. This is the critical attack rate beyond which treatment at the rate specified cannot control the pandemic.

9.4 Seasonal Influenza Epidemics

There are annual influenza outbreaks, usually in the fall and winter. Currently, the annual influenza outbreak is the subtype H3N2, coexisting with the subtype H1N1. Influenza A viruses are divided into subtypes identified by two proteins on the surface of the virus, namely hemagglutinin (HA) and neuraminidase (NA). The designation H3N2 means an influenza A virus that has an HA3 protein and an NA2 protein. An influenza virus undergoes antigenic drift, rapid minor genetic variation in a currently circulating subtype. Recovery from an influenza virus confers immunity against reinfection, but antigenic drift means that a later contact with an influenza virus of the same subtype will mean only partial immunity against infection. Until now, we have studied influenza epidemics in isolation, but if there is a seasonal outbreak of a strain that has been occurring for several seasons part of the population will have some cross-immunity protecting against another infection. Models for recurring seasonal outbreaks taking this into account have been analyzed by Andreasen et al. [3, 4]. Thus we now assume that at the beginning of a seasonal influenza outbreak individuals have a level of cross-immunity depending on the most recent outbreak, up to a maximum of n seasons.

We assume that cross-immunity reduces both susceptibility to infection and infectivity in case of infection. At the beginning of a seasonal outbreak, we assume that all hosts are susceptible and we let N_i denote the number of hosts whose last infection occurred i seasons ago and we include in N_n all who have never been infected. We are dividing the total population into n sub-populations (N_1, N_2, \dots, N_n) representing the distribution of cross immunities at the start of a seasonal epidemic. The (constant) total population size is

$$N = \sum_{i=1}^n N_i.$$

For simplicity, we assume that during a seasonal epidemic the model is a simple *SIR* model, ignoring the latent period and the existence of asymptomatic cases.

Thus for each class i we have I_i infectious and R_i recovered individuals. The transmission dynamics are given by the system

$$\begin{aligned} S'_i &= -\sigma_i S_i \beta \sum_{j=1}^n \tau_j I_j \\ I'_i &= \sigma_i S_i \beta \sum_{j=1}^n \tau_j I_j - \alpha I_i \\ R'_i &= \alpha I_i. \end{aligned} \tag{9.13}$$

Here α is the recovery rate, β is the transmission coefficient in the absence of cross-immunity, σ_i is the relative susceptibility of individuals whose last infection occurred i seasons ago, and τ_i is the relative infectivity of individuals whose last infection occurred i seasons ago. It is reasonable to assume that cross-immunity fades with time, so that

$$0 \leq \sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_n = 1, \quad 0 \leq \tau_1 \leq \tau_2 \leq \cdots \leq \tau_n = 1.$$

If we define the total infectivity

$$\varphi(t) = \sum_{j=1}^n \tau_j I_j,$$

we may rewrite the model (9.13) as

$$\begin{aligned} S'_i &= -\sigma_i \beta S_i \varphi \\ I'_i &= \sigma_i \beta S_i \varphi - \alpha I_i. \end{aligned} \tag{9.14}$$

There is a disease-free equilibrium $S_i = N_i$, $I_i = 0$ ($i = 1, 2, \dots, n$).

The next generation matrix is

$$K = \frac{\beta}{\alpha} \begin{bmatrix} \sigma_1 N_1 \tau_1 & \sigma_2 N_1 \tau_1 & \cdots & \sigma_n N_1 \tau_1 \\ \sigma_2 N_2 \tau_1 & \sigma_2 N_2 \tau_2 & \cdots & \sigma_n N_2 \tau_2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_n N_n \tau_1 & \sigma_n N_n \tau_2 & \cdots & \sigma_n N_n \tau_n \end{bmatrix}.$$

Now, if P is the diagonal matrix with diagonal entries $\sigma_i N_i$, $1 \leq i \leq n$, the matrix $P^{-1} K P$ similar to K has every row β/α multiplied by $(\sigma_1 \tau_1 N_1, \sigma_2 \tau_2 N_2, \dots, \sigma_n \tau_n N_n)$ and therefore has rank 1. This implies that all but one of the eigenvalues of K are zero, and the remaining eigenvalue is equal to the trace of K . From this we conclude that the spectral radius of K is equal to the trace of K and the control reproduction number of the cross-immunity model (9.14) is

$$\mathcal{R}_S = \frac{\beta}{\alpha} \sum_{i=1}^n \sigma_i \tau_i N_i. \quad (9.15)$$

If there were no cross-immunity, the basic reproduction number would be

$$\mathcal{R}_0 = \frac{\beta N}{\alpha}.$$

Because $(S_i + I_i)' = -\alpha I_i$, we can deduce as in Sect. 2.4 that

$$\lim_{t \rightarrow \infty} I_i(t) = 0, \quad \lim_{t \rightarrow \infty} S_i(t) = S_i(\infty),$$

and

$$\alpha \int_0^\infty I_i(t) dt = N_i - S_i(\infty),$$

using $S_i(0) + I_i(0) = N_i$. Integration of the first equation of (9.14) gives

$$\begin{aligned} \log \frac{N_i}{S_i(\infty)} &= \beta \sigma_i \int_0^\infty \varphi(t) dt \\ &= \beta \sigma_i \sum_{j=1}^n \tau_j \int_0^\infty I_j(t) dt \\ &= \frac{\beta}{\alpha} \sigma_i \sum_{j=1}^n \tau_j [N_j - S_j(\infty)] \\ &= \sigma_i \sum_{j=1}^n \tau_j \frac{\beta N_j}{\alpha} \left[1 - \frac{S_j(\infty)}{N_j} \right]. \end{aligned} \quad (9.16)$$

We can write this final size relation in terms of the single unknown $S_n(\infty)$ because from (9.14) and $\sigma_n = 1$ we have

$$\frac{S'_i}{S_i} = \sigma_i \frac{S'_n}{S_n},$$

and integration gives

$$\frac{S_i(\infty)}{N_i} = \left[\frac{S_n(\infty)}{N_n} \right]^{\sigma_i}. \quad (9.17)$$

Substitution of (9.17) into (9.16) gives a relation for $S_n(\infty)$, namely

$$\log \frac{N_n}{S_n(\infty)} = \sigma_n \sum_{j=1}^n \tau_j \frac{\beta N_j}{\alpha} \left[1 - \left(\frac{S_n(\infty)}{N_n} \right)^{\sigma_j} \right].$$

Then (9.17) gives solutions for $S_i(\infty)$, $i = 1, \dots, n-1$.

9.4.1 Season to Season Transition

The above analysis traces the development of a seasonal epidemic. We assume that at the beginning of the next season's epidemic, the final compartment sizes give the separation of the population into new cross-immunity groups ($N_1^*, N_2^*, \dots, N_n^*$). The individuals who were infected during the previous season's epidemic form the new group N_1^* , and the remaining individuals from N_i move to N_{i+1}^* . Thus

$$\begin{aligned} N_1^* &= N - \sum_{i=1}^n S_i(\infty) \\ N_j^* &= S_{j-1}(\infty), \quad j = 2, \dots, n-1 \\ N_n^* &= S_{n-1}(\infty) + S_n(\infty). \end{aligned}$$

This relation describes the transition from one seasonal epidemic to the next. The severity of one season's epidemic affects the cross-immunity for the next season's epidemic. If the seasonal epidemic 1 year is severe, then because there is more cross-immunity the next year it is reasonable to expect a less severe epidemic the next year.

9.5 Pandemic Influenza

In some years, there is an immunological change that produces a new subtype. Such a change is called an antigenic shift, and is usually a result of recombination of gene segments from viruses circulating in humans with virus segments from avian viruses. Since this subtype is new, humans have little or no immunity against it, and the result often is a pandemic. For example, the H1N1 subtype emerged in the pandemic influenza of 1918 and circulated until 1957. In the pandemic of 1957 a new subtype, H2N2, emerged and replaced H1N1. In the pandemic of 1968, H3N2 replaced the prevailing subtype H2N2. In 1977, H1N1 was reintroduced and has been circulating along with H3N2 since then. A natural question is whether a new pandemic strain will replace the currently circulating subtype or coexist with it. To study this question, we must model a pandemic following a seasonal influenza outbreak and then model the next seasonal outbreak. This is necessary because there may be some cross-immunity between the seasonal and pandemic subtypes. The interplay between seasonal and pandemic influenza has been studied by Asaduzzaman et al. [8].

9.5.1 A Pandemic Outbreak

We begin by assuming that there has been a seasonal influenza epidemic and that at the end of this epidemic the population of total size N is divided into cross-immunity sub-populations with respect to the seasonal epidemics of sizes (N_1, N_2, \dots, N_n) . We assume that there is a cross-immunity between the pandemic strain and the seasonal strain, for individuals in the population who have recovered from the seasonal strain in the previous n seasonal outbreaks reducing susceptibility by a factor ρ , $(0 \leq \rho \leq 1)$. Thus $\rho = 0$ corresponds to full cross-immunity and $\rho = 1$ corresponds to no cross-immunity. Then the model for pandemic influenza is

$$\begin{aligned} S'_i &= -\rho \tilde{\beta} S_i I, \quad i = 1, 2, \dots, n-1 \\ S'_n &= -\tilde{\beta} S_n I \\ I' &= \tilde{\beta} \left[\rho(S_1 + S_2 + \dots + S_{n-1}) + S_n \right] I - \alpha I. \end{aligned} \tag{9.18}$$

If there were no cross-immunity, the basic reproduction number would be

$$\tilde{\mathcal{R}}_0 = \frac{\tilde{\beta} N}{\alpha}, \tag{9.19}$$

where $\tilde{\beta}$ is the pandemic contact rate.

An analysis very similar to that of Sect. 9.5 shows that the pandemic reproduction number is

$$\mathcal{R}_P = \frac{\tilde{\beta}}{\alpha} \left[\rho(N_1 + N_2 + \dots + N_{n-1}) + N_n \right] = \frac{\rho \tilde{\beta} N}{\alpha} + \frac{(1-\rho)\tilde{\beta} N_n}{\alpha},$$

an increasing function of ρ . If $\mathcal{R}_P > 1$, there will be a pandemic.

We assume that the pandemic would invade if there were no cross-immunity, that is,

$$\tilde{\mathcal{R}}_0 > 1.$$

Also, again much as in Sect. 9.5, we obtain a final size relation

$$\begin{aligned} \log \frac{N_n}{S_n(\infty)} &= \frac{\tilde{\beta} N}{\alpha} \left[1 - \frac{\sum_{i=1}^n S_i(\infty)}{N} \right] \\ \frac{S_i(\infty)}{N_i} &= \left[\frac{S_n(\infty)}{N_n} \right]^\rho, \quad i = 1, \dots, n-1. \end{aligned} \tag{9.20}$$

This relation gives the cross-immunity distribution for the pandemic. When we model the seasonal epidemic following the pandemic, we will need this result in order to distinguish between individuals susceptible to the seasonal influenza

virus having cross-immunity to the pandemic strain and individuals without cross-immunity.

9.5.2 Seasonal Outbreaks Following a Pandemic

In order to model a seasonal outbreak following a pandemic, because of cross-immunity between the seasonal and pandemic viruses, we must divide each of the subgroups (N_1, N_2, \dots, N_n) into two parts, depending on whether the members of the group were infected by the pandemic. Now, we let M_i denote the individuals whose most recent seasonal infection was i seasons ago and who escaped the pandemic, and let \tilde{M}_i denote the individuals whose most recent seasonal infection was i seasons ago and who were infected in the pandemic. Then $N_i = M_i + \tilde{M}_i$. We denote by S_i the number of individuals in M_i susceptible to the seasonal influenza (who escaped the pandemic influenza) and by \tilde{S}_i the number of individuals in M_i susceptible to the seasonal influenza (infected by the pandemic). Taking into account the cross-immunity of individuals in M_i and again defining the total infectivity

$$\varphi(t) = \sum_{j=1}^n \tau_j I_j,$$

we obtain the model

$$\begin{aligned} S'_i &= -\sigma_i \beta S_i \varphi \\ \tilde{S}'_i &= -\rho \sigma_i \tilde{S}_i \varphi \\ I'_i &= \sigma_i \beta (S_i + \rho \tilde{S}_i) \varphi - \alpha I_i. \end{aligned} \tag{9.21}$$

The calculation of the reproduction number is again much the same as in Sect. 9.4, and yields the result

$$\mathcal{R} = \frac{\beta}{\alpha} \sum_{i=1}^n \sigma_i \tau_i (M_i + \rho \tilde{M}_i).$$

In order to evaluate this, we need to determine M_i and \tilde{M}_i from the final size relation (9.20). We have $M_i = S_i(\infty)$, $\tilde{M}_i = N_i - S_i(\infty)$, so that

$$M_i + \rho \tilde{M}_i = M_i(1 - \rho) + \rho N_i,$$

with $S_i(\infty)$ given by (9.20). If we let $x = \frac{S_n(\infty)}{N_n}$, then (9.20) gives

$$\begin{aligned}
-\log x &= \frac{\tilde{\beta}}{\alpha} [N - \sum_{i=1}^n S_i(\infty)] \\
&= \frac{\tilde{\beta}N}{\alpha} \left[1 - \frac{N_1 S_n^\rho + N_2 S_n^\rho + \cdots + N_n S_n^\rho}{N N_n^\rho} \right] \\
&= \frac{\tilde{\beta}N}{\alpha} [1 - x^\rho].
\end{aligned} \tag{9.22}$$

Consider the function

$$f(x) = \log x + \frac{\tilde{\beta}N}{\alpha} [1 - x^\rho] = \log x + \tilde{\mathcal{R}}_0 [1 - x^\rho],$$

where $\tilde{\mathcal{R}}_0$ is given by (9.19). A solution x of (9.22) is epidemiologically meaningful only if $x < 1$; otherwise, the pandemic does not invade. The value $x = 1$ is always a root of (9.22); the pandemic invades if and only if there is a second root less than 1. The function $f(x)$ is negative for x close to 0. If $\rho \tilde{\mathcal{R}}_0 < 1$, $f(x)$ is monotone increasing for $0 < x \leq 1$ and therefore there is no second root of (9.22) less than 1. Thus, if $\rho \tilde{\mathcal{R}}_0 < 1$, the pandemic does not invade. On the other hand, if $\rho \tilde{\mathcal{R}}_0 > 1$, $f'(1) < 0$, and there must be a second root of (9.22). The pandemic invades if and only if $\rho \tilde{\mathcal{R}}_0 > 1$.

Now, the reproduction number is

$$\begin{aligned}
\mathcal{R} &= \frac{\beta}{\alpha} \sum_{i=1}^n \sigma_i \tau_i [(1 - \rho) N_i x^\rho + \rho N_i] \\
&= \frac{\beta}{\alpha} (1 - \rho) \sum_{i=1}^n \sigma_i \tau_i N_i x^\rho + \frac{\beta}{\alpha} \rho \sum_{i=1}^n \sigma_i \tau_i N_i.
\end{aligned} \tag{9.23}$$

If \mathcal{R} , with x given as a function of ρ by (9.22), is greater than 1, then the pandemic and seasonal strains will coexist, but if $\mathcal{R} < 1$, then the pandemic strain will replace the seasonal strain.

We assume that the seasonal epidemic would take place if there were no pandemic strain present, that is,

$$\mathcal{R}_S > 1.$$

For $\rho = 1$ (no cross-immunity) we have

$$\mathcal{R} = \mathcal{R}_S > 1,$$

and this means that the pandemic and seasonal strains coexist. For $\rho = 1/\tilde{\mathcal{R}}_0$, $x = 1$, and

$$\mathcal{R} = \frac{\beta}{\alpha} \sum_{i=1}^n \sigma_i \tau_i N_i = \mathcal{R}_S > 1,$$

and the pandemic and seasonal strains coexist.

Numerical simulations indicate that $\mathcal{R} < 1$ for some values of ρ between $1/\tilde{\mathcal{R}}_0$ and 1, which implies that for some values of ρ the pandemic strain will replace the seasonal strain.

9.6 The Influenza Pandemic of 2009

In the spring of 2009 a new strain (H1N1) of influenza A was developed, apparently first in Mexico, and spread rapidly through much of the world. Initially, it was thought that there was no resistance to this strain, although it developed later that people who were old enough to have been exposed to similar strains in the 1950s appeared to be less susceptible than younger people. It also appeared initially that this strain had a high case fatality rate, but it was learned later that since many cases were mild enough not to be reported the early data were weighted towards severe cases. The case fatality rate was actually lower than for most seasonal influenza strains.

Management of the H1N1 influenza pandemic of 2009 made use of mathematical models and the experience gained from previous epidemics, but also exposed some gaps between a well-developed mathematical theory of epidemics and real-life epidemics, notably in the acquisition of reliable data early in the pandemic, understanding of spatial spread of a pandemic, and the development of multiple epidemic waves.

There are important differences in the kinds of models of value to different types of scientists. There are *strategic* disease transmission models aimed at the understanding of broad general principles, and *tactical* models with the goal of helping make decisions in specific and detailed situations on short-term policies. A model that predicts how many people will become ill in an anticipated epidemic is quite different from a model that helps to identify which subgroups of a population should have highest priority for preventive vaccination with an uncertain prediction of how much vaccine will be available in a given time frame.

9.6.1 A Tactical Influenza Model

There was a second wave of infections in the 2009 H1N1 influenza pandemic, just as in several previous influenza pandemics. As soon as possible, during the first wave, work began on the development of a vaccine matched to the virus to be used to combat the expected second wave [41]. Since a vaccine requires at least 6 months

for development, and the second wave could begin within 6 months of the first wave, there was an urgent need to prepare a vaccine distribution strategy.

In this section, we describe the model of [18] as an example of a tactical model. This model was used by the BC Center for Disease Control in making planning decisions for vaccination distribution during the second wave of the 2009 H1N1 influenza pandemic. Because of the need to make such decisions rapidly, the development and application of the model was more urgent than the writing of the paper that described the work, and use of the results preceded the write-up. In a pandemic, there is still much to be done after the model has been developed and applied, and frequently there is no time to explore basic theoretical questions that may arise in the study. Ideally, these questions can be studied in more depth after the urgency of coping with a pandemic has passed, and this is an opportunity to return to more general strategic models.

Infection rates in an influenza epidemic depend strongly on the (demographic) age of the individuals in contact. This dependence of transmission and infection rates on age (the age profile) may vary significantly between locations and from year to year for seasonal epidemics. Also, the age profile in a pandemic is probably unlike that for seasonal epidemics. For this reason, real-time planning during an epidemic must make use of surveillance data obtained as soon as possible after the epidemic has begun. The data gathered during the first wave in the spring of 2009 were used to project the age profile to be expected in the second wave. A compartmental model with 6 compartments in each of the 40 population subgroups, covering 8 age classes and 5 activity levels using this age profile and existing estimates of the Vancouver contact network structure, was developed to project the results of different vaccination strategies.

The analysis of the model was carried out using numerical simulations, because comparisons of different strategies, including numerical estimates of disease cases and vaccine quantities required, were needed quickly and because general theoretical analysis of this high-dimensional system would have been too complicated. The numerical simulations indicated that a good estimate of the epidemic peak would be essential for making policy decisions on vaccination strategies. In a pandemic situation, delays in vaccine production may mean that vaccination cannot be started until an epidemic is already underway, and the model suggested that an early start to vaccine distribution makes a big difference in the effectiveness of the vaccination program. This raises a general theoretical question of the relation between the starting time of vaccination and the effectiveness of the vaccination program. Future theoretical study of models of this type would be very useful for planning vaccination strategies for epidemics in the future.

9.6.2 *Multiple Epidemic Waves*

During the 1918 “Spanish flu” pandemic, North America and much of Western Europe experienced two waves of infection with the second wave more severe than

the first [20, 21]. Unlike seasonal influenza epidemics, which occur at predictable times each year, influenza pandemics are often shifted slightly from the usual “influenza season” and have multiple waves of varying severity. This suggests a modeling question that does not arise with seasonal influenza. During a first wave of a pandemic it should be possible to isolate virus samples and begin development of a vaccine matched to the virus in time to allow vaccination against the virus that could help to manage a second wave. One question that arises is prediction of the timing of a second wave. This is not yet possible because there is not yet a satisfactory explanation of the causes of a second wave. One suggestion that has been made is that transmissibility of virus varies seasonally, and this has been used to try to predict whether an endemic disease will exhibit seasonal outbreaks [44, 45]. The same approach can be used to formulate an *SIR* epidemic model with a periodic contact rate that can exhibit two epidemic waves [12]. However, the behavior of such a model depends strongly on the timing of the epidemic. If we assume that the contact rate is highest in the winter, lowest in the summer, and varies sinusoidally with a period of 1 year, we could use a simple *SIR* model with a variable contact rate. With N as the (constant) total population size, this suggests a model

$$\begin{aligned} S' &= -\beta(t) \frac{SI}{N} \\ I' &= \beta(t) \frac{SI}{N} - \alpha I, \end{aligned} \tag{9.24}$$

where

$$\beta(t) = \beta \left[1 + c \cos \left(\frac{\pi(t + t_0)}{180} \right) \right],$$

with parameter values

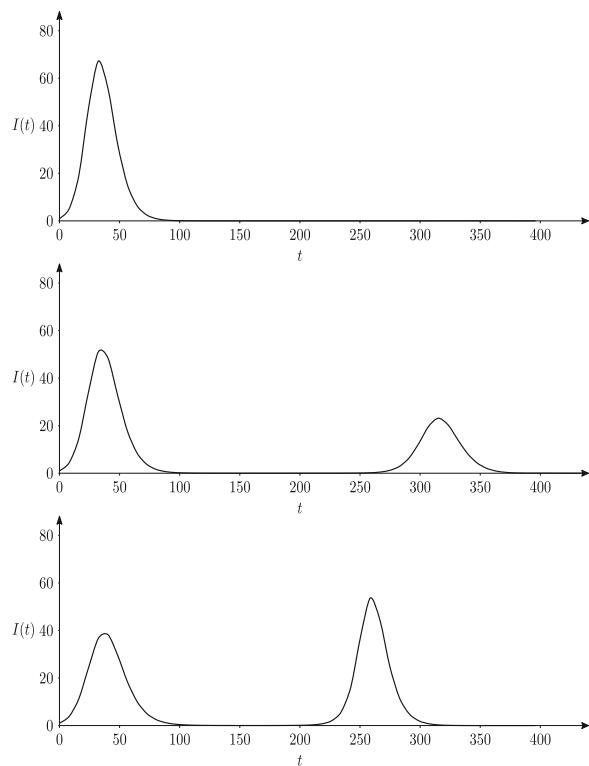
$$\begin{aligned} \alpha &= 0.25, & \beta &= 0.45, & c &= 0.45, \\ N &= 1000, & S_0 &= 999, & I_0 &= 1, & t_0 &= 85, 90, 95. \end{aligned}$$

The choice of t_0 determines where in the oscillation of $\beta(t)$ the epidemic begins, because at the start of the epidemic ($t = 0$)

$$\beta(0) = \beta \left[1 + c \cos \left(\frac{\pi t_0}{180} \right) \right].$$

Thus t_0 is the number of days after the maximum transmissibility that the epidemic begins. By numerical integration of (9.24), we obtain the following three epidemic curves (giving the number of infective individuals as a function of time), with the same parameters except that by varying t_0 the starting date of the epidemic is moved 5 days later from one curve to the next.

Fig. 9.4 Top: A one-wave epidemic curve, $t_0 = 85$.
 Middle: A two-wave epidemic curve, first wave more severe, $t_0 = 90$.
 Bottom: A two-wave epidemic curve, second wave more severe, $t_0 = 95$



An interpretation of these curves is that if the epidemic begins when the contact rate is decreasing and is close to its minimum value and the contact rate is relatively small over the course of the epidemic, the wave may end while there are still enough susceptibles to support a second wave when the contact rate increases, as in Fig. 9.4 (middle and bottom plots). However, if the epidemic begins earlier it may continue until enough individuals are infected that a second wave cannot be supported even when the contact rate becomes large, as in the top plot of Fig. 9.4.

Simulations indicate that for the model (9.24) with the parameter values used here there is a small window of starting times corresponding to the interval $90 \leq t_0 \leq 110$ for which a second wave is possible. The nature of the epidemic curves indicates that the behavior depends critically on timing and this means that such a model is not suitable for precise predictions. Note that there may be a single wave or two waves, and if there are two waves either one may be more severe. Not only does the prediction depend on the timing of the introduction of infection, but this timing is stochastic, and by its very nature unpredictable. It depends on mutations of the virus and importations of new cases.

We have here a strategic model that predicts the possibility of a second wave, but this is not sufficient to be confident about using it to advise policy, because we do not have enough evidence to be sure about the cause of a second wave. Before

we could make predictions we need more evidence about factors that could cause a second wave and more detailed tactical models, including sensitivity analysis. Another factor that has been suggested as a possible explanation for a second epidemic wave is coinfection with other respiratory infections that might increase susceptibility, and some experimental data would be needed that might confirm or deny that either seasonal variation in transmission or coinfection, or both, could explain multiple pandemic waves. In the H1N1 influenza pandemic of 2009, there were some concerns about the development of drug resistance as a consequence of antiviral treatment. While this did not appear to have widespread consequences, in a more severe disease outbreak in which more patients receive antiviral treatment there could be major effects. The modeling of drug resistance effects in an influenza pandemic has begun [1, 5, 36, 42, 43], but much more needs to be understood. A full analysis of the development of drug resistance will require nested models including immunological in-host aspects as well as effects on the population level.

Data from the 1918 pandemic indicate clearly that behavioral response, both individual and public health measures, had significant effects on the outcome of the epidemic [9]. Incorporating behavioral responses is a new aspect of epidemic modeling, and there is much to be learned about the factors that influence behavioral responses when a disease outbreak begins.

9.6.3 Parameter Estimation and Forecast of the Fall Wave

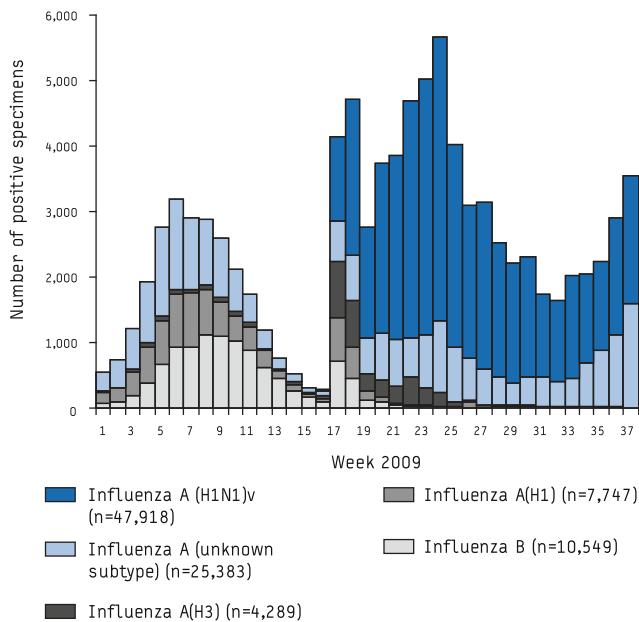
The model predictions included in this section are presented in [47]. With the recognition of a new, potentially pandemic strain of influenza A(H1N1) in April 2009, the laboratories at the US CDC and the World Health Organization (WHO) dramatically increased their testing activity from week 17 onwards (week ending 2 May 2009), as can be seen in Fig. 9.5. In this analysis, we use the extrapolation of a model fitted to the confirmed influenza A(H1N1)v case counts during summer 2009 to predict the behavior of the pandemic during autumn 2009.

The following model with seasonally forced transmission is used:

$$\begin{aligned} S' &= -\beta(t) \frac{SI}{N} \\ I' &= \beta(t) \frac{SI}{N} - \alpha I, \end{aligned} \tag{9.25}$$

where $N = 305,000,000$ denote the total population in the United States (US). The R equation is omitted as it can be determined by $R = N - S - I$. The transmission rate is chosen to be

$$\beta(t) = \beta_0 + \beta_1 \cos(\pi t / 180), \tag{9.26}$$



US CDC: United States Centers for Disease Control and Prevention; NREVSS: National Respiratory and Enteric Virus Surveillance System; WHO: World Health Organization.

Fig. 9.5 Influenza-positive tests reported to the US CDC by US WHO/NREVSS-collaborating laboratories, national summary, United States, 2009 until 26 September

where β_0 and β_1 are constants to be estimated from data. Assume that $I(t_0) = 1$, where t_0 is the initial time, which is another parameter to be estimated from data. The parameter α is chosen so that the infective period is $1/\alpha = 3$ days.

The three parameters (β_0 , β_1 , t_0) were estimated using the data shown in Fig. 9.5 on influenza-positive tests reported to the US CDC by US WHO/NREVSS-collaborating laboratories, national summary, United States, 2009 until 26 September. To avoid bias due to increased testing for H1N1 around week 16, only data from week 21 to 33 (from 24 May to 22 August 2009) were used. From the past experience with influenza, the lower and upper bounds were chosen to be $\beta_0 \in (0.92\alpha, 2.52\alpha)$ and $\beta_1 \in (0.05\alpha, 0.8\alpha)$, and $t_0 \in (-8, 10)$ (weeks relative to the beginning of 2009). The best estimates were determined by fitting the model to data, and the parameter values that provided the best Pearson chi-square statistics are listed in Table 9.3.

Using the estimated parameter values, forecast of the time and size of the fall wave was produced by simulating the model under two scenarios: one is without vaccination and the other included the planned CDC vaccination program, which would begin with six to seven million doses being delivered by the end of the first full week in October (week 40), with 10–20 million doses being delivered

Table 9.3 Estimates of parameters for the 2009 H1N1 model

Parameter	Value	95% confidence interval
β_0	1.56	(1.43, 1.77)
β_1	0.54	(0.39, 0.54)
t_0	24 February	(8 February, 7 March)

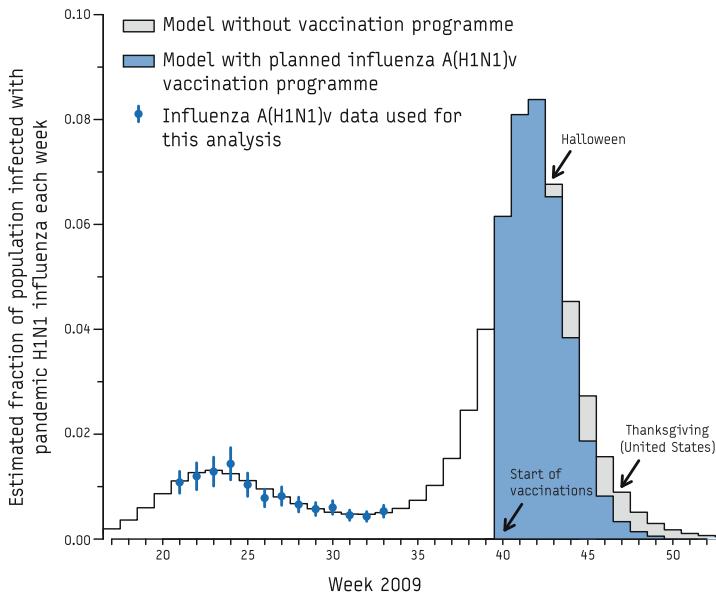


Fig. 9.6 Prediction for the 2009 H1N1 pandemic in the US using model (9.25)

weekly thereafter. It is assumed that for healthy adults, full immunity to H1N1 influenza is achieved about 2 weeks after vaccination with one dose of vaccine [30, 37]. For simulations of the model when vaccinations are incorporated, the number of susceptible individuals in the simulations was decreased according to the appropriate proportion of immune due to vaccination. The simulation outcomes are presented in Fig. 9.6.

In Fig. 9.6, the curves associated with the darker and lighter areas correspond to the cases of with and without vaccination. The model predicts that in the absence of vaccine, the peak wave of infection will occur near the end of October in week 42 (95% CI: week 39, 43). By the end of 2009, the model predicts that a total of 63% of the population will have been infected (95% CI: 57%, 70%). For the case when the planned vaccination program is considered, the model results suggest that a relative reduction of about 6% in the total number of people infected with influenza A(H1N1)v virus by the end of the year 2009 (95% CI: 1%, 17%).

The most striking feature of the model is that it accurately predicted the peak time of the pandemic. According to CDC 2009 H1N1 confirmed case count data (see [17]), the peak of the fall wave was reached at the end of October (which is between weeks 42 and 43, see the left-hand plot in Fig. 9.7), which is consistent with

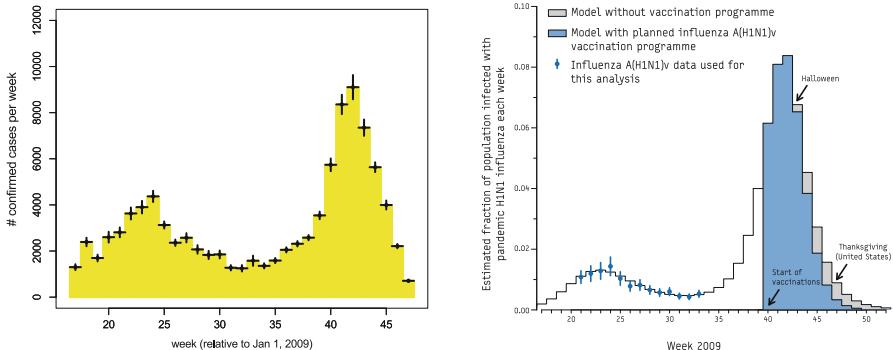


Fig. 9.7 The left figure illustrates the CDC 2009 confirmed H1N1 count data (the error bars represent variations calculated by the proposers that account for US regional variability in the timing of the pandemic). The right figure shows predictions by our model in [47]

our model result. It is worth noting that the model used in the analysis is a simple SIR model with a seasonally forced infection rate. Although further examinations are certainly needed to study the applicability of the modeling approach to general scenarios, the model results in this analysis demonstrated clearly the advantage and capability of mathematical models in understanding disease dynamics.

9.7 *An SIQR Model with Multiple Strains with Cross-Immunity

It was during the *Black Death* in the twentieth century in Venice that a system of quarantine was first put in operation. It demanded ships to lay at anchor for “40 days” (in Latin “*quaranti giorni*”) before sailors and guests could come on land. Quarantine is often thought of as a policy that separates individuals who may have been exposed to a contagious agent regardless of symptoms. Isolation is, generally speaking, considered a severe form of quarantine, often put in place in response to high morbidity and mortality. Quarantine was used extensively in the treatment and control of tuberculosis in Europe first and later, at end of the nineteenth century, in the USA. The emergence of SARS in 2003 re-instated the world’s interests in the concepts of Isolation and Quarantine (Q & I), the only initial methods available of disease control [11, 19].

The concepts of Q & I have multiple working meanings and uses. Hence, selecting a definition depends on the disease, the suspected level of risk that it poses to others, the means and modes of transmission, and the system’s knowledge and experience with the infectious agent. Regardless of the definition used, one of the challenges associated with Q & I strategies is that there is hardly any reliable assessments of their population level efficacy. There are no effective

quantitative frameworks that account for direct and indirect economic losses and/or the costs associated with the implementation of Q & I strategies (but see [14, 27–29, 33, 34, 38, 40]).

Critical to any method of assessment comes from the fact that dynamic models that include the Q & I classes disease must be prepared to account for their sometimes destabilizing impact on the disease dynamics (sustained oscillations). The introduction of Q & I classes can generate the kind of dynamics where assessing the effectiveness of interventions may be difficult. Feng et al. showed in [24, 25], for example, that the incorporation of a quarantine or isolation class (Q) was enough to destabilize the unique disease endemic equilibrium in a susceptible–infected–recovered (SIR) model. These results were confirmed by Hethcote and collaborators [31] using alternative SIQR modeling frameworks.

Tracing exposed individuals, assuming that a test exists that determines if an individual is infected or not, and their contacts, would make it possible to quarantine or isolate diagnosed infectives, a first step towards assessing the impact of Q & I [15]. Models are used to address questions like: What impact will placing a fixed proportion of individuals living in the “neighborhood” of an index case in quarantine have on disease control? We observe that when large numbers are involved, the costs and challenges become immense. Should isolated individuals be kept at their homes or moved to designated quarantine facilities?

In [39] a two-strain influenza model is studied to investigate competitive outcomes (mediated by cross-immunity) that result from the interactions between two strains of influenza A in a population where sick individuals may be quarantined or isolated. The inclusion of a quarantined class makes the model analysis much more challenging due to the presence of sustained oscillations and possible other bifurcations. To make the analysis more transparent, an SIQR model with a single strain is presented first.

9.7.1 *An SIQR Model with a Single Infectious Class

In this section, results in [24] and [25] are revisited in order to highlight the impact of the inclusion of quarantine and/or isolation classes in generating sustained periodic solutions from SIQR epidemic models.

Let $S(t)$, $I(t)$, $Q(t)$, and $R(t)$ denote the susceptible, infective, quarantined, and recovered classes, respectively, and let $N = S + I + Q + R$ denote the total population size. Model parameters are: μ , the per-capita rates for birth and death; γ , the rate at which infected individuals are isolated (quarantined); δ , recovery rate; and β , the transmission rate. The SIQR model can be formulated as follows:

$$\begin{aligned}
\frac{dS}{dt} &= \mu N - \beta S \frac{I}{N-Q} - \mu S, \\
\frac{dI}{dt} &= \beta S \frac{I}{N-Q} - (\mu + \gamma) I, \\
\frac{dQ}{dt} &= \gamma I - (\mu + \delta) Q, \\
\frac{dR}{dt} &= \delta Q - \mu R
\end{aligned} \tag{9.27}$$

with appropriate initial conditions.

What makes the above model “distinct” is that the incidence rate now accounts for the possibility that a large number of individuals (those in the Q class) do not participate (by request, mandate, or personal decision) in the transmission process. Hence, the “random mixing” of infected proportion with other individuals is $\frac{I}{N-Q}$ rather than $\frac{I}{N}$.

The basic reproduction number for the above SIQR model is

$$\mathcal{R}_0 = \frac{\beta}{\mu + \gamma}. \tag{9.28}$$

It was shown in [25] that the disease-free equilibrium is globally asymptotically stable when $\mathcal{R}_0 < 1$. However, for $\mathcal{R}_0 > 1$, the behavior of model solutions is very different from the standard SIR model. That is, the unique endemic equilibrium, denoted by E^* , can become unstable due to the appearance of stable periodic solution via a Hopf bifurcation [32]. Notice from (9.28) that \mathcal{R}_0 is independent of the isolation period $1/\delta$. A bifurcation analysis using δ as the bifurcation parameter shows that E^* is asymptotically stable when the quarantine period is either very large or very small.

To facilitate the analysis, consider the following equivalent system with re-scaled parameters:

$$\begin{aligned}
\frac{dI}{d\tau} &= \left(1 - \frac{I+R}{N-Q}\right)I - (\nu + \theta)I, \\
\frac{dQ}{d\tau} &= \theta I - (\nu + \alpha)Q, \\
\frac{dR}{d\tau} &= \alpha Q - \nu R,
\end{aligned} \tag{9.29}$$

where $\tau = \beta t$ and

$$\nu = \frac{\mu}{\beta}, \quad \theta = \frac{\gamma}{\beta}, \quad \alpha = \frac{\delta}{\beta}.$$

The S equation has been eliminated as N is constant and $S = N - I - Q - R$. Note that the scaled parameter representing the isolation period is α . Based on the system (9.29), the following result on the possibility of periodic solutions via Hopf bifurcation was established in [25]:

Theorem 9.1 *There is a function, $\alpha_c(\nu)$, defined for small $\nu > 0$,*

$$\alpha_c(\nu) = \theta^2(1 - \theta) + O(\nu^{1/2}),$$

such that there are two critical values α_{c1} and α_{c2} with the following properties:

- (a) *The endemic equilibrium is locally asymptotically stable if $1/\alpha < 1/\alpha_{c1}(\nu)$ or $1/\alpha > 1/\alpha_{c2}(\nu)$ and unstable if $1/\alpha_{c1}(\nu) < 1/\alpha < 1/\alpha_{c2}(\nu)$, as long as $1/\alpha$ is close to the critical values.*
- (b) *Hopf bifurcations occur at $\alpha = \alpha_{ci}(\nu)$ ($i = 1, 2$), leading to stable periodic solutions near the bifurcation points.*

Further, at the bifurcation point corresponding to α_{c1} , the length of periods can be approximated by the formula

$$T \approx \frac{2\pi}{(1 - \theta)^{1/2}\nu^{1/2}} \approx \frac{2\pi}{(\theta y^*)^{1/2}},$$

where $y^* = I^*/(N - Q^*)$ is the proportion of infectious individuals at the endemic equilibrium scaled by the active population, $N - Q$.

With parameter values relevant to scarlet fever, numerical simulations of the model confirm the occurrence of two Hopf bifurcation points, which are illustrated in Fig. 9.8. The right plot in Fig. 9.8 shows two Hopf bifurcation points (labeled by HB) occurring for values of the average quarantine periods in two distinct ranges near $1/\alpha_{ci}$ ($i = 1, 2$). The plot on the left shows the enlarged portion near the lower bifurcation point $1/\alpha_{c1}$, with the curves labeled with SP representing the maximum and minimum of the stable periodic solutions. The solid and dashed curves labeled by SSS and USS represent stable steady state and unstable steady state (showing the I^* component of the endemic equilibrium E^*), respectively.

It was pointed out in [24, 25] that only the region near the lower critical point $1/\alpha_{c1}$ is relevant for childhood diseases, and that the data on the length of reported isolation periods during the 1897–1978 scarlet fever epidemics in England and Wales [2] were very close to the range that supports periodic solutions for model (9.29) near $1/\alpha_{c1}$.

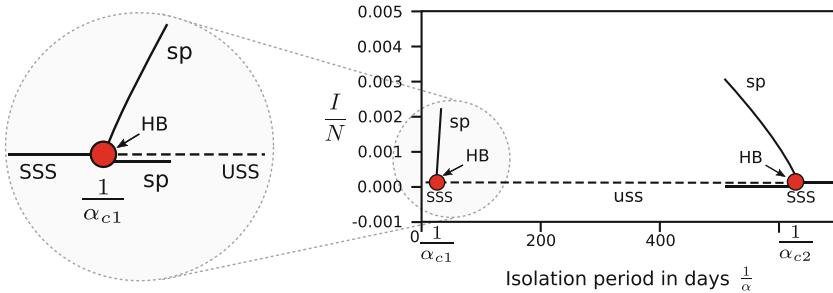


Fig. 9.8 The plot on the right is a bifurcation diagram based on numerical simulations of the system (9.29) with parameter values relevant to scarlet fever. The solid and dashed curves (labeled by SSS and USS) represent the fraction of the I components when the endemic equilibrium E^* is stable and unstable, respectively, depending on the value of the isolation period $1/\alpha$. The right plot shows two Hopf bifurcation points, α_{c1} and α_{c2} , labeled by HB. An enlarged version of the right HB is shown on the left. The solid curves labeled with sp illustrate the maximum and minimum of the stable periodic solutions

Several extensions of the model (9.27) have been considered including [23, 31, 49]. For example, the model in [23] reads

$$\begin{aligned} \frac{dS}{dt} &= \mu N - \beta S \frac{I}{N - \sigma Q} - \hat{\beta} S \frac{(1 - \sigma)Q}{N - \sigma Q} - \mu S, \\ \frac{dI}{dt} &= \beta S \frac{I}{N - \sigma Q} + \hat{\beta} S \frac{(1 - \sigma)Q}{N - \sigma Q} - (\gamma + \kappa + \mu)I, \\ \frac{dQ}{dt} &= \gamma I - (\delta + \mu)Q, \\ \frac{dR}{dt} &= \kappa I + \delta Q - \mu R. \end{aligned} \quad (9.30)$$

In (9.30), individuals in the I class can recover without going through the Q class. In addition, the degree of effectiveness (i.e., reducing the movement of quarantined individuals) is considered through the use of the parameter σ with $\sigma = 1$ and 0 corresponding to a perfect and ineffective quarantine/isolation, respectively. It is shown in [23] that the likelihood for the existence of sustained oscillations depends on σ .

Because Hopf bifurcations are local properties, and because the bifurcation diagram generated by numerical simulations in Fig. 9.8 seems to suggest that the periods of the periodic solutions for parameters away from the Hopf bifurcation points become very large, it indicates the possibility for a homoclinic bifurcation. This is explored in [51]. It was shown in [51] that a center manifold reduction of the system (9.27) at a bifurcation point has the normal form $x' = y$, $y' = axy + bx^2y + O(4)$, indicating a bifurcation of codimension greater than two. Here, x and y are variables after several changes of variables from the system (9.29). By considering an unfolding of the normal form, it was shown that a perturbed system can indeed generate a homoclinic bifurcation. This is illustrated in Fig. 9.9.

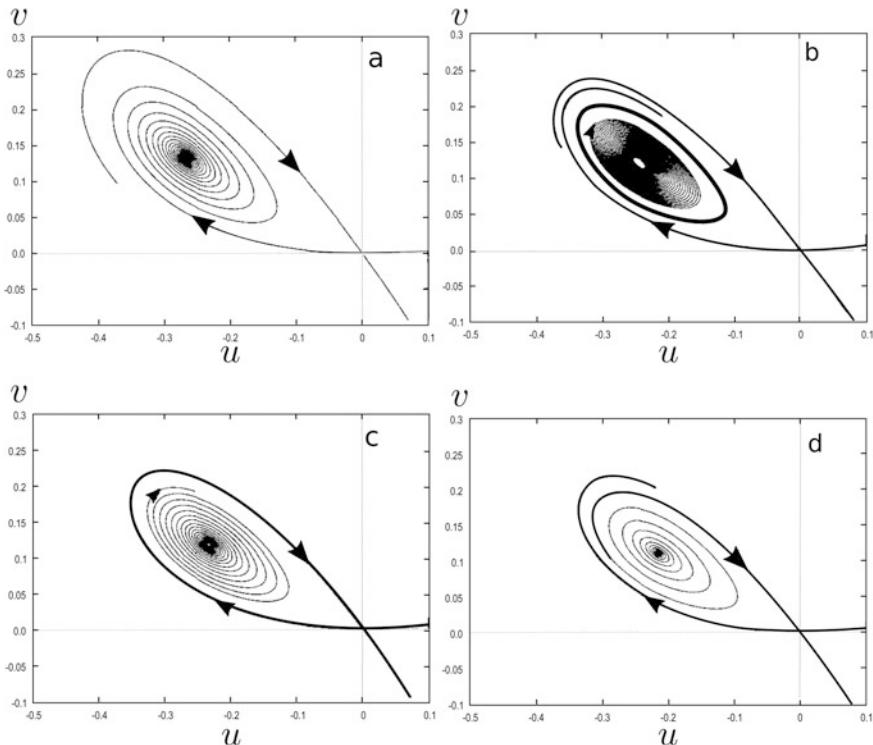


Fig. 9.9 Numerical simulations of a perturbed system of the normal form reduction from the system (9.29). The four plots correspond to four sets of parameter values near the homoclinic bifurcation. As the parameter values change, the interior equilibrium E^* changes from stable to unstable leading to the appearance of a stable periodic solution (see plots (a) and (b)), and that a homoclinic bifurcation occurs as parameter values change from (b) to (d), in which case, a homoclinic solution exists (see (c))

The four sub-figures a–d in Fig. 9.9 demonstrate trajectories of the perturbed system for four sets of parameter values. It shows first that the stability of E^* switches from stable to unstable with the appearance of a stable periodic solution (see a and b). As parameter values continue to change from b to c, the system undergoes a homoclinic bifurcation, in which case, the stable periodic solution expends its magnitude and eventually disappears after reaching the saddle point (see b–d) and a homoclinic orbit appears (see c).

9.7.2 *The Case of Two Strains with Cross-Immunity

The SIQR model (9.27) can be extended to include two pathogen strains. The description of the two-strain model requires the division of the population into ten different classes: susceptibles (S), infected with strain i (I_i , primary infection),

isolated with strain i (Q_i), recovered from strain i (R_i , as a result of primary infection), infected with strain i (V_i , secondary infection), given that the population had recovered from strains $j \neq i$, and recovered from both strains (W). The population is assumed to mix randomly, except that the mixing is impacted by the process of quarantine/isolation [15, 25, 31]. Using the flow diagram in Fig. 9.10, we arrive at the model

$$\begin{aligned} \frac{dS}{dt} &= \Lambda - \sum_{i=1}^2 \beta_i S \frac{(I_i + V_i)}{A} - \mu S, \\ \frac{dI_i}{dt} &= \beta_i S \frac{(I_i + V_i)}{A} - (\mu + \gamma_i + \delta_i) I_i, \\ \frac{dQ_i}{dt} &= \delta_i I_i - (\mu + \alpha_i) Q_i, \\ \frac{dR_i}{dt} &= \gamma_i I_i + \alpha_i Q_i - \beta_j \sigma_{ij} R_i \frac{(I_j + V_j)}{A} - \mu R_i, \quad j \neq i \\ \frac{dV_i}{dt} &= \beta_i \sigma_{ij} R_j \frac{(I_i + V_i)}{A} - (\mu + \gamma_i) V_i, \quad j \neq i, \quad i, j = 1, 2 \\ \frac{dW}{dt} &= \sum_{i=1}^2 \gamma_i V_i - \mu W, \\ A &= S + W + \sum_{i=1}^2 (I_i + V_i + R_i), \end{aligned} \tag{9.31}$$

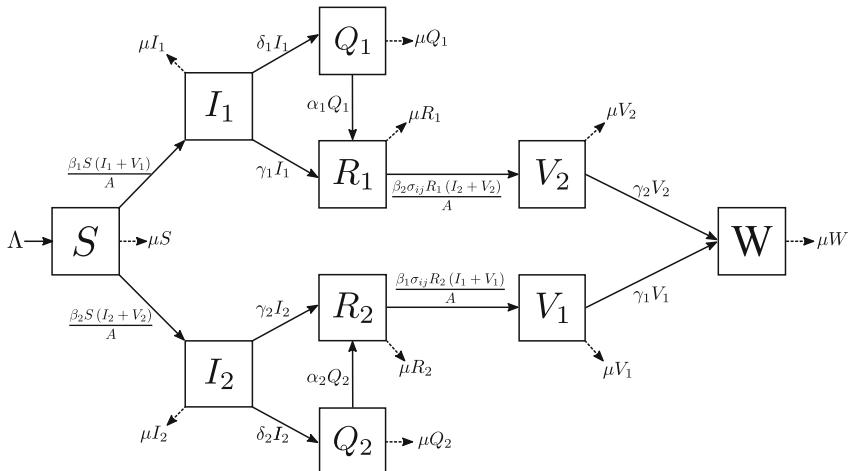


Fig. 9.10 Schematic diagram of the dynamics in host exposed to two co-circulating influenza strains. Λ is the rate at which individuals are born into the population, β_i denotes the transmission coefficient for strain i , μ is the per-capita mortality rate, δ_i is the per-capita isolation rate for strain i , γ_i denotes the per-capita recovery rate from strain i , α_i is the per-capita rate at which individuals leave the isolated class as a result of infection with strain i , and σ_{ij} is the relative susceptibility to strain j for an individual that has been infected with and recovered from strain i ($i \neq j$). $\sigma_{ij} = 0$ corresponds to total cross-immunity, while $\sigma_{ij} = 1$ indicates no cross-immunity between strains. The protection is said to be strong if $0 \leq \sigma_{ij} \ll 1$, and weak if $0 \ll \sigma_{ij} \leq 1$

where A denotes the population of non-isolated individuals and $\frac{\beta_i S(I_i + V_i)}{A}$ models the rate at which susceptibles become infected with strain i . That is, the i th ($i \neq j$) incidence rate is assumed to be proportional to both the number of susceptibles and the available modified proportion of i -infectious individuals, $\frac{(I_i + V_i)}{A}$. The parameter σ_{ij} is a measure of the cross-immunity provided by a prior infection with strain i to exposure with strain j ($i \neq j$). It is assumed that $\sigma_{ij} \in [0, 1]$. Model (1) includes the models in [13, 16]. The omission of the Q classes in earlier work precludes the possibility of sustained oscillations (see [13, 16]) in the absence of population (age) structure.

System (9.31) has at least four equilibria. Analysis of the local stability of the trivial equilibrium (absence of disease) helps identify conditions under which the “flu” can invade. Hence, we first focus on establishing the conditions that make it impossible (at least locally) for both strains to invade a disease-free population, simultaneously. In the analytical results obtained here it is assumed that $\sigma_{12} = \sigma_{21} = \sigma$.

The basic reproduction number for the two strains is

$$\mathcal{R}_i = \frac{\beta_i}{\mu + \gamma_i + \delta_i}.$$

Using the perturbation technique based on the fact that μ is a much smaller parameter than other parameters, we can identify two curves in the $(\mathcal{R}_1, \mathcal{R}_2)$ plane that determine the stability of the non-trivial equilibria. Let $f(\mathcal{R}_1)$ and $g(\mathcal{R}_2)$ be two functions defined by

$$f(\mathcal{R}_1) = \frac{\mathcal{R}_1}{1 + \sigma(\mathcal{R}_1 - 1) \left(1 + \frac{\delta_2}{\mu + \gamma_2}\right) \left(1 - \frac{\mu(\mu + \alpha_1)}{(\mu + \gamma_1)(\mu + \alpha_1) + \alpha_1 \delta_1}\right)} \quad (9.32)$$

and

$$g(\mathcal{R}_2) = \frac{\mathcal{R}_2}{1 + \sigma(\mathcal{R}_2 - 1) \left(1 + \frac{\delta_1}{\mu + \gamma_1}\right) \left(1 - \frac{\mu(\mu + \alpha_2)}{(\mu + \gamma_2)(\mu + \alpha_2) + \alpha_2 \delta_2}\right)}. \quad (9.33)$$

Let $\mathcal{R}_i^* = \mathcal{R}_i$ ($i = 1, 2$) evaluated at $\mu = 0$. Then the following result holds.

Theorem 9.2 *There exists a function $\alpha_{1c}(\mu)$ defined for small $\mu > 0$ by*

$$\alpha_{1c}(\mu) = \frac{\delta_1}{\mathcal{R}_1^*} \left(1 - \frac{1}{\mathcal{R}_1^*}\right) + O(\mu^{1/2}),$$

with the following properties: (i) The boundary endemic equilibrium E_1 is locally asymptotically stable if $\mathcal{R}_2 < f(\mathcal{R}_1)$ and $\alpha_1 < \alpha_{1c}(\mu)$, and unstable if $\mathcal{R}_2 > f(\mathcal{R}_1)$ or $\alpha_1 > \alpha_{1c}(\mu)$. (ii) When $\mathcal{R}_2 < f(\mathcal{R}_1)$, periodic solutions arise at $\alpha_1 = \alpha_{1c}(\mu)$ via Hopf bifurcation for small enough $\mu > 0$. The period can be approximated by

$$T = \frac{2\pi}{|\Im\omega_{2,3}|} \approx \frac{2\pi}{((\gamma_1 + \delta_1)(\mathcal{R}_1^* - 1))^{\frac{1}{2}} \mu^{1/2}}.$$

Because we have focused on the symmetric case, an analogous result for the second boundary equilibrium E_2 can be stated immediately. That is, the boundary endemic equilibrium E_2 is locally asymptotically stable if $\mathcal{R}_1 < g(\mathcal{R}_2)$ and $\alpha_2 < \alpha_{2c}(\mu)$. It becomes unstable if $\mathcal{R}_1 > g(\mathcal{R}_2)$ or $\alpha_2 > \alpha_{2c}(\mu)$. A summary of the stability results as presented in Theorem 9.2 for strain 1 is obtained for strain 2 by replacing the parameter indices 1's with 2's and replacing $f(\mathcal{R}_1)$ with $g(\mathcal{R}_2)$. Functions $f(\mathcal{R}_1)$ and $g(\mathcal{R}_2)$ help determine the stability and coexistence regions for strains 1 and 2. In fact, changes in the regions of stability for a single and both strains can be illustrated by varying the coefficient of cross-immunity. For instance, from (9.32) we can compute the value of σ at which

$$f'(\mathcal{R}_1) \equiv \left. \frac{\partial f(\mathcal{R}_1, \sigma)}{\partial \mathcal{R}_1} \right|_{\sigma_1^*} = 0, \quad (9.34)$$

namely

$$\sigma_1^* = \frac{1}{\left(1 + \frac{\delta_2}{\mu + \gamma_2}\right) \left(1 - \frac{\mu(\mu + \alpha_1)}{(\mu + \gamma_1)(\mu + \alpha_1) + \alpha_1 \delta_1}\right)}.$$

Hence, for all $\mathcal{R}_1 > 1$,

$$f'(\mathcal{R}_1) > (< \text{ or } =) 0, \quad f(\mathcal{R}_1) > (< \text{ or } =) 1 \quad \text{if } \sigma < (> \text{ or } =) \sigma_1^*.$$

These properties can be easily checked by noticing from (9.32) that

$$f(\mathcal{R}_1) = \frac{\mathcal{R}_1}{1 + \frac{\sigma}{\sigma_1^*}(\mathcal{R}_1 - 1)} \quad \text{and} \quad f'(\mathcal{R}_1) = \frac{1 - \frac{\sigma}{\sigma_1^*}}{\left(1 + \frac{\sigma}{\sigma_1^*}(\mathcal{R}_1 - 1)\right)^2}.$$

Using the symmetry between two strains, we can show that similar properties hold for another threshold value σ_2^* (interchanging the subscripts 1 and 2 in the expression of σ_1^*) and a function $\mathcal{R}_1 = g(\mathcal{R}_2)$. The properties of f and g are illustrated in Fig. 9.11. The first two plots in Fig. 9.11 are for the special case when the two strains have identical parameter values ($\sigma_1^* = \sigma_2^* = \sigma^*$), whereas the last two plots are for the case $\sigma_1^* \neq \sigma_2^*$. $\mathcal{R}_2 < f(\mathcal{R}_1)$ is a necessary condition for the stability of strain 1 (either a stable boundary endemic equilibrium E_1 or the equilibrium associated with strain 1 oscillations). Hence, E_2 is unstable when $\mathcal{R}_2 > f(\mathcal{R}_1)$. Similarly, E_1 is unstable when $\mathcal{R}_1 > g(\mathcal{R}_2)$. Hence, coexistence is expected when $\mathcal{R}_2 > f(\mathcal{R}_1)$ and $\mathcal{R}_1 > g(\mathcal{R}_2)$.

Figure 9.12 depicts the stability of the equilibria and periodic solutions when $(\mathcal{R}_1, \mathcal{R}_2)$ lies in Regions I and III. It illustrates how the stabilities of equilibria and

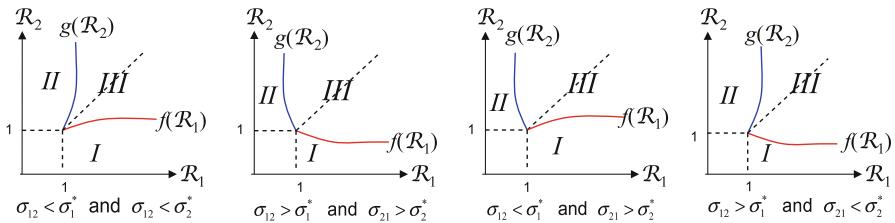


Fig. 9.11 Bifurcation diagrams in the $(\mathcal{R}_1, \mathcal{R}_2)$ plane for various combination of σ_1 and σ_2 values in relation to the threshold levels of σ_1^* and σ_2^* (cross-immunity). Regions I–III denote the existence and stability of E_1 , E_2 , and coexistence, respectively

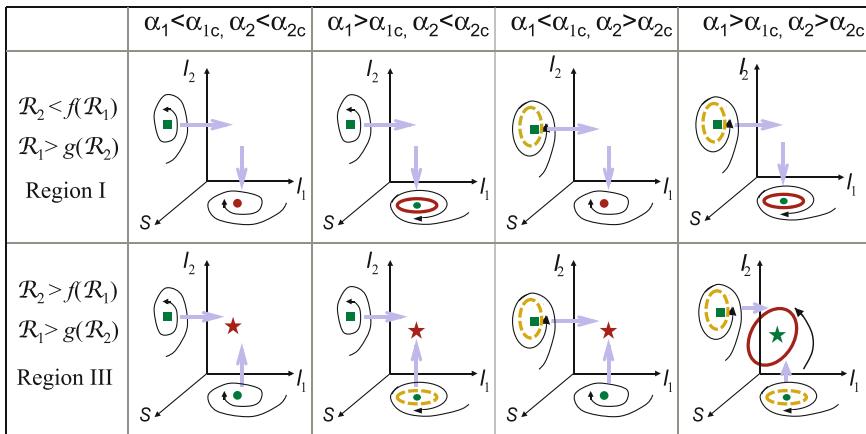


Fig. 9.12 Depiction of the stability properties of equilibria and periodic solutions for parameter values in different regions. A solid circle represents a stable boundary (strain 1 only) equilibrium. A solid square represents an unstable boundary (strain 2 only) equilibrium. A star represents a stable interior (coexistence of both strains) equilibrium. A solid (dashed) closed orbit represents a stable (unstable) period solution

periodic orbits change their stability when the parameters α_1 and α_2 change their values crossing the critical points α_{ic} ($i = 1, 2$).

The focus of this study is on the time evolution of influenza A in a non-fixed landscape driven by tight coevolutionary interactions (that is, interactions where the fate of the host and the parasite are intimately connected) between human hosts and competing strains. The process is mediated by intervention (behavioral changes) and cross-immunity. In other words, the nature of the invading landscape (susceptible host) changes dynamically from behavioral changes (isolation, short time scale) and past immunological experience (cross-immunity, long time scale).

The “partial” herd immunity generated by past history of invasions on the host population can have a huge impact on the quantitative dynamics of the “flu” at the population level. The assumption that $\sigma_{12} = \sigma_{21} = \sigma$ for $i \neq j$ naturally results in a dynamic landscape that is not too different (in the oscillatory regime) than

the one observed on single-strain models with isolation [25, 31]. That is, a lack of heterogeneity in cross-immunity results in a system “more or less” driven (in the oscillatory regime) by the process of isolation.

The results show that multiple strain coexistence is highly likely for antigenically distinct (weak cross-immunity) strains and not for antigenically similar under symmetric cross-immunity (“competitive exclusion” principle [10]). As the levels of cross-immunity weaken, the likelihood of sub-threshold coexistence increases. However, “full” understanding of the evolutionary implications that result from human host and influenza virus interactions may require the study of systems that incorporate additional mechanisms such as seasonality in transmission rates, age structure, individual differences in susceptibility or infectiousness, and the possibility of coinfections. Thacker [46] notes that the observed seasonality of influenza in temperate zones may be the key to observed patterns of recurrent epidemics. Superinfection may also be a mechanism worth consideration, even though studies in [46] show that it is only moderately possible for young individuals to become infected with two different strains in one “flu” season.

9.8 Exercises

1. Write the basic influenza model of Sect. 9.2 as an age of infection model and calculate its basic reproduction number.
2. Write the vaccination model of Sect. 9.2.1 as an age of infection model and calculate its reproduction number. [Warning: In order to do this, you will need to formulate a heterogeneous mixing age of infection model because vaccinated and unvaccinated individuals have different infectivities and susceptibilities.]

In the analysis of seasonal and pandemic influenza in this chapter, we have used simple *SIR* models. However, influenza has a more complicated structure, with exposed periods and asymptomatics. The obvious way to include this more complicated structure would be to use age of infection models.

3. Formulate an age of infection model analogous to the seasonal epidemic model (9.13), and calculate its reproduction number and final size relation.
4. Formulate an age of infection model analogous to the pandemic influenza model (9.18), and calculate its reproduction number and final size relation.
5. Formulate an age of infection model analogous to the combined seasonal/pandemic influenza model (9.21) and calculate its reproduction number.
6. Consider the model (9.14) in Sect. 9.4. Choose parameters to make the basic reproduction number equal to 1.5, take $\sigma_i = 1 - q^i$, where $q = 0.75$, and $\tau_i = 1$. Simulate the model and determine the final sizes.
7. Use the final sizes obtained from Exercise 6 as initial sizes and repeat the simulation. Repeat this several times to see if the final sizes approach limits.

8. Use the limiting final sizes obtained in Exercise 7 as initial sizes and simulate the model (9.18) of Sect. 9.5 with parameters giving a basic reproduction number of 2. Determine the final size of the pandemic.
9. Simulate the model (9.21) of Sect. 9.5 for various values of ρ between 0 and 1. For which values do seasonal and pandemic strains coexist and for which values does the pandemic strain replace the seasonal strain? (References: [3, 4, 8].)

References

1. Alexander, M.E., C.S. Bowman, Z. Feng, M. Gardam, S.M. Moghadas, G. Röst, J. Wu & P. Yan (2007) Emergence of drug resistance: implications for antiviral control of pandemic influenza, *Proc. Roy. Soc. B* **274**: 1675–1684.
2. Anderson, G. W., R.N. Arnestin, and M.R. Lester (1962) Communicable Disease Control, Macmillan, New York.
3. Andreasen, V. (2003) Dynamics of annual influenza A epidemics with immuno-selection, *J. Math. Biol.* **46**: 504–536.
4. Andreasen, V., J. Lin, & S.A. Levin (1997) The dynamics of cocirculating influenza strains conferring partial cross-immunity, *J. Math. Biol.* **35**: 825–842.
5. Arino, J., C.S. Bowman & S.M. Moghadas (2009) Antiviral resistance during pandemic influenza: implications for stockpiling and drug use, *BMC Infectious Diseases* **9**: 1–12.
6. Arino, J., F. Brauer, P. van den Driessche, J. Watmough & J. Wu (2006) Simple models for containment of a pandemic, *J. Roy. Soc. Interface* **3**: 453–457.
7. Arino, J., F. Brauer, P. van den Driessche, J. Watmough & J. Wu (2008) A model for influenza with vaccination and antiviral treatment, *Theor. Pop. Biol.* **253**: 118–130.
8. Asaduzzaman, S.M., J. Ma, & P. van den Driessche (2015) The coexistence or replacement of two subtypes of influenza, *Math. Biosc.*, **270**:1–9.
9. Bootsma, M.C.J. & N.M. Ferguson (2007) The effect of public health measures on the 1918 influenza pandemic in U.S. cities, *Proc. Nat. Acad. Sci.* **104**: 7588–7593.
10. Bremermann, H. J. and H.R. Thieme (1989) A competitive exclusion principle for pathogen virulence, *J. Math. Biol.* **27**:179–190.
11. Brown, D. (2003) A Model of Epidemic Control, The Washington Post, Saturday, May 3, 2003; Page A 07. <http://www.washingtonpost.com/ac2/>.
12. Brauer, F., P. van den Driessche, & J. Watmough (2010) Seasonal and pandemic influenza: Strategic and tactical models, *Can. Appl. Math. Q.* **19**: 139–150.
13. Castillo-Chavez, C (1987) Cross-immunity in the dynamics of homogenous and heterogeneous populations, **87** (37), Mathematical Sciences Institute, Cornell University.
14. Castillo-Chavez, C. and G. Chowell (2011) Special Issue: Mathematical Models, Challenges, and Lessons Learned from the 2009 A/H1N1 Influenza Pandemic, *Math. Biosc. Eng.* **8**(1): 246 pages, (<http://aims sciences.org/journals/displayPapers1.jsp?pubID=411>)
15. Castillo-Chavez, C., C. Castillo-Garsow, and A.A. Yakubu (2003) Mathematical models of isolation and quarantine. *JAMA*, **290**: 2876–2877.
16. Castillo-Chavez, C., H.W. Hethcote, V. Andreasen, S.A. Levin, and W.M. Liu (1989) Epidemiological models with age structure, proportionate mixing, and cross-immunity, *J. Math. Biol.* **27**: 233–258.
17. CDC (2011) Weekly Influenza Surveillance Report 2009-2010 Influenza Season, <http://www.cdc.gov/flu/weekly/weeklyarchives2009-2010/weekly20.htm>.
18. Conway, J.M., A.R.Tuite, D.N. Fisman, N. Hupert, R. Meza, B. Davoudi, K. English, P. van den Driessche, F. Brauer, J. Ma, L.A. Meyers, M. Smieja, A. Greer, D.M. Skowronski, D.L. Buckeridge, J. Kwong, J. Wu, S.M. Moghadas, D. Coombs, R.C. Brunham, & B. Pourbohloul

- (2011) Vaccination against 2009 pandemic H1N1 in a population dynamic model of Vancouver, Canada: timing is everything, *Biomed Central Public Health* **11**: 932.
19. Chowell, G., P.W. Fenimore, M.A. Castillo-Garsow, and C. Castillo-Chavez (2003) SARS Outbreaks in Ontario, Hong Kong and Singapore: the role of diagnosis and isolation as a control mechanism, *J. Theor. Biol.* **24**(1): 1–8.
20. Chowell, G., C.E. Ammon, N.W. Hengartner, & J.M. Hyman (2006) Transmission dynamics of the great influenza pandemic of 1918 in Geneva, Switzerland: Assessing the effects of hypothetical interventions, *J. Theor. Biol.* **241**: 193–204.
21. Chowell, G., C.E. Ammon, N.W. Hengartner, & J.M. Hyman (2007) Estimating the reproduction number from the initial phase of the Spanish flu pandemic waves in Geneva, Switzerland, *Math. Biosc. & Eng.* **4**: 457–479.
22. Elveback, L.R., J.P. Fox, E. Ackerman, A. Langworthy, M. Boyd & L. Gatewood (1976) An influenza simulation model for immunization studies, *Am. J. Epidemiol.* **103**: 152–165.
23. Erdem, M., M. Safan, and C. Castillo-Chavez (2017) Mathematical analysis of an SIQR influenza model with imperfect quarantine, *Bull. Math. Biol.* **79**: 1612–1636.
24. Feng, Z. (1994) Multi-annual outbreaks of childhood diseases revisited the impact of isolation, Ph.D. Thesis, Arizona State University.
25. Feng, Z. and H.R.Thieme (1995) Recurrent outbreaks of childhood diseases revisited: the impact of isolation, *Math. Biosc.* **128**: 93–130.
26. Gardam, M., D. Liang, S.M. Moghadas, J. Wu, Q. Zeng & H. Zhu (2007) The impact of prophylaxis of healthcare workers on influenza pandemic burden, *J. Roy. Soc. Interface*, **4**: 727–734.
27. Fenichel, E.P., C. Castillo-Chavez, M.G. Cuddia, G. Chowell, P.A. Gonzalez Parra, G.J. Hickling, G. Holloway, R. Horan, B. Morin, C. Perrings, M. Springborn, L. Velazquez, and C. Villalobos (2011) Adaptive human behavior in epidemiological models, *Proc. Nat. Acad. Sci.* **108**: 6306–6311.
28. Gonzalez-Parra, P., S. Lee, C. Castillo-Chavez, and L. Velazquez (2011) A note on the use of optimal control on a discrete time model of influenza dynamics, *Math. Biosc. Eng.* **8**: 193–205.
29. Gonzalez-Parra, P. A., L. Velazquez, M.C. Villalobos, and C. Castillo-Chavez (2010) Optimal control applied to a discrete influenza model. Conference Proceedings Book of the XXXVI International Operation Research Applied to Health Services, Book ISBN 13: 9788856825954 edited by Franco Angeli Edition.
30. Hannoun, C., F. Megas, and J. Piercy (2004) Immunogenicity and protective efficacy of influenza vaccination, *Virus Research* **103**: 133–138.
31. Hethcote, H.W., Z. Ma, and S. Liao (2002) Effects of quarantine in six endemic models for infectious diseases, *Math. Biosc.*, **180**: 141–160.
32. Hopf, E. (1942) Abzweigung einer periodischen Lösungen von einer stationären Lösung eines Differentialsystems, Berlin Math-Phys. Sachsische Akademie der Wissenschaften, Leipzig, **94**: 1–22.
33. Lee, S., G. Chowell, and C. Castillo-Chavez (2010) Optimal control of influenza pandemics: The role of antiviral treatment and isolation, *J. Theor. Biol.* **265**: 136–150.
34. Lee, S., R. Morales, and C. Castillo-Chavez (2011) A note on the use of influenza vaccination strategies when supply is limited, *Math. Biosc. Eng.* **8**: 179–191.
35. Longini, I.M., M.E. Halloran, A. Nizam, & Y. Yang (2004) Containing pandemic influenza with antiviral agents, *Am. J. Epidemiol.* **159**: 623–633.
36. Lipsitch, M., T. Cohen, M. Murray & B.R. Levin (2007) Antiviral resistance and the control of pandemic influenza *PLoS Med.* **4**: e15.
37. Manuel, O., M Pascual, K. Hoschler, S. Giulieri, K. Ellefsen, P. Bart, J. Venet, T. Calandra, and M. Cavassini (2011) Humoral response to the influenza A H1N1/09 monovalent AS03-adjuvanted vaccine in immunocompromised patients, *Clinical Infectious Diseases* **52**: 248–256.
38. Mubayi, A., C. Kribs-Zaleta, M. Martcheva, and C. Castillo-Chavez (2010) A cost-based comparison of quarantine strategies for new emerging diseases, *Math. Biosc. Eng.* **7**: 687–717.

39. Nuno, M., Z. Feng, M. Martcheva, and C. Castillo-Chavez (2005) Dynamics of two-strain influenza with isolation and partial cross-immunity, SIAM J. Appl. Math. **65**: 964–982.
40. Prosper, O., O. Saucedo, D. Thompson, G. Torres-Garcia, X. Wang, and C. Castillo-Chavez (2011) Control strategies for concurrent epidemics of seasonal and H1N1 influenza, Math. Biosc. Eng. **8**: 147–177.
41. Public Health Agency of Canada, Vaccine Development Process, URL <http://www.phac-aspc.gc.ca/influenza/pandemic-eng.php>.
42. Qiu, Z. & Z. Feng (2010) Transmission dynamics of an influenza model with vaccination and antiviral treatment, Bull. Math. Biol. **72**: 1–33.
43. Stillianakis, N.I., A.S. Perelson & F.G. Hayden (1998) Emergence of drug resistance during an influenza epidemic: insights from a mathematical model, J. Inf. Diseases, **177**: 863–873.
44. Olinsky, R., A. Huppert & L. Stone (2008) Seasonal dynamics and threshold governing recurrent epidemics, J. Math. Biol. **56**: 827–839.
45. Stone, L., R. Olinsky & A. Huppert (2007) Seasonal dynamics of recurrent epidemics, Nature **446**: 533–536.
46. Thacker, S. B. (1986) The persistence of influenza A in human populations, Epidemiologic Reviews, **8**: 129–142.
47. Towers, S., and Z. Feng (2009) Pandemic H1N1 influenza: Predicting the course of pandemic and assessing the efficacy of the planned vaccination programme in the United States, Eurosurveillance, **14**(41): Article 2.
48. van den Driessche, P. & J. Watmough (2002) Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission, Math. Biosc. **180**:29–48.
49. Vivas-Barber, A., C. Castillo-Chavez, and E. Barany (2014) Dynamics of an “SAIQR” influenza model, Biomath, **3**: 1–13.
50. Welliver, R., A.S. Monto, O. Carewicz, E. Schatteman, M. Hassman J. Hedrick, H.C. Jackson, L. Huson, P. Ward & J.S. Oxford (2001) Effectiveness of oseltamivir in preventing influenza in household contacts: a randomized controlled trial, JAMA **285**: 748–754.
51. Wu, L. and Z. Feng (2000) Homoclinic bifurcation in an SIQR model for childhood disease, J. Differential Equations, **168**: 150–167.

Chapter 10

Models for Ebola



Another important infectious disease is Ebola virus disease (EVD). Ebola hemorrhagic fever is a very infectious disease with a case fatality rate of more than 70%. It was first identified in 1976 in the Democratic Republic of Congo and there have been more than a dozen serious outbreaks since then. The most serious outbreak to date occurred in 2014 in Guinea, Liberia, and Sierra Leone and caused more than 10,000 deaths. Response to this epidemic included the development of vaccines to combat the disease, currently being used to combat the most recent outbreak in the Democratic Republic of Congo.

A distinctive feature of the Ebola virus is much disease transmission occurs through contact with bodily fluids at funerals of Ebola victims. Many mathematical models have been used to study its disease transmission dynamics. Most of these studies have focused on estimating the basic and effective reproduction numbers of EVD, assessing the rate of growth of an epidemic outbreak, evaluating the effect of control measures on the spread of EVD, and conducting more theoretical investigations on how model assumptions may affect model outcomes (see, for example, [1, 3, 7, 8, 11, 18, 23, 25–27, 29, 31, 35, 36, 40]). Although some of these models have provided useful information and better understanding of EVD dynamics and evaluation of control programs, most of these models failed to provide reasonable projections for the 2014 outbreak in West Africa. It is important to examine the reasons for this, including the underlying assumptions made in these models. In this chapter, we describe several models for Ebola.

10.1 Estimation of Initial Growth and Reproduction Numbers

Estimation of the basic and effective reproduction numbers for EVD has been conducted for both the 2014 outbreaks in West Africa and some of the outbreaks in the past (see, for example, [1, 11, 23, 25, 36]). In [11], a standard *SEIR* model

is used for estimating \mathcal{R}_0 and exploring the effect of timing of the intervention on the epidemic final size. Let $S(t)$, $E(t)$, $I(t)$, and $R(t)$ denote the number of susceptible, exposed, infective, and removed individuals at time t (the dot denotes time derivatives), and let $C(t)$ denote the cumulative number of Ebola cases from the time of onset of symptoms. Assume that exposed individuals undergo an average incubation period (asymptomatic and uninfected) of $1/k$ days before progressing to the infective class I . Infective individuals move to the R -class (death or recovered) at the per-capita rate γ . The model reads

$$\begin{aligned} S' &= -\beta S(t)I(t)/N \\ E' &= \beta S(t)I(t)/N - kE(t) \\ I' &= kE(t) - \gamma I(t) \\ R' &= \gamma I(t) \\ C' &= kE(t). \end{aligned} \tag{10.1}$$

The model (10.1) predicts initial exponential growth of the number of infectives, but in fact the initial growth rate is less than exponential. One way to modify the model to allow this is to assume behavioral changes including education of hospital personnel and community members on the use of strict barrier nursing techniques (i.e., protective clothing and equipment, patient management), and the rapid burial or cremation of patients who die from the disease. It is assumed that the net effect is a reduced transmission rate β from β_0 to $\beta_1 < \beta_0$. To take into consideration that the impact of the intervention is not instantaneous, the transmission rate is assumed to decrease gradually from β_0 to β_1 according to

$$\beta(t) = \begin{cases} \beta_0 & t < \tau \\ \beta_1 + (\beta_0 - \beta_1)e^{-q(t-\tau)} & t \geq \tau \end{cases}$$

where τ is the time at which interventions start and q controls the rate of the transition from β_0 to β_1 . Another interpretation of the parameter q can be given in terms of $t_h = \frac{\log(2)}{q}$, the time to achieve $\beta(t) = \frac{\beta_0+\beta_1}{2}$.

The basic reproduction number \mathcal{R}_0 corresponds to β_0 . If \mathcal{R}_0 can be estimated from data, then β_0 can be determined using the relation $\mathcal{R}_0 = \beta_0/\gamma$. For estimating \mathcal{R}_0 , consider the E and I equations in (10.1) and the corresponding Jacobian matrix J at the disease-free equilibrium

$$J = \begin{pmatrix} -k & \beta \\ k & -\gamma \end{pmatrix}.$$

The characteristic equation is given by

$$r^2 + (k + \gamma)r + (\gamma - \beta)k = 0, \tag{10.2}$$

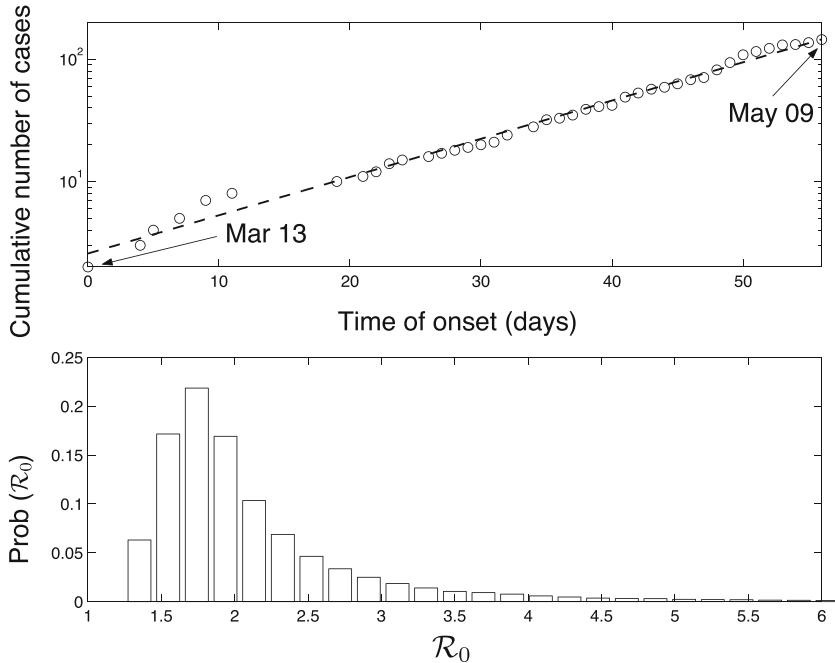


Fig. 10.1 (Top) Cumulative number of cases (log-lin scale) during the exponential growth phase of the Congo 1995 epidemic as identified by the date of start of interventions (09 May 1995 [24]). The model-free initial growth rate of the number of cases for Congo 1995 is 0.07 (linear regression); (bottom) estimated distribution of \mathcal{R}_0 from our uncertainty analysis (see text). \mathcal{R}_0 lies in the interquartile range (IQR) (1.66–2.28) with a median of 1.89. Notice that 100% of the weight lies above $\mathcal{R}_0 = 1$

and the dominant eigenvalue r represents the early-time and per-capita free growth of the outbreak. Replacing the β in (10.2) by $\gamma\mathcal{R}_0$ and solving for \mathcal{R}_0 , we obtain

$$\mathcal{R}_0 = 1 + \frac{r^2 + (k + \gamma)r}{k\gamma}.$$

Using the time series $y(t)$ (before intervention) of the cumulative number of cases and assuming exponential growth ($y(t) \propto e^{rt}$) an estimate of r can be obtained, as shown in Fig. 10.1 (top). The estimate of the initial rate of growth r for the Congo 1995 epidemic is $r = 0.07 \text{ day}^{-1}$. Based on this fixed r and Monte Carlo sampling of size 10^5 from the distributed epidemic parameters ($1/k$ and $1/\gamma$) [5], a distribution of \mathcal{R}_0 can be obtained as demonstrated in Fig. 10.1 (bottom), which shows that the distribution lies in the interquartile range (IQR) (1.66–2.28) with a median of 1.89.

Similar estimates for Congo 1995 data can be obtained. The estimated parameter values are listed in Table 10.1.

Table 10.1 Parameter definitions and baseline estimates obtained from the best fit of the model equations (10.1) to the epidemic-curve data of the Congo 1995 and Uganda 2000 outbreaks

Parameter	Definition	Congo 1995		Uganda 2000	
		Estim.	SD	Estim.	SD
β_0	Pre-interventions transmission rate (days ⁻¹)	0.33	0.06	0.38	0.24
β_1	Post-interventions transmission rate (days ⁻¹)	0.09	0.01	0.19	0.13
t_h	Time to achieve $\frac{\beta_0+\beta_1}{2}$ (days)	0.71	(0.02, 1.39)	0.11	(0, 0.87)
$1/k$	Mean incubation period (days)	5.30	0.23	3.35	0.49
$1/\gamma$	Mean infectious period (days)	5.61	0.19	3.50	0.67

The parameters are $0 < \beta < 1$, $0 < q < 100$, $1 < 1/k < 21$, $3.5 < 1/\gamma < 10.7$

We use the model (10.1) to evaluate intervention strategies, including surveillance and placement of suspected cases in quarantine for 3 weeks (the maximum estimated length of the incubation period). The effectiveness of interventions can be quantified in terms of the reproduction number \mathcal{R}_p after interventions are implemented. For the case of Congo $\mathcal{R}_p = 0.51$ (SD 0.04) and $\mathcal{R}_p = 0.66$ (SD 0.02) for Uganda. Furthermore, the time to achieve a transmission rate of $\frac{\beta_0+\beta_1}{2}$ (t_h) is 0.71 (95% CI (0.02, 1.39)) days and 0.11 (95% CI (0, 0.87)) days for the cases of Congo and Uganda, respectively, after the time at which interventions begin.

Using the parameter values estimated from early growth, the model (10.1) can be used to simulate the Ebola outbreaks in Congo (1995) and Uganda (2000). Figure 10.2 illustrates results via Monte Carlo simulations of the stochastic model corresponding to (10.1) [33], which is constructed by considering three events: *exposure*, *infection*, and *removal*. The transition rates are defined as

Event	Effect	Transition rate
Exposure	$(S, E, I, R) \rightarrow (S-1, E+1, I, R)$	$\beta(t)SI/N$
Infection	$(S, E, I, R) \rightarrow (S, E-1, I+1, R)$	kE
Removal	$(S, E, I, R) \rightarrow (S, E, I-1, R+1)$	γI

Figure 10.2 illustrates that there is very good agreement between the mean of the stochastic simulations and the reported cases. The empirical distribution of the epidemic final sizes for the cases of Congo 1995 and Uganda 2000 is given in Fig. 10.3.

The epidemic final size is sensitive to the start time of interventions τ . Numerical solutions (deterministic model) show that the epidemic final size grows exponentially fast with the initial time of interventions. For instance, for the case of Congo, the model predicts that there would have been 20 more cases if interventions had started 1 day later, as shown in Fig. 10.4.

As for most outbreaks, the initial growth of the outbreaks in Congo 1995 and Uganda 2000 is exponential. However, it is discussed that for the 2014 outbreaks in West Africa there are significant differences in the growth patterns of EVD cases at

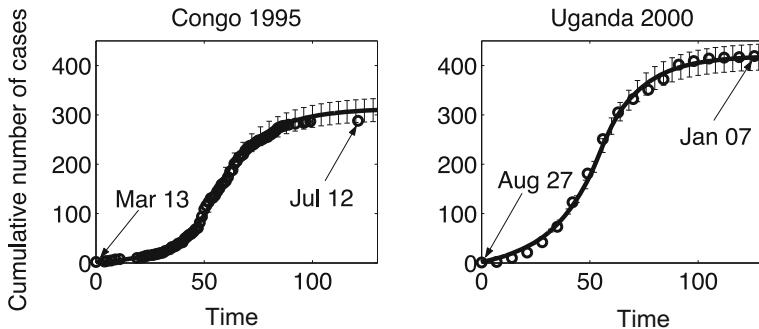


Fig. 10.2 Comparison of the cumulative number of Ebola cases during the Congo 1995 and Uganda 2000 Ebola outbreaks, as a function of the time of onset of symptoms. Circles are the data. The solid line is the average of 250 Monte Carlo replicates and the error bars represent the standard error around the mean from the simulation replicates using our parameter estimates (Table 10.1). For the case of Congo 1995, simulations were begun on 13 Mar 1995. A reduction in the transmission rate β due to the implementation of interventions occurs on 09 May 1995 (day 56) [24]. For the case of Uganda 2000, simulations start on 27 August 2000 and interventions take place on 22 October 2000 (day 56) [41]

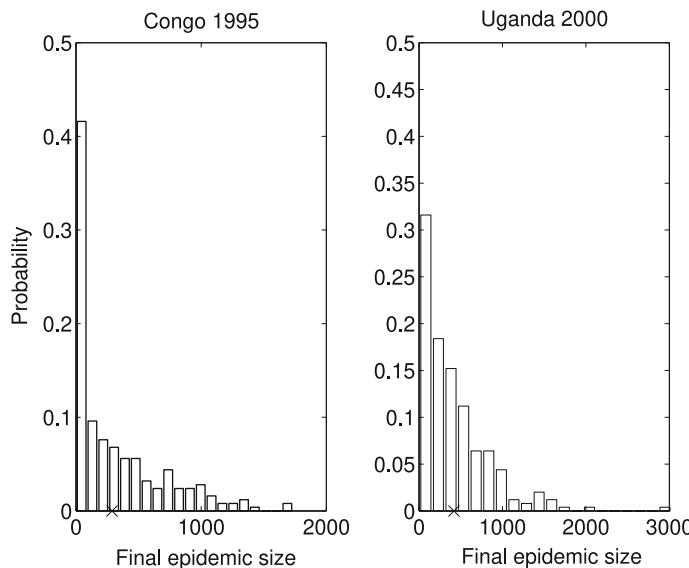


Fig. 10.3 The epidemic final size distributions for the cases of Congo 1995 and Uganda 2000 obtained from 250 Monte Carlo replicas. Crosses (X) represent the epidemic final size from data

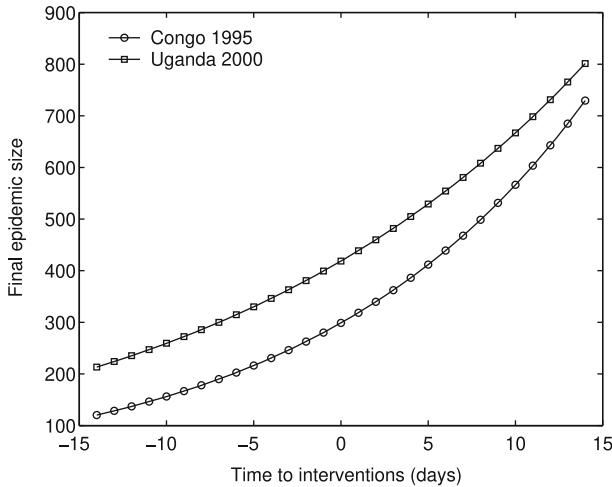


Fig. 10.4 Sensitivity of the final epidemic final size to the starting time of interventions. The negative numbers represent number of days before the actual reported intervention date and positive numbers represent a delay after the actual reported intervention date ($\tau = 0$)

the scale of the country, district, and other sub-national administrative divisions. It is illustrated that the cumulative number of EVD cases in a number of administrative areas of Guinea, Sierra Leone, and Liberia is best approximated by polynomial rather than exponential growth over several generations of EVD. It is also observed that, when data are aggregated nationally, or across the broader West Africa region, total case counts show periods of approximate exponential growth.

Temporal evolution of the effective reproduction number of Ebola is studied in [36]. In this study, a simple SEIR model with standard incidence is used in combination with the limited existing data to determine whether the transmission rate of Ebola has been changing over time in West Africa. To this end, piecewise exponential curves were fit to the time series of outbreak data to estimate the temporal evolution of the effective reproduction number of the disease. Instead of \mathcal{R}_0 , the study focuses on assessing the time evolution of the effective reproduction number denoted by \mathcal{R}_{eff} , which is a dynamic estimate of the average number of secondary cases per infectious case in a population composed of both susceptible and non-susceptible individuals during the course of an outbreak. The SEIR model and its linearization about this temporary “equilibrium” are used to determine the predicted local rate of exponential rise of the epidemic curve, ρ_{eff} . For the SEIR model, this is related to \mathcal{R}_{eff} by

$$\mathcal{R}_{eff} = \left(1 + \frac{\rho_{eff}}{\gamma}\right) \left(1 + \frac{\rho_{eff}}{\kappa}\right), \quad (10.3)$$

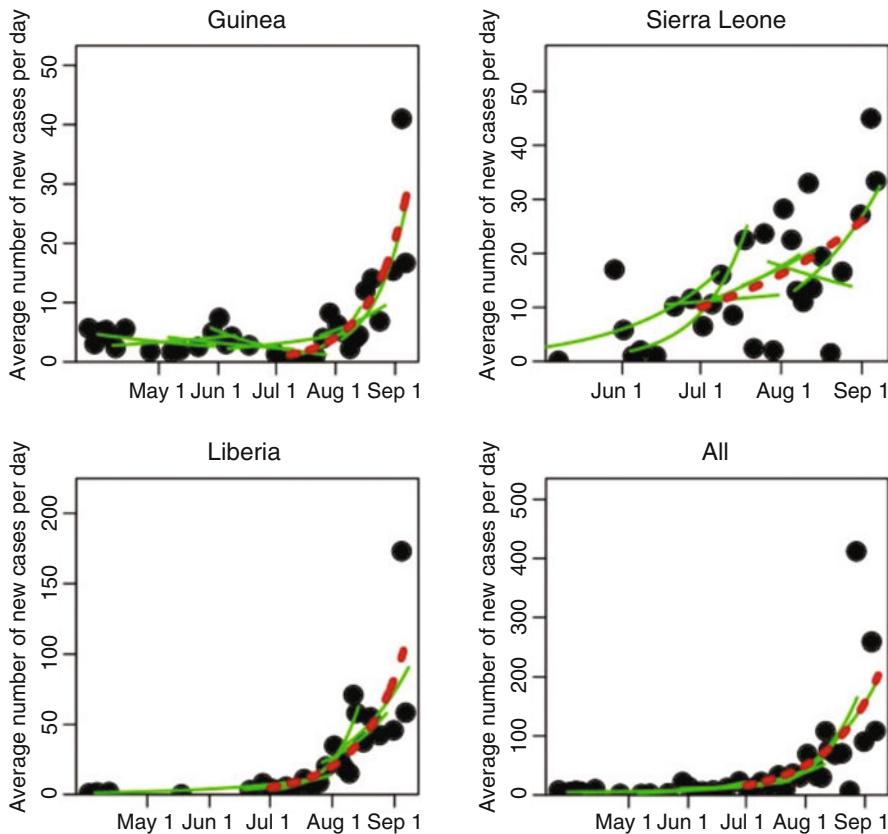


Fig. 10.5 Time series of incidence of EVD cases in West Africa

where $1/\kappa$ and $1/\gamma$ are the average incubation and infective periods of the disease, respectively. With estimates of ρ_{eff} from piecewise exponential rise fits to the incidence data from an outbreak (along with estimates of the incubation and infective periods of the disease), ρ can be estimated (Fig. 10.5), and then Eq. (10.3) can be used to obtain estimates of the temporal behavior of \mathcal{R}_{eff} (Fig. 10.6), in essence approximating the temporal behavior with a piece-wise step-function.

Figure 10.5 shows time series of recorded average number of new EVD cases per day during the initial phase of the 2014 West African outbreak, for Guinea, Sierra Leone, and Liberia (dots). The green lines show a selection of the piecewise exponential fits to the data (not all fits are shown to clarify the presentation); a moving window of groups of ten contiguous points are taken at a time, and the rate of exponential rise estimated for those ten points. The results for the estimations of the exponential rise for the full set of piece-wise fits are shown in Fig. 10.6. Shown in red is the fitted exponential rise from July 1st onwards.

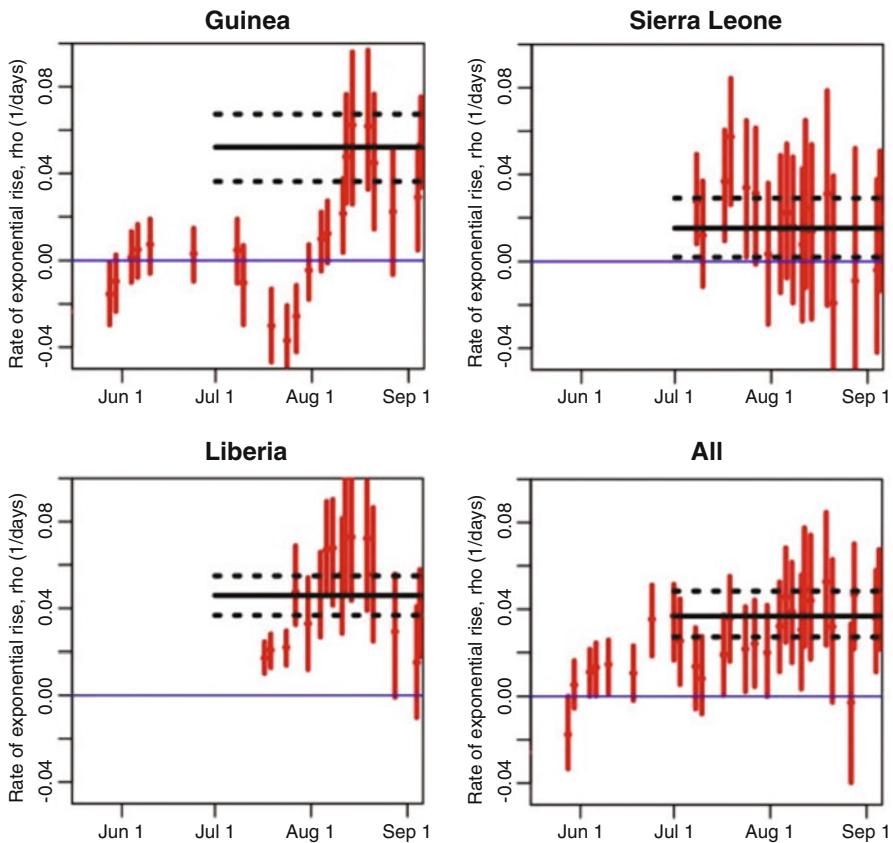


Fig. 10.6 Estimated rates of exponential rise from piece-wise exponential fits

In Fig. 10.6, estimated rates of exponential rise from piece-wise exponential fits to the average daily EVD incidence data, as shown in Fig. 10.5; a moving window of groups of ten contiguous incidence data time series points are taken at a time, and the rate of exponential rise estimated for those ten points. The dates shown on the x axis are the last date in each contiguous set of ten points, and the vertical error bars denote the 95% confidence interval. The horizontal black line shows the estimated rate of rise of an exponential fit to the incidence time series from July 1st to September 8th, with the black dotted lines indicating the 95% confidence interval.

10.1.1 Early Detection

The effect of early detection on Ebola control is studied in [10]. The model considers six epidemiological classes: susceptible individuals (S), latent undetectable individuals (E_1), latent detectable individuals (E_2), infectious symptomatic individuals (I),

isolated individuals (J), and individuals removed from isolation after recovery or disease-induced death (P), $P = R + D$, where R is the recovered class and D is the death-induced class. Susceptible individuals become infected and latent through contact with an infectious individual at the per-capita rate $(I + lJ) = N$, where β is the mean transmission rate per day, l is defined as the relative transmissibility of isolated individuals, i.e., it is a measure of the effectiveness of isolation of infective individuals, and N is the total population size. Latent undetectable individuals (E_1) enter the latent detectable class E_2 at a rate k_1 , and become infectious symptomatic at a rate k_2 . A fraction of the latent detectable individuals are diagnosed (i.e., by RT-PCR), $p_T = f_T = (f_T + k_2)$, and become isolated. We assume that the latent detectable class represents individuals with a viral load above the detection limit of the specific diagnostic test. Infective individuals are isolated at a rate α , or they are removed after recovery or disease-induced death at a rate γ . Similarly, individuals are removed from isolation after recovery or disease-induced death, but at a rate γ_r . The model reads

$$\begin{aligned} S' &= -\beta S \frac{I + lJ}{N} \\ E'_1 &= \beta S \frac{I + lJ}{N} - k_1 E_1 \\ E'_2 &= k_1 E_1 - k_2 E_2 - f_T E_2 \\ I' &= k_2 E_2 - (\alpha + \gamma) I \\ J' &= \alpha I + f_T E_2 - \gamma_r J \\ R' &= \gamma(1 - \delta) I + \gamma_r(1 - \delta) J \\ D' &= \gamma \delta I + \gamma_r \delta J \\ N &= S + E_1 + E_2 + I + J + R. \end{aligned} \tag{10.4}$$

The analysis suggests that the impact of early diagnosis of pre-symptomatic infections is strongly dependent on the effectiveness of isolation of infective individuals in health-care settings. For instance, with an isolation effectiveness of 50% and with an average time of 3 days from the onset of symptoms to isolation, the attack rate (total number of Ebola cases/population size) remains essentially unchanged as the rate of pre-symptomatic case detection increases (Fig. 10.7). In contrast, early detection of pre-symptomatic individuals can have a significant impact on the transmission dynamics of Ebola if the effectiveness of isolating infective cases is at least 60%. Even at this level of isolation, at least 50% of pre-symptomatic cases would need to be detected in the community, a scenario difficult to achieve with limited resources. When the effectiveness of isolation is increased to 65%, detecting about 25% of pre-symptomatic cases is predicted to lead to epidemic control, i.e., the effective reproduction number is reduced below the epidemic threshold.

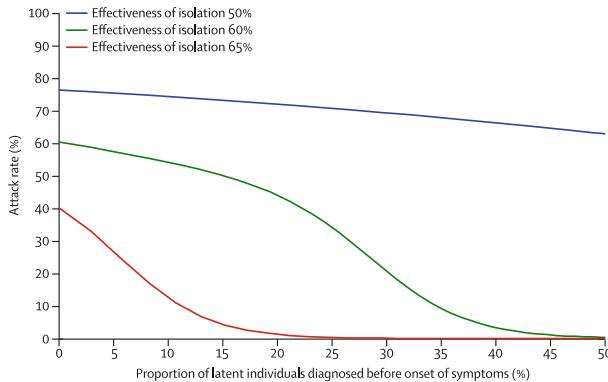


Fig. 10.7 Predictions of the effect of diagnosing pre-symptomatic individuals on the attack rate of the Ebola epidemic. The mean time from onset of symptoms to isolation was set at 3 days

10.2 Evaluations of Control Measures

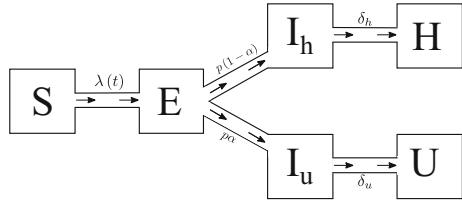
One of the typical characteristics for EVD is that significant transmissions can occur by infectious individuals after death but before burial. Also, no effective drugs or vaccine were available for the outbreak in West Africa. The main control measures include isolation, hospitalization, contact tracing, and safe burial. Several mathematical models have been used to assess the effectiveness of these control measures (e.g., [3, 7, 26, 27, 35, 40]). Most of the models in these studies use SEIR type of models with various modifications. One of the EVD models on which many other models are based is the one considered in Legrand et al. [26], which will be discussed in more detail in the next section.

A model considered in [7] has the following form (with modified notation):

$$\begin{aligned}
 S' &= -\beta S(I_h + I_u)/N \\
 E' &= \beta S(I_h + I_u)/N - \alpha E \\
 I'_h &= p\alpha E - \delta_h I_h \\
 I'_u &= (1 - p)\alpha E - \delta_u I_u \\
 H' &= \delta_h I_h,
 \end{aligned} \tag{10.5}$$

where N is the total population size and is assumed to be constant. In this model, two separate compartments for infective individuals are considered, namely infective individuals which will be hospitalized/reported (I_h) and infectious individuals which will not be hospitalized and unreported (I_u). The H compartment represents the cumulative hospitalized/reported cases (so it includes those who are recovered after being hospitalized). The model (10.5) also assumes that individuals in both I_h and I_u have the same transmission rate β , p is the fraction of hospitalized cases, $1/\delta_h$ is time from infectiousness (symptom) onset until hospitalization/isolation

Fig. 10.8 A transition diagram for model (10.5)



and reporting, and $1/\delta_u$ is the mean infective period for an unhospitalized case. A transition diagram of the model is depicted in Fig. 10.8.

For model (10.5), the reproduction number is given by

$$\mathcal{R}_e = \mathcal{R}_0 \left(p \frac{\delta_u}{\delta_h} + 1 - p \right),$$

where $\mathcal{R}_0 = \beta/\delta_u$ is the basic reproduction number. An extension of the model (10.5) is also considered in [7] to take into consideration contact tracing.

The following deterministic model is considered in [25].

$$\begin{aligned} S' &= -\mathcal{R}_0 \delta S I / N, \\ E'_1 &= \mathcal{R}_0 \delta S I / N - m\alpha E_1, \\ E'_i &= m\alpha(E_{i-1} - E_i), \quad i = 2, \dots, m \\ I' &= m\alpha E_m - \delta I, \\ R' &= \delta I. \end{aligned} \tag{10.6}$$

In model (10.6), the stage distribution of the latent stage is assumed to be gamma with the shape parameter equal to m , which leads to a division of the exposed class E into m sub-classes with mean duration $1/(m\alpha)$. By examining the model fit to the weekly case reports in Guinea, Liberia, and Sierra Leone from the WHO situation report dated from 1 October 2014 (<http://www.who.int/csr/disease/ebola/situation-reports/en/>), the authors pointed out that fitting of such deterministic models to cumulative incidence data can lead to bias and pronounced underestimation of the uncertainty associated with model parameters.

The models (10.5) and (10.6) ignored the special characteristic associated with the fact that significant transmissions can occur by those who are dead but not yet buried. This is considered in the Legrand model, which is discussed in detail next.

10.3 The Legrand Model and Underlying Assumptions

Many mathematical models have been used for the recent epidemics of Ebola in West Africa. However, the success of these models in the case of the 2014 Ebola outbreak in West Africa was very limited. As pointed out in [8], “mathematical models have failed to accurately project the outbreak’s course.” Although various reasons may explain why “on-the-ground data contradict the projections of published models,” including incomplete and unreliable data on Ebola epidemiology (especially in the hardest-hit areas) and lack of empirical data on how disease-control measures quantitatively affect Ebola transmission, it is important to examine the appropriateness of assumptions made in the models on which the projections are based. This is the objective of this section. There have been various modeling approaches, including deterministic and stochastic models, or relatively simple models consisting of ordinary differential equations (ODEs) and more complicated agent-based models, among others. Many of the ODE models are variations of the model studied by the Legrand model (10.7). It has been pointed out that some of the assumptions made in the Legrand model may not have clear justifications (e.g., [35]). Thus, it is important to examine the critical assumptions made in this model and better understand their possible impact on model outcomes.

10.3.1 The Legrand Model

The Legrand et al. model [26] consists of a system of ordinary differential equations with six compartments representing the epidemiological classes of susceptible (S), exposed (E), infective (I), hospitalized (H), dead but not yet buried (D), and removed (R). The transition diagram of the model is depicted in Fig. 10.9.

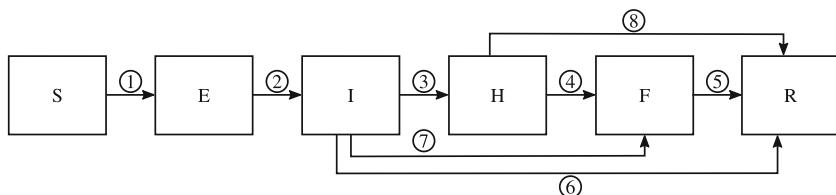


Fig. 10.9 A transition diagram for the model in Legrand et al. [26]

The model presented by Legrand et al. [26] reads

$$\begin{aligned}\frac{dS}{dt} &= -\frac{1}{N} S(\beta_I I + \beta_H H + \beta_D D) \\ \frac{dE}{dt} &= \frac{1}{N} S(\beta_I I + \beta_H H + \beta_D D) - \alpha E \\ \frac{dI}{dt} &= \alpha E - (\gamma_h \theta_1 + \gamma_i(1 - \theta_1)(1 - \delta_1) + \gamma_d(1 - \theta_1)\delta_1)I \\ \frac{dH}{dt} &= \gamma_h \theta_1 I - (\gamma_{dh}\delta_2 + \gamma_{ih}(1 - \delta_2))H \\ \frac{dD}{dt} &= \gamma_d(1 - \theta_1)\delta_1 I + \gamma_{dh}\delta_2 H - \gamma_f D \\ \frac{dR}{dt} &= \gamma_i(1 - \theta_1)(1 - \delta_1)I + \gamma_{ih}(1 - \delta_2)H + \gamma_f D.\end{aligned}\tag{10.7}$$

The parameters β_I , β_H , and β_D denote the transmission rates in the I , H , and D classes, respectively; let $1/\alpha$ be the mean latent period; and let $1/\gamma_f$ be the mean time between death and burial.

The three key parameters that are in the Legrand model (10.7), θ_1 , δ_1 , and δ_2 , do not have direct biological meaning but are computed based on the probabilities of hospitalization and of disease-induced mortality with or without hospitalization. For example, the fraction of infective people hospitalized is

$$p = \frac{\gamma_h \theta_1}{\gamma_h \theta_1 + \gamma_i(1 - \theta_1)(1 - \delta_1) + \gamma_d(1 - \theta_1)\delta_1},\tag{10.8}$$

and the probabilities of death with (f_h) and without hospitalization (f_i) are given by

$$\begin{aligned}f_h &= \frac{\gamma_{dh}\delta_2}{\gamma_{dh}\delta_2 + \gamma_{ih}(1 - \delta_2)} \\ f_i &= \frac{\gamma_d\delta_1}{\gamma_i(1 - \delta_1) + \gamma_d\delta_1}.\end{aligned}\tag{10.9}$$

If we assume $f_i = f_h = f$, then θ_1 , δ_1 , and δ_2 can be determined in terms of p and f using (10.8) and (10.9).

In addition, the Legrand model imposes the following constraints:

$$\frac{1}{\gamma_i} = \frac{1}{\gamma_h} + \frac{1}{\gamma_{ih}} \quad \text{and} \quad \frac{1}{\gamma_d} = \frac{1}{\gamma_h} + \frac{1}{\gamma_{dh}},\tag{10.10}$$

and assumes that hospitalization does not affect the time from onset to recovery or from onset to death. Other assumptions made in model (10.7) are associated with exponential waiting times. It is assumed that, after entering the infective class I , individuals can leave due either to hospitalization (entering H) or recovery without being hospitalized (entering R from I) or death without being hospitalized (entering D from I) with average waiting times $1/\gamma_h$, $1/\gamma_i$, $1/\gamma_d$, respectively. Or equivalently, after onset, individuals enter the H , R , and D classes at constant rates γ_h , γ_i , and γ_d , respectively. Their model assumes that the overall rate of leaving the I class, denoted by Δ , is a weighted average of the three rates γ_h , γ_i , and γ_d as

$$\Delta = \theta_1 \gamma_h + (1 - \theta_1) \delta_1 \gamma_d + (1 - \theta_1)(1 - \delta_1) \gamma_i, \quad (10.11)$$

where θ_1 is the proportion of cases hospitalized, and δ_1 is a coefficient that is determined such that

$$\frac{\delta_1 \gamma_d}{\delta_1 \gamma_d + (1 - \delta_1) \gamma_i}$$

is equal to case fatality (i.e., the proportion of cases that die).

10.3.2 A Simpler System Equivalent to the Legrand Model

It is shown in [18] that the Legrand model (10.7) is equivalent to the following model:

$$\begin{aligned} \frac{dS}{dt} &= -\frac{1}{N} S(\beta_I I + \beta_H H + \beta_D D), \\ \frac{dE}{dt} &= \frac{1}{N} S(\beta_I I + \beta_H H + \beta_D D) - \alpha E, \\ \frac{dI}{dt} &= \alpha E - \gamma I, \\ \frac{dH}{dt} &= p\gamma I - \omega H, \\ \frac{dD}{dt} &= (1-p)f\gamma I + f\omega H - \gamma_f D, \\ \frac{dR}{dt} &= (1-p)(1-f)\gamma I + (1-f)\omega H, \end{aligned} \quad (10.12)$$

where

$$\begin{aligned} \frac{1}{\gamma} &= p \frac{1}{\gamma_{IH}} + (1-p)f \frac{1}{\gamma_{ID}} + (1-p)(1-f) \frac{1}{\gamma_{IR}}, \\ \frac{1}{\omega} &= f \frac{1}{\omega_{HD}} + (1-f) \frac{1}{\omega_{HR}}. \end{aligned} \quad (10.13)$$

Using the following connections between parameters in (10.12) and the Legrand model (10.7):

$$\gamma_{IR} = \gamma_i, \quad \gamma_{IH} = \gamma_h, \quad \gamma_{ID} = \gamma_d,$$

the conditions in (10.13) can be written as

$$\begin{aligned} \frac{1}{\gamma} &= p \frac{1}{\gamma_h} + (1-p)f_i \frac{1}{\gamma_d} + (1-p)(1-f_i) \frac{1}{\gamma_i}, \\ \frac{1}{\omega} &= f_h \frac{1}{\gamma_{dh}} + (1-f_h) \frac{1}{\gamma_{ih}}, \end{aligned} \quad (10.14)$$

and the constraints (10.10) becomes

$$\frac{1}{\gamma_{IR}} = \frac{1}{\gamma_{IH}} + \frac{1}{\omega_{HR}}, \quad \frac{1}{\gamma_{ID}} = \frac{1}{\gamma_{IH}} + \frac{1}{\omega_{HD}}. \quad (10.15)$$

The only minor difference between the two models is where to move the buried (whether or not to R), which does not affect the dynamic behavior of the model.

10.4 *Models with Various Assumptions on Stage Transition Times

In the Legrand model (10.7) or the equivalent model (10.12), it is not clear what underlying assumptions have been made regarding the distributions of waiting times for epidemiological processes including the time from onset to recovery (transition from I to R), to hospitalization (transition from I to H), and to death (transition from I to D). For ease of reference, we refer to these three transitions as IR, IH, and ID, respectively. In addition, the two possible transitions for hospitalized individuals, recovery or death, are denoted by HR and HD. In this section, we derive three integro-differential equations models under different assumption on those transition times and compare the difference in the model outcomes.

Let T_P , T_L , and T_M denote random variables for the waiting times associated with IR, IH, and ID, and let the associated survival functions be denoted by $P(t)$, $L(t)$, and $M(t)$, respectively. The mean duration of these transitions are, respectively, $\mathbb{E}[T_P]$, $\mathbb{E}[T_L]$, and $\mathbb{E}[T_D]$. Similarly, let \mathcal{D}_{HR} and \mathcal{D}_{HD} denote the mean duration from hospitalization to recovery or death, respectively. For ease of comparison between models presented in this paper, we list in Table 10.2 some of the quantities that play common roles and have clear biological meaning in these models. Several of these quantities should have values that are independent of model assumptions, including the mean duration (absent intervention) from onset to recovery $\mathbb{E}[T_P]$, the probability of hospitalization p , and the probability of death f .

Table 10.2 Definition of symbols commonly used in the models in this section

Symbol	Definition
T_P, T_L, T_M	Random variables for the waiting times in I before moving to R, H, D , respectively
X_I	Random variable for the overall time spent in the I compartment
X_H	Random variable for the overall time spent in the H compartment
$P_i(s)$	Probability that a living individual remains infectious s units of time since onset for models I, II, III when $i = 1, 2, 3$, respectively. That is, $\mathbb{P}[T_{P_i} > s] = P_i(s)$
$L_i(s)$	Probability of a living individual not being hospitalized s units of time since onset for models I, II when $i = 1, 2$, respectively. That is, $\mathbb{P}[T_{L_i} > s] = L_i(s)$
$M_1(s)$	Probability of surviving the disease s units of time since onset for model I. That is, $\mathbb{P}[T_{M_1} > s] = M_1(s)$
$Q_3(s)$	Probability of not having recovered s time units after being hospitalized for model III
$\mathbb{E}[T_P]$	Mean duration from onset to recovery (absent intervention or death)
$\mathbb{E}[T_L]$	Mean duration from onset to hospitalization (given hospitalized and not dead)
$\mathbb{E}[T_M]$	Mean duration between onset and death (absent intervention or recovery)
$\mathbb{E}[X_I]$	Mean duration in the I compartment (hospitalization and death included)
$\mathbb{E}[X_H]$	Mean duration in the H compartment (death included)
\mathcal{D}_{HR}	Mean duration from hospitalization to recovery
\mathcal{D}_{HD}	Mean duration from hospitalization to death
γ_{IR}	$= 1/\mathbb{E}[T_P]$
γ_{IH}	$= 1/\mathbb{E}[T_L]$
γ_{ID}	$= 1/\mathbb{E}[T_M]$
ω_{HR}	$= 1/\mathcal{D}_{HR}$, per-capita rate of transition from H to R if the transition is exponential
ω_{HD}	$= 1/\mathcal{D}_{HD}$, per-capita rate of transition from H to D if the transition is exponential
p	Proportion hospitalized (dependent on control effort)
f	Probability of death (with or without hospitalization)
γ	$= 1/\mathbb{E}[X_I]$, per-capita rate of exiting I if X_I is exponential

If the transitions IR, IH, and ID are assumed to be independent in the Legrand model and the waiting times are all exponentially distributed with mean durations $1/\gamma_{IR}$, $1/\gamma_{IH}$, and $1/\gamma_{ID}$, respectively, then the mean overall time spent in the I compartment is

$$\mathbb{E}[\min\{T_P, T_L, T_M\}] = \int_0^\infty P(t)L(t)M(t)dt = \frac{1}{\gamma_{IR} + \gamma_{IH} + \gamma_{ID}}.$$

Thus, the overall rate of exiting I is $\gamma_{IR} + \gamma_{IH} + \gamma_{ID}$, which is not a weighted average given by Δ in (10.11) for the Legrand model. This implies that the Legrand model has made different assumptions on these transitions.

In the formulation of the integro-differential equations models, we will adopt probabilistic terminology to facilitate the interpretation of these models, and focus on the following three scenarios:

- (I) Assume that the three transitions IR, IH, and ID are *independent* and the waiting times are described by the survival functions $P_1(s)$, $L_1(s)$, and $M_1(s)$, respectively, where s represents the time-since-onset. It is also assumed that hospitalization does not affect the time from onset to recovery or death.
- (II) The two transitions IR and ID are *combined* and described by a single survival function $P_2(s)$, with a fraction $1 - f$ of the exiting individuals recovering (and the fraction f dying). The transition IH is independent of IR and ID and the waiting time is described by the survival function $L_2(s)$. Similar to model I, it is also assumed that hospitalization does not affect the time from onset to recovery or probability of death.
- (III) All three transitions (IH, IR, and ID) are combined and described by a single survival function $P_3(s)$, with a fraction p of the exiting individuals being hospitalized and a fraction $1 - f$ (respectively, f) of the non-hospitalized individuals recovering (respectively, dying). The two transitions HR and HD are combined and the waiting time is described by a single survival function $Q_3(s)$ with a fraction $1 - f$ (or f) of the exiting individuals recovering (or dying). P_3 and Q_3 are assumed to be independent. Unlike models I and II, in which the time from onset to hospitalization is tracked due to the independent stage distributions, in model III a constraint must be imposed so that the time between onset and hospitalization plus the time between hospitalization and recovery (or death) equals the time between onset and recovery (or death).

To focus on the general waiting time for the infective stage and its influence on model formulation when hospitalization is considered, assume simpler distributions for other stages including the latent stage and the duration between death and burial. That is, the E and D stages are assumed to have exponential distributions with constant rates α and γ_f . As the models are derived under arbitrary distributions for the waiting times of key disease stages, they consist of systems of integro-differential equations. It shows that these systems reduce to ODE systems when the arbitrary stage distributions are replaced by gamma or exponential distributions. Detailed derivations of the systems of integral equations are provided in Feng et al. (2016).

The model I, which corresponds to the scenario (I) described above, assumes independent IR, IH, and ID processes. Let T_{P_1} , T_{L_1} , and T_{M_1} denote random variables for the independent waiting times of IR, IH, and ID, respectively, that are described by the following survival functions:

- $P_1(s)$: Probability that a living individual remains infectious s units of time since onset (governing both IR and HR).
- $L_1(s)$: Probability of a living individual not being hospitalized s units of time since onset (governing the IH transition).
- $M_1(s)$: Probability of surviving the disease s units of time since onset (governing both ID and HD).

Figure 10.10 depicts the transitions between epidemiological classes for the model under scenario (I). All variables and parameters have the same meanings as before unless otherwise stated. The diagram in (a) depicts transitions between compartments when stage durations for the IR, IH, and ID transitions are arbitrarily described by the survival functions $P_1(t)$, $L_1(t)$, and $M_1(t)$. The dotted rectangle around the I and H compartments indicates that individuals in these two compartments are being tracked for their time-since-onset using the same survival

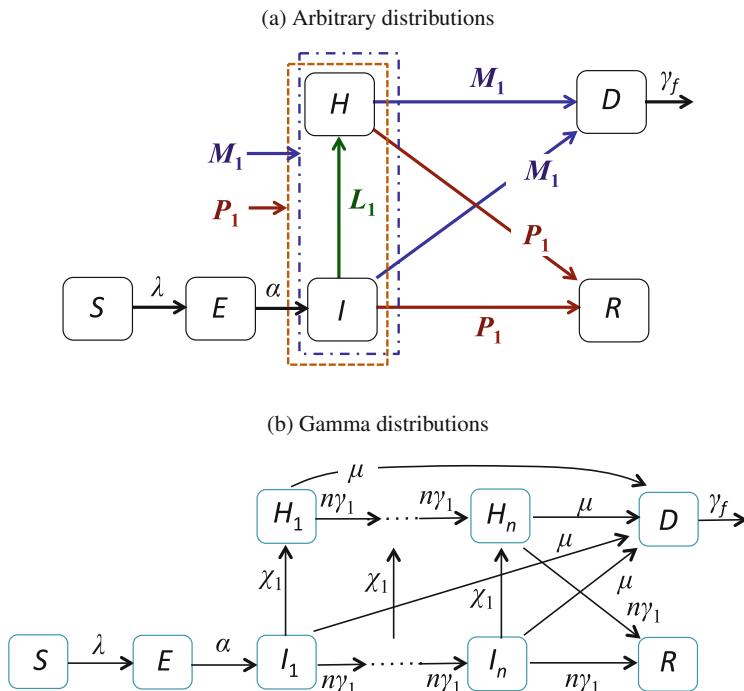


Fig. 10.10 Transition diagram for Model I when T_{P_1} , T_{L_1} , and T_{M_1} are arbitrary (a), or gamma/exponential (b). The corresponding survival functions are $P_1(t)$ (red), $L_1(t)$ (green), and $M_1(t)$, respectively

function $P_1(t)$, i.e., the time elapsed in I before entering H is taken into account when determining the time between entering H and recovery. The diagram in (b) illustrates the effect of the “linear chain trick” when $P_1(t)$ follows a gamma distribution, and $L_1(t)$ and $M_1(t)$ follow exponential survival functions.

The model with the general stage distributions P_1 , L_1 , and M_1 consists of the following system of integro-differential equations:

$$\frac{dS}{dt} = -\lambda(t)S, \quad \frac{dE}{dt} = \lambda(t)S - \alpha E,$$

$$\begin{aligned} I(t) &= \int_0^t \alpha E(s) P_1(t-s) L_1(t-s) M_1(t-s) ds + I(0) P_1(t) L_1(t) M_1(t), \\ H(t) &= \int_0^t \alpha E(s) P_1(t-s) M_1(t-s) [1 - L_1(t-s)] ds \\ &\quad + I(0) P_1(t) M_1(t) [1 - L_1(t)], \end{aligned} \tag{10.16}$$

$$\begin{aligned} D(t) &= \int_0^t \left[\int_0^\tau \alpha E(s) P_1(\tau-s) g_{M_1}(\tau-s) ds + I(0) P_1(\tau) g_{M_1}(\tau) \right] e^{-\gamma_f(t-\tau)} d\tau, \\ R(t) &= \int_0^t \left[\int_0^\tau \alpha E(s) g_{P_1}(\tau-s) M_1(\tau-s) ds + I(0) g_{P_1}(\tau) M_1(\tau) \right] d\tau, \end{aligned}$$

where $\lambda(t)$ is given by

$$\lambda(t) = \frac{\beta_I I + \beta_H H + \beta_D D}{N}, \tag{10.17}$$

The initial condition is $(S(0), E(0), I(0), H(0), D(0), R(0)) = (S_0, E_0, I_0, 0, 0, 0)$, where S_0 and E_0 are positive constants. Notice that in system (10.16), the probability distributions for T_{P_1} , T_{L_1} , and T_{M_1} are arbitrary.

Consider the case when T_{P_1} follows a gamma distribution with shape and rate parameters $(n, n\gamma_1)$ (where $n \geq 1$ is an integer), and T_{L_1} and T_{M_1} follow exponential distributions with parameters χ_1 and μ , respectively (which are gamma distributions with shape parameter 1). That is,

$$\begin{aligned} P_1(t) &= G_{n\gamma_1}^n(t) = \sum_{j=1}^n \frac{(n\gamma_1 t)^{j-1} e^{-n\gamma_1 t}}{(j-1)!}, \\ L_1(t) &= G_{\chi_1}^1(t) = e^{-\chi_1 t}, \\ M_1(t) &= G_\mu^1(t) = e^{-\mu t}. \end{aligned} \tag{10.18}$$

Then, it is shown in [18] that (10.16) is equivalent to the following system of ODEs:

$$\begin{aligned}
 \frac{dS}{dt} &= -\frac{1}{N}S\left(\beta_I \sum_{j=1}^n I_j + \beta_H \sum_{j=1}^n H_j + \beta_D D\right), \\
 \frac{dE}{dt} &= \frac{1}{N}S\left(\beta_I \sum_{j=1}^n I_j + \beta_H \sum_{j=1}^n H_j + \beta_D D\right) - \alpha E, \\
 \frac{dI_1}{dt} &= \alpha E - (n\gamma_1 + \chi_1 + \mu)I_1, \\
 \frac{dI_j}{dt} &= n\gamma_1 I_{j-1} - (n\gamma_1 + \chi_1 + \mu)I_j, \quad \text{for } j = 2, \dots, n, \\
 \frac{dH_1}{dt} &= \chi_1 I_1 - (n\gamma_1 + \mu)H_1, \\
 \frac{dH_j}{dt} &= \chi_1 I_j + n\gamma_1 H_{j-1} - (n\gamma_1 + \mu)H_j, \quad \text{for } j = 2, \dots, n, \\
 \frac{dD}{dt} &= \mu \sum_{j=1}^n I_j + \mu \sum_{j=1}^n H_j - \gamma_f D, \\
 \frac{dR}{dt} &= n\gamma_1 I_n + n\gamma_1 H_n.
 \end{aligned} \tag{10.19}$$

In the special case when $n = 1$ (i.e., P_1 is also an exponential distribution), the model (10.19) simplifies to

$$\begin{aligned}
 \frac{dS}{dt} &= -\frac{1}{N}S(\beta_I I + \beta_H H + \beta_D D), \\
 \frac{dE}{dt} &= \frac{1}{N}S(\beta_I I + \beta_H H + \beta_D D) - \alpha E, \\
 \frac{dI}{dt} &= \alpha E - (\gamma_1 + \chi_1 + \mu)I, \\
 \frac{dH}{dt} &= \chi_1 I - (\gamma_1 + \mu)H, \\
 \frac{dD}{dt} &= \mu I + \mu H - \gamma_f D, \\
 \frac{dR}{dt} &= \gamma_1 I + \gamma_1 H.
 \end{aligned} \tag{10.20}$$

Note that in model (10.20) the per-capita transition rates from I to R and from H to R are both equal to γ_1 , from which we have $\gamma_{IR} = \omega_{HR}$. This means that the constraints in (10.15) for model (10.12) cannot be satisfied. Thus, model I cannot be equivalent to the Legrand model. This suggests that Legrand et al. did *not* assume that the three transitions of IR, IH, and ID were described by independent exponential distributions and that the overall waiting time in the I compartment was the minimum of these three exponential waiting times.

When P_1 , L_1 , and M_1 are exponential with the respective parameters γ_1 , χ_1 , and μ , because of the assumption in model I that the three transitions IR, IH, and ID are independent, the overall waiting time in the I compartment is also exponential with the rate constant $\gamma_1 + \chi_1 + \mu$. From the definition of these parameters, we can link them to the general parameters (i.e., independent of model assumptions) in Table 10.2. There might be multiple ways of making the connections. One example is the following: For example,

$$\frac{1}{\gamma_1} = \frac{1}{\gamma_{IR}} = \mathbb{E}[T_{P_1}], \quad f = \frac{\mu}{\gamma_1 + \mu}, \quad p = \frac{\chi_1}{\gamma_1 + \chi_1 + \mu}, \quad (10.21)$$

where γ_{IR} , f , and p are parameters that are independent of models. From the relations in (10.21) we can express the rates γ_1 , χ_1 , and μ in terms of only γ_{IR} , p , and f :

$$\gamma_1 = \gamma_{IR}, \quad \chi_1 = \gamma_{IR} \frac{p}{(1-f)(1-p)}, \quad \mu = \gamma_{IR} \frac{f}{1-f}. \quad (10.22)$$

Reproduction Numbers for Models (10.16), (10.19), and (10.20)

Based on the biological meaning of \mathcal{R}_C , we obtain the following expression for $\mathcal{R}_{C1}^{\text{general}}$ for model (10.16) under general stage distributions:

$$\begin{aligned} \mathcal{R}_{C1}^{\text{general}} &= \beta_I \mathbb{E}(\min\{T_{P_1}, T_{L_1}, T_{M_1}\}) \\ &\quad + \beta_H [\mathbb{E}(\min\{T_{P_1}, T_{M_1}\}) - \mathbb{E}(\min\{T_{P_1}, T_{L_1}, T_{M_1}\})] + \beta_D \frac{1}{\gamma_f} p_M, \end{aligned} \quad (10.23)$$

where

$$\mathbb{E}(\min\{T_{P_1}, T_{L_1}, T_{M_1}\}) = \int_0^\infty P_1(t)L_1(t)M_1(t)dt$$

represents the average time spent in the I compartment, and

$$\mathbb{E}(\min\{T_{P_1}, T_{M_1}\}) = \int_0^\infty P_1(t)M_1(t)dt$$

represents the average total time spent in the I and H compartments.

For the model (10.19) with gamma distribution,

$$\begin{aligned} \mathcal{R}_{C1}^{\text{Gamma}} &= \frac{\beta_I}{\chi_1 + \mu} \left[1 - \left(\frac{n\gamma_1}{n\gamma_1 + \chi_1 + \mu} \right)^n \right] \\ &\quad + \frac{\beta_H}{\mu} \left[\frac{\chi_1}{\chi_1 + \mu} + \frac{\mu}{\chi_1 + \mu} \left(\frac{n\gamma_1}{n\gamma_1 + \chi_1 + \mu} \right)^n - \left(\frac{n\gamma_1}{n\gamma_1 + \mu} \right)^n \right] \\ &\quad + \frac{\beta_D}{\gamma_f} \left[1 - \left(\frac{n\gamma_1}{n\gamma_1 + \mu} \right)^n \right]. \end{aligned} \tag{10.24}$$

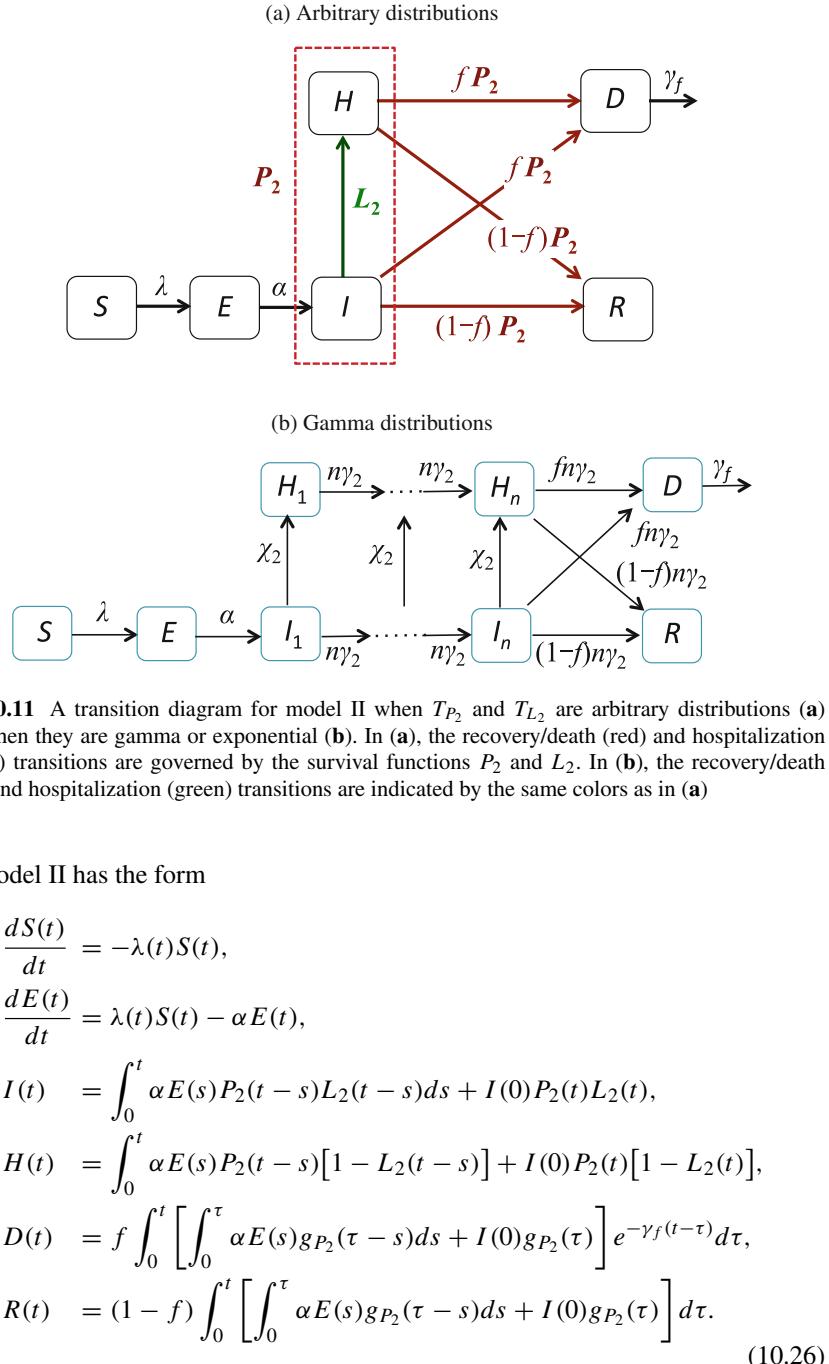
For the model (10.20) with exponential distribution,

$$\begin{aligned} \mathcal{R}_{C1}^{\text{Exp}} &= \frac{\beta_I}{\gamma_1 + \chi_1 + \mu} + \frac{\beta_H}{\gamma_1 + \mu} \frac{\chi_1}{\gamma_1 + \chi_1 + \mu} + \frac{\beta_D}{\gamma_f} \frac{\mu}{\gamma_1 + \mu} \\ &= \frac{\beta_I}{\gamma_1 + \chi_1 + \mu} + \frac{\beta_H p}{\gamma_1 + \mu} + \frac{\beta_D f}{\gamma_f}. \end{aligned} \tag{10.25}$$

Next, consider different assumptions on the transition processes. In model I, the three transitions IR, IH, and ID are assumed to be independent, in which case the three transitions “compete” for individuals in the I class. Another scenario is to consider only two independent transitions, one being hospitalization and the other combining recovery and death for those who are not hospitalized. In this case, we have two survival probability functions:

- $P_2(s)$: Probability of still being infectious *and* alive s time units after onset (governing all four transitions IR, ID, HR, and HD).
- $L_2(s)$: Probability of not being hospitalized s time units after onset (governing the IH transition).

Assume that, for those who exit I without being hospitalized, a fixed fraction $1 - f$ (or f) will recover (or die) an assumption of Legrand model. Assume also that individuals in I and H classes have the same probability of death (f), also as assumed in the Legrand model. The transition diagram is depicted in Fig. 10.11a.



The function $\lambda(t)$ is the same as in model I and given in (10.17). We can reduce the integral equations in (10.26) to the ODEs given below:

$$\begin{aligned}
 \frac{dS}{dt} &= -\frac{1}{N}S\left(\beta_I \sum_{j=1}^n I_j + \beta_H \sum_{j=1}^n H_j + \beta_D D\right), \\
 \frac{dE}{dt} &= \frac{1}{N}S\left(\beta_I \sum_{j=1}^n I_j + \beta_H \sum_{j=1}^n H_j + \beta_D D\right) - \alpha E, \\
 \frac{dI_1}{dt} &= \alpha E - (n\gamma_2 + \chi_2)I_1, \\
 \frac{dI_j}{dt} &= n\gamma_2 I_{j-1}(t) - (n\gamma_2 + \chi_2)I_j, \quad \text{for } j = 2, \dots, n, \\
 \frac{dH_1}{dt} &= \chi_2 I_1 - n\gamma_2 H_1, \\
 \frac{dH_j}{dt} &= \chi_2 I_j + n\gamma_2 H_{j-1} - n\gamma_2 H_j, \quad \text{for } j = 2, \dots, n, \\
 \frac{dD}{dt} &= f n\gamma_2 I_n + f n\gamma_2 H_n - \gamma_f D, \\
 \frac{dR}{dt} &= (1-f)n\gamma_2 I_n + (1-f)n\gamma_2 H_n.
 \end{aligned} \tag{10.27}$$

The reproduction numbers \mathcal{R}_{C2} for model II also have different forms than those for model I. In the case of general distributions,

$$\begin{aligned}
 \mathcal{R}_{C2}^{\text{general}} &= \beta_I \mathbb{E}(\min\{T_{P_2}, T_{L_2}\}) + \beta_H [\mathbb{E}(T_{P_2}) - \mathbb{E}(\min\{T_{P_2}, T_{L_2}\})] \\
 &\quad + \beta_D \frac{1}{\gamma_f} [f(1-p_{H2}) + fp_{H2}].
 \end{aligned} \tag{10.28}$$

When P_2 is a gamma distribution,

$$\begin{aligned}
 \mathcal{R}_{C2}^{\text{Gamma}} &= \beta_I \frac{1}{\chi_2} \left[1 - \left(\frac{n\gamma_2}{n\gamma_2 + \chi_2} \right)^n \right] \\
 &\quad + \beta_H \left[\frac{1}{\gamma_2} - \frac{1}{\chi_2} \left[1 - \left(\frac{n\gamma_2}{n\gamma_2 + \chi_2} \right)^n \right] \right] + \beta_D \frac{f}{\gamma_f}.
 \end{aligned} \tag{10.29}$$

When P_2 is exponential, the model (10.27) becomes

$$\begin{aligned}\frac{dS}{dt} &= -\frac{1}{N}S(\beta_I I + \beta_H H + \beta_D D), \\ \frac{dE}{dt} &= \frac{1}{N}S(\beta_I I + \beta_H H + \beta_D D) - \alpha E, \\ \frac{dI}{dt} &= \alpha E - (\gamma_2 + \chi_2)I, \\ \frac{dH}{dt} &= \chi_2 I - \gamma_2 H, \\ \frac{dD}{dt} &= f\gamma_2 I + f\gamma_2 H(t) - \gamma_f D(t), \\ \frac{dR}{dt} &= (1-f)\gamma_2 I + (1-f)\gamma_2 H.\end{aligned}\tag{10.30}$$

The formula for \mathcal{R}_{C2} in (10.28) simplifies to

$$\mathcal{R}_{C2}^{\text{Exp}} = \frac{\beta_I}{\gamma_2 + \chi_2} + \frac{p\beta_H}{\gamma_2} + \frac{f\beta_D}{\gamma_f}.\tag{10.31}$$

As in model I, there can be multiple choices for linking the parameter γ_2 to the common parameters. For example,

$$\frac{1}{\gamma_2} = (1-f)\frac{1}{\gamma_{IR}} + f\frac{1}{\gamma_{ID}},$$

where $1/\gamma_{ID}$ denotes the average time from onset to death. Also, $p = \chi_2/(\gamma_2 + \chi_2)$. Thus,

$$\gamma_2 = \frac{1}{(1-f)/\gamma_{IR} + f/\gamma_{ID}}, \quad \chi_2 = \frac{\gamma_2 p}{1-p}.\tag{10.32}$$

Another set of possible assumptions that are different from models I and II is to consider two independent distributions for the waiting times in I and H compartments, denoted by T_{P_3} and T_{Q_3} , with survival functions

- $P_3(s)$: Probability of remaining in the I class s units of time since onset (governing the transitions IR, IH, and ID).
- $Q_3(s)$: Probability remaining in the H class s units of time after being hospitalized (governing the HR and HD transitions).

Let $p_3(t) = -P'_3(t)$ and $q_3(t) = -Q'_3(t)$ denote the probability density functions. P_3 describes the waiting time for the combined transitions, IR, IH, and ID. Assume that among the individuals exiting the I class, fractions of p , $(1-p)f$, $(1-p)(1-f)$ will be hospitalized, non-hospitalized and dead, and non-hospitalized and recovered, respectively ($0 \leq p, f < 1$). Q_3 describes the waiting time for the combined two

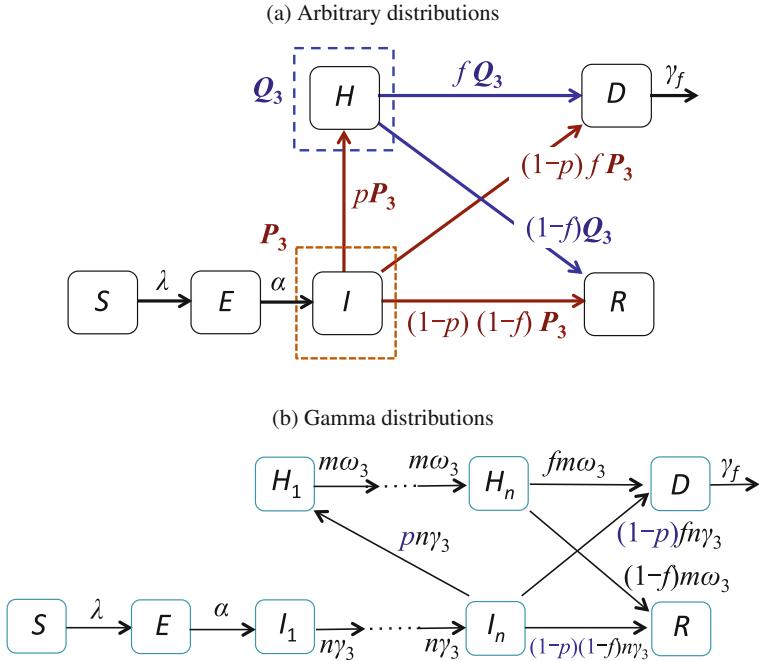


Fig. 10.12 A transition diagram for model III when T_{P_3} and T_{Q_3} are arbitrary distributions (a) and when they are gamma or exponential (b)

transitions, HR and HD, and we assume that fractions $1-f$ and f of the hospitalized individuals recover or die, respectively. A transition diagram is shown in Fig. 10.12.

In this case, model III consists of the following system of integro-differential equations:

$$\begin{aligned}
 S'(t) &= -\lambda(t)S(t), & E'(t) &= \lambda(t)S(t) - \alpha E(t), \\
 I(t) &= \int_0^t \alpha E(s)P_3(t-s)ds + I(0)P_3(t), \\
 H(t) &= \int_0^t p \left[\int_0^s \alpha E(\tau)p_3(s-\tau)d\tau + I(0)p_3(s) \right] Q_3(t-s)ds \\
 D'(t) &= (1-p)f \left[\int_0^t \alpha E(s)p_3(t-s)ds + I(0)p_3(t) \right] \\
 &\quad + f \int_0^t \left[p \int_0^s \alpha E(\tau)p_3(s-\tau)d\tau + pI(0)p_3(s) \right] q_3(t-s)ds - \gamma_f D(t), \\
 R'(t) &= (1-p)(1-f) \left[\int_0^t \alpha E(s)p_3(t-s)ds + I(0)p_3(t) \right] \\
 &\quad + (1-f) \int_0^t \left[p \int_0^s \alpha E(\tau)p_3(s-\tau)d\tau + pI(0)p_3(s) \right] q_3(t-s)ds.
 \end{aligned} \tag{10.33}$$

Note that it is easier in this case to write equations for D' and H' than for D and H .

Assume that T_{P_3} and T_{Q_3} follow gamma distributions with shape parameters n and m respectively, i.e., the survival functions are given by

$$\begin{aligned} P_3(t) &= G_{n\gamma_3}^n(t) = \sum_{j=1}^n \frac{[n\gamma_3(t-s)]^{j-1} e^{-n\gamma_3(t-s)}}{(j-1)!}, \\ Q_3(t) &= G_{m\omega_3}^m(t) = \sum_{j=1}^m \frac{[m\omega_3(t-s)]^{j-1} e^{-m\omega_3(t-s)}}{(j-1)!}. \end{aligned} \quad (10.34)$$

Then, the system (10.33) reduces to the following system of ODEs:

$$\begin{aligned} \frac{dS}{dt} &= -\frac{1}{N} S \left(\beta_I \sum_{j=1}^n I_j + \beta_H \sum_{j=1}^m H_j + \beta_D D \right), \\ \frac{dE}{dt} &= \frac{1}{N} S \left(\beta_I \sum_{j=1}^n I_j + \beta_H \sum_{j=1}^m H_j + \beta_D D \right) - \alpha E, \\ \frac{dI_1}{dt} &= \alpha E - n\gamma_3 I_1, \\ \frac{dI_k}{dt} &= n\gamma_3 I_{k-1} - n\gamma_3 I_k, \quad k = 2, \dots, n \\ \frac{dH_1}{dt} &= p n \gamma_3 I_n - m \omega_3 H_1, \\ \frac{dH_k}{dt} &= m \omega_3 H_{k-1} - m \omega_3 H_k, \quad k = 2, \dots, m \\ \frac{dD}{dt} &= (1-p)f n \gamma_3 I_n + f m \omega_3 H_n - \gamma_f D, \\ \frac{dR}{dt} &= (1-p)(1-f)n\gamma_3 I_n + (1-f)m\omega_3 H_m. \end{aligned} \quad (10.35)$$

A transition diagram under the gamma distributions for T_{P_3} and T_{Q_3} , for the ODE model (10.35) is shown in Fig. 10.12b. We observe a major difference between this figure and Fig. 10.10b or Fig. 10.11b in the recovery rates from I_n and H_m , which have different values here. A similar difference exists in the transition rates from I_j to I_{j+1} ($j = 1, \dots, n-1$) and from H_j to H_{j+1} ($j = 1, \dots, m-1$).

In the special case when $n = m = 1$ (i.e., P_3 and Q_3 are exponential), the ODE model (10.35) simplifies to

$$\begin{aligned}\frac{dS}{dt} &= -\frac{1}{N}S(\beta_I I + \beta_H H + \beta_D D), \\ \frac{dE}{dt} &= \frac{1}{N}S(\beta_I I + \beta_H H + \beta_D D) - \alpha E, \\ \frac{dI}{dt} &= \alpha E - \gamma_3 I, \\ \frac{dH}{dt} &= p\gamma_3 I - \omega_3 H, \\ \frac{dD}{dt} &= (1-p)f\gamma_3 I + f\omega_3 H - \gamma_f D, \\ \frac{dR}{dt} &= (1-p)(1-f)\gamma_3 I + (1-f)\omega_3 H.\end{aligned}\tag{10.36}$$

Notice that, if we ignore the last term in the R equation in the equivalent Legrand model (10.12) (this term indicates that the R class includes those buried), the model (10.36) is identical to the model (10.12) when the subscript “3” is dropped; that is, when γ_3 and ω_3 are defined as follows:

$$\begin{aligned}\frac{1}{\gamma_3} &= \frac{p}{\gamma_{IH}} + \frac{(1-p)f}{\gamma_{ID}} + \frac{(1-p)(1-f)}{\gamma_{IR}}, \\ \frac{1}{\omega_3} &= \frac{f}{\omega_{HD}} + \frac{1-f}{\omega_{HR}},\end{aligned}\tag{10.37}$$

together with the constraints

$$\frac{1}{\gamma_{IR}} = \frac{1}{\gamma_{IH}} + \frac{1}{\omega_{HR}}, \quad \frac{1}{\gamma_{ID}} = \frac{1}{\gamma_{IH}} + \frac{1}{\omega_{HD}}.\tag{10.38}$$

The reproduction number \mathcal{R}_{C3} for model III can be derived using the same approach as for models I and II. Because the derivations are similar, we omit the details, and present the formula only for the special case when P_3 and Q_3 are both exponential, i.e., $P_3 = G_{\gamma_3}^1$ and $Q_3 = G_{\omega_3}^1$ (see (10.34)). In this case, the formula for \mathcal{R}_{C3} for the ODE model (10.35) is independent of m and n , and is given by

$$\mathcal{R}_{C3}^{\text{Exp}} = \frac{\beta_I}{\gamma_3} + p\frac{\beta_H}{\omega_3} + f\frac{\beta_D}{\gamma_f},\tag{10.39}$$

where γ_3 and ω_3 are given in (10.37).

Note that the main difference between models I, II, and III lies in the assumptions on the underlying biological processes, particularly the sojourn distributions for

various stage transitions, which are described by functions $L(t)$, $P(t)$, $M(t)$, and $Q(t)$. The fact that the Legrand model can only be obtained from model III, not from models I and II, identifies the specific assumptions made in the Legrand model in terms of these sojourn distributions. For example, our analyses suggest the following assumptions made in the Legrand model:

- (a) The overall sojourn in the I stage is assumed to be exponentially distributed with the average duration $1/\gamma$, which is further assumed to be the specific weighted average of $1/\gamma_{IR}$, $1/\gamma_{IH}$, and $1/\gamma_{ID}$ as given in (10.13), where $1/\gamma_{IR}$, $1/\gamma_{IH}$, and $1/\gamma_{ID}$ are the respective average stage duration of the IR, IH, and ID transitions. However, from model I we see that if the IR, IH, and ID transitions follow independent exponential distributions with parameters γ_{IR} , γ_{IH} , and γ_{ID} , then the overall sojourn in the I stage is exponentially distributed with the parameter $\gamma = \gamma_{IR} + \gamma_{IH} + \gamma_{ID}$ with the average duration

$$\frac{1}{\gamma} = \frac{1}{\gamma_{IR} + \gamma_{IH} + \gamma_{ID}},$$

which differs from (10.13).

- (b) The distributions for the I and H stages are independent (see $P_3(t)$ and $Q_3(t)$). This implies that the average time spent in the H stage before recovery or death ($1/\omega_3$) does not depend on the average time spent in the I stage before recovery or being hospitalized or dying ($1/\gamma_3$). Under this assumption, the time spent in H before recovery ($1/\gamma_{HR}$) does not take into consideration the time spent in I before being hospitalized ($1/\gamma_{IH}$). Because of this independence, the model needs to impose a constraint to link these two durations (see (10.10)).

Difference in Evaluations by Models I, II, and III

Among the three ODE models (10.20), (10.30), and (10.36), which are reduced from the models I, II, and III with general distributions, the only model that can match the Legrand model is (10.36), for which the assumptions include: (i) the waiting times of the three transitions IR, IH, and ID are *not* independent and (ii) the overall waiting time in the I compartment is a weighted average of the mean durations ($1/\gamma_{IR}$, $1/\gamma_{IH}$, and $1/\gamma_{ID}$) for the three transitions with weights determined by the probabilities of hospitalization p and death f , as described in (10.14). By examining the ODE model (10.20), we found that, if the waiting times of the three transitions IR, IH, and ID are independent and exponentially distributed (with parameters γ_1 , χ_1 , and μ), then the overall waiting time in I should be an exponential distribution with the parameter $\gamma_1 + \chi_1 + \mu$. That is, the average overall waiting time should be $1/(\gamma_1 + \chi_1 + \mu)$, *not* a weighted average such as the ones in (10.14).

Formulas for the control reproduction numbers \mathcal{R}_{Ci} ($i = 1, 2, 3$) for the three general models provide a means of examining the influence of assumptions on model outcomes. For example, consider the three control reproduction numbers \mathcal{R}_{Ci} ($i = 1, 2, 3$), which are given in (10.25), (10.31), and (10.39) corresponding to

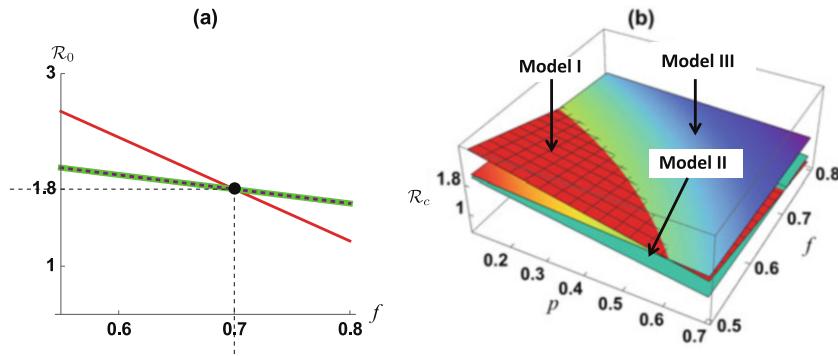


Fig. 10.13 Plots of the basic reproduction numbers (a) and control reproduction numbers (b) for the three models. In (a), \mathcal{R}_0 is plotted as a function of f for model I (thin solid), model II (dashed), and model III (thick solid). The parameter values are chosen such that all three \mathcal{R}_{0i} have the same value 1.8 at $f = 0.7$ (note that $p = 0$). In (b), \mathcal{R}_{Ci} is plotted as a function of p and f for models I–III. Other parameter values are given in the text

the three ODE models (10.20), (10.30), and (10.36), respectively. In the absence of hospitalization (i.e., $p = 0$), these \mathcal{R}_{Ci} reduce to the corresponding basic reproduction numbers \mathcal{R}_{0i} ($i = 1, 2, 3$). Figure 10.13 illustrates the difference between the basic and control reproduction numbers of the three models for a given set of parameter values, mostly based on the Ebola outbreak in West Africa in 2014. We fix all parameters except β_i ($i = I, H, D$) and f . Then, for a fixed value of $f_0 = 0.7$, we estimate β_i ($i = I, H, D$) from a given value of \mathcal{R}_0 (assumed to be the same for all three models). If we further assume that $\beta_H = \beta_D = 0.3\beta_I$, then we can get a unique value for β_I for fixed \mathcal{R}_0 . Once we have the value of β_i , we have three functions of f , $\mathcal{R}_{0i}(f)$ ($i = 1, 2, 3$). For Fig. 10.13a, we used the common value of $\mathcal{R}_0(f_0) = 1.8$. The three curves are for model I (the thin solid curve), model II (the dashed curve), and model III (the thick curve). For the selected set of parameter values, the \mathcal{R}_0 curves for models II and III overlap. The decreasing property of these curves represents the fact that higher disease mortality decreases $\mathcal{R}_{0i}(f)$, which is expected because the assumption that $\beta_D < \beta_I$. An interesting observation is that the dependence of the basic reproduction number on disease death f is more dramatic in model I than models II and III, particularly for smaller f values. For smaller values of f , model I tends to generate the highest \mathcal{R}_0 , while for larger f values, model III provides higher \mathcal{R}_0 . Other parameter values used are (time in days): $1/\gamma_{IR} = 18$, $1/\gamma_f = 2$, $1/\alpha = 9$. Parameters such as μ , χ_i ($i = 1, 2$) are calculated based on their relationships with the common parameters. For model III, additional parameter values include $1/\gamma_{IH} = 7$, $1/\gamma_{ID} = 8$, which can be used to determine γ_3 and ω_3 from (10.37).

When control is considered ($p > 0$), the dependence of \mathcal{R}_{Ci} on p and f is illustrated in Fig. 10.13b. We observe that, for the given set of parameter values, model I (the darker surface with mesh) generates higher \mathcal{R}_C values than models

II (the lighter surface) and III (the darker surface with no mesh) for smaller p and f , while model III provides the higher values for larger values of p and/or f . The differences in \mathcal{R}_0 and \mathcal{R}_C between the three models indicate that model predictions about and evaluations of the effectiveness of control measures could be very different as well. Figure 10.14 shows numerical simulation results of the three ODE models (10.20), (10.30), and (10.36), which are reduced from the models I, II, and III, and presented in columns 1, 2, and 3, respectively. The A, B, and C panels correspond to three sets of (p, f) values: $(p, f) = (0, 0.7)$ (top panel), $(p, f) = (0.3, 0.5)$ (middle panel), and $(p, f) = (0.4, 0.7)$ (bottom panel). The top panel (A1–A3) is for the case of no hospitalization ($p = 0$). We observe that models II and III generate similar epidemic curves (fractions of infected individuals $(E + I + H)/N$), including peak sizes, times to peak, duration of epidemic (which lead to similar epidemic final sizes). Model I shows a higher peak size and an earlier time to peak. The middle panel (B1–B3) is for the case when the hospitalization is $p = 0.3$, and we observe that model I has the highest peak size while model II has the lowest. This is in agreement with the relative magnitudes of the control

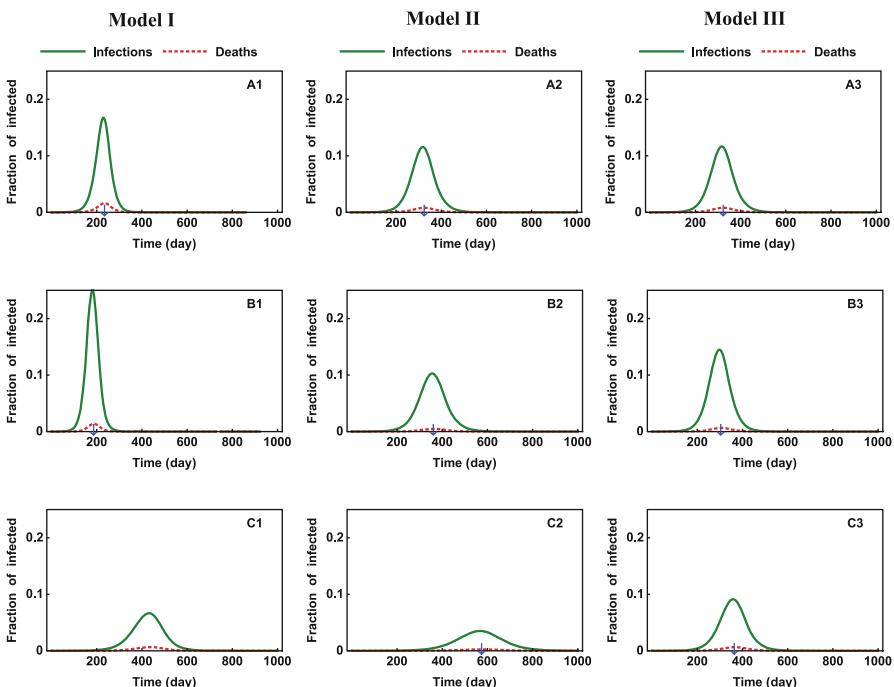


Fig. 10.14 Numerical simulations of the three ODE models (10.20), (10.30), and (10.36), which are reduced from the models I, II, and III, respectively. The fractions of infected individuals $(E + I + H)/N$ and death (D/N) are plotted over a time period of 1000 days. Three sets of (p, f) values are used: $(p, f) = (0, 0.7)$ (top row), $(p, f) = (0.3, 0.5)$ (middle row), and $(p, f) = (0.4, 0.7)$ (bottom row). Parameter values are the same as in Fig. 10.13

reproduction numbers \mathcal{R}_{Ci} , as the point $(p, f) = (0.3, 0.5)$ lies in the region where $\mathcal{R}_{C1} > \mathcal{R}_{C3} > \mathcal{R}_{C2}$ (see Fig. 10.13). For the bottom panel (C1–C3), because the point $(p, f) = (0.4, 0.7)$ lies in the region where $\mathcal{R}_{C3} > \mathcal{R}_{C1} > \mathcal{R}_{C2}$, we observe that model III generates the highest peak size while model II again has the lowest.

10.5 Slower than Exponential Growth

It has been standard practice in analyzing disease outbreaks to formulate a dynamical system as a deterministic compartmental model, then to use observed early outbreak data to fit parameters to the model, and finally to analyze the dynamical system to predict the course of the disease outbreak and to compare the effects of different management strategies. In general, such models predict an initial stochastic stage (while the number of infectious individuals is small), followed by a period of exponential growth. Measurement of this early exponential growth rate is an essential step in estimating contact rate parameters for the model. A thorough description of the analysis of compartmental models may be found in [21].

However, instances have been noted where the growth rate of an epidemic is clearly slower or faster than exponential. For example, [13], the 2013–2015 epidemic in West Africa has been viewed as a composition of locally asynchronous outbreaks at local levels displaying sub-exponential growth patterns during several generations. Specifically, if $I(t)$ is the number of infectious individuals at time t , a graph of $\log I(t)$ against t is a straight line if the growth rate is exponential, and for some disease outbreaks this has not been true. One of the earliest examples [16] concerns the growth of HIV/AIDS in the USA, and a possible explanation might be the mixture of short-term and long-term contacts. This could be a factor in other diseases where there are repeated contacts in family groups and less frequent contacts outside the home.

During the 2013–2015 Ebola epidemic in West Africa, in the country of Liberia, which has a total population size of 4,300,000, there were 10,678 suspected, probable, and confirmed cases of disease as of September 3, 2015 [32]. An SIR model for the whole country would have predicted more than 1,500,000 cases. This discrepancy cannot be explained by the assumption of control measures. In order to make plausible predictions of the effects of large-scale epidemics, it is necessary to use a phenomenological model based on observations in the early stages of the epidemic rather than to try to fit a mechanistic model. Some phenomenological models that have been used to good effect have been the generalized Richards model, the “generalized growth model,” and the IDEA model.

10.5.1 The Generalized Richards Model

Perhaps the first attempt to fit early epidemic data is the Richards model [34]. This is a modification of the logistic population growth model, described in [22]

$$I'(t) = rI \left[1 - \left(\frac{I}{K} \right)^a \right].$$

In this model, I represents the cumulative number of infected individuals at time t , K is the carrying capacity or total case number of the outbreak, r is the per-capita growth rate of the infected population, and a is an exponent of deviation from the standard logistic model. The basic premise of the model is that the incidence curve has a single turning point t_m . The analytic solution of the model is

$$I(t) = \frac{K}{[1 + e^{-r(t-t_m)}]^{1/a}}.$$

10.5.2 The Generalized Growth Model

It has been pointed out [12–15, 39] that a so-called general growth model of the form

$$C'(t) = rC(t)^p,$$

where $C(t)$ is the number of disease cases occurring up to time t , and p , $0 \leq p \leq 1$, is a “deceleration of growth” parameter, has exponential solutions if $p = 1$ but solutions with polynomial growth if $0 < p < 1$. This is not and does not claim to be a mechanistic epidemic model, but it has proved to be remarkably successful for fitting epidemic growth and predicting the course of an epidemic. For example, it has provided much better estimates of epidemic final size than the exponential growth assumption for the Ebola epidemic of 2014. However, it assumes a sustained increase in the number of disease cases and cannot capture the later decline in the number of new infections. Such phenomenological models are particularly likely to be suitable in situations where it is difficult to construct a mechanistic approach because of multiple transmission routes, interactions of spatial influences, or other aspects of uncertainty.

10.5.3 The IDEA Model

Another direction that would be well worth further exploration would be contact rates decreasing in time because of individual behavioral changes in response to a disease outbreak. A contact rate which is a decreasing function of time can certainly lead to early epidemic growth slower than exponential. A step in this direction has been initiated in a discrete model [19] that has been applied to an Ebola model in [20].

The IDEA model [19, 20, 37] is a discrete model that assumes damping of recruitment of new infections because of spontaneous or planned behavioral changes. It is assumed that the time interval in the model is equal to the duration of infection, so that the number of infective individuals at each stage is equal to the number of new infections. The resulting model is

$$I_t = \left[\frac{\mathcal{R}_0}{(1+d)^t} \right]^t.$$

Here, d is a discount factor describing the recruitment damping.

A variety of epidemiological situations in which slower than exponential epidemic growth might be possible have been described. Ultimately, the challenge for epidemiological modeling would be to determine which of these situations allow slower than exponential growth by deriving and analyzing mechanistic models to describe each of these situations. This is an important new direction for epidemic modeling. Some suggestions include metapopulation models with spatial structure including cross-coupling and mobility, clustering in spatial structure, dynamic contacts, agent-based models with differences in infectivity and susceptibility of individuals, and reactive behavioral changes early in a disease outbreak. It may well turn out that slower than exponential growth may be ruled out in some cases but is possible in others. For example, heterogeneity of mixing in a single location can be modeled by an autonomous dynamical system and the linearization theory of dynamical systems at an equilibrium shows that early epidemic growth for such a system is always exponential. On the other hand, metapopulation models may well allow many varieties of behaviors.

An important broader question is the matter of what information influences the behavior of people during a disease outbreak, and how to include this in a model. A recent book [28] describes some studies in this direction.

10.5.4 Models with Decreasing Contact Rates

There is anecdotal evidence that in an outbreak of a disease considered very serious there is early action to decrease the risk of being infected by decreasing contact rates. This early action may begin even before government efforts to combat the disease.

Since Ebola is a very serious disease, with case fatality rates of 70% or more, it is reasonable to assume a decreasing contact rate. This has been suggested in [1, 2]. In [1], an *SIR* model is assumed with a contact rate that decreases exponentially in time, and observations are used to estimate the rate of decrease. In [6], such a model is used to estimate the final size of an Ebola epidemic over a country using early growth rate data and this yields early final size estimates that are much closer to the eventual outbreak data than estimates assuming a constant contact rate. For example, for the 2014–5 Ebola outbreak in Guinea, a country with a total population size of 10,589,000 the assumption of a constant contact rate corresponding to $\mathcal{R}_0 = 1.5$ led to an estimate of more than 9,000,000 cases over the whole country, while a decreasing contact rate assumption led to an estimate of about 27,000 disease cases. The actual number of Ebola cases in Guinea in this outbreak was fewer than 4,000. The models described here are for an entire country and are not intended to replace more detailed models needed for disease management describing the progress of the disease in individual villages.

10.6 Project: Slower than Exponential Growth

We have suggested in Sect. 10.5.4 that slower than exponential growth of the number of infectious individuals may be explained by assuming a time-dependent decrease in the contact rate. The question of what modeling assumptions can lead to slower than exponential growth is a complex one [13]. Another possible explanation might be a decrease in contact rate depending on the current state of the system. If the model system remains autonomous, growth will remain exponential so long as the contact rate is a differentiable function of I so that the theory of linearization at a disease-free equilibrium remains valid. However, if the rate of new infections is $a \frac{S}{N} f(I)$ with $f(I)$ not differentiable at $I = 0$, different behavior is possible.

Consider an *SIR* model in which the equation for I is

$$I' = \beta \frac{S}{N} I^\alpha - \gamma I,$$

with $0 < \alpha < 1$. Initially, so long as $S \approx N$, we approximate this equation by

$$I' = \beta I^\alpha - \gamma I. \quad (10.40)$$

Question 7 Show that the basic reproduction number is

$$\mathcal{R}_0 = \frac{\beta}{\gamma}.$$

Question 8 Make the change of dependent variable $u = \log I$ and derive the differential equation satisfied by u .

Exponential growth of I corresponds to linear growth of u . It is clear that if $\alpha = 1$, u does grow linearly.

Question 9 Determine whether the rate of growth of u is less than linear for $\alpha < 1$, either by explicit solution of the equation for u or by simulations with $\beta > \gamma$ (so that $\mathcal{R}_0 > 1$) and a range of values of α .

10.7 Project: Movement Restrictions as a Control Strategy

Cordons Sanitaire or “sanitary barriers” are designed to prevent the movement, in and out, of people and goods from particular areas. The effectiveness of the use of *cordons sanitaire* has been controversial. This policy was last implemented nearly 100 years ago [9]. In desperate attempts to control disease, Ebola-stricken countries enforced public health officials decided to use this medieval control strategy in the EVD hot-zone, that is, the region of confluence of Guinea, Liberia, and Sierra Leone [30]. In this project, a framework that allows, in the simplest possible setting, the possibility of assessing the potential impact of the use of a *Cordon Sanitaire* during a disease outbreak is introduced. We consider an *SIR* epidemic model in two patches, one of which has a significantly larger contact rate, with short-term travel between the two patches. The total population resident in each patch is constant. We follow a Lagrangian perspective, that is, we keep track of each individual’s place of residence at all times [4, 17]. This is in contrast to an Eulerian perspective, which describes migration between patches.

Thus we consider two patches, with total resident population sizes N_1 and N_2 , respectively, each population being divided into susceptibles, infectives, and removed members. S_i and I_i denote the number of susceptibles and infectives, respectively, who are residents in Patch i , regardless of the patch in which they are present.

Residents of Patch i spend a fraction p_{ij} of their time in Patch j , with

$$p_{11} + p_{12} = 1, \quad p_{21} + p_{22} = 1.$$

The contact rate in Patch i is β_i , and we assume $\beta_1 > \beta_2$.

Each of the $p_{11}S_1$ susceptibles from Group 1 present in Patch 1 can be infected by infectives from Group 1 and from Group 2 present in Patch 1. Similarly, each of the $p_{12}S_1$ susceptibles present in Patch 2 can be infected by infectives from Group 1 and from Group 2 present in Patch 2.

Question 1 Show that the model equations are

$$\begin{aligned} S'_i &= -p_{i1}S_i \left[p_{11}\frac{I_1}{N_1} + p_{12}\frac{I_2}{N_2} \right] - p_{i2}S_i \left[p_{21}\frac{I_1}{N_1} + p_{22}\frac{I_2}{N_2} \right] \\ I'_i &= p_{i1}S_i \left[p_{11}\frac{I_1}{N_1} + p_{21}\frac{I_2}{N_2} \right] + p_{i2}S_i \left[p_{21}\frac{I_1}{N_1} + p_{22}\frac{I_2}{N_2} \right] - \gamma I_i, \quad i = 1, 2. \end{aligned}$$

Question 2 Use the next generation matrix [38] to calculate the basic reproduction number.

Question 3 Use the approach described in Chaps. 4 and 5 to determine the final size relations.

Imposition of a *Cordon Sanitaire* amounts to replacing the fractions of normal travel between groups by

$$p_{11} = p_{22} = 1, \quad p_{12} = p_{21} = 0.$$

Question 4 Compare the reproduction numbers and final sizes with and without a *Cordon Sanitaire*, using numerical simulations with various parameter value choices.

The approach suggested here can be used with more detailed models for a specific disease, such as Ebola, in which transmission from one village to another through temporary visits is a factor [17].

10.8 Project: Effect of Early Detection

The following model for Ebola is considered in [10]:

$$\begin{aligned} \frac{dS}{dt} &= -\beta S \frac{I + lJ}{N}, \\ \frac{dE_1}{dt} &= \beta S \frac{I + lJ}{N} - k_1 E_1, \\ \frac{dE_2}{dt} &= k_1 E_1 - k_2 E_2 - f_T E_2, \\ \frac{dI}{dt} &= k_2 E_2 - (\alpha + \gamma)I, \\ \frac{dJ}{dt} &= \alpha I + f_T E_2 - \gamma_r J, \\ \frac{dR}{dt} &= \gamma(1 - \delta)I + \gamma_r(1 - \delta)J, \\ \frac{dD}{dt} &= \gamma \delta I + \gamma_r \delta J, \end{aligned} \tag{10.41}$$

where $N = S + E_1 + E_2 + I + J + R$. The variables E_1 and E_2 denote the numbers of latent undetectable and detectable individuals, respectively, I is the number of infectious individuals, J is the number of isolated infective individuals, R and D are numbers of recovered and dead due to the disease. For the parameters, β is the transmission rate, l is the relative transmissibility of isolated individuals, k is the rate of entering the detectable class, k_2 is the rate of becoming infective, f_T is the rate of being diagnosed, α is the rate of isolation, γ is the recovery rate, and γ_r is the rate at which individuals are removed from isolation after recovery or disease death.

References

1. Althaus, C.L. (2014) Estimating the reproduction number of Ebola Virus (EBOV) during the 2014 outbreak in West Africa, PLoS Currents: Edition 1. <https://doi.org/10.1371/current.outbreaks.91afb5e0f279e7f29e7056095255b288>.
2. Barbarossa, M.V., A Denes, G. Kiss, Y. Nakata, G. Rost, & Z. Vizi (2015) Transmission dynamics and final epidemic size of Ebola virus disease dynamics with varying interventions, PLoS One 10(7): e0131398. <https://doi.org/10.1371/journal.pone.0131398>.
3. Bellan, S.E., J.R.C. Pulliam, J. Dushoff, and L.A. Meyers (2014) Ebola control: effect of asymptomatic infection and acquired immunity, The Lancet **384**:1499–1500.
4. Bichara, D., Y. Kang, C. Castillo-Chavez, R. Horan and C. Perrings (2015) *SIS* and *SIR* epidemic models under virtual dispersal, Bull. Math. Biol. **77**: 2004–2034.
5. Blower, S.M. and H. Dowlatabadi (1994) Sensitivity and uncertainty analysis of complex models of disease transmission: an HIV model, as an example. Int Stat Rev. **2**: 229–243.
6. Brauer, F. (2019) The final size of a serious epidemic, Bull. Math. Biol. <https://doi.org/10.1017/s11538-018-00549-x>
7. Browne, C.J., H. Gulbudak, and G. Webb (2015) Modeling contact tracing in outbreaks with application to Ebola, J. Theor. Biol. **384**: 33–49.
8. Butler, D. (2014) Models overestimate Ebola cases, Nature **515**: 18. <https://doi.org/10.1038/515018a>.
9. Byrne, J.P. (2008) Encyclopedia of Pestilence, Pandemics, and Plagues, **1** ABC-CLIO, 2008.
10. Chowell, D., C. Castillo-Chavez, S. Krishna, X. Qiu, and K.S. Anderson (2015) Modelling the effect of early detection of Ebola, Lancet Infectious Diseases, February, DOI: [http://dx.doi.org/10.1016/S1473-3099\(14\)71084-9](http://dx.doi.org/10.1016/S1473-3099(14)71084-9).
11. Chowell, G., N.W. Hengartner, C. Castillo-Chavez, P.W. Fenimore, and J.M. Hyman (2004) The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda, J. Theor. Biol. **229**: 119–126.
12. Chowell, G. and J. M. Hyman (2016) Mathematical and Statistical Modeling for Emerging and Re-emerging Infectious Diseases. Springer, 2016.
13. Chowell G., C. Viboud, J.M. Hyman, L. Simonsen (2015) The Western Africa Ebola virus disease epidemic exhibits both global exponential and local polynomial growth rates. PLOS Current Outbreaks, January 21.
14. Chowell, G. and C. Viboud (2016) Is it growing exponentially fast?—impact of assuming exponential growth for characterizing and forecasting epidemics with initial near-exponential growth dynamics Infectious Dis. Modelling **1**: 71–78.
15. Chowell, G., C. Viboud, L. Simonsen, & S. Moghadas (2016) Characterizing the reproduction number of epidemics with early sub-exponential growth dynamics, J. Roy. Soc. Interface: <https://doi.org/10.1098/rsif.2016.0659>.

16. Colgate, S.A., E. A. Stanley, J. M. Hyman, S. P. Layne, and C. Qualls (1989) Risk behavior-based model of the cubic growth of acquired immunodeficiency syndrome in the United States. *Proc. Nat. Acad. Sci.*, *textbf{86}*: 4793–4797
17. Espinoza, B., V. Moreno, D. Bichara, and C. Castillo-Chavez (2016) Assessing the efficiency of movement restriction as a control strategy of Ebola, In Mathematical and Statistical Modeling for Emerging and Re-emerging Infectious Diseases, Springer International Publishing: pp. 123–145.
18. Feng, Z., Z. Zheng, N. Hernandez-Ceron, J.W. Glasser, and A.N. Hill (2016) Mathematical models of Ebola - Consequences of underlying assumptions, *Math. Biosc.* **277**: 89–107.
19. Fisman, D.N., T.S. Hauck, A.R. Tuite, & A.L. Greer (2013) An IDEA for short term outbreak projection: nearcasting using the basic reproduction number, *PLOS One* **8**: 1–8.
20. Fisman, D.N., E. Khoo, & A.R. Tuite (2014) Early epidemic dynamics of the west Africa 2014 Ebola outbreak: Estimates derived with a simple two-parameter model, *PLOS currents* **6**: September 8.
21. Hethcote, H.W. (2000) The mathematics of infectious diseases, *SIAM Review* **42**: 599–653.
22. Hsieh, Y.-H. (2009) Richards model: a simple procedure for real-time prediction of outbreak severity In *Modeling and Dynamics of Infectious Diseases*, pages 216–236. World Scientific.
23. Khan, A., M. Naveed, M. Dur-e-Ahmad, and M. Imran (2015) Estimating the basic reproductive ratio for the Ebola outbreak in Liberia and Sierra Leone, *Infect. Dis. Poverty* **4**.
24. Khan, A.S, F.K. Tshioko, D.L. Heymann, et al (1999) The reemergence of Ebola hemorrhagic fever, Democratic Republic of the Congo, 1995, *J. Inf. Dis.* **179**: S76–86.
25. King, K., et al (2015) Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to Ebola, *Proc. R. Soc. B* **282**: 20150347. <http://dx.doi.org/10.1098/rspb.2015.0347>
26. Legrand, J., R.F. Grais, P.Y. Boelle, A.J. Valleron, and A. Flahault (2007) Understanding the dynamics of Ebola epidemics, *Epidemiol. Infect* **135**: 610–621.
27. Lewnard, J.A., M.L.N. Mbah, J.A. Alfaro-Murillo, et al. (2014) Dynamics and control of Ebola virus transmission in Montserrado, Liberia: a mathematical modelling analysis, *Lancet Infect. Dis.* **14**: 1189–1195.
28. Manfredi, P. & A. d'Onofrio (eds.) (2013) Modeling the Interplay between Human Behavior and the Spread of Infectious Diseases, Springer - Verlag, New York-Heidelberg-Dordrecht-London.
29. Meltzer M.I., C.Y. Atkins, S. Santibanez, et al. (2014) Estimating the future number of cases in the Ebola epidemic: Liberia and Sierra Leone, 2014–2015. *MMWR Surveill. Summ.* **63**: 2014–2015.
30. McNeil Jr., D.G. Jr. 9@014) NYT: Using a Tactic Unseen in a Century, Countries Cordon Off Ebola-Racked Areas, August 12, 2014.
31. Nishiura, H and G. Chowell (2014) Early transmission dynamics of Ebola virus disease (EVD), west Africa, March to August 2014, *Euro Surveill.* **19**: 1–6.
32. Nyenswah, T.G., F. Kateh, L. Bawo, M. Massaquoi, M. Gbanyan, M. Fallah, T. K. Nagbe, K. K. Karsor, C. S. Wesseh, S. Sieh, et al. (2016) Ebola and its control in Liberia, 2014–2015, *Emerging Infectious Diseases* **22**: 169.
33. Renshaw, E. (1991) Modelling Biological Populations in Space and Time. Cambridge University Press, Cambridge, 1991.
34. Richards, F. (1959) A flexible growth function for empirical use, *J. Experimental Botany* **10**: 290–301.
35. Rivers, C.M. et al. (2014) Modeling the impact of interventions on an epidemic of Ebola in Sierra Leone and Liberia, *PLoS Current Outbreaks*, November 6, 2014.
36. Towers, S., O. Patterson-Lomba, and C. Castillo-Chavez (2014) Temporal variations in the effective reproduction number of the 2014 West Africa Ebola Outbreak, *PLoS Currents: Outbreaks* **1**.
37. Tuite, A.R. & D.N. Fisman (2016) The IDEA model: A single equation approach to the Ebola forecasting challenge, *Epidemics*, <http://dx.doi.org/10.1016/j.epidem.2016.09.001>

38. van den Driessche, P. & J. Watmough (2002) Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math. Biosc.*, **180**: 29–48.
39. Viboud, C., L. Simonsen, and G. Chowell (2016) A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks, *Epidemics* **15**: 27–37.
40. Webb G., C. Browne, X. Huo, O. Seydi, M. Seydi, and P. Magal (2014) A model of the 2014 Ebola epidemic in west Africa with contact tracing. *PLoS Currents*, <https://doi.org/10.1371/currents.outbreaks.846b2a31ef37018b7d1126a9c8adf22a>.
41. World Health Organization (WHO) (2001) Outbreak of Ebola hemorrhagic fever, Uganda, August 2000 - January 2001. *Weekly epidemiological record* **2001**;76:41–48.

Chapter 11

Models for Malaria



Malaria is one of the most important diseases transmitted by vectors. The vectors for many vector-transmitted diseases are mosquitoes or other insects which tend to be more common in warmer climates. One influence of climate change in coming years may be to extend the regions where mosquitoes can thrive and thus to cause the spread of vector-transmitted diseases geographically.

11.1 A Malaria Model

As we have remarked earlier, many of the important underlying ideas of mathematical epidemiology arose in the study of malaria begun by Sir R.A. Ross [11]. Malaria is one example of a disease with vector transmission, the infection being transmitted back and forth between vectors (mosquitoes) and hosts (humans). It kills nearly 1,000,000 people annually, mostly children and mostly in poor countries in Africa.

We begin with a basic vector transmission model, namely the model (6.2) of Sect. 6.2 reduced by assuming both hosts and vectors satisfy *SIR* structure. For simplicity, we assume that there are no disease deaths of either hosts or vectors so that the two populations have constant total sizes N_h , N_v , respectively. The resulting model is

$$\begin{aligned} S'_h &= \Lambda_h - \beta_h S_h \frac{I_v}{N_v} - \mu_h S_h \\ I'_h &= \beta_h S_h \frac{I_v}{N_v} - (\mu_h + \gamma_h) I_h \\ S'_v &= \Lambda_v - \beta_v S_v \frac{I_h}{N_h} - \mu_v S_v \\ I'_v &= \beta_v S_v \frac{I_h}{N_h} - (\mu_v + \gamma_v) I_v. \end{aligned} \tag{11.1}$$

There is always a disease-free equilibrium $(N_h, 0, N_v, 0)$, and there may also be an endemic equilibrium with the infective population sizes of both species positive. At an endemic equilibrium,

$$\begin{aligned}\beta_h S_h I_v &= (\gamma_h + \mu_h) I_h N_v \\ \beta_v S_v I_h &= (\gamma_v + \mu_v) I_h \\ \Lambda_h &= S_h \left(\mu_h + \beta_h \frac{I_v}{N_v} \right) N_h \\ \Lambda_v &= S_v \left(\mu_v + \beta_v \frac{I_h}{N_h} \right).\end{aligned}$$

In applying the next generation matrix approach to determine the basic reproduction number, we obtain

$$F = \begin{bmatrix} 0 & \beta_h \frac{N_h}{N_v} \\ \beta_v \frac{N_v}{N_h} & 0 \end{bmatrix}, \quad V = \begin{bmatrix} \gamma_h + \mu_h & 0 \\ 0 & \gamma_v + \mu_v \end{bmatrix}.$$

This leads to

$$FV^{-1} = \begin{bmatrix} 0 & \frac{\beta_h \frac{N_h}{N_v}}{\gamma_v + \mu_v} \\ \frac{\beta_v \frac{N_v}{N_h}}{\gamma_h + \mu_h} & 0 \end{bmatrix}.$$

This would imply that

$$\mathcal{R}_0 = \sqrt{\frac{\beta_h \beta_v}{(\gamma_v + \mu_v)(\gamma_h + \mu_h)}}.$$

However, this approach views the transition from host to vector to host as two generations, and it may be more reasonable to say that

$$\mathcal{R}_0 = \frac{\beta_h \beta_v}{(\gamma_v + \mu_v)(\gamma_h + \mu_h)}. \quad (11.2)$$

With either choice, the transition is at $\mathcal{R}_0 = 1$. Since a single infectious host infects

$$\frac{\beta_v \frac{N_v}{N_h}}{\mu_h + \gamma_h}$$

vectors in a completely susceptible vector population, and each of these infects

$$\frac{\beta_h \frac{N_h}{N_v}}{\mu_v + \gamma_v}$$

hosts in a completely susceptible host population. Thus, the number of secondary host infections caused by an infective host (and also the number of secondary vector infections caused by an infective vector) is given by (11.2), and this is a valid choice for \mathcal{R}_0 . The other choice is also a valid choice and it is essential to check which form is used in a given work.

By linearizing the system (11.1) at an equilibrium, it is possible to show that the disease-free equilibrium is asymptotically stable if and only if $\mathcal{R}_0 < 1$ and that the endemic equilibrium exists only if $\mathcal{R}_0 > 1$ and is asymptotically stable. Because (11.1) is a four-dimensional system, this requires showing that the roots of a fourth degree polynomial have negative real parts, and this is technically complicated.

For malaria, with humans as hosts and mosquitoes as vectors, we modify the vector transmission model (6.2) in two ways. Since infected mosquitoes remain infected for life and do not recover, we take $\gamma_v = 0$. Also,

$$\beta_h = bT_{hv}, \quad \beta_v = bT_{vh},$$

where b is the biting rate, the number of bites made by a mosquito in unit time, T_{hv} is the probability that a bite by an infected mosquito will infect a susceptible human, and T_{vh} is the probability that a bite by a susceptible mosquito of an infected human will infect the mosquito. This gives the model

$$\begin{aligned} S'_h &= \Lambda_h - bT_{hv}S_h \frac{I_v}{N_v} - \mu_h S_h \\ I'_h &= bT_{hv}S_h \frac{I_v}{N_v} - (\mu_h + \gamma_h)I_h \\ S'_v &= \Lambda_v - bT_{vh}S_v \frac{I_h}{N_h} - \mu_v S_v \\ I'_v &= bT_{vh}S_v \frac{I_h}{N_h} - \mu_v I_v. \end{aligned} \tag{11.3}$$

This is basically the original model of Ross [11], with modifications added by MacDonald [8–10]. It is known as the Ross–MacDonald model, and it has remained the basic malaria model ever since its development.

If we write the basic reproduction number in terms of the mosquito biting rate, we obtain

$$\mathcal{R}_0 = \frac{b^2 T_{hv} T_{vh}}{\mu_v (\gamma_h + \mu_h)}. \tag{11.4}$$

Thus, the basic reproduction number depends on the ratio of vector population size to host population size, increases with the vector population size but decreases with the host population size. For the infection to spread, there must be enough vectors that frequently a single host may be bitten at least twice—once by an infected vector to infect the host and a second time to transmit infection to another vector.

Macdonald [10] speaks of a stability index, aT_{vh}/μ_v , which is very difficult to estimate in any particular situation, but gives some indication of the qualitative behavior of a malaria model. Large values of the stability index indicate “stable malaria,” which suggests that malaria would be endemic, while small values of the stability index in a region suggest that it is more likely that there will be epidemic outbreaks. Estimates of the stability index in various regions have ranged from 0.5 to 5.0.

Dr. Ross was awarded the second Nobel Prize in Medicine for his demonstration of the dynamics of the transmission of malaria between mosquitoes and humans. His work received immediate acceptance in the medical community, but his deduction that malaria could be controlled by controlling mosquitoes was dismissed on the grounds that it would be impossible to rid a region of mosquitoes completely and that in any case, mosquitoes would soon re-invade the region. Ross then formulated a mathematical model [11] predicting that malaria outbreaks could be avoided if the mosquito population could be reduced below a critical threshold level, and this is in fact, the first instance of the idea of the basic reproduction number in modeling disease transmission.

Field trials supported Ross’s conclusions and led to sometimes brilliant successes in malaria control. A notable example is the draining of swamps in the Galilee region in Israel to reduce mosquito habitat. Unfortunately, the Garki project provides a dramatic counterexample. This project worked to eradicate malaria from a region temporarily. However, people who have recovered from an attack of malaria have a temporary immunity against reinfection. Thus, elimination of malaria from a region leaves the inhabitants of this region without immunity when the campaign ends, and the result can be a serious outbreak of malaria.

While the model (11.3) describes the basic properties of malaria quite well, it ignores important aspects that could be added. For example, including an incubation period for the mosquito population would give considerably better quantitative agreement with data. Another extension would include seasonality in mosquito population density.

The model (11.3) assumes that infected humans are not subject to further infection, but there is evidence to indicate that “superinfection” may be an important phenomenon. MacDonald’s 1957 model [10] included successive infections waiting to express themselves when the previous infection ends. While this addition does not affect the qualitative behavior, it does affect quantitative results.

Immunity to malaria acquired by repeated exposure to malaria has the effect, in regions in which malaria is endemic, of causing most cases of malaria to be in children. Modeling this would require inclusion of age structure, as well as including acquired immunity that is boosted by reinfection. One way to model this would be

to assume that there is an immune period of fixed length following recovery in the absence of exposure but that if a person is exposed again during a fixed period following the loss of immunity the immunity is not lost [7, p. 179]. A description of various extensions of the basic Ross–MacDonald model may be found in [3, 13].

Controlling malaria by controlling mosquito populations is a fine theoretical solution. However, mosquitoes adapt very rapidly to insecticides and control methods must be revised frequently in practice. Malaria remains one of the diseases causing the largest number of deaths worldwide.

11.2 Some Model Refinements

The model (11.3) is a basic model which has been the standard model for many years. As we have pointed out in the preceding section, there are many possible refinements. In this section, we give a little more detail of some of these refinements.

11.2.1 Mosquito Incubation Periods

The basic malaria model (11.3) gives a good description of malaria infection dynamics, but some of its quantitative predictions are quite different from observations. Some of these differences can be resolved by taking into account the incubation period for mosquitoes [10, 11]. The most elementary model including an incubation period assumes an exposed period with an exponential rate of moving from exposed to infectious mosquitoes,

$$\begin{aligned} S'_h &= \Lambda_h - bT_{hv}S_h \frac{I_v}{N_v} - \mu_h S_h \\ I'_h &= bT_{hv}S_h \frac{I_v}{N_v} - (\mu_h + \alpha_h)I_h \\ S'_v(t) &= \Lambda_v - bT_{vh}S_v \frac{I_h}{N_h} - \mu_v S_v(t) \\ E'_v(t) &= bT_{vh}S_v \frac{I_h}{N_h} - (\mu + \kappa_v)E_v \\ I'_v(t) &= \kappa E_v - \mu_v I_v. \end{aligned} \tag{11.5}$$

To determine the basic reproduction number of the model (11.5), we use the next generation matrix method with disease states (I_h , E_v , I_v). We have

$$F = \begin{bmatrix} 0 & 0 & bT_{hv} \frac{N_h}{N_v} \\ bT_{vh} \frac{N_v}{N_h} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad V = \begin{bmatrix} \mu_h + \gamma_h & 0 & 0 \\ 0 & \mu_v + \kappa_v & 0 \\ 0 & 0 & \mu_v \end{bmatrix}$$

This leads to

$$FV^{-1} = \begin{bmatrix} 0 & bT_{vh}\frac{N_h}{N_v}\frac{\kappa_v}{\mu_h(\mu_v+\kappa_v)} & bT_{hv}\frac{N_h}{N_v}\frac{1}{\mu_v} \\ bT_{vh}\frac{N_v}{N_h}\frac{1}{\mu_h+\gamma_h} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

from which we derive

$$\mathcal{R}_0 = b^2 T_{vh} T_{hv} \frac{\kappa_v}{\mu_h(\mu_h + \gamma_h)(\mu_v + \kappa_v)}.$$

In fact, the model (11.3) was modified by assuming that infection in mosquitoes has an incubation period of fixed length τ [10, 11]. This gives a model

$$\begin{aligned} S'_h &= \Lambda_h - bT_{hv}S_h\frac{I_v}{N_v} - \mu_h S_h \\ I'_h &= bT_{hv}S_h\frac{I_v}{N_v} - (\mu_h + \gamma_h)I_h \\ S'_v(t) &= \Lambda_v - bT_{vh}S_v(t)\frac{I_h}{N_h} - \mu S_v(t) \\ E'_v(t) &= bT_{vh}S_v(t)\frac{I_h}{N_h} - e^{-\mu\tau}bT_{vh}S_v(t-\tau)\frac{I_h}{N_h} \\ I'_v(t) &= e^{-\mu\tau}bT_{vh}S_v(t-\tau)\frac{I_h}{N_h} - \mu_v I_v. \end{aligned} \tag{11.6}$$

For this model,

$$\mathcal{R}_0 = e^{-\mu\tau} \frac{b^2 T_{hv} T_{vh}}{\mu_v(\gamma_h + \mu_h)}.$$

As in the preceding section, it would be possible to use the square root as the basic reproduction number.

11.2.2 Boosting of Immunity

Another property of malaria is that the immunity obtained by recovery from infection is not permanent. Recovery brings temporary immunity and this immunity is boosted by exposure to further infection. This may be described by the following model. We divide the host population into three compartments, S the number of uninfected members, I the number of infected individuals, and R the number of removed individuals with immunity. We assume that susceptible individuals become infected at a rate h . They then recover at a rate r to become immune. There is a rate of reversion from immune to susceptible γ , chosen so that the average period of

immunity corresponds to an assumption that immunity lasts until the occurrence of a gap of τ years without exposure [1–3].

Then

$$\gamma(h) = \frac{he^{-h\tau}}{1 - e^{-h\tau}}. \quad (11.7)$$

To establish the relation (11.7), we let $p = e^{-h\tau}$ be the probability of infection after time τ and $q = 1 - p$ be the probability of infection after time τ . Then $\gamma(h)$, the rate of loss of immunity, is hp/q , and this establishes (11.7).

The loss of immunity has been observed frequently. In regions where there have been mosquito control programs that reduced malaria cases, one effect of terminating a control program has been an increase in malaria cases because of the loss of immunity since there was less boosting of immunity. The mean period of immunity may be estimated from observation, and the relation (11.7) gives the rate of reversion to susceptibility as a function of the rate of developing infection.

11.2.3 Alternative Forms for the Force of Infection

Several forms for the force-of-infection for malaria models are considered in [4]. We denote the transmission rates from humans to mosquitoes and from mosquitoes to humans by

$$\lambda_h = T_{hv} b_h(N_h, N_v) \frac{I_v}{N_v} \quad \text{and} \quad \lambda_v = T_{vh} b_v(N_h, N_v) \frac{I_h}{N_h}, \quad (11.8)$$

respectively, where

$$b_h = \frac{b(N_h, N_v)}{N_h}, \quad b_v = \frac{b(N_h, N_v)}{N_v},$$

and $b(N_h, N_v)$ denotes the total number of mosquito bites on humans given by

$$b = \frac{\sigma_v N_v \sigma_h N_h}{\sigma_v N_v + \sigma_h N_h} = \frac{\sigma_v \sigma_h}{\sigma_v (N_v/N_h) + \sigma_h} N_v. \quad (11.9)$$

Depending on the sizes of the human and mosquito populations, the biting rates b_h and b_v can take several limiting forms, as summarized in Table 11.1.

Table 11.1 Number of mosquito bites on humans in the force-of-infection functions (11.8)

	b_h	b_v	b
General model	$\frac{\sigma_v \sigma_h N_v}{\sigma_v (N_v/N_h) + \sigma_h}$	$\frac{\sigma_v \sigma_h N_h}{\sigma_v (N_v/N_h) + \sigma_h}$	$\frac{\sigma_v N_v \sigma_h N_h}{\sigma_v (N_v/N_h) + \sigma_h}$
As $N_h \rightarrow \infty$ or $N_v \rightarrow 0$	$\frac{\sigma_v N_v}{N_h}$	σ_v	$\sigma_v N_v$
As $N_h \rightarrow 0$ or $N_v \rightarrow \infty$	σ_h	$\frac{\sigma_h N_h}{N_v}$	$\sigma_h N_h$

11.3 *Coupling Malaria Epidemiology and Sickle-Cell Genetics

Although the population dynamics of malaria and the population genetics of the sickle-cell genes occur on very different time scales, it is straightforward to develop an appropriate model relating these. The high mortality associated with malaria has led to strong historical selection for resistance, and hence for single major genes conferring resistance in heterozygotes, despite the associated burden borne by homozygotes. The foundation is the classical Ross–MacDonald model for the spread of malaria, expanded to include the relevant genetic structure of the host.

Let S_{h1} denote the density of uninfected humans of genotype AA . Similarly, S_{h2} denotes the population density of genotype AS . Furthermore, let I_{h1} and I_{h2} represent the population densities of infected individuals of each genotype. We ignore SS individuals; high mortality rates from sickle-cell disease are typical in countries with high transmission rates of falciparum malaria, so these individuals rarely reach reproductive maturity. An extended model including the SS individuals can be studied using similar methods but it is very difficult to interpret the threshold conditions due to the complexity of the model. Finally, let z be the fraction of mosquitoes that are transmitting malaria. The fraction of the AS individuals in the population is

$$w = \frac{S_{h2} + I_{h2}}{N_h}$$

where $N_h = S_{h1} + I_{h1} + S_{h2} + I_{h2}$ is the total human population density. The frequency of the S gene is denoted by $q = w/2$ and the frequency of the A gene is denoted by $p = 1 - q$.

Let $B(N)$ denote the human per-capita birth rate, possibly density dependent (e.g., logistic growth), with a constant per-capita natural mortality of μ_h . To couple ecology and evolution, we make two assumptions. First, we assume that the ratio of mosquitoes to humans is a constant, c . This is a standard assumption in the modeling of malaria (ever since the original Ross–MacDonald model). Any other assumption about variability in the ratio of mosquitoes to humans (N_v/N_h) would need to be justified.

Second, we assume that the fraction of each genotype born into the population P_i is given by

$$P_1 = p^2, \quad P_2 = 2pq.$$

The transmission of malaria between humans and mosquitoes is governed by some basic epidemiological parameters. The human biting rate is denoted by b , and average life of an infected mosquito is $1/\mu_v$. The probability that a human develops a parasitemia from a bite is denoted T_{hv_i} ; we assume that $T_{hv_1} \geq T_{hv_2}$. The disease induced death rate is denoted by ρ_i , and we assume that $\rho_1 \gg \rho_2$. In addition, we consider that AS individuals may die faster than AA individuals from causes other than malaria, and the excess rate of mortality for AS individuals is v . The probability that a mosquito acquires plasmodium from biting an individual of type i is denoted by T_{vh_i} . The average time until a victim of malaria recovers, denoted by $1/\gamma_{hi}$, may be different in AA and AS individuals.

The changes in population density of each genotype with each infection status are described by a set of five coupled ordinary differential equations:

$$\begin{aligned} S'_{hi} &= P_i B(N)N - m_i S_{hi} - bT_{hv_i}czS_{hi} + \gamma_{hi}I_{hi}, \\ I'_{hi} &= bT_{hv_i}czS_{hi} - (m_i + \gamma_{hi} + \rho_{hi})I_{hi}, \\ z' &= (1-z)\left(bT_{vh_1}\frac{I_{h1}}{N_h} + bT_{vh_2}\frac{I_{h2}}{N_h}\right) - \mu_v z, \end{aligned} \quad (11.10)$$

where $m_1 = \mu_h$ and $m_2 = \mu_h + v$. More detailed analysis of the model (11.10) can be found in [5, 6].

It is both mathematically convenient and biologically relevant to introduce new variables for prevalence of malaria infections in each genotype, $x_i = u_i/N$ and $y_i = v_i/N$, as well as the frequency of the S -gene, $w = x_2 + y_2 = 2q$. The equations in the new variables are derived from the original (11.10) using the chain rule. We note for clarification that

$$x_1 + y_1 + x_2 + y_2 = 1 \quad \text{and} \quad x_1 + y_1 = 1 - w.$$

We also introduce notation to reduce the number of parameters, $\beta_{hi} = bT_{hv_i}c$, $\beta_{vi} = bT_{vh_i}$, $i = 1, 2$. Then, we obtain the following system equivalent to (11.10) in the terms that describe important epidemiological, demographic, and population genetic quantities, y_1 , y_2 , z , w , and N_h :

$$\begin{aligned}
y'_1 &= \beta_{h1}z(1 - w - y_1) - (m_1 + \gamma_{h1} + \rho_{h1})y_1 - y_1N'_h/N_h, \\
y'_2 &= \beta_{h2}z(w - y_2) - (m_2 + \gamma_{h2} + \rho_{h2})y_2 - y_2N'_h/N_h, \\
z' &= (1 - z)(\beta_{v1}y_1 + \beta_{v2}y_2) - \mu_v z, \\
w' &= P_2B(N) - \rho_{h2}y_2 - m_2w - wN'_h/N_h, \\
N'_h &= N_h ((P_1 + P_2)B(N) - m_1(1 - w) - m_2w - \rho_{h1}y_1 - \rho_{h2}y_2).
\end{aligned} \tag{11.11}$$

Although most of the equations assume a general birth function $B(N)$, our detailed mathematical analysis for the specific case in which $B(N)$ is a density dependent per-capita birth function, $B(N) = B_0(1 - N/K)$, where B_0 is a constant (the maximum birth rate when population size is small) and K is approximately the density dependent reduction in birth rate.

The relevant parameters vary across many orders of magnitude. For example, the demographic parameters (B_0 and m_i) and the genetic parameters (ρ_{hi}) are on the order of 1/decades, and the malaria disease parameters (β_{hi} , γ_{hi} , β_{vi} , and μ_v) are on the order of 1/days. Hence, although the malaria disease dynamics and the changes in genetic composition are two coupled processes, the former occurs on a much faster time scale than the latter. Let $m_i = \epsilon \tilde{m}_i$, $\rho_{hi} = \epsilon \tilde{\rho}_{hi}$, and $B_0 = \epsilon \tilde{B}_0$ with $\epsilon > 0$ being small. We can use this fact to simplify the mathematical analysis of the full model with the use of singular perturbation techniques, which allows us to separate the time scales of the different processes. By letting $\epsilon = 0$, we obtain the following system for the fast dynamics:

$$\begin{aligned}
y'_1 &= \beta_{h1}z(1 - y_1 - w) - \gamma_{h1}y_1, \\
y'_2 &= \beta_{h2}z(w - y_2) - \gamma_{h2}y_2, \\
z' &= (1 - z)(\beta_{v1}y_1 + \beta_{v2}y_2) - \mu_v z,
\end{aligned} \tag{11.12}$$

which describes the epidemics of malaria for a given distribution of genotypes determined by w . Here, on the fast time scale, w is considered as a parameter. On the fast time scale, the basic reproduction number of malaria disease can be calculated as the leading eigenvalue of the next generation matrix:

$$\mathcal{R}_0 = \mathcal{R}_1(1 - w) + \mathcal{R}_2w, \tag{11.13}$$

where $\mathcal{R}_i = \frac{\beta_{hi}\beta_{vi}}{\gamma_{hi}\mu_v}$, $i = 1, 2$ involves parameters associated with malaria transmission between mosquitoes and humans of genotype i . In fact, \mathcal{R}_i (or $\sqrt{\mathcal{R}_i}$) is the basic reproduction number when the population consists of entirely humans of genotype i . It can be shown that, when $\mathcal{R}_0 < 1$, the disease-free equilibrium of the system (11.12) is locally asymptotically stable; and when $\mathcal{R}_0 > 1$, the system (11.12) has a unique non-trivial equilibrium $E^* = (y_1^*, y_2^*, z^*)$ given by

$$y_1^* = \frac{Q_{h1}z^*}{1 + Q_{h1}z^*}(1 - w), \quad y_2^* = \frac{Q_{h2}z^*}{1 + Q_{h2}z^*}w, \quad (11.14)$$

where $Q_{hi} = \beta_{hi}/\gamma_{hi}$ for $i = 1, 2$, and z^* is a solution of the equation

$$k_0z^2 + k_1z + k_2 = 0, \quad (11.15)$$

with

$$\begin{aligned} k_0 &= T_{h1}T_{h2} + \mathcal{R}_1T_{h2}(1 - w) + \mathcal{R}_2T_{h1}w, \\ k_1 &= T_{h1} + T_{h2} + \mathcal{R}_1(1 - T_{h2})(1 - w) + \mathcal{R}_2(1 - T_{h1})w, \\ k_2 &= 1 - \mathcal{R}_0. \end{aligned} \quad (11.16)$$

It can be shown that Eq. (11.15) has a positive solution only if $\mathcal{R}_0 > 1$. When $\mathcal{R}_0 < 1$, note that $k_2 = 1 - \mathcal{R}_0 > 0$ and $k_1 > 0$; it follows that Eq. (11.15) has no positive solution. Hence, E^* is not biologically feasible.

If $\mathcal{R}_0 > 1$, then $k_2 = 1 - \mathcal{R}_0 < 0$. It is easy to show that $k_0 > 0$ as $0 < w < 1$. Hence, Eq. (11.15) has a unique positive solution z^* . Let $h(z)$ denote the function of z given by the left-hand side of (11.15). Notice that $h(0) = k_2 < 0$, $h(1) = 1 + T_{h1}T_{h2} + T_{h1} + T_{h2} > 0$, and $h(z^*) = 0$. Hence, $0 < z^* < 1$. From (11.14) we also have that $0 < y_i^* < 1$, $i = 1, 2$. It follows that an endemic equilibrium $E^* = (y_1^*, y_2^*, z^*)$ exists and is unique.

For the stability of E^* , it can be shown that all eigenvalues of the Jacobian matrix at E^* have negative real part if and only if the following expression:

$$\lambda =: \sqrt{\left[\frac{\beta_{h1}(1 - w - y_1^*)}{\beta_{v1}y_1^* + \beta_{v2}y_2^* + \mu_v} \right] \left[\frac{\beta_{v1}(1 - z^*)}{\beta_{h1}z^* + \gamma_{h1}} \right] + \left[\frac{\beta_{h2}(w - y_2^*)}{\beta_{v1}y_1^* + \beta_{v2}y_2^* + \mu_v} \right] \left[\frac{\beta_{v2}(1 - z^*)}{\beta_{h2}z^* + \gamma_{h2}} \right]}$$

is less than 1 (see [5] for more details). Using the following Equalities:

$$z_1^* = \frac{\beta_{v1}y_1^* + \beta_{v2}y_2^*}{\beta_{v1}y_1^* + \beta_{v2}y_2^* + \mu_v} = \frac{\gamma_{h1}y_1^*}{\beta_{h1}(1 - w - y_1^*)},$$

$$z_2^* = \frac{\beta_{v1}y_1^* + \beta_{v2}y_2^*}{\beta_{v1}y_1^* + \beta_{v2}y_2^* + \mu_v} = \frac{\gamma_{h2}y_2^*}{\beta_{h2}(w - y_2^*)},$$

and noticing that $\beta_{hi}z^* + \gamma_{hi} > \gamma_{hi}$, $i = 1, 2$, and that $0 < z^* < 1$, we get

$$\lambda^2 < \frac{\gamma_{h1}y_1^*}{\beta_{v1}y_1^* + \beta_{v2}y_2^*} \frac{\beta_{v1}(1 - z^*)}{\gamma_{h1}} + \frac{\gamma_{h2}y_2^*}{\beta_{v1}y_1^* + \beta_{v2}y_2^*} \frac{\beta_{v2}(1 - z^*)}{\gamma_{h2}} = 1 - z^* < 1.$$

It follows that $\lambda < 1$ and that E^* is locally asymptotically stable.

Using the re-scaled time $\tau = \epsilon t$, we can re-write the full system (11.11) as

$$\begin{aligned} \epsilon \frac{dy_1}{d\tau} &= \beta_{h1}z(1 - y_1 - w) - \gamma_{h1}y_1 - \epsilon y_1((\tilde{m}_1 - \tilde{m}_2)w \\ &\quad + \tilde{\rho}_{h1}(1 - y_1) - \tilde{\rho}_{h2}y_2 + (P_1 + P_2)\tilde{B}(N_h)), \\ \epsilon \frac{dy_2}{d\tau} &= \beta_{h2}z(w - y_2) - \gamma_{h2}y_2 - \epsilon y_2((\tilde{m}_1 - \tilde{m}_2)(w - 1) \\ &\quad - \tilde{\rho}_{h1}y_1 + \tilde{\rho}_{h2}(1 - y_2) + (P_1 + P_2)\tilde{B}(N_h)), \\ \epsilon \frac{dz}{d\tau} &= (1 - z)(\beta_{v1}y_1 + \beta_{v2}y_2) - \mu_v z, \\ \epsilon \frac{dw}{d\tau} &= ((1 - w)P_2 - wP_1)\tilde{B}(N_h) + (\tilde{m}_1 - \tilde{m}_2)w(1 - w) \\ &\quad + \tilde{\rho}_{h1}wy_1 - \tilde{\rho}_{h2}(1 - w)y_2, \\ \frac{dN_h}{d\tau} &= N_h((P_1 + P_2)\tilde{B}(N_h) - \tilde{m}_1(1 - w) - \tilde{m}_2w - \tilde{\rho}_{h1}y_1 - \tilde{\rho}_{h2}y_2). \end{aligned} \tag{11.17}$$

This system has a two-dimensional slow manifold:

$$M = \{(y_1, y_2, z, w, N_h) : y_1 = y_1^*(w, N_h), y_2 = y_2^*(w, N_h), z = z^*(w, N_h)\},$$

which is normally hyperbolically stable as it consists of a set of such equilibria of the fast system (11.12). Here y_1^* and y_2^* are given in (11.14). The slow dynamics on M is described by the equations

$$\begin{aligned} \frac{dw}{d\tau} &= ((1 - w)P_2 - wP_1)\tilde{B}(N_h) + (\tilde{m}_1 - \tilde{m}_2)w(1 - w) \\ &\quad + \tilde{\rho}_{h1}wy_1^* - \tilde{\rho}_{h2}(1 - w)y_2^*, \\ \frac{dN_h}{d\tau} &= N_h((P_1 + P_2)\tilde{B}(N_h) - \tilde{m}_1(1 - w) - \tilde{m}_2w - \tilde{\rho}_{h1}y_1^* - \tilde{\rho}_{h2}y_2^*). \end{aligned} \tag{11.18}$$

Define the fitness of the S -gene by

$$\mathcal{F} = \left(\frac{1}{w} \frac{dw}{d\tau} \right),$$

which represents the initial per-capita growth of the S -gene. Then the following formula can be derived:

$$\left(\frac{1}{w} \frac{dw}{d\tau} \right) \Big|_{w=0} = (\tilde{m}_1 + W_1\tilde{\rho}_{h1}) - (\tilde{m}_2 + W_2\tilde{\rho}_{h2}), \tag{11.19}$$

where

$$W_1 = \frac{Q_{h1}(\mathcal{R}_1 - 1)}{(1 + Q_{h1})\mathcal{R}_1}, \quad W_2 = \frac{Q_{h2}(\mathcal{R}_1 - 1)}{(1 + Q_{h1})\mathcal{R}_1 + Q_{h1} - Q_{h2}}. \quad (11.20)$$

Let

$$\sigma_i = \tilde{m}_i + W_i \tilde{\rho}_{hi}. \quad (11.21)$$

Then $\sigma_i \geq 0$ is the total per-capita death rate of type i individuals weighted by W_i , which depends only on malaria epidemiological parameters. The biological interpretation of \mathcal{F} suggests that, when the S -gene is initially introduced into a population, it may or may not establish itself depending on whether the fitness is positive or negative, which is equivalent to whether $\sigma_2 < \sigma_1$ or $\sigma_2 > \sigma_1$. This is indeed confirmed by both analytical and numerical studies of the slow system. Figure 11.1a shows a bifurcation diagram of the slow dynamics with σ_1 and σ_2 being the bifurcation parameters. In Fig. 11.1, \tilde{B}^* is a constant larger than the maximum per-capita birth rate \tilde{B} ($= B/\epsilon$); $\sigma_2 = h(\sigma_1)$ is a decreasing function satisfying

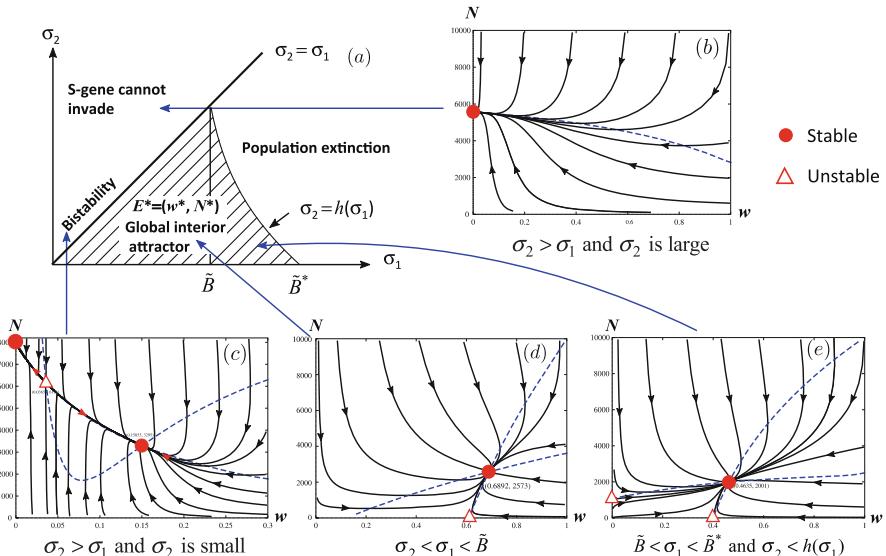


Fig. 11.1 Phase portraits (a) of the slow system in the (σ_1, σ_2) plane. Plots (b)–(e) illustrate the phase portraits of the slow system for (σ_1, σ_2) in different regions

$h(\tilde{B}) = \tilde{B}$ and $h(\tilde{B}^*) = 0$. Some of the results from this bifurcation diagram are summarized as follows:

Case 1: $\sigma_2 < \sigma_1$ (Positive Fitness)

- (a) If $\sigma_1 \leq \tilde{B}$, or $\tilde{B} < \sigma_1 < \tilde{B}^*$ and $\sigma_2 < h(\sigma_1)$, then there is a unique interior equilibrium $E_* = (w_*, N_{h*})$ that is globally asymptotically stable (g.a.s.),
- (b) If $\tilde{b} < \sigma_1 < \tilde{B}^*$ and $\sigma_2 > h(\sigma_1)$, then the population will be wiped out (due to the death rates being too much higher than the “birth” rate) with the fraction of AS individuals tending to a positive constant as $t \rightarrow \infty$.

Case 2: $\sigma_2 > \sigma_1$ (Negative Fitness)

The fraction of AS individuals will tend to zero as $t \rightarrow \infty$, whereas the total population size will tend to either K (when σ_2 is small) or zero (when σ_2 is large).

In either case, the system (11.18) has neither periodic solutions nor homoclinic loops.

An analytic proof of these results can be found in [6]. We point out that Case 1(b) is due to the standard incidence form of infection rate used in the z equation. Similar scenarios have been observed in other population models, and such scenarios may not be present if the mass action form is used. The standard incidence form is more appropriate if the number of contacts is relatively constant, independent of density. Figure 11.1b–e demonstrates some numerical calculations of solutions of the system (11.18) for (σ_1, σ_2) in different regions. It shows a couple of possible scenarios when $\sigma_1 < \sigma_2$. It is interesting to notice that it is possible for the system to have two locally asymptotically stable equilibria. One is the boundary equilibrium at which $N_h > 0$ and $w = 0$, and the other is one of the two interior equilibria. The regions of attraction of the two stable equilibria are divided by the separatrix formed by the stable manifold of the unstable interior equilibrium. This type of bi-stability can occur in several different ways. There are also cases when the system (11.18) has none, or one, or two equilibria on the positive w -axis. It shows that, if $w(0)$ is small, i.e., the initial population size of AS individuals is small, then the S-gene will be extinct due to a negative fitness. However, if for some reason (e.g., immigration of AS individuals) w suddenly becomes large (large enough to be on the right side of the separatrix), then the S-gene will be able to establish itself, even though the fitness is negative.

These results can be used to study questions related to the evolution of associated traits. We see from above that whether or not the S-gene can invade and establish itself in a population is determined by whether the fitness coefficient is positive or negative. Recall that the fitness is given by the difference $\sigma_1 - \sigma_2$, where the total death rate σ_i is a sum of weighted death rates m_i and ρ_{hi} (see (11.21)) with the weight W_i given by (11.20). The quantity W_i contains all the malaria transmission parameters.

In [12], an extension of model (11.10) is considered by including a class with asymptomatic malaria. It assumes that among all malaria infections, the proportion k is assumed to be symptomatic, while a proportion, $1 - k$, is asymptomatic. It is assumed that a proportion f of individuals who experience symptomatic infections receives antimalarial therapy. It is known that sickle-cell traits protect individuals against symptomatic but not against asymptomatic infections [14, 15]. Thus, we let $1 - \zeta$ denote the level of protection against symptomatic infections among the sub-population of individuals with sickle-cell traits. Furthermore, $1/\gamma_h$ is used to denote the average infective period of symptomatic patients, and the malaria-induced death rate is denoted by ρ_{hi} ($\rho_{h1} \gg \rho_{h2}$). Let M_{hi} denote symptomatically infected or treated individuals of type i , and let A_{hi} denote asymptotically infected hosts of type i . For treated individuals, the average infective period of symptomatic patients can be shortened to $1/\gamma_{hT}$ and disease mortality is reduced to ρ_{hT} . With or without treatment, infections do not confer permanent immunity, allowing repeated infections.

The model reads

$$\begin{aligned}
\frac{dS_{h1}}{dt} &= P_1 B(N_h) N_h - \beta_h S_{h1} z + \gamma_h (I_{h1} + A_{h1}) + \gamma_{hT} M_{h1} - \mu_{h1} S_{h1}, \\
\frac{dS_{h2}}{dt} &= P_2 B(N_h) N_h - [\zeta k + (1 - k)] \beta_h S_{h2} z + \gamma_h (I_{h2} \\
&\quad + A_{h2}) + \gamma_{hT} M_{h2} - \mu_{h2} S_{h2}, \\
\frac{dI_{h1}}{dt} &= k(1 - f) \beta_h S_{h1} z - (\gamma_h + \mu_{h1} + \rho_{h1}) I_{h1}, \\
\frac{dI_{h2}}{dt} &= \zeta k(1 - f) \beta_h S_{h2} z - (\gamma_h + \mu_{h2} + \rho_{h2}) I_{h2}, \\
\frac{dM_{h1}}{dt} &= kf \beta_h S_{h1} z - (\gamma_{hT} + \mu_{h1} + \rho_{hT1}) M_{h1}, \\
\frac{dM_{h2}}{dt} &= \zeta kf \beta_h S_{h2} z - (\gamma_{hT} + \mu_{h2} + \rho_{hT2}) M_{h2}, \\
\frac{dA_{h1}}{dt} &= (1 - k) \beta_h S_{h1} z - (\gamma_h + \mu_{h1}) A_{h1}, \\
\frac{dA_{h2}}{dt} &= (1 - k) \beta_h S_{h2} z - (\gamma_h + \mu_{h2}) A_{h2}, \\
\frac{dz}{dt} &= \frac{\beta_v (I_{h1} + I_{h2} + A_{h1} + A_{h2} + \delta M_{h1} + \delta M_{h2})(1 - z)}{N_h} - \mu_v z,
\end{aligned} \tag{11.22}$$

where $N_h = \sum_{i=1}^2 S_{hi} + I_{hi} + M_{hi} + A_{hi}$ and $B(N_h) = B_0(1 - N_h/K)$.

Similarly to the analysis of the system (11.10), it can be shown that on the fast time scale, the basic reproduction number of malaria disease can be calculated as the leading eigenvalue of the next generation matrix:

$$\begin{aligned}\mathcal{R}_c &= (1-g) \left(\frac{\beta_v \beta_h}{\mu_v \gamma_h} \right) + g \left(\frac{\beta_v \beta_h (1 - k(1-\zeta))}{\mu_v \gamma_h} \right) \\ &= (1-g)\mathcal{R}_1 + g\mathcal{R}_2.\end{aligned}\quad (11.23)$$

Here \mathcal{R}_i ($i = 1, 2$) is defined as $\mathcal{R}_i = Q_v Q_{hi}$, where

$$Q_v = \frac{\beta_v}{\mu_v}, \quad Q_{h1} = \frac{\beta_h}{\gamma_h}, \quad \text{and} \quad Q_{h2} = \frac{\beta_h (1 - k(1-\zeta))}{\gamma_h}.$$

It can be shown that E_0 is locally asymptotically stable if $\mathcal{R}_c < 1$ and unstable if $\mathcal{R}_c > 1$.

The model can be used to explore the influence of S-gene frequency on the malaria epidemic. To assess how the frequency of the S-gene g influences the overall endemic level of malaria, we examined how the reproduction number of malaria (\mathcal{R}_c) changes when g is varied (Fig. 11.2). In general, the disease prevalence increases with \mathcal{R}_c . It is also worth noting that $\mathcal{R}_c = (1-g)\mathcal{R}_1 + g\mathcal{R}_2$ where \mathcal{R}_1 and \mathcal{R}_2 are contributions from non-carriers and carriers of sickle-cell disease, respectively (see (11.23)). As a result, \mathcal{R}_1 is greater than \mathcal{R}_2 in the presence of the S-gene, and \mathcal{R}_c decreases with the frequency of the S-gene (g). This proves that the overall prevalence of malaria infection decreases with an increasing number of individuals who have the sickle-cell trait. However, this effect is stronger on symptomatic infections than on asymptomatic infections (Fig. 11.2). Nevertheless, the relative prevalence of asymptomatic infections among sickle-

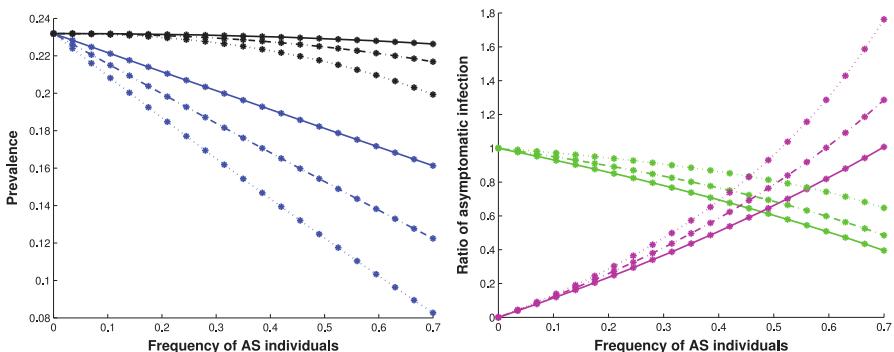


Fig. 11.2 Impact of sickle-cell trait on the epidemiology of malaria as the frequency of AS individuals (g) was varied (solid: $\zeta = 0.6$; dashed: $\zeta = 0.4$; dotted: $\zeta = 0.2$). The left figure plots the prevalence of symptomatic ($(I_{h1}^* + I_{h2}^*)/N_h^*$, blue) and asymptomatic ($(A_{h1}^* + A_{h2}^*)/N_h^*$, black) malaria infection for different ζ . The figure on the right plots the relative prevalence of asymptomatic AA individuals and asymptomatic AS individuals compared to all symptomatic individuals (shown as $A_{h1}^*/(I_{h1}^* + I_{h2}^*)$ in green and $A_{h2}^*/(I_{h1}^* + I_{h2}^*)$ in magenta)

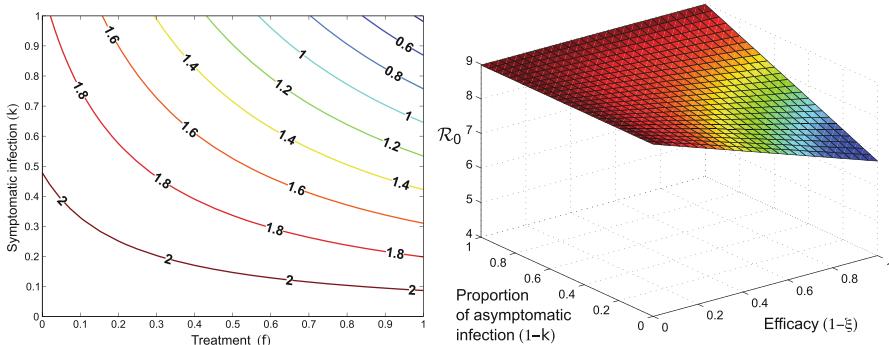


Fig. 11.3 The left figure is a plot of the level curves of the control reproduction number (\mathcal{R}_c) when the proportion of symptomatic infection (k) and the proportion of treatment (f) are varied. The figure on the right is a plot of \mathcal{R}_0 vs relative efficacy against symptomatic infection among individuals of type 2 compared to individuals of type 1, and the proportion of asymptomatic infection

cell carriers increased with the frequency of the S-gene(Fig. 11.2). Specifically, the relative prevalence of asymptomatic infections with sickle-cell trait, compared to symptomatic infections among both the carriers and non-carriers of sickle-cell disease, increased from 0.12 to 0.59 as the frequency of the S-gene increased from 0.10 to 0.40. This increase occurred partially because the sickle-cell trait does not provide protection against asymptomatic malaria. This pattern was more pronounced when the relative susceptibility of sickle-cell carriers to malaria (ζ) decreased (Fig. 11.2).

Figure 11.2 demonstrates how various factors may influence the disease dynamics. The effects on \mathcal{R}_c can be examined using the formula given in (11.23). This is shown in Fig. 11.3, which shows how \mathcal{R}_c changes as the probability of treatment (f) or the proportion of symptomatic infection (k) varies. We observe that the value of \mathcal{R}_c was more sensitive to the changes in the proportion of k than to changes in the treatment probability f . Furthermore, when the proportion of symptomatic infection (k) is relatively low, increasing treatment rate is unlikely to be effective in lowering the burden of malaria. Given that the asymptomatic infection of malaria is common in malaria-endemic regions, our result indicates that the effect of treatment might be limited. In addition, strong selection of sickle-cell traits (i.e., lower value of ζ) is likely to reduce the impact of treatment on reducing the prevalence of malaria.

References

1. Aron, J.L. (1982) Dynamics of acquired immunity boosted by exposure to infection, *Math. Biosc.* **64**: 249–259.
2. Aron, J.L. (1988) Mathematical modeling of immunity to malaria (1988) *Math. Biosc.* **90**: 385–396.

3. Aron, J.L. & R.M. May (1982) The population dynamics of malaria, in Population Dynamics of Infectious Diseases, R.M. Anderson, ed. Chapman and Hall, London, pp. 139–179.
4. Chitnis, N., J.M.Cushing, & J.M. Hyman (2006) Bifurcation analysis of a mathematical model for malaria transmission, SIAM J. App. Math. **67**: 24–45.
5. Feng, Z., D.L. Smith, E.F. McKenzie & S.A. Levin (2004) Coupling ecology and evolution: malaria and the S-gene across time scales, Math. Biosc. **189**: 1–19.
6. Feng, Z., Y. Yi, & H. Zhu (2004) Fast and slow dynamics of malaria and the S-gene frequency, Journal of Dynamics and Differential Equations, **16**: 869–896.
7. Karlin, S. & H.M. Taylor (1975) A First Course in Stochastic Processes, 2nd ed., Academic Press, New York.
8. MacDonald, G. (1950) The analysis of infection rates in diseases in which superinfection occurs, Tropical Diseases Bull. **47**: 907–915.
9. MacDonald, G. (1952) The analysis of equilibrium in malaria, Tropical diseases Bull. **49**: 813–828.
10. MacDonald, G. (1957) The Epidemiology and Control of Malaria, Oxford University Press, London.
11. Ross, R. (1911) The Prevention of Malaria, 2nd ed., (with Addendum), John Murray, London.
12. Shim, E., Z. Feng, C. Castillo-Chavez (2012) Differential impact of sickle cell trait on symptomatic and asymptomatic malaria, Math. Biosc. & Eng., **9**: 877–898.
13. Teboh-Ewungkem, M.I., G.A. Ngwa & C.N. Ngonghala (2013) Models and proposals for malaria: A review Math. Pop. Studies **20**: 57–81.
14. Vafa M., M. Troye-Blomberg, J. Anchang, A. Garcia & F. Migot-Nabias (2008) Multiplicity of Plasmodium falciparum infection in asymptomatic children in Senegal: relation to transmission, age and erythrocyte variants, Malaria J. **7**: 17.
15. Williams T.N. (2006) Human red blood cell polymorphisms and malaria, Curr. Opin. Microbiol. **9**: 388–394.

Chapter 12

Dengue Fever and the Zika Virus



12.1 Dengue Fever

While there have been cases of probable dengue fever more than 1000 years ago, the first recognized dengue epidemics occurred in Asia, Africa, and North America in the 1780s. There have been frequent outbreaks since then, and the number of reported cases has been increasing rapidly recently. According to the World Health Organization, approximately 50,000,000 people worldwide are infected with dengue. Symptoms may include fever, headaches, joint and muscle pain, and nausea, but many cases are very mild. There is no cure for dengue fever, but most patients recover with rest and fluids. There are at least four different strains of dengue fever, and there is some cross-immunity between strains. Dengue fever is transmitted by the mosquito *aedes aegypti*, and most control strategies are aimed at mosquito control.

Dengue, a re-emerging vector-borne disease, is caused by members of the genus *Flavivirus* in the family *Flaviviridae* with four active antigenically distinct serotypes, DENV-1, DENV-2, DENV-3, and DENV-4 [15]. The pathogenicity of dengue can range from asymptomatic, mild dengue fever (DF), to dengue hemorrhagic fever (DHF), and dengue shock syndrome (DSS) [15, 23]. Although infection with a dengue serotype does not usually protect against other serotypes, it is believed that secondary infections with a heterologous serotype increase the probability of DHF and DSS [10, 22]. According to the World Health Organization, 40% of the global population is at risk for dengue infection with an estimate of 50–100 million infections yearly including 500,000 cases of DHF. It has been estimated that about 22,000 deaths, mostly children under 15 years of age, can be attributed to DHF [46]. In the United States, approximately 5% or more of the Key West population in Florida was exposed to dengue during the 2009–2010 outbreak [11] while the Hawaii Department of Health reported 190 cases

during the 2015 outbreak on Oahu, the first outbreak since 2011. Since dengue is not endemic in Hawaii, health authorities have suggested that the recent outbreak may have been started by infected visitors [25]. Dengue is highly prevalent and endemic in Southeast Asia, which has experienced a 70% increase in cases since 2004 [29]; Mexico, also an endemic country, reported over a million cases of DF and more than 17,000 cases of DHF [21, 33] during the 2002 outbreak. Dengue is transmitted primarily by the vector *Ae. aegypti*, which is now found in most countries in the tropics [24, 37]. The secondary vector, *Ae. albopictus*, has a range reaching farther north than *Ae. aegypti* with eggs better adapted to subfreezing temperatures [26, 33]. Differences in susceptibility and transmission of dengue infection [3, 27, 42] raise the possibility that some serotypes are either more successful at invading a host population, or more pathogenic, or both [30]. DENV-2 is the most associated with dengue outbreaks involving DHF and DSS cases [32, 38, 48], followed by DENV-1 and DENV-3 viruses [4, 24, 32]. While infection with any of the four dengue serotypes could lead to DHF, the rapid displacement of DENV-2 American by DENV-2 Asian genotype has been linked to major outbreaks with DHF cases in Cuba, Jamaica, Venezuela, Colombia, Brazil, Peru, and Mexico [31, 32, 38, 39, 41, 48]. A possible mechanism involved in the dispersal and persistence of DENV-2 in nature is vertical transmission (transovarial transmission) via *Ae. aegypti*. Advances in molecular biology have been used to show that vertical transmission involving *Ae. aegypti* and *Ae. albopictus* is possible in captivity and in the wild [3, 8, 12, 20, 34, 40]. Thus, assessing transmission dynamics and pathogenicity between the DENV-2 American and Asian genotypes' differences is one of the priorities associated with the study of the epidemiology of dengue. In short, dengue has an increasing recurrent presence putting a larger percentage of the global population at risk of dengue infection, a situation that has become the norm due to the growth of travel and tourism between endemic and non-endemic regions.

The potential role of vertical transmission in dengue endemic regions or in fluctuating environments has been explored in [1, 18, 36]. The role of host movement has also been explored in the context of dengue [2] in a formulation that does not account for the effective population size.

The model (6.2) in Chapter 6 is a generic model for vector-transmitted diseases. For any specific disease, it is necessary to modify this model to incorporate properties of the disease not included in the generic model.

Suppose we assume that the mosquito population is in equilibrium, so that the mosquito birth rate is $\beta_v N_v$. We assume also vertical transmission for mosquitoes. The birth rate of mosquitoes is $\mu_v N_v$, of which $\mu_v I_v$ are born to infective mother mosquitoes, and we assume that a fraction q of these are born infective. Then a model describing the dynamics of DENV-2 is given by the following system of differential equations:

$$\begin{aligned}
S'_h &= \mu_h N_h - \beta_h S_h \frac{I_v}{N_v} - \mu_h S_h \\
E'_h &= \beta_h S_h \frac{I_v}{N_v} - (\eta_h + \mu_h) E_h \\
I'_h &= \eta_h E_h - (\gamma + \mu_h) I_h \\
S'_v &= \mu_v (N_v - q I_v) - \beta_v S_v \frac{I_h}{N_h} - \mu_v S_v \\
E'_v &= \beta_v S_v \frac{I_h}{N_h} - (\eta_v + \mu_v) E_v \\
I'_v &= q \mu_v I_v + \eta_v E_v - \mu_v I_v.
\end{aligned} \tag{12.1}$$

This model is the same as the basic vector transmission model (6.2) in Chapter 6 except that the birth rate of susceptible hosts is now $\mu_h N_h$ and vertical transmission of hosts is added. In the absence of selection, that is, differences in birth and death rate and in the absence of vertical transmission, the model (12.1) turns out to be equivalent to a model considered by Chowell et al. in [13]. Model (12.1) is well defined supporting a sharp threshold property, namely, the disease dies out if the basic reproduction number \mathcal{R}_0 is less than unity and persists whenever $\mathcal{R}_0 > 1$.

12.1.1 Calculation of the Basic Reproduction Number

We calculate the basic reproduction number for the model (12.1) in two stages, as we did for the model (6.2) in Sect. 6.2.

In the first stage, an infective mosquito infects humans at a rate $\beta N_h / N_v$ for a time $1/\mu_v$, producing $\beta N_h / N_v \mu_v$ infected humans per mosquito.

In the second stage, an infective human infects mosquitoes, at a rate $\beta_v N_v / N_h$ for a time $1/(\mu_h + \gamma)$. This produces $\beta_v N_v / N_h (\gamma + \mu_h)$ infected mosquitoes, of whom a fraction $\eta_v / (\eta_v + \mu_v)$ proceeds to become infective.

The net result of these two stages is

$$\mathcal{R}_v = \frac{\beta_v N_v}{N_h} \frac{1}{\mu_h + \gamma} \frac{\eta_v}{\eta_v + \mu_v} \frac{\eta_h}{\eta_h + \mu_h} \frac{1}{\mu_v} = \beta_h \beta_v \frac{1}{\mu_h + \gamma} \frac{\eta_v}{\eta_v + \mu_v} \frac{\eta_h}{\eta_h + \mu_h} \frac{1}{\mu_v} \tag{12.2}$$

infected vectors. In addition, an infective mosquito produces

$$\mathcal{R}_d = q \mu_v \tag{12.3}$$

infective mosquitoes through vertical transmission, giving a total basic reproduction number

$$\begin{aligned}\mathcal{R}_0 &= \mathcal{R}_v + \mathcal{R}_d \\ &= \beta_h \beta_v \frac{1}{\mu_h + \gamma} \frac{\eta_v}{\eta_v + \mu_v} \frac{\eta_h}{\eta_h + \mu_h} \frac{1}{\mu_v} + q \mu_v.\end{aligned}\tag{12.4}$$

We could also calculate the basic reproduction number by using the next generation matrix approach [45]. If we interpret only human infections as new infections and consider vector infections as transitions, we would obtain the same result.

12.2 A Model with Asymptomatic Infectives

Many cases of dengue are very mild and may not be reported. We can incorporate this in a model by assuming that a fraction p of exposed members become infective while the remainder of the exposed class go to an asymptomatic stage with lower infectivity and possibly more rapid recovery. We have already described a model with such transitions in influenza models in Sect. 9.2. We consider a model including this structure, namely

$$\begin{aligned}S'_h &= \mu_h N_h - \beta_h S_h \frac{I_v}{N_v} - \mu_h S_h \\ E'_h &= \beta_h S_h \frac{I_v}{N_v} - (\eta_h + \mu_h) E_h \\ I'_h &= p \eta_h E_h - (\gamma + \mu_h) I_h \\ A'_h &= (1 - p) \eta_h E_h - (\kappa + \mu_h) A_h \\ S'_v &= \mu_v (N_v - q I_v) - \beta_v S_v \frac{I_h + \delta A_h}{N_h} - \mu_v S_v \\ E'_v &= \beta_v S_v \frac{I_h + \delta A_h}{N_h} - (\eta_v + \mu_v) E_v \\ I'_v &= q \mu_v + \eta_v E_v - \mu_v I_v.\end{aligned}\tag{12.5}$$

Here, δ is the infectivity reduction factor for asymptomatics, and κ is the recovery rate for asymptomatics.

12.2.1 Calculation of the Basic Reproduction Number

We calculate the basic reproduction number for the model (12.5) in two stages, as we did for the model (12.1) in the previous section.

In the first stage, an infective mosquito infects humans at a rate $\beta N_h/N_v$ for a time $1/\mu_v$, producing $\beta N_h/N_v \mu_v$ infected humans per mosquito. A fraction $\eta_h/(\eta_h + \mu_h)$ of these proceed to an infective stage, with

$$p\beta_h \frac{N_h}{\mu_v N_v} \frac{\eta_h}{\eta_h + \mu_h}$$

going to I_h and

$$(1-p)\beta_h \frac{N_h}{\mu_v N_v} \frac{\eta_h}{\eta_h + \mu_h}$$

going to A_h . In the second stage, an infective human infects mosquitoes, at a rate $\beta_v N_v/N_h$ for a time $1/(\mu_h + \gamma)$. An asymptomatic infects mosquitoes at a rate $\delta \beta_v N_v/N_h$ for a time $1/(\mu_h + \kappa)$. A fraction $\eta_v/(\eta_v + \mu_v)$ of each of these groups develop into infective mosquitoes. Thus, the second stage produces

$$\beta_v \frac{N_v}{N_h} \frac{\eta_v}{\eta_v + \mu_v} \left[\frac{p}{\mu_h + \gamma} + \frac{\delta(1-p)}{\mu_h + \kappa} \right]$$

infected mosquitoes.

The net result of these two stages is

$$\mathcal{R}_v = \beta_h \frac{N_h}{\mu_v N_v} \frac{\eta_h}{\eta_h + \mu_h} \beta_v \frac{N_v}{N_h} \frac{\eta_v}{\eta_v + \mu_v} \left[\frac{p}{\mu_h + \gamma} + \frac{\delta(1-p)}{\mu_h + \kappa} \right] \quad (12.6)$$

infected mosquitoes. In addition, an infective mosquito produces

$$\mathcal{R}_d = q\mu_v \quad (12.7)$$

infective mosquitoes through vertical transmission, giving a total basic reproduction number

$$\mathcal{R}_0 = \mathcal{R}_v + \mathcal{R}_d. \quad (12.8)$$

Again, we could also calculate the basic reproduction number by using the next generation matrix approach [45]. If we interpret only human infections as new infections and consider vector infections as transitions, we would obtain the same result, but if we interpreted both human and vector infections we would obtain a different version of the basic reproduction number

$$\mathcal{R}^* = \frac{1}{2} \left[\mathcal{R}_d + \sqrt{\mathcal{R}_d^2 + 4\mathcal{R}_v} \right].$$

12.3 The Zika Virus

The Zika virus, a mosquito borne arbovirus, was first identified in Uganda in 1947. Similar to the dengue and chikungunya viruses, Zika is primarily spread by the mosquito, *Aedes aegypti*. Recent outbreaks of Zika disease have occurred in Yap Island in the Pacific in 2007 [16] and in French Polynesia in 2013–2014 [28]. Since 2015, Zika has spread through much of South America, Central America, and the Caribbean, where the *Aedes aegypti* species is endemic, recently reaching pandemic levels in 2016. While the reasons for the explosive spread of the disease in the Americas are still unclear, the rapid urbanization in countries with under-developed infrastructure for surveillance and vector control probably plays a role.

Zika disease is usually asymptomatic, and typically is mild even with a clinical presentation. Its symptoms are similar to those of dengue and chikungunya virus infections. However, the disease has been linked to an apparent increased risk of the neurological disorder Guillain–Barré syndrome, and also to neonate microcephaly. The latter is of particular concern, because pregnant women may not even know they have been infected, and the damage to their unborn infants may result in subsequent lifelong disabilities. There is currently no vaccine or specific treatment for Zika infection, leaving control of the vector populations and avoidance through the use of mosquito repellents as the only means to control the spread of the disease.

For the Zika virus, it has been established that in addition to vector transmission of infection there may also be direct transmission through sexual contact. The Zika virus is the first example of an infection that can be transferred both directly and through a vector, and it is important to include direct transmission (in this case sexual transmission) in a model. Estimation of the basic reproduction number is particularly difficult for Zika because of the difficulty in estimating the relative importance of transmission through vectors and direct transmission through sexual contact. A vaccine is being developed for the Zika virus, and analysis of the question of whether this vaccine can control an outbreak is contained in [44].

12.4 A Model with Vector and Direct Transmission

We approach the question of formulating a model for the Zika virus by beginning with the basic vector transmission model (6.2) and adding direct host to host transmission. We add to the model (6.2) a term $\alpha S_h \frac{I_h}{N_h}$ describing a rate α of movement from S to E . Also, we consider a single outbreak model, and omit demographic terms and vertical transmission in the host population.

This leads to the following model [9]:

$$\begin{aligned} S'_h &= -\beta_h S_h \frac{I_v}{N_v} - \alpha S_h \frac{I_h}{N_h} \\ E'_h &= \beta_h S_h \frac{I_v}{N_v} + \alpha S_h \frac{I_h}{N_h} - \eta_h E_h \end{aligned}$$

$$\begin{aligned} I'_h &= \eta_h E_h - \gamma I_h \\ S'_v &= \mu_v N_v - \mu_v S_v - \beta_v S_v \frac{I_h}{N_h} \\ E'_v &= \beta_v S_v \frac{I_h}{N_h} - (\mu_v + \eta_v) E_v \\ I'_v &= \eta_v E_v - \mu_v I_v. \end{aligned} \tag{12.9}$$

The rate α is an average over the human population; if transmission is possible only from male to female this is incorporated into α .

To calculate the basic reproduction number \mathcal{R}_0 , we use the same direct approach as that used in Section 6.2. If there is sexual transmission, this operates independent of the host–vector interaction, and produces α cases in unit time for a time $1/\gamma$, adding a simple term α/γ to the reproduction number

$$\mathcal{R}_0 = \beta_h \beta_v \frac{\eta_v}{\mu_v \gamma (\mu_v + \eta_v)} + \frac{\alpha}{\gamma}. \tag{12.10}$$

We define

$$\mathcal{R}_v = \beta_h \beta_v \frac{\eta_v}{\mu_v \gamma (\mu_v + \eta_v)}, \tag{12.11}$$

the vector transmission reproduction number, and

$$\mathcal{R}_d = \frac{\alpha}{\gamma}, \tag{12.12}$$

the direct transmission reproduction number, so that

$$\mathcal{R}_0 = \mathcal{R}_v + \mathcal{R}_d.$$

If we use the next generation matrix approach, using the same approach as that used in Sect. 12.1.1, we form the matrix product $K_L = FV^{-1}$ with

$$F = \begin{bmatrix} 0 & \alpha & 0 & \beta_h \frac{N_h}{N_v} \\ 0 & 0 & 0 & 0 \\ 0 & \beta_v \frac{N_v}{N_h} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad V = \begin{bmatrix} \eta_h & 0 & 0 & 0 \\ -\eta_h & \gamma & 0 & 0 \\ 0 & 0 & \mu_v + \eta_v & 0 \\ 0 & 0 & -\eta_v & \mu_v \end{bmatrix}.$$

Then the next generation matrix with large domain is

$$K_L = \begin{bmatrix} \frac{\alpha}{\gamma} & \frac{\alpha}{\gamma} & \beta_h \frac{N_h}{N_v} \frac{\eta_v}{\mu_v (\mu_v + \eta_v)} & \beta_h \frac{N_h}{N_v} \frac{1}{\mu_v} \\ 0 & 0 & 0 & 0 \\ \beta_v \frac{N_v}{N_h} \frac{1}{\gamma} & \beta_v \frac{N_v}{N_h} \frac{1}{\gamma} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The next generation matrix K is the 2×2 matrix

$$K = \begin{bmatrix} \frac{\alpha}{\gamma} & \beta_h \frac{N_h}{N_v} \frac{\eta_v}{\mu_v(\mu_v + \eta_v)} \\ \beta_v \frac{N_v}{N_h} \frac{1}{\gamma} & 0 \end{bmatrix}.$$

The positive eigenvalue of this matrix is

$$\begin{aligned} \lambda &= \frac{\alpha}{2\gamma} + \frac{1}{2} \sqrt{\frac{\alpha^2}{\gamma^2} + 4\mathcal{R}_v} \\ &= \frac{1}{2} \left[\mathcal{R}_d + \sqrt{\mathcal{R}_d^2 + 4\mathcal{R}_v} \right]. \end{aligned}$$

We may calculate that $\lambda = 1$ if and only if

$$\mathcal{R}_v + \mathcal{R}_d = 1.$$

We now have two potential expressions for the basic reproduction number, namely $\mathcal{R}_v + \mathcal{R}_d$, with \mathcal{R}_v and \mathcal{R}_d given by (12.11) and (12.12) respectively, and

$$\mathcal{R}^* = \frac{1}{2} \left[\mathcal{R}_d + \sqrt{\mathcal{R}_d^2 + 4\mathcal{R}_v} \right].$$

Different expressions are possible for the next generation matrix and these may lead to different expressions for the basic reproduction number. This is shown in [14].

The expression $\mathcal{R}_v + \mathcal{R}_d$ appears to us to be a more natural form than \mathcal{R}^* , and we choose to use this for the basic reproduction number. It can be obtained from the following expression for the next generation matrix. We consider only human infections as new infections, and take

$$F = \begin{bmatrix} 0 & \alpha & 0 & \beta_h \frac{N_h}{N_v} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad V = \begin{bmatrix} \eta_h & 0 & 0 & 0 \\ -\eta_h & \gamma & 0 & 0 \\ 0 & -\beta_v \frac{N_v}{N_h} & \mu_v + \eta_v & 0 \\ 0 & 0 & -\eta_v & \mu_v \end{bmatrix}.$$

Then

$$V^{-1} = \begin{bmatrix} \frac{1}{\eta_h} & 0 & 0 & 0 \\ [2pt] \frac{1}{\gamma} & \frac{1}{\gamma} & 0 & 0 \\ \beta_v \frac{N_v}{N_h} \frac{1}{\gamma(\mu_v + \eta_v)} & \beta_v \frac{N_v}{N_h} \frac{\eta_v}{\gamma(\mu_v + \eta_v)} & \frac{1}{\mu_v + \eta_v} & 0 \\ \beta_v \frac{N_v}{N_h} \frac{\eta_v}{\mu_v \gamma(\mu_v + \eta_v)} & \beta_v \frac{N_v}{N_h} \frac{\eta_v}{\mu_v \gamma(\mu_v + \eta_v)} & \frac{\eta_v}{\mu_v(\mu_v + \eta_v)} & \frac{1}{\mu} \end{bmatrix}.$$

Since only the first row of F has non-zero entries, the same is true of FV^{-1} , and from this we can deduce that the only non-zero eigenvalue of FV^{-1} is the entry in the first row, first column of FV^{-1} , and this is

$$\frac{\alpha}{\gamma} + \beta_h \beta_v \frac{\eta_v}{\mu_v \gamma (\mu_v + \eta_v)} = \mathcal{R}_d + \mathcal{R}_v = \mathcal{R}_0.$$

If we use this somewhat unorthodox approach to the next generation matrix for the model (6.2), we obtain the form \mathcal{R}_v , with no square root, for the reproduction number. We now have two viable expressions for the basic reproduction number, namely \mathcal{R}^* and \mathcal{R}_0 , both derived from a next generation matrix approach but with different separations. We have chosen to use \mathcal{R}_0 for the basic reproduction number because it is more readily interpreted as a number of secondary infections. Other sources, including [19], use \mathcal{R}^* . In studying data for epidemic models that include vector transmission it is absolutely vital to specify exactly which form is being used for the basic reproduction number.

12.4.1 The Initial Exponential Growth Rate

In order to determine the initial exponential growth rate from the model, a quantity that can be compared with experimental data, we linearize the model (12.9) about the disease-free equilibrium $S = N_h$, $E_h = I_h = 0$, $S_v = N_v$, $E_v = I_v = 0$. If we let $y = N_h - S$, $z = N_v - S_v$, we obtain the linearization

$$\begin{aligned} y' &= \beta_h N_h \frac{I_v}{N_v} + \alpha I_h \\ E'_h &= \beta_h N_h \frac{I_v}{N_v} + \alpha I_h - \eta_h E_h \\ I'_h &= \eta_h E_h - \gamma I_h \\ z' &= -\mu_v z + \beta_v N_v \frac{I_h}{N_h} \\ E_v &= \beta_v N_v \frac{I_h}{N_h} - (\mu_v + \eta_v) E_v \\ I'_v &= \eta_v E_v - \mu_v I_v. \end{aligned} \tag{12.13}$$

The corresponding characteristic equation is

$$\det \begin{bmatrix} -\lambda & 0 & \alpha & 0 & 0 & \beta_h \frac{N_h}{N_v} \\ 0 & -(\lambda + \eta_h) & \alpha & 0 & 0 & \beta_h \frac{N_h}{N_v} \\ 0 & \eta_h & -(\lambda + \gamma) & 0 & 0 & 0 \\ 0 & 0 & \beta_v \frac{N_v}{N_h} & -(\lambda + \mu_v) & 0 & 0 \\ 0 & 0 & \beta_v \frac{N_v}{N_h} & 0 & -(\lambda + \mu_v + \eta_v) & 0 \\ 0 & 0 & 0 & 0 & \eta_v & -(\lambda + \mu_v) \end{bmatrix} = 0.$$

We can reduce this equation to a product of two factors and a fourth degree polynomial equation

$$\lambda(\lambda + \mu_v) \det \begin{bmatrix} -(\lambda + \eta_h) & \alpha & 0 & \beta_h \frac{N_h}{N_v} \\ \eta_h & -(\lambda + \gamma) & 0 & 0 \\ 0 & \beta_v \frac{N_v}{N_h} & -(\lambda + \mu_v + \eta_v) & 0 \\ 0 & 0 & \eta_v & -(\lambda + \mu_v) \end{bmatrix} = 0.$$

The initial exponential growth rate is the largest root of this fourth degree equation, which reduces to

$$\begin{aligned} g(\lambda) &= (\lambda + \eta_h)(\lambda + \gamma)(\lambda + \mu_v + \eta_v)(\lambda + \mu_v) - \beta_h \beta_v \eta_v \eta_v \\ &\quad - \eta_h \alpha (\lambda + \mu_v)(\lambda + \mu_v + \eta_v) = 0. \end{aligned} \quad (12.14)$$

The largest root of this equation is the initial exponential growth rate, and this may be measured experimentally. If the measured value is ρ , then from (12.14) we obtain

$$\begin{aligned} (\rho + \eta_h)(\rho + \gamma)(\rho + \mu_v + \eta_v)(\rho + \mu_v) - \beta_h \beta_v \eta_h \eta_v \\ - \eta_h \alpha (\rho + \mu_v)(\rho + \mu_v + \eta_v) = 0. \end{aligned} \quad (12.15)$$

From (12.15) we can see that $\rho = 0$ corresponds to $\mathcal{R}_0 = 1$, confirming that our calculated value of \mathcal{R}_0 has the proper threshold behavior.

Equation (12.15) determines the value of $\beta_h \beta_v$, and we may then calculate \mathcal{R}_0 , provided we know the value of α . However, this presents a major problem. In [19] it is suggested that the contribution of sexual disease transmission is small, based on estimates of sexual activity and the probability of disease transmission. Since the probability of sexual transmission of a disease depends strongly on the particular disease, this estimate is quite uncertain. Estimates based on a possible imbalance between male and female disease prevalence are also quite dubious. Most Zika cases are asymptomatic or quite light but the risks of serious birth defects means that diagnosis of Zika is much more important to women than to men. If there are more female than male cases, it is not possible to distinguish between additional cases caused by sexual contact and cases identified by higher diagnosis rates. To the best of our knowledge, there is not yet a satisfactory resolution of this problem.

What would be required would be another quantity which can be determined experimentally and can be expressed in terms of the model parameters. In the absence of further information, all we can accomplish is to estimate reproduction numbers for various choices of α and $\beta_h \beta_v$ that satisfy (12.15). We use the parameter values [43] obtained for the 2015 Zika outbreak in Barranquilla, Colombia, including an analysis of the exponential rise in confirmed Zika cases identified by the Colombian SIVIGILA surveillance system up to the end of December, 2015.

$$\kappa = 1/7 \quad \gamma = 1/5 \quad \eta_v = 1/9.5 \quad \mu_v = 1/13,$$

Table 12.1 Reproduction number values

α	$\beta_h \beta_v$	\mathcal{R}_d	\mathcal{R}_v	\mathcal{R}_0	\mathcal{R}^*	S_∞
0	0.243	0	4.86	4.86	2.185	14
0.1	0.184	0.5	3.69	4.19	2.187	24
0.2	0.125	1.0	2.51	3.51	2.16	45
0.3	0.0665	1.5	1.335	2.835	2.13	79
0.4	0.0076	2.0	0.152	2.152	2.074	166
0.413	0	2.065	0	2.065	2.065	185

and the estimated measurement $\rho = 0.073$. With these values we have

$$11\beta_h \beta_v + 6.48\alpha = 2.676.$$

To satisfy this equation, we must have $0 \leq \alpha \leq 0.413$. We then calculate \mathcal{R}_0 and \mathcal{R}^* for several values of α in this range, assuming population sizes of 1000 humans and 4000 mosquitoes. We obtain the results summarized in Table 12.1.

We observe that \mathcal{R}^* is not very sensitive to changes in the direct contact rate while \mathcal{R}_0 is quite sensitive to changes in α . We have also shown the results of simulations of the model (12.9) showing how the epidemic size depends on α . These simulations suggest that the epidemic final size does vary considerably, and without some way of estimating how many disease cases arise from direct contact we are unable to estimate the epidemic final size.

12.5 A Second Zika Virus Model

A model for the Zika virus with somewhat more detail than the model (12.9) has been described in [19]. This model includes the assumption that many Zika cases are asymptomatic, and has two infectious stages acute and convalescent. The model takes the form

$$\begin{aligned} S'_h &= -abS_h \frac{I_v}{N_h} - \beta S_h \frac{\kappa E_h + I_{h1} + \tau I_{h2}}{N_h} \\ E'_h &= \theta \left[abS_h \frac{I_v}{N_h} + \beta S_h \frac{\kappa E_h + I_{h1} + \tau I_{h2}}{N_h} \right] - v_h E_h \\ I'_{h1} &= v_h E_h - \gamma_{h1} I_{h1} \\ I'_{h2} &= \gamma_{h1} I_{h1} - \gamma_{h2} I_{h2} \\ A'_h &= (1 - \theta) \left[bS_h \frac{I_v}{N_h} + \beta S_h \frac{\kappa E_h + I_{h1} + \tau I_{h2}}{N_h} \right] - \gamma_h A_h \\ R'_h &= \gamma_{h2} + \gamma_h A_h \\ S'_v &= \mu_v N_v - \mu_v S_v - ac \frac{S_v \eta A_h + I_{h1}}{N_h} \\ E'_v &= ac S_v \frac{\eta A_h + I_{h1}}{N_h} - (\mu_v + v_v) E_v \\ I'_v &= v_v E_v - \mu_v I_v. \end{aligned} \tag{12.16}$$

Table 12.2 Model parameters

Parameter	Description	Value
a	Mosquito biting rate	0.5
b	Transmission probability, vector to human	0.4
c	Transmission probability, human to vector	0.5
η	Transmission probability, asymptomatic humans to vector	0.1
β	Transmission probability, human to human	0.1
κ	Relative transmission probability, exposed to infective	0.6
τ	Relative transmission probability, convalescent to asymptomatic	0.3
θ	Proportion of symptomatic infections	18
m	Ratio of mosquitoes to humans	5
$1/v_h$	Human incubation period (days)	5
$1/v_v$	Mosquito incubation period (days)	10
$1/\gamma_{h1}$	Acute phase duration (days)	5
$1/\gamma_{h2}$	Convalescent period duration (days)	20
$1/\gamma_h$	Asymptomatic infection duration (days)	7
$1/\mu_v$	Mosquito lifetime (days)	14

The reproduction numbers for this model are

$$\mathcal{R}_d = \frac{\beta\kappa\theta}{v_h} + \frac{\beta\theta}{\gamma_{h1}} + \frac{\beta\tau\theta}{\gamma_{h2}}, \quad \mathcal{R}_v = \left(\frac{a^2 b \eta c \theta}{v_h \mu_v} + \frac{a^2 b c \theta}{\gamma_{h1} \mu_v} \right) \frac{\eta_v}{\eta_v + \mu_v}.$$

Parameter values were chosen to fit vector transmission data for Brazil, El Salvador, and Colombia up to February 2016. For direct (sexual) transmission of infection, parameters were chosen with an assumed probability of transmission of infection per sexual contact, but this assumption might be questionable (Table 12.2).

With these parameter values, this model yielded estimates of

$$\mathcal{R}_d = 0.136, \quad \mathcal{R}_v = 3.842,$$

so that

$$\mathcal{R}_0 = 3.973, \quad \mathcal{R}^* = 2.055.$$

However, there are indications that the number of cases of Zika due to sexual contacts may be considerably higher. Perhaps, the value of β should be larger.

12.6 Project: A Dengue Model with Two Patches

When dengue fever invades a location, it tends to move from one patch to another. To describe this, we consider a model for DENV-2 in two separate patches, with movement between the patches. The single patch model (12.1) is the building block for the two-patch model. Within each patch, in the absence of host mobility, dengue dynamics are modeled via the system (12.1). We consider an epidemic model in two patches, one of which has a significantly larger contact rate, with short term travel between the two patches. The total population resident in each patch is constant. We follow a Lagrangian perspective, that is, we keep track of each individual's place of residence at all times [6, 17]. It is assumed that vectors do not move between patches since the vectors *Ae. aegypti* and *Ae. albopictus* do not travel more than few tens of meters over their lifetime [2, 47]; moving 400–600 meters at most [7, 35], respectively. In short, we neglect vector dispersal.

Thus we consider two patches, with total resident population sizes N_1 and N_2 respectively, each population being divided into susceptibles, infectives, and removed members. S_i and I_i denote the number of susceptibles and infectives respectively who are residents in Patch i , regardless of the patch in which they are present.

The host residents of Patch 1, population size $N_{h,1}$, spends, on average, p_{11} proportion of its time in their own Patch 1 and p_{12} proportion of its time visiting Patch 2. Residents of Patch 2, population of size $N_{h,2}$, spend p_{22} proportion of their time in Patch 2 while spending $p_{21} = 1 - p_{22}$ visiting Patch 1. Thus, at time t , the *effective population* in Patch 1 is $p_{11}N_{h,1} + p_{21}N_{h,2}$ and the *effective population* in Patch 2 is $p_{12}N_{h,1} + p_{22}N_{h,2}$. The susceptible population of Patch 1 ($S_{h,1}$) could be infected by a vector in either Patch 1 ($I_{v,1}$) or Patch 2 by ($I_{v,2}$), depending on which patch they are located in at the time of infection. Thus, the dynamics of the susceptible population in Patch 1 are given by

$$\dot{S}_{h,1} = \mu_h N_{h,1} - \beta_h S_{h,1} \sum_{j=1}^2 a_j p_{1j} \frac{I_{v,j}}{p_{1j} N_{h,1} + p_{2j} N_{h,2}} - \mu_h S_{h,1}. \quad (12.17)$$

The *effective infectious* population in Patch 1 is $p_{11}I_{h,1} + p_{21}I_{h,2}$ and, consequently, the proportion of infectious individuals in Patch 1, is

$$\frac{p_{11}I_{h,1} + p_{21}I_{h,2}}{p_{11}N_{h,1} + p_{21}N_{h,2}}.$$

The dynamics of susceptible mosquitoes in Patch 1 are modeled as follows:

$$\dot{S}_{v,i} = \mu_v (N_{v,i} - q I_{v,i}) - \beta_v S_{v,1} \frac{p_{11}I_{h,1} + p_{21}I_{h,2}}{p_{11}N_{h,1} + p_{21}N_{h,2}} - \mu_v S_{v,1}. \quad (12.18)$$

Question 1 Show that the dynamics of DENV-2, with the host moving between patches, is given by the system

$$\begin{aligned}
 S'_{h,i} &= \mu_h N_{h,i} - \beta_h S_{h,i} \sum_{j=1}^2 a_j p_{ij} \frac{I_{v,j}}{p_{1j} N_{h,1} + p_{2j} N_{h,2}} - \mu_h S_{h,i} \\
 E'_{h,i} &= \beta_h S_{h,i} \sum_{j=1}^2 a_j p_{ij} \frac{I_{v,j}}{p_{1j} N_{h,1} + p_{2j} N_{h,2}} - (\mu_h + \eta_h) E_{h,i} \\
 I'_{h,i} &= \eta_h E_{h,i} - (\mu_h + \gamma_i) I_{h,i} \\
 S'_{v,i} &= q\mu_v (N_{v,i} - qI_{v,i}) - \beta_v S_{v,i} \frac{\sum_{j=1}^2 p_{ji} I_{h,j}}{\sum_{k=1}^2 p_{ki} N_{h,k}} - \mu_v S_{v,i} \\
 E'_{v,i} &= \beta_v S_{v,i} \frac{\sum_{j=1}^2 p_{ji} I_{h,j}}{\sum_{k=1}^2 p_{ki} N_{h,k}} - (\mu_v + \eta_v) E_{v,i} \\
 I'_{v,i} &= \eta_v E_{h,i} + (1-q)\mu_v I_{v,i}, \quad i = 1, 2.
 \end{aligned} \tag{12.19}$$

Question 2 Determine the basic reproduction number of the model (12.19).

Question 3 Obtain a pair of final size relations for the model (12.19).

The results of this project, together with appropriate data, can be used to estimate the spread of dengue from one community to another [5].

12.7 Exercises

- 1.* Use the next generation matrix to calculate the basic reproduction number of the model (12.1).
- 2.* Use the next generation matrix to calculate the basic reproduction number of the model (12.5).
3. If \mathcal{R}_v and \mathcal{R}_d are two non-negative numbers, show that $\frac{1}{2} \left[\mathcal{R}_d + \sqrt{\mathcal{R}_d^2 + 4\mathcal{R}_v} \right]$ is equal to 1 if and only if $\mathcal{R}_d + \mathcal{R}_v = 1$.

References

1. Adams, B. and M. Boots (2010) How important is vertical transmission in mosquitoes for the persistence of dengue? Insights from a mathematical model, *Epidemics* 2: 1–10
2. Adams, B. and D.D. Kapan (2009) Man bites mosquito: understanding the contribution of human movement to vector-borne disease dynamics, *PLoS One* 4(8); e6373. <https://doi.org/10.1371/journal.pone.0006763>.

3. Arunachalam, N., S.C. Tewari, V. Thenmozhi, R. Rajendran, R. Paramasivan, R. Manavalan, K. Ayanar and B.K. Tyagi (2008) Natural vertical transmission of dengue viruses by Aedes aegypti in Chennai, Tamil Nadu, India, Indian J Med Res **127**: 395–397.
4. Balmaseda, A., S.N. Hammond, L. Perez, Y. Tellez, S.I. Saborio, J.C. Mercado, J.C., R. Cuadra, J. Rocha, M.A. Perez, S. Silva et al (2006) Serotype-specific differences in clinical manifestations of dengue, Am. J. of tropical medicine and hygiene **74**: 449.
5. Bichara, D., S.A. Holechek, J. Velasco-Castro, A.L. Murillo and C. Castillo-Chavez (2016) On the dynamics of dengue virus type 2 with residence times and vertical transmission, to appear
6. Bichara, D., Y. Kang, C. Castillo-Chavez, R. Horan and C. Perrings (2015) SIS and SIR epidemic models under virtual dispersal, Bull. Math. Biol. **77** (2015):2004–2034.
7. Bonnet, D.D. and D.J. Worcester (1946) The dispersal of Aedes albopictus in the territory of Hawaii, Am J. of tropical medicine and hygiene, **4**: 465–476.
8. Bosio, C.F., R.E.X.E. Thomas, P.R. Grimstad and K.S. Rai (1992) Variation in the efficiency of vertical transmission of dengue-1 virus by strains of Aedes albopictus (Diptera: Culicidae, J. Medical Entomology **29**: 985–989.
9. Brauer, F., S.M. Towers, A. Mubayi, & C. Castillo-Chavez (2016) Some models for epidemics of vector - transmitted diseases, Infectious Disease Modelling **1**: 79–87.
10. Burke, D.S., A. Nisalak, D.E. Johnson and R.M.N. Scott (1988) A prospective study of dengue infections in Bangkok, Am. J. of Tropical Medicine and Hygiene **38**: 172.,
11. CDC (2010) CDC, Report Suggests Nearly 5 Percent Exposed to Dengue Virus in Key West, Url = <http://www.cdc.gov/media/pressrel/2010/r100713.htm>.
12. Cecilio, A.B. E.S. Campanelli, K.P.R. Souza, L.B. Figueiredo and M.C. Resende (2009) Natural vertical transmission by Stegomyia albopicta as dengue vector in Brazil, Brazilian Journal of Biology **69**: 123–127.
13. Chowell, G., P. Diaz-Duenas, J.C. Miller, A. Alcazar-Velasco, J.M. Hyman, P.W. Fenimore, and C. Castillo-Chavez (2007) Estimation of the reproduction number of dengue fever from spatial epidemic data, Math. Biosc. **208**: 571–589.
14. Cushing, J.M. and O. Diekmann (2016) The many guises of \mathcal{R}_0 (a didactic note), J. Theor. Biol. **404**: 295–302.
15. Deubel, V., R.M. Kinney and D.W. Trent (1988) Nucleotide sequence and deduced amino acid sequence of the nonstructural proteins of dengue type 2 virus, Jamaica genotype: comparative analysis of the full-length genome. Virology **65**: 234–244.
16. Duffy, M.R., T.H. Chen, W.T. Hancock, A.M. Powers, J.L. Kool, R.S. Lanciotti, et al. (2009) Zika virus outbreak on Yap Island, federated states of Micronesia, NEJM **360**:2536–2543.
17. Espinoza, B., Moreno, V., Bichara, D., Castillo-Chavez, C. (2016) Assessing the Efficiency of Movement Restriction as a Control Strategy of Ebola. In Mathematical and Statistical Modeling for Emerging and Re-emerging Infectious Diseases, 2016, 123–145, Springer International Publishing.
18. Esteva, L. and C. Vargas (2000) Influence of vertical and mechanical transmission on the dynamics of dengue disease, Math. Biosc. **167**: 51–64.
19. Gao, D., Y. Lou, D. He, T.C. Porco, Y. Kuang, G. Chowell, S. Ruan (2016) Prevention and control of Zika as a mosquito - borne and sexually transmitted disease: A mathematical modeling analysis, Sci. Rep. **6**, 28070, <https://doi.org/10.1038/srep28070>.
20. Gunther, J., J.P. Martínez-Muñoz, D.G. Pérez-Ishiwara and J. Salas-Benito (2007) Evidence of vertical transmission of dengue virus in two endemic localities in the state of Oaxaca, Mexico, Intervirology **50**: 347–352.
21. Guzman, G, G. Kouri, Gustavo (2003) Dengue and dengue hemorrhagic fever in the Americas: lessons and challenges. J. Clinical Virology **27**: 1–13.
22. Halstead, S.B., S. Nimmannya and S.N. Cohen (1970) Observations related to pathogenesis of dengue hemorrhagic fever. IV. Relation of Yale Journal of Biology and Medicine **42**: 311.
23. Halstead, S. B, N.T.Lan, T.T. Myint, T.N. Shwe, A. Nisalak S. Kalyanarooj, S. Nimmannya, S. Soegijanto, D.W. Vaughn and T.P. Endy (2002) Dengue hemorrhagic fever in infants: research opportunities ignored, Emerging infectious diseases **8**: 1474–1479.

24. Harris, E., E. Videa, L. Perez, E. Sandoval, Y. Tellez, M. Perez, R. Cuadra, J. Rocha, W. Idiaquez, and R.E. Alonso et al (2000) Clinical, epidemiologic, and virologic features of dengue in the 1998 epidemic in Nicaragua, Am. J. of tropical medicine and hygiene **63**: 5.
25. State of Hawaii (2015) Dengue outbreak 2015, Department of Health <http://www.health.hawaii.gov/docd/dengue-outbreak-2015/>.
26. Hawley, W.A., P. Reiter, R.S. Copeland, C.B. Pumpuni and G.B. Craig Jr (1987) Aedes albopictus in North America: probable introduction in used tires from northern Asia, Science **236**: 1114.
27. Knox, T.B., B.H. Kay, R.A. Hall and P.A. Ryan (2003) Enhanced vector competence of Aedes aegypti (Diptera: Culicidae) from the Torres Strait compared with mainland Australia for dengue 2 and 4 viruses, J. Medical Entomology **40**: 950–956.
28. Kucharski, A.J., S. Funk, R.M. Egge, H-P. Mallet, W.J. Edmunds and E.J. Nilles (2016) Transmission dynamics of Zika virus in island populations: a modelling analysis of the 2013–14 French Polynesia outbreak, PLOS Neglected Tropical Diseases DOI 101371
29. Kwok, Y. (2010) Dengue Fever Cases Reach Record Highs, in *Time*, Sept. 24, 2010.
30. Kyle, J.L. and E. Harris (2008) Global spread and persistence of dengue, Ann. Rev. of Microbiology.
31. Lewis, J.A., G.J. Chang, R.S. Lanciotti, R.M. Kinney, L.W. Mayer and D.W. Trent (1993) Phylogenetic relationships of dengue-2 viruses, Virology **197**: 216–224.
32. Montoya, Y.S. Holechek, O. Cáceres, A. Palacios, J. Burans, C. Guevara, F. Quintana, V. Herrera, E. Pozo, E. Anaya, et al (2003) Circulation of dengue viruses in North-Western Peru, 2000–2001, Dengue Bulletin **27**
33. Morens, D.M. and A.S. Fauci (2008) Dengue and hemorrhagic fever: A potential threat to public health in the United States JAMA **299**: 214.
34. Murillo, D., S.A. Holechek, A.L. Murillo, F. Sanchez, and C. Castillo-Chavez (2014) Vertical Transmission in a Two-Strain Model of Dengue Fever, Letters in Biomathematics **1**: 249–271.
35. Niebylski, M.L. and G.B. Craig Jr (1994) Dispersal and survival of Aedes albopictus at a scrap tire yard in Missouri, Journal of the American Mosquito Control Association, **10**: 339–343.
36. Nishiura, H. (2006) Mathematical and statistical analyses of the spread of dengue, Dengue Bulletin **30**: 51.
37. Reiter, P. and D.J. Gubler (1997) Surveillance and control of urban dengue vectors, Dengue and dengue hemorrhagic fever : 45–60.
38. Rico-Hesse, R., L.M. Harrison, R.A. Salas, D. Tovar, A. Nisalak, C. Ramos, J. Boshell, M.T.R. de Mesa, H.M.R. Nogueira and A.T. Rosa (1997) Origins of dengue-type viruses associated with increased pathogenicity in the Americas, Virology **230**: 344–251.
39. Rico-Hesse, R., L.M. Harrison, A. Nisalak, D.W. Vaughn, S. Kalayanarooj, S. Green, A.L. Rothman and F.A. Ennis (1998) Molecular evolution of dengue type 2 virus in Thailand, Am.J. of Tropical Medicine and Hygiene **58**: 96.
40. Rosen, L., D.A. Shroyer, R.B. Tesh, J.E. Freier and J.C. Lien (1983) Transovarial transmission of dengue viruses by mosquitoes: Aedes albopictus and Aedes aegypti Am. J. tropical medicine and hygiene **32**: 1108–1119.
41. Sittisombut, N., A. Sistyanarain, M.J. Cardosa, M. Salminen, S. Damrongdachakul, S. Kalayanarooj, S. Rojanasuphot, J. Supawadee and N. Maneekarn (1997) Possible occurrence of a genetic bottleneck in dengue serotype 2 viruses between the 1980 and 1987 epidemic seasons in Bangkok, Thailand, Am. J. of Tropical Medicine and Hygiene **57**: 100.
42. Tewari, S.C., V. Thenmozhi, C.R. Katholi, R. Manavalan, A. Munirathinam and A. Gajana (2004) Dengue vector prevalence and virus infection in a rural area in south India, Tropical Medicine and International Health **9**: 499–507.
43. Towers, S., F. Brauer, C. Castillo-Chavez, A.K.I. Falconer, A. Mubayi, C.M.E. Romero-Vivas (2016) Estimation of the reproduction number of the 2015 Zika virus outbreak in Barranquilla, Columbia, Epidemics **17**: 50–55.
44. Valega-Mackenzie, W. and K. Rios-Soto (2018) Can vaccination save a Zika virus epidemic?, Bull.Math. Biol. **80**:598–625.

45. van den Driessche, P. and J. Watmough (2002) Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission, *Math. Biosc.* **180**:29–48.
46. WHO (2009) Dengue and dengue hemorrhagic fever. Fact Sheet
47. World Health Organization (2015) Dengue Control, <http://www.who.int/denguecontrol/mosquito/en/>.
48. Zhang, C., M.P. Mammen Jr, P. Chinnawirotpisan, C. Klungthong, P. Rodpradit, A. Nisalak, D.W. Vaughn, S. Nimmannitya, S. Kalayanarooj and E.C. Holmes, E.C. (2006) Structure and age of genetic diversity of dengue virus type 2 in Thailand, *Journal of General Virology* **87**: 873.

Part III

More Advanced Concepts

Chapter 13

Disease Transmission Models with Age Structure



13.1 Introduction

Age is one of the most important characteristics in the modeling of populations and infectious diseases. Because age groups frequently mix heterogeneously it may be appropriate to include age structure in epidemiological models. While there are other aspects of heterogeneity in disease transmission models, such as behavioral and spatial heterogeneity, age structure is one of the most important aspects of heterogeneity in disease modeling. As references for the mathematical details, we suggest [1, Chapters 9 and 13], [21, Part III], [23, Chapter 19], [28, Section 10.9], [32, Chapter 22]. Some applications may be found in [2, 4, 9, 13, 15, 21, 24, 25, 27, 34, 35].

Childhood diseases, such as measles, chicken pox, and rubella, are spread mainly by contacts between children of similar ages. More than half of the deaths attributed to malaria are in children under 5 years of age due to their weaker immune systems. This suggests that in models for disease transmission in an age structured population it may be advisable to allow the contact rates between two members of the population to depend on the ages of both members.

Sexually transmitted diseases (STDs) are spread through partner interactions with pair-formations, and the pair formations are age-dependent in most cases. For example, most AIDS cases occur in the group of young adults. More instances of age dependence may be found in [1].

A description of age structured disease transmission requires an introduction to age structured population models. In fact, the first models for age structured populations [26] were designed for the study of disease transmission in such populations. Therefore, we begin with a brief description of some aspects of age structured population models.

In this chapter, we discuss some structured models that have been used in ecology and epidemiology. Some topics are included in projects. References to the literature indicate where more detailed discussions can be found. We have chosen applications that do not demand a sophisticated mathematical background and that yet give a feeling for some active areas of research.

13.2 Linear Age Structured Models

It was a physician, Lieut. Col. A.G. McKendrick (1926), who first introduced age into the description of the dynamics of a one-sex population. The linear theory is based on the McKendrick equation [26], which is most often referred to as the Von Foerster (1959) equation since McKendrick's earlier work was not well known. Fortunately, the contributions of McKendrick have been made available widely to the scientific community [21]. The McKendrick model assumes that the female population can be described by a density function $\rho(a, t)$ of chronological age a and time t , both to be measured on the same scale. Then, $N(a, t) = \int_0^a \rho(\sigma, t) d\sigma$ is the number of individuals of ages in the interval $[0, a]$ at time t , and $P(t) = \int_0^\infty \rho(\sigma, t) d\sigma$ is the total population size at time t .

Let $\beta(a) \geq 0$ denote the per-capita age specific fertility rate. Then the total number of births at time t is

$$B(t) = \int_0^\infty \beta(a) \rho(a, t) da, \quad (13.1)$$

and the total number of births in the interval $[t, t + h](h > 0)$ is given by

$$\int_t^{t+h} \int_0^\infty \beta(a) \rho(a, t) da ds = \int_t^{t+h} B(s) ds.$$

Let $\mu(a) \geq 0$ denote the per-capita age specific death rate. Then the probability that an individual will survive from birth to age a is

$$\pi(a) = e^{-\int_0^a \mu(\alpha) d\alpha}.$$

Note that, for each $s \in [t, t + h]$, the total number of individuals of age less than or equal to $a+s$ who die at time $t+s$ is $\int_0^{a+s} \mu(\sigma) \rho(\sigma, t+s) d\sigma$. Thus, the total number of individuals of age less than or equal to $a+u-t$ who die at time $u \in [t, t+h]$ is given by (letting $u = s+t$):

$$\int_t^{t+h} \int_0^{a+u-t} \mu(\sigma) \rho(\sigma, u) d\sigma du = \int_0^h \int_0^{a+s} \mu(\sigma) \rho(\sigma, t+s) d\sigma ds.$$

Assume also that the only changes in the population size from time t to $t + h$ are due to births and deaths. Then

$$N(a + h, t + h) - N(a, t) = \int_t^{t+h} B(s)ds - \int_0^h \int_0^{a+s} \mu(\sigma)\rho(\sigma, t + s)d\sigma ds. \quad (13.2)$$

The next step is to divide (13.2) by h and let $h \rightarrow 0$. In this process, we will need to make use of differentiation of an integral. The basic rule for this is that if

$$F(a, x) = \int_0^a f(\xi, x)d\xi,$$

then the partial derivatives of F are given by

$$F_a(a, x) = f(a, x), \quad F_x(a, x) = \int_0^a f_x(\xi, x)d\xi.$$

Since

$$N(a + h, t + h) - N(a, t) = [N(a + h, t + h) - N(a, t + h)] + [N(a, t + h) - N(a, t)],$$

the limit as $h \rightarrow 0$ of the left side of (13.2) divided by h is

$$N_a(a, t) + N_t(a, t) = \rho(a, t) + \int_0^a \rho_t(\alpha, t)d\alpha.$$

Thus, we obtain the following from the left-hand side:

$$\begin{aligned} & \lim_{h \rightarrow 0} \frac{N(a + h, t + h) - N(a, t + h) + N(a, t + h) - N(a, t)}{h} \\ &= N_t(a, t) + N_a(a, t) = \int_0^h \rho_t(\sigma, t)d\sigma + \rho(a, t). \end{aligned} \quad (13.3)$$

The limit as $h \rightarrow 0$ of the right-hand side of (13.2) divided by h , again using differentiation of the integral, leads to

$$B(t) - \int_0^a \mu(\sigma)\rho(\sigma, t)d\sigma. \quad (13.4)$$

Combining (13.3)–(13.4), we obtain

$$\int_0^a \rho_t(\sigma, t)d\sigma + \rho(a, t) = B(t) - \int_0^a \mu(\sigma)\rho(\sigma, t)d\sigma. \quad (13.5)$$

Differentiation of (13.5) with respect to a yields

$$\rho_t(a, t) + \rho_a(a, t) = -\mu(a)\rho(a, t). \quad (13.6)$$

Assume that the initial age distribution as $\rho(a, 0) = \rho_0(a)$ and that the gender ratio is one to one. Note that $\rho(0, t)$ also represents the total number of births at time t , i.e.,

$$\rho(0, t) = B(t).$$

Then, we arrive at the following partial differential equation with initial and boundary conditions:

$$\begin{aligned} \rho_t(a, t) + \rho_a(a, t) &= -\mu(a)\rho(a, t), \\ \rho(0, t) &= \int_0^\infty \beta(a)\rho(a, t)da = B(t), \\ \rho(a, 0) &= \rho_0(a). \end{aligned} \quad (13.7)$$

13.3 The Method of Characteristics

System (13.7) can be solved by the method of characteristics. That is, we can solve for $\rho(a, t)$ along the characteristics lines $a = t + c$, where c is a constant. Let ω_c denote the cohort function tracking through time of those individuals whose age is c at $t = 0$, which is defined as:

$$\omega_c(t) = \rho(t + c, t), \quad \text{for } t > t_c$$

where $t_c = \max(0, -c)$. Note that

$$\frac{d\omega_c}{dt} = \frac{\partial\rho}{\partial a} \frac{da}{dt} + \frac{\partial\rho}{\partial t} \frac{dt}{dt} = \frac{\partial\rho}{\partial a} + \frac{\partial\rho}{\partial t},$$

thus, (13.7) can be rewritten as

$$\frac{d\omega_c(t)}{dt} = -\mu(t + c)\omega_c(t), \quad t \geq t_c,$$

whose solution is

$$\omega_c(t) = \omega_c(t_c) \exp \left[- \int_{t_c}^t \mu(s + c)ds \right], \quad \text{for } t > t_c. \quad (13.8)$$

Therefore, the solution to (13.7) can be written as:

$$\rho(a, t) = \begin{cases} \rho_0(a - t) \exp\left[-\int_0^t \mu(s + a - t)ds\right], & a \geq t \\ B(t - a) \exp\left[-\int_0^a \mu(s)ds\right], & a < t. \end{cases} \quad (13.9)$$

Note that $B(t)$ is given in (13.1), which depends on $\rho(a, t)$.

13.4 Equivalent Formulation as an Integral Equation Model

Substituting (13.9) into (13.1) yields

$$B(t) = \int_0^t \beta(a) B(t - a) e^{-\int_0^a \mu(s)ds} da + \int_t^\infty \rho_0(a - t) e^{-\int_0^t \mu(a - t + s)ds} da. \quad (13.10)$$

Let

$$F(t) = \int_t^\infty \rho_0(a - t) e^{-\int_0^t \mu(a - t + s)ds} da, \quad K(a) = \beta(a) e^{-\int_0^a \mu(s)ds} = \beta(a) \pi(a).$$

Then, substituting the expressions for F and K into (13.10) and making a change of variable $a = t - u$, we obtain the *renewal equation*:

$$B(t) = F(t) + \int_0^t K(t - u) B(u) du, \quad (13.11)$$

which was initially derived by Sharpe and Lotka [31].

Using the notation of convolution, we can rewrite Eq. (13.11) in the form

$$B(t) = F(t) + (K * B)(t). \quad (13.12)$$

Let $\hat{f}(p) = \int_0^\infty e^{-ps} f(s) ds$ denote the Laplace transform of the function $f(t)$. Then, from Eq. (13.12) we have

$$\hat{B}(p) = \hat{F}(p) + \hat{B}(p) \hat{K}(p).$$

Solving for $\hat{B}(p)$ we get

$$\hat{B}(p) = \frac{\hat{F}(p)}{1 - \hat{K}(p)}.$$

Because $\hat{K}(p)$ and $\hat{F}(p)$ are analytic functions, the properties of $\hat{B}(p)$ are determined by the roots of $1 - \hat{K}(p)$, or solutions of

$$\hat{K}(p) = \int_0^\infty \beta(a)\pi(a)e^{-pa}da = 1 \quad (13.13)$$

for $p \in \mathbb{C}$. Equation (13.13) is called the *Lotka characteristic equation* for the renewal equation (13.11).

For $p \in \mathbb{R}$, from $\frac{d\hat{K}}{dp} < 0$ we know that $\hat{K}(p)$ decreases monotonically. Note that, since $\hat{K}(p) \rightarrow 0$ as $p \rightarrow \infty$, Eq.(13.13) has a unique solution, which we denote by p^* , with the property that

$$\begin{aligned} p^* &> 0 \text{ if } \hat{K}(0) > 1 \\ p^* &< 0 \text{ if } \hat{K}(0) < 1. \end{aligned}$$

For $p = u + iv \in \mathbb{C}$, note that $\hat{K}(p) = \int_0^\infty \beta(a)\pi(a)e^{-(u+iv)a}da = 1$. Thus,

$$\int_0^\infty \beta(a)\pi(a)e^{-ua}\cos(va)da = 1, \quad \int_0^\infty \beta(a)\pi(a)e^{-ua}\sin(va)da = 0.$$

From $|\cos(va)| \leq 1$ and $\hat{K}'(u) < 0$ for all $u \in \mathbb{R}$, we know that $u \leq p^*$. Therefore, p^* is the dominant root (greatest modulus) of $\hat{K}(p)$. That is, all solutions of $\hat{K}(p) = 1$ have real part less than or equal to p^* .

It can be shown that

$$B(t) = B_0 e^{p^* t} (1 + \mathcal{Q}(t)), \quad (13.14)$$

where $B_0 > 0$ and $\lim_{t \rightarrow \infty} \mathcal{Q}(t) = 0$ (see, e.g., [22]).

Therefore, the sign of p^* determines the asymptotic behavior of the solution $B(t)$, and thus of $\rho(a, t)$. From (13.9) and (13.14), we know that $\rho(a, t)$ has the following asymptotic property:

$$\lim_{t \rightarrow \infty} \rho(a, t) = \lim_{t \rightarrow \infty} B_0 e^{p^*(t-a)} \pi(a). \quad (13.15)$$

Define

$$\mathcal{R} = \hat{K}(0) = \int_0^\infty \beta(a)\pi(a)da,$$

which is the expected number of offspring for each individual during a lifetime. The results above show that $B(t) \rightarrow 0 (\infty)$ as $t \rightarrow \infty$ if and only if $\mathcal{R} < 1 (> 1)$.

Recall that $P(t) = \int_0^\infty \rho(a, t)da$ denotes the total number of individuals at time t . A *stable age distribution* (or *persistent age distribution*) is defined to be a solution $\rho(a, t)$ for which $\rho(a, t)/P(t)$ is independent of time t . It follows from (13.15) that

asymptotically every solution given in (13.9) tends to a stable age distribution as $t \rightarrow \infty$.

13.5 Equilibria and the Characteristic Equation

For system (13.7), a steady-state, or an equilibrium distribution, $\rho^*(a)$, is a time-independent solution. That is, $\rho^*(a)$ satisfies the equations:

$$\begin{aligned} \frac{d\rho^*(a)}{da} &= -\mu(a)\rho^*(a), \\ \rho^*(0) &= \int_0^\infty \beta(a)\rho^*(a)da. \end{aligned} \quad (13.16)$$

For the linear system (13.16), the only equilibrium is $\rho^*(a) = 0$, and the linearization at this equilibrium is the same as (13.7). For the stability of $\rho^*(a) = 0$, consider the separable solution of the form

$$\rho(a, t) = A(a)e^{pt} \quad (13.17)$$

with $A(a) > 0$. The stability of $\rho^*(a)$ is determined by the sign of p . From the second equation in (13.7) we have

$$A(0) = \int_0^\infty \beta(a)A(a)da =: B_1. \quad (13.18)$$

Substitution of (13.17) into the first equation in (13.7) yields

$$\frac{dA(a)}{da} = -(p + \mu(a))A(a).$$

Solving for $A(a)$ and using $A(0) = B_1$ we have:

$$A(a) = B_1\pi(a)e^{-pa}, \quad (13.19)$$

where $\pi(a) = e^{-\int_0^a \mu(s)ds}$. Using the definition of B_1 in (13.18) with $A(a)$ replaced by the expression in (13.19), we obtain:

$$B_1 = \int_0^\infty \beta(a)A(a)da = \int_0^\infty B_1\beta(a)\pi(a)e^{-pa}da.$$

Since $B_1 \neq 0$, dividing both sides of the above equation by B_1 we obtain the following equation which p must satisfy:

$$\int_0^\infty \beta(a)\pi(a)e^{-pa}da = 1. \quad (13.20)$$

System (13.20) is the same characteristic equation as (13.13).

Note that the left-hand side of (13.20) is the same as the function $\hat{K}(p)$ defined in (13.13). Thus, $p > 0 (=, < 0)$ if and only if $\mathcal{R} = \hat{K}(0) > 1 (=, < 1)$. Therefore, the steady-state $\rho^*(a) = 0$ is locally asymptotically stable if and only if $\mathcal{R} < 1$.

13.6 The Demographic Model with Discrete Age Groups

Partition the age interval into a finite number n of subintervals $[a_0, a_1], [a_1, a_2], \dots, [a_{n-1}, a_n]$, where $a_0 = 0$ and $a_n \leq \infty$. Denote the number of individuals with ages in interval $[a_{i-1}, a_i]$ by $H_i(t)$, so that $H_i(t) = \int_{a_{i-1}}^{a_i} \rho(t, a)da$, $i = 1, \dots, n$. Then integrating the partial differential equation in (13.7) from a_0 to a_1 , we have

$$\frac{dH_1(t)}{dt} + \rho(t, a_1) - \rho(t, a_0) + \int_{a_0}^{a_1} \mu(a, P)\rho(t, a)da = 0. \quad (13.21)$$

Assume that individuals with ages in each interval have the same vital rates such that $\beta(a, P) = \beta_i$, $\mu(a, P) = \mu_i$, for a in $[a_{i-1}, a_i]$, $i = 1, \dots, n$. Here β_i and μ_i are age-independent, but may be density-dependent. Then, in the age interval $[0, a_1]$, we have

$$\rho(t, 0) = \sum_1^n \beta_i H_i(t), \quad \int_{a_0}^{a_1} \mu \rho(t, a)da = \mu_1 H_1(t),$$

which leads to

$$\frac{dH_1}{dt} = \sum_1^n \beta_i H_i - (m_1 + \mu_1)H_1. \quad (13.22)$$

Here m_1 is the progression rate from groups 1 to 2, defined by $m_1 = \rho(t, a_1)/H_1(t)$, and we assume it is time-independent.

Integrating the partial differential equation in (13.24) from a_{i-1} to a_i for $2 \leq i \leq n$, we have

$$\frac{dH_i}{dt} = m_{i-1} H_{i-1} - (m_i + \mu_i) H_i, \quad i = 2, \dots, n, \quad (13.23)$$

where m_i is the age progression rate from groups i to $i+1$, and we let $m_n = 0$. Then the system in (13.7) is reduced to a system of n ordinary differential equations.

13.7 Nonlinear Age Structured Models

In the models considered in previous sections, the birth and death rates were linear, and this implies that total population size either grows exponentially (if $q > 0$), dies out exponentially (if $q < 0$), or remains constant (if $q = 0$). In studying models without age structure, we considered situations in which populations have a carrying capacity that is approached as $t \rightarrow \infty$. In order to allow this possibility for age-structured models we must assume some non-linearity. We now consider the possibility of birth and death moduli of the form $\beta(a, P(t))$ and $\mu(a, P(t))$, depending on total population size. A variant, which can be developed by analogous methods, would be to allow the birth modulus (and possibly also the death modulus) to depend on the density $\rho(a, t)$. We will consider only the continuous case because the methods of linear algebra methods used to treat the linear discrete case have no direct adaptation to the nonlinear discrete model.

If the birth and death moduli are allowed to depend on total population size, the description of the model must include $P(t)$. Thus, the full model is now

$$\begin{aligned} \rho_a(a, t) + \rho_t(a, t) + \mu(a, P(t))\rho(a, t) &= 0 \\ B(t) &= \rho(0, t) = \int_0^\infty \beta(a, P(t))\rho(a, t)da \\ P(t) &= \int_0^\infty \rho(a, t)da \\ \rho(a, 0) &= \phi(a). \end{aligned} \tag{13.24}$$

We can transform the problem just as in the linear case by integrating along characteristics. Using the notation

$$\mu^*(\alpha) = \mu(\alpha, P(t - a + \alpha)),$$

the same calculations lead to

$$\rho(a, t) = \begin{cases} B(t - a)e^{-\int_0^a \mu^*(\alpha)d\alpha} & \text{for } t \geq a \\ \phi(a - t)e^{-\int_{a-t}^a \mu^*(\alpha)d\alpha} & \text{for } t < a. \end{cases} \tag{13.25}$$

Substituting these expressions into (13.24), we obtain a pair of functional equations for $B(t)$ and $P(t)$, whose solution gives an explicit solution for $\rho(a, t)$, namely

$$\begin{aligned} B(t) &= b(t) + \int_0^t \beta(a, P(t))e^{-\int_0^a \mu^*(\alpha)d\alpha} B(t - a)da, \\ P(t) &= p(t) + \int_0^t e^{-\int_0^a \mu^*(\alpha)d\alpha} B(t - a)da, \end{aligned}$$

where

$$\begin{aligned} b(t) &= \int_t^\infty \beta(a, P(t))\phi(a-t)e^{-\int_{a-t}^a \mu^*(\alpha)d\alpha} da, \\ p(t) &= \int_t^\infty \phi(a-t)e^{-\int_{a-t}^a \mu^*(\alpha)d\alpha} da. \end{aligned}$$

It is reasonable to assume that $\int_0^\infty \phi(a)da < \infty$ and that the functions $\beta(a, P)$ and $\mu(a, P)$ are continuous and non-negative. Under these hypotheses it is easy to verify that $b(t)$ and $p(t)$ are continuous and non-negative and tend to zero as $t \rightarrow \infty$ with $b(0) > 0$ and $p(0) > 0$. It is possible to prove that a unique continuous non-negative solution exists on $0 \leq t < \infty$ if $\sup_{a \geq 0, P \geq 0} \beta(a, P) < \infty$. This model is due to Gurtin and MacCamy [18].

If $\rho(a, t) = \rho(a)$ is an equilibrium age distribution, then

$$P(t) = \int_0^\infty \rho(a)da =: P \quad \text{and} \quad B(t) = \int_0^\infty \beta(a, P)\rho(a)da =: B$$

are constant. In this case, the McKendrick equation becomes an ordinary differential equation

$$\rho'(a) + \mu(a, P)\rho(a) = 0, \quad \rho(0) = B,$$

whose solution is

$$\rho(a) = Be^{-\int_0^a \mu(\alpha, P)d\alpha}.$$

Let

$$\pi(a, P) = e^{-\int_0^a \mu(\alpha, P)d\alpha} \tag{13.26}$$

denote the probability of survival from birth to age a when the population size is the constant P . Then

$$\rho(a) = B\pi(a, P). \tag{13.27}$$

From $B = \int_0^\infty \beta(a, P)\rho(a)da$ and (13.27), we obtain

$$B = B \int_0^\infty \beta(a, P)\pi(a, P)da.$$

Thus, for an equilibrium age distribution with birth rate B , the population size P must satisfy the equation

$$\mathcal{R}(P) = \int_0^\infty \beta(a, P)\pi(a, P)da = 1, \tag{13.28}$$

in which case the birth rate is given by

$$B = \frac{P}{\int_0^\infty \pi(a, P)da}.$$

The quantity $\mathcal{R}(P)$ defined in (13.28) is called the reproduction number, which represents the expected number of offspring that an individual has over a lifetime when the total population size is P .

Example 1 Suppose that β is independent of age and is a function of P only. Then

$$\begin{aligned} B(t) &= \int_0^\infty \beta(P(t))\rho(a, t)da = \beta(P(t)) \int_0^\infty \rho(a, t)da \\ &= P(t)\beta(P(t)). \end{aligned} \quad (13.29)$$

If we define $g(P) = P\beta(P)$, then the problem is reduced to a single functional equation for $P(t)$:

$$P(t) = p(t) + \int_0^t e^{-\int_0^a \mu^*(\alpha)d\alpha} g(P(t-a))da$$

together with this explicit formula for $B(t)$ given in (13.29).

13.8 Age-Structured Epidemic Models

Consider an age-structured *SIR* model. Let $s(t, a)$, $i(t, a)$, $r(t, a)$ represent the age densities of susceptible, infective, and removed members, respectively, so that the total numbers of individuals in the respective classes are

$$S(t) = \int_0^\infty s(a, t)da, \quad I(t) = \int_0^\infty i(a, t)da, \quad R(t) = \int_0^\infty r(a, t)da.$$

The age density of the total population is

$$\rho(a, t) = s(a, t) + i(a, t) + r(a, t),$$

and the total population size is

$$P(t) = S(t) + I(t) + R(t) = \int_0^\infty \rho(a, t)da.$$

Then, the model is described by the following system of equations:

$$\begin{aligned} s_t(a, t) + s_a(a, t) &= -(\mu(a) + \lambda(a, t))s(a, t) \\ i_t(a, t) + i_a(a, t) &= \lambda(a, t)s(a, t) - (\mu(a) + \gamma(a))i(a, t), \\ r_t(a, t) + r_a(a, t) &= -\mu(a)r(a, t) + \gamma(a)i(a, t), \end{aligned} \quad (13.30)$$

where $\lambda(a, t)$ is the per-capita infection rate, which may have various forms depending on the assumption about mixing. One possible form focuses on *intracohort* mixing:

$$\lambda(a, t) = c(a)i(a, t),$$

corresponding to the assumption that infection can be transmitted only between individuals of the same age. Another possible form is the *intercohort* mixing :

$$\lambda(a, t) = \int_0^\infty b(a, \alpha)i(\alpha, t)d\alpha,$$

where $b(a, \alpha)$ denotes the rate of infection from contacts between an infective of age α and a susceptible of age a . For intercohort mixing, it is common to assume that the mixing function has a separable form:

$$b(a, \alpha) = b_1(a)b_2(\alpha).$$

Other parameter functions in (13.30) are $\mu(a)$ and $\gamma(a)$, which represent the age-dependent per-capita natural death and recovery rates. The initial conditions are

$$s(a, 0) = s_0(a), \quad i(a, 0) = i_0(a), \quad r(a, 0) = 0, \quad (13.31)$$

where $s_0(a)$ and $i_0(a)$ are the initial distributions of susceptibles and infectives, respectively. The boundary conditions involve the birth or renewal condition. Under the assumption that all newborns are susceptible, the total number of newborns is

$$s(0, t) = \int_0^\infty \beta(a, P(t))\rho(a, t)da, \quad (13.32)$$

with $i(0, t) = r(0, t) = 0$.

13.8.1 *Age-Dependent Vaccination in Epidemic Models

In many cases, age-structured models are more appropriate to consider than models without an age structure. For example, if transmission is highly age-dependent due to a higher activity in certain age groups, it might be useful to identify the age groups to target for more effective disease control strategies such as vaccination. The model and results presented below are taken from [10]. It concerns age-dependent vaccination strategies for TB.

Divide the population into susceptible, vaccinated, latent, infective, and treated classes, where $s(t, a)$, $v(t, a)$, $l(t, a)$, $i(t, a)$, and $j(t, a)$ denote the associated density functions with these respective epidemiological age-structure classes, and

$n(t, a) = s(t, a) + v(t, a) + l(t, a) + i(t, a) + j(t, a)$. Assume that all newborns are susceptible and that the mixing between individuals is proportional to their age-dependent activity level. Assume also that an individual may become infected only through contact with infectious individuals, that vaccination is partially effective (i.e., vaccinated individuals can become infected again but with a reduced susceptibility), that only susceptibles will be vaccinated. The dynamics are governed by the following initial boundary value problem:

$$\begin{aligned} \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) s(t, a) &= -\beta(a)c(a)B(t)s(t, a) - \mu(a)s(t, a) - \psi(a)s(t, a), \\ \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) v(t, a) &= \psi(a)s(t, a) - \mu(a)v(t, a) - \delta\beta(a)c(a)B(t)v(t, a), \\ \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) l(t, a) &= \beta(a)c(a)B(t)(s(t, a) + \sigma j(t, a) + \delta v(t, a)) \\ &\quad - (k + \mu(a))l(t, a), \\ \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) i(t, a) &= kl(t, a) - (r + \mu(a))i(t, a), \\ \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) j(t, a) &= ri(t, a) - \sigma\beta(a)c(a)B(t)j(t, a) - \mu(a)j(t, a), \end{aligned} \tag{13.33}$$

where

$$B(t) = \int_0^\infty \frac{i(t, a')}{n(t, a')} p(t, a, a') da',$$

with initial and boundary conditions:

$$\begin{aligned} s(t, 0) &= \Lambda, \quad v(t, 0) = l(t, 0) = i(t, 0) = j(t, 0) = 0, \\ s(0, a) &= s_0(a), \quad v(0, a) = v_0(a), \quad l(0, a) = l_0(a), \quad i(0, a) = i_0(a), \\ j(0, a) &= j_0(a). \end{aligned}$$

The function p represents the mixing between susceptible individuals of age a and infectious individual of age a' , and is assumed to have the form of proportionate mixing:

$$p(t, a, a') = \frac{c(a')n(t, a')}{\int_0^\infty c(u)n(t, u)du}.$$

Λ is the recruitment/birth rate (assumed constant); $\beta(a)$ is the age-specific (average) probability of becoming infected through contact with infectious individuals; $c(a)$ is the age-specific per-capita contact/activity rate; $\mu(a)$ is the age-specific per-capita natural death rate (all of these functions are assumed to be continuous and to be

zero beyond some maximum age); k is the per-capita rate at which individuals leave the latent class by becoming infectious; r is the per-capita treatment rate; σ and δ are the reductions in risk due to prior exposure to TB and vaccination, respectively, $0 \leq \sigma \leq 1$, $0 \leq \delta \leq 1$; and $p(t, a, a')$ gives the probability that an individual of age a has contact with an individual of age a' given that it has a contact with a member of the population. Proportionate mixing has also been used in other studies including [3, 8, 11, 16, 20].

Using the approach of [11], we assume that $p(t, a, a') = p(t, a')$ as explicitly described above. The initial age distributions are assumed to be known and to be zero beyond some maximum age. The model (13.33) is well-posed and the proof is similar to that found in [8].

Notice that $n(t, a)$ satisfies the following equations:

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) n(t, a) = -\mu(a)n(t, a), \\ n(t, 0) = \Lambda, \quad n(0, a) = n_0(a) = s_0(a) + v_0(a) + l_0(a) + i_0(a) + j_0(a).$$

Using the method of characteristic curves we can solve for n explicitly:

$$n(t, a) = n_0(a) \frac{\mathcal{F}(a)}{\mathcal{F}(a-t)} H(a-t) + \Lambda \mathcal{F}(a) H(t-a),$$

where

$$\mathcal{F}(a) = e^{-\int_0^a \mu(s) ds}, \\ H(s) = 1, \quad s \geq 0; \quad H(s) = 0, \quad s < 0.$$

Hence,

$$n(t, a) \rightarrow \Lambda \mathcal{F}(a), \\ p(t, a) \rightarrow \frac{c(a) \mathcal{F}(a)}{\int_0^\infty c(b) \mathcal{F}(b) db} =: p_\infty(a), \quad t \rightarrow \infty. \quad (13.34)$$

Introducing the fractions

$$u(t, a) = \frac{s(t, a)}{n(t, a)}, \quad w(t, a) = \frac{v(t, a)}{n(t, a)}, \quad x(t, a) = \frac{l(t, a)}{n(t, a)}, \quad y(t, a) = \frac{i(t, a)}{n(t, a)}, \\ z(t, a) = \frac{j(t, a)}{n(t, a)},$$

we get a simplified system of (13.33):

$$\begin{aligned}
\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) u(t, a) &= -\beta(a)c(a)B(t)u(t, a) - \psi(a)u(t, a), \\
\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) w(t, a) &= \psi(a)u(t, a) - \delta\beta(a)c(a)B(t)w(t, a), \\
\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) x(t, a) &= \beta(a)c(a)B(t)(u(t, a) + \delta w(t, a) + \sigma z(t, a)) - kx(t, a), \\
\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) y(t, a) &= kx(t, a) - ry(t, a), \\
\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) z(t, a) &= ry(t, a) - \sigma\beta(a)c(a)B(t)z(t, a), \\
B(t) &= \int_0^\infty y(t, a)p(t, a)da, \\
p(t, a) &= \frac{c(a)n(t, a)}{\int_0^\infty c(u)n(t, u)du}, \\
u(t, 0) = 1, \quad w(t, 0) = x(t, 0) = y(t, 0) = z(t, 0) &= 0.
\end{aligned} \tag{13.35}$$

Let $\mathcal{F}_\psi(a)$ denote the probability that a susceptible individual has not been vaccinated at age a . Then

$$\mathcal{F}_\psi(a) = e^{-\int_0^a \psi(b)db}.$$

The system (13.35) has the infection-free steady state

$$\begin{aligned}
u(a) &= \mathcal{F}_\psi(a), \quad w(a) = 1 - \mathcal{F}_\psi(a), \\
x(a) = y(a) = z(a) &= 0, \quad n(a) = \Lambda\mathcal{F}(a).
\end{aligned} \tag{13.36}$$

To study the local stability of the infection-free equilibrium, we linearize (13.35) about (13.36) and consider exponential solutions of the form

$$x(t, a) = X(a)e^{\lambda t}, \quad y(t, a) = Y(a)e^{\lambda t}, \quad B(t) = B_0e^{\lambda t} + O(e^{2\lambda t}),$$

where

$$B_0 = \int_0^\infty Y(a)p_\infty(a)da \tag{13.37}$$

is a constant and $p_\infty(a)$ is as in (13.34). Then the linear parts of the x and y equations in (13.35) are of the form

$$\begin{aligned}
\lambda X(a) + \frac{d}{da}X(a) &= \beta(a)c(a)B_0\mathcal{F}_\psi(a) - kX(a), \\
\lambda Y(a) + \frac{d}{da}Y(a) &= kX(a) - rY(a),
\end{aligned}$$

where

$$\mathcal{V}_\psi(a) = \mathcal{F}_\psi(a) + \delta(1 - \mathcal{F}_\psi(a)). \quad (13.38)$$

An expression for $Y(a)$ can be obtained by solving the above system:

$$Y(a) = B_0 \int_0^a \frac{k}{r-k} \beta(\alpha) c(\alpha) (e^{(\lambda+k)(\alpha-a)} - e^{(\lambda+r)(\alpha-a)}) \mathcal{V}_\psi(\alpha) d\alpha. \quad (13.39)$$

From (13.37) and (13.39) we get

$$B_0 = B_0 \int_0^\infty \int_0^a \frac{k}{r-k} p_\infty(a) \beta(\alpha) c(\alpha) (e^{(\lambda+k)(\alpha-a)} - e^{(\lambda+r)(\alpha-a)}) \mathcal{V}_\psi(\alpha) d\alpha da. \quad (13.40)$$

By changing the order of integration, introducing $\tau = a - \alpha$, and dividing both sides by B_0 (because $B_0 \neq 0$) in (13.40) we get the characteristic equation

$$1 = \int_0^\infty \int_0^\infty \frac{k}{r-k} p_\infty(\alpha + \tau) \beta(\alpha) c(\alpha) (e^{-(\lambda+k)\tau} - e^{-(\lambda+r)\tau}) \mathcal{V}_\psi(\alpha) d\tau d\alpha \\ := G(\lambda). \quad (13.41)$$

Now we are ready to define the net reproduction number as $\mathcal{R}(\psi) = G(0)$; i.e.,

$$\mathcal{R}(\psi) = \int_0^\infty \int_0^\infty \frac{k}{r-k} p_\infty(\alpha + \tau) \beta(\alpha) c(\alpha) (e^{-k\tau} - e^{-r\tau}) \mathcal{V}_\psi(\alpha) d\tau d\alpha. \quad (13.42)$$

Noticing that

$$G'(\lambda) < 0, \lim_{\lambda \rightarrow \infty} G(\lambda) = 0, \lim_{\lambda \rightarrow -\infty} G(\lambda) = \infty,$$

we know that (13.41) has a unique negative real solution λ^* if, and only if, $G(0) < 1$, or $\mathcal{R}(\psi) < 1$. Also, (13.41) has a unique positive (zero) real solution if $\mathcal{R}(\psi) > 1$ ($\mathcal{R}(\psi) = 1$). To show that λ^* is the dominant real part of roots of $G(\lambda)$, we let $\lambda = x + iy$ be an arbitrary complex solution to (13.41). Note that

$$1 = G(\lambda) = |G(x + iy)| \leq G(x),$$

indicating that $\Re \lambda \leq \lambda^*$. It follows that the infection-free steady state is locally asymptotically stable if $\mathcal{R}(\psi) < 1$, and unstable if $\mathcal{R}(\psi) > 1$. This establishes the following result.

Theorem 13.1 Let $\mathcal{R}(\phi)$ be the reproduction number defined in (13.42).

- (a) The infection-free steady state (13.36) is locally asymptotically stable if $\mathcal{R}(\psi) < 1$ and unstable if $\mathcal{R}(\psi) > 1$.
- (b) There exists an endemic steady state of (13.35) when $\mathcal{R}(\psi) > 1$.

13.8.2 Pair-Formation in Age-Structured Epidemic Models

Sexually transmitted diseases (STDs) spread through sexual activities between partners. The pair-formation, or mixing, is one of the key terms in modeling of STDs [19, 20]. Assume that the function describing pair formation in the model is separable, which makes the mathematical analysis more tractable. However, it has been shown that the assumption of a separable pair formation function is equivalent to assuming a total proportionate or random partnership formation [5–8, 12]. We briefly explain it as follows.

Let $\rho(t, a, a')$ be the pair formation or mixing, which is the proportion of partners with age a' that an individual of age a has at time t . Let $r(t, a)$ be the average number of partners that an individual of age a has per unit of time, and let $T(t, a)$ be the total number of individuals of age a at time t . Then the function $\rho(t, a, a')$ has the following properties:

- 1) $0 \leq \rho(t, a, a') \leq 1$,
- 2) $\int_0^\infty \rho(t, a, a')da' = 1$,
- 3) $\rho(t, a, a')r(t, a)T(t, a) = \rho(t, a', a)r(t, a')T(t, a')$,
- 4) $r(t, a)T(t, a)r(t, a')T(t, a') = 0 \implies \rho(t, a, a') = 0$.

Properties 1) and 2) follow from the fact that $\rho(t, a, a')$ is a proportion so that it is always between zero and one, and its total sum equals one. Property 3) comes from the fact that the total number of pairs of individuals of age a with individuals of age a' needs to be equal to the total number of pairs of individuals of age a' with individuals of age a . Moreover, if there are no active individuals, then there is no pair formation, which leads to property 4).

13.8.2.1 Total Proportionate Mixing

Suppose that the pair formation is a separable function such that

$$\rho(t, a, a') = \rho_1(t, a)\rho_2(t, a'). \quad (13.43)$$

It follows from property 2 that

$$\int_0^\infty \rho(t, a, a') da' = \int_0^\infty \rho_1(t, a) \rho_2(t, a') da' = \rho_1(t, a) \int_0^\infty \rho_2(t, a') da' = 1,$$

for all t . Hence

$$\rho_1(t, a) = \frac{1}{\int_0^\infty \rho_2(t, a') da'}$$

is independent of a . Denote it by $L(t)$. Then

$$\rho(t, a, a') = L(t) \rho_2(t, a'). \quad (13.44)$$

It follows from property 3) and (13.44) that

$$L(t) \rho_2(t, a') r(t, a) T(t, a) = \rho(t, a', a) r(t, a') T(t, a'). \quad (13.45)$$

Integrating (13.45) with respect to a from 0 to ∞ yields

$$L(t) \rho_2(t, a') \int_0^\infty r(t, a) T(t, a) da = r(t, a') T(t, a'). \quad (13.46)$$

Hence

$$L(t) \rho_2(t, a') = \frac{r(t, a') T(t, a')}{\int_0^\infty r(t, a) T(t, a) da}, \quad (13.47)$$

which implies that $\rho(t, a, a')$ satisfies

$$\rho(t, a, a') = \frac{r(t, a') T(t, a')}{\int_0^\infty r(t, a) T(t, a) da}. \quad (13.48)$$

Notice that the right-hand side in (13.48) is the fraction of the total partners of age a' in the population, or the availability of partners of age a' . A pair formation or mixing function satisfying (13.45) is called a total proportionate mixing. Such a mixing depends completely on the availability of partners and is a kind of random mixing. While it may be appropriate to assume a proportionate mixing or random mixing in special cases such as modeling of HIV/AIDS for homosexual men, in general, it is necessary to assume the pair-formation or mixing function to be non-separable. It is possible to show that proportionate mixing is the only separable pair formation [8].

13.9 A Multi-Age-Group Malaria Model

Consider a human population in which malaria spreads. Divide the human population into four classes: susceptible, exposed, infective, and recovered who are recovered and also immune from re-infection. Denote the densities by $s(a, t)$, $e(a, t)$, $i(a, t)$, and $r(a, t)$, respectively. We further divide the human population into n age groups with group i consisting of ages $a \in [a_{i-1}, a_i]$, such that the number of individuals in the i th group in the respective epidemiological classes are $S_i(t) = \int_{a_{i-1}}^{a_i} s(a, t)da$, $E_i(t) = \int_{a_{i-1}}^{a_i} e(a, t)da$, $I_i(t) = \int_{a_{i-1}}^{a_i} i(a, t)da$, and $R_i(t) = \int_{a_{i-1}}^{a_i} r(a, t)da$ for $i = 1, 2, \dots, n$. Then, based on the similar derivation in Sect. 13.6, we obtain the following system of ordinary differential equations for the disease transmission in humans:

$$\begin{aligned} S'_1(t) &= B(t) - (\mu_1 + c_1)S_1 - \lambda_1(t)S_1, \\ S'_j(t) &= c_{j-1}S_{j-1} - \lambda_j(t)S_j - (\mu_j + c_j)S_j, \quad j = 2, \dots, n, \\ E'_1(t) &= \lambda_1(t)S_1 - (\mu_1 + \epsilon_1 + c_1)E_1, \\ E'_j(t) &= \lambda_j(t)S_j + c_{j-1}E_{j-1} - (\mu_j + \epsilon_j + c_j)E_j, \quad j = 2, \dots, n, \\ I'_1(t) &= \epsilon_1 E_1 - (\mu_1 + \gamma_1 + \omega_1 + c_1)I_1, \\ I'_j(t) &= \epsilon_j E_j + c_{j-1}I_{j-1} - (\mu_j + \gamma_j + \omega_j + c_j)I_j, \quad j = 2, \dots, n, \\ R'_1(t) &= \gamma_1 I_1 - (\mu_1 + c_1)R_1, \\ R'_j(t) &= c_{j-1}R_{j-1} + \gamma_j I_j - (\mu_j + c_j)R_j, \quad j = 2, \dots, n, \end{aligned} \tag{13.49}$$

where $B(t)$ is the recruitment rate of the susceptible class, μ_i the age specific natural death rate, ω_i the age specific disease induced death rates, η_i the age progression rate, ϵ_i the age specific disease progression rate, and γ_i the age specific recovery rate.

The infection rates $\lambda_j(t)$ for humans are related to the vector (mosquito) population and are given by

$$\lambda_j(t) = \frac{bN_v(t)}{N(t)}\beta_j \frac{I_v(t)}{N_v(t)} = \frac{b\beta_j I_v(t)}{N(t)}, \quad j = 1, \dots, n, \tag{13.50}$$

where b is the number of bites on humans per mosquito in unit time, N_v the total mosquito population, $N = \sum_{j=1}^n (S_j + E_j + I_j + R_j)$ the total size of human population, I_v the number of infective mosquitoes, and β_j the probability of infection per bite for group i .

Let S_v and E_v denote the numbers of susceptible and exposed mosquitoes, and $N_v = S_v + E_v + I_v$. The dynamics of the mosquito population are described by the equations

$$\begin{aligned} S'_v(t) &= M_v - \lambda_v S_v - \mu_v S_v, \\ E'_v(t) &= \lambda_v(N_v - E_v - I_v) - (\mu_v + \epsilon_v), \\ I'_v(t) &= \epsilon_v E_v - \mu_v I_v, \end{aligned} \quad (13.51)$$

where M_v is an input flow of susceptible mosquitoes, μ_v is the natural death rate of mosquitoes, ϵ_v is the disease progression rate for exposed mosquitoes, and λ_v is the infection rate for mosquitoes given by

$$\lambda_v(t) = b \sum_{j=1}^n \left(\frac{\beta_{v_j} I_j(t)}{N(t)} \right). \quad (13.52)$$

Here β_{v_j} are the infection rate of mosquitoes from biting infected humans in group j . Readers are referred to [30] for model analysis.

13.10 *Project: Another Malaria Model

In this project, we consider an SIS system for the human host and an SI system for the mosquitoes. Let $s(a, t)$ and $i(a, t)$ denote the densities for susceptible and infective hosts, and let $n(a, t) = s(a, t) + i(a, t)$. The system of PDEs for hosts reads

$$\begin{aligned} s_t(a, t) + s_a(a, t) &= -\lambda(a, t)s(a, t) - \mu(a)s(a, t) + \gamma(a)i(a, t), \\ i_t(a, t) + i_a(a, t) &= \lambda(a, t)s(a, t) - (\gamma(a) + \mu(a))i(a, t), \end{aligned} \quad (13.53)$$

where

$$\lambda(a, t) = b\beta(a) \frac{I_v(t)}{N(t)}$$

with boundary and initial conditions

$$s(0, t) = \int_0^\infty \rho(a)n(a, t)da, \quad i(0, t) = 0, \quad s(a, 0) = s_0(a), \quad i(a, 0) = i_0(a).$$

The function $\rho(a)$ denotes the age dependent birth rate. The equations for mosquitoes are

$$\begin{aligned} S'_v(t) &= M_v - \lambda_v(t)S_v(t) - \mu_v S_v(t), \\ I'_v(t) &= \lambda_v(t)S_v(t) - \mu_v I_v(t), \end{aligned} \quad (13.54)$$

where

$$\lambda_v(t) = b \int_0^\infty \beta_v(a) \frac{i(a, t)}{N(t)} da,$$

with initial conditions $S_v(0) = S_{v0} \geq 0$, $I_v(0) = I_{v0} \geq 0$. All other parameters have the same meaning as in (13.51).

For the questions in this project, assume that the host population has reached the stable age distribution with constant total population size, i.e.,

$$\tilde{n}(a) = \beta\pi(a), \quad \pi(a) = e^{-\int_0^a \mu(u)du}, \quad \tilde{N} = \int_0^\infty \tilde{n}(a)da = \beta \int_0^\infty \pi(a)da,$$

where β is a positive constant and the constraint $\int_0^\infty \rho(a)\tilde{n}(a, t)da = 1$ holds. Assume also that the total mosquito population has reached the steady state $\tilde{N}_v = M_v/\mu_v$.

Question 1

(a) Show that the basic reproduction number has the form

$$\mathcal{R}_0 = \int_0^\infty b\beta_v(a) \frac{\mathcal{I}(a)}{\tilde{N}} \tilde{N}_v da, \quad (13.55)$$

where

$$\mathcal{I}(a) = \int_0^a b\beta(\sigma) \frac{\beta\pi(\sigma)}{\tilde{N}} \frac{1}{\mu_v} e^{-\int_\sigma^a (\gamma(\tau) + \mu(\tau))d\tau} d\sigma. \quad (13.56)$$

(b) Provide the biological interpretation for the quantities $\mathcal{I}(a)$ and \mathcal{R}_0 .

Question 2 Show that the disease-free equilibrium of the system (13.53)–(13.54) is locally asymptotically stable if $\mathcal{R}_0 < 1$, and it is unstable if $\mathcal{R}_0 > 1$.

Question 3 Assume that all age-dependent parameters are constants in the age intervals $[a_{k-1}, a_k]$, $k = 1, \dots, m$, i.e., $\rho(a) = \rho_k$, $\beta(a) = \beta$, $\gamma(a) = \gamma_k$, $\mu(a) = \mu_k$. Let $n(t, a) = n(a) = \beta\pi(a)$ with constant total population sizes \tilde{N} . Let

$$\begin{aligned} S_k(t) &= \int_{a_{k-1}}^{a_k} s(a, t)da, \quad I_k(t) = \int_{a_{k-1}}^{a_k} i(a, t)da, \quad N_k = \int_{a_{k-1}}^{a_k} n(a)da, \\ \lambda_k(t) &= b\beta \frac{I_k(t)}{\tilde{N}}, \quad \lambda_v(t) = b \sum_{k=1}^m \beta_v \frac{I_k(t)}{\tilde{N}}, \quad k = 1, 2, \dots, m. \end{aligned}$$

Then the ODE system corresponding to the system (13.53)–(13.54) is

$$\begin{aligned} S'_1(t) &= \Lambda - (\alpha_1 + \lambda_1(t) + \mu_1)S_1(t) + \gamma_1 I_1(t), \\ S'_k(t) &= -(\alpha_k + \lambda_k + \mu_k)S_k + \gamma_k I_k + \alpha_{k-1}S_{k-1}(t), \quad k = 2, \dots, m, \\ I'_1(t) &= \lambda_1(t)S_1(t) - (\gamma_1 + \mu_1 + \alpha_1)I_1(t), \\ I'_k(t) &= \lambda_k S_k - (\gamma_k + \mu_k + \alpha_k)I_k + \alpha_{k-1}I_{k-1}(t), \quad k = 2, \dots, m, \\ I'_v(t) &= \lambda_v(t)(\tilde{N}_v - I_v) - \mu_v I_v, \end{aligned} \tag{13.57}$$

where $\alpha_k = \pi(a_k)/\int_{a_{k-1}}^{a_k} \pi(a)da$, $k = 1, \dots, m-1$, and $\alpha_m = 0$.

- (a) Show that the disease-free equilibrium (DFE), E_0 , of system (13.57) is given by

$$E_0 = \left(\frac{\beta}{\alpha + \mu_1}, \dots, \frac{\beta \prod_{j=1}^{k-1} \alpha_j}{\prod_{j=1}^k (\alpha_j + \mu_j)}, \dots, \frac{\beta \prod_{j=1}^{n-1} \alpha_j}{\prod_{j=1}^n (\alpha_j + \mu_j)}, 0, \dots, 0, \tilde{N}_v, 0 \right). \tag{13.58}$$

Note that the first m components of E_0 describe the age distribution of the population at the DFE, i.e.,

$$N_k = \frac{\beta \prod_{j=1}^{k-1} \alpha_j}{\prod_{j=1}^k (\alpha_j + \mu_j)}, \quad k = 1, 2, \dots, m \quad (\alpha_m = 0). \tag{13.59}$$

- (b) Derive the basic reproduction number $\tilde{\mathcal{R}}_0$ for the ODE system (13.57).
(c) Show that E_0 is locally asymptotically stable if $\tilde{\mathcal{R}}_0 < 1$, and unstable if $\tilde{\mathcal{R}}_0 > 1$.

Question 4 Verify the result in part (c) of Question 3 by numerical simulations. Consider the case of $m = 2$ age groups (e.g., children and adults), and assume that $\tilde{\mathcal{R}}_0$ can be varied by changing the value of β . To adopt different ratio of children to adults, the value of μ_2 can be determined by the ratio N_1/N_2 . One set of possible parameter values can be found in Table 13.1.

- (a) Choose the values and β_k , $k = 1, 2$, such that $\tilde{\mathcal{R}}_0 < 1$ and plot $I_1(t)$ and $I_2(t)$ vs. time.
(b) Repeat part (a) but use different values of β_k for which $\tilde{\mathcal{R}}_0 > 1$.
(c) Assume that drug treatment can increase the recovery rate γ_k . Plot $\tilde{\mathcal{R}}_0$ as a function of γ_1 or γ_2 or both.
(d) Consider various scenarios based on treatment in children or adults alone or both. Plot $I_1(t)$ and $I_2(t)$ vs. time and observe how they change with γ_k .

References: [14, 17, 29, 33].

Table 13.1 Parameter values and ranges for numerical simulations

Parameters	Two age groups	References
Fixed values		
M_v	150	
μ_v	0.03	[17, 29]
β	$0.45N(\mu_1 + \alpha_1)$	
Values with ranges		
b	0.2 (0.1, 0.25)	[14, 17]
Age-dependent	Age group 1	Age group 2
μ_k	0.0002	$\alpha_1 \bar{N}_1 / \bar{N}_2$
α_k	1/15/365	0
β_k	0.03	0.015
β_{v_k}	0.3	0.24
γ_k	1/100	1/80
		Reference

13.11 Project: A Model Without Vaccination

Consider the model in the previous section but without vaccination, i.e.,

$$\begin{aligned} s_t(t, a) + s_a(t, a) &= -\beta(a)c(a)B(t)s(t, a) - \mu(a)s(t, a), \\ l_t(t, a) + l_a(t, a) &= \beta(a)c(a)B(t)(s(t, a) + \sigma j(t, a)) - (k + \mu(a))l(t, a), \\ i_t(t, a) + i_a(t, a) &= kl(t, a) - (r + \mu(a))i(t, a), \\ j_t(t, a) + j_a(t, a) &= ri(t, a) - \sigma\beta(a)c(a)B(t)j(t, a) - \mu(a)j(t, a), \end{aligned} \quad (13.60)$$

where

$$B(t) = \int_0^\infty \frac{i(t, a')}{n(t, a')} p(t, a') da',$$

with initial and boundary conditions:

$$\begin{aligned} s(t, 0) &= \Lambda, \quad l(t, 0) = i(t, 0) = j(t, 0) = 0, \\ s(0, a) &= s_0(a), \quad l(0, a) = l_0(a), \quad i(0, a) = i_0(a), \quad j(0, a) = j_0(a). \end{aligned}$$

The function p is the same as defined in the previous section. Assume that the total population is at the stable age distribution, i.e., $n(t, a) = \Lambda \mathcal{F}(a) =: N(a)$ and let $p(t, a) = p_\infty(a) = c(a)N(a) / \int_0^\infty c(b)N(b)db$.

(a) Introduce the fractions

$$u(t, a) = \frac{s(t, a)}{n(t, a)}, \quad x(t, a) = \frac{l(t, a)}{n(t, a)}, \quad y(t, a) = \frac{i(t, a)}{n(t, a)}, \quad z(t, a) = \frac{j(t, a)}{n(t, a)}.$$

Derive a system for the fractions u , x , y , and z .

- (b) Derive a formula for the basic reproduction number \mathcal{R}_0
- (c) Show that the DFE, $(u(a), x(a), y(a), z(a)) = (1, 0, 0, 0)$, is locally asymptotically stable when $\mathcal{R}_0 < 1$ and unstable when $\mathcal{R}_0 > 1$.
- (d) An endemic steady state of the system is a time-independent solution

$$(u^*(a), x^*(a), y^*(a), z^*(a)), \quad \text{with } x^* > 0.$$

Find an endemic steady state.

13.12 Project: An SIS Model with Age of Infection Structure

Consider the following system which describes an SIS model with infectiousness dependent on age-of-infection a :

$$\begin{cases} S'(t) = \mu N - [\mu + \lambda(t)] S(t) + \int_0^\infty \delta \gamma(a) i(t, a) da \\ i_t(t, a) + i_a(t, a) = -[\mu + \delta \gamma(a)] i(t, a) \\ i(0, t) = \lambda(t) S(0), \quad S(0) = S_0 > 0, \quad i(0, a) = i_0(a) \geq 0 \end{cases} \quad (13.61)$$

where

$$\lambda(t) = \frac{\beta}{N} \int_0^\infty p(a) i(t, a) da$$

and

$$N(t) = S(t) + I(t), \quad I(t) = \int_0^\infty i(t, a) da.$$

The function $p(a) \in (0, 1)$ describes the fraction of infected individuals who are infectious at infection age a , β is the transmission coefficient for infectious individuals, $\gamma(a)$ is the age-dependent recovery rate, and δ is a factor representing the increased recovery rate due to treatment.

- (a) Denote the incidence function by $v(t) = \lambda(t)S(t)$. Solve the i equation using the method of characteristics and express the solution as a function of v .
- (b) Derive a system of integral equations using variables $N(t)$ and $v(t)$ that is equivalent to the system (13.61).
- (c) Derive a formula for the reproduction number \mathcal{R}_0 .
- (d) Show that an endemic steady state E exists when $\mathcal{R}_0 > 1$, and determine how the disease level at the steady-state E depends on \mathcal{R}_0 .
- (e) Examine the dependence of E and \mathcal{R}_0 on the treatment effort δ .

13.13 Exercises

1. Consider the model (13.33) in Sect. 13.8.1.

(a) Show that

$$\lim_{t \rightarrow \infty} n(t, a) = \Lambda \mathcal{F}(a)$$

$$\lim_{t \rightarrow \infty} p(t, a) = \frac{c(a) \mathcal{F}(a)}{\int_0^\infty c(b) \mathcal{F}(b) db} := p_\infty(a),$$

where

$$\mathcal{F}(a) = e^{-\int_0^a \mu(s) ds}.$$

(b) Show that at the disease-free steady state E_0

$$\frac{s(a)}{n(a)} = \mathcal{F}_\psi(a), \quad \frac{v(a)}{n(a)} = 1 - \mathcal{F}_\psi(a), \quad (13.62)$$

$$l(a) = i(a) = j(a) = 0, \quad n(a) = \Lambda \mathcal{F}(a),$$

where

$$\mathcal{F}_\psi(a) = e^{-\int_0^a \psi(b) db}.$$

(c) Let \mathcal{R}_0 denote the basic reproduction number. Show that

$$\mathcal{R}_0 = \int_0^\infty \int_0^\infty \frac{k}{r - k} p_\infty(\alpha + \tau) \beta(\alpha) c(\alpha) (e^{-k\tau} - e^{-r\tau}) d\tau d\alpha.$$

References

1. Anderson, R.M. & R.M. May (1991) Infectious Diseases of Humans. Oxford University Press (1991)
2. Anderson, R.M. & R.M. May (1979) Population biology of infectious diseases I, *Nature* **280**: 361–367.
3. Anderson, R.M., and R.M. May (1984) Spatial, temporal, and genetic heterogeneity in host populations and the design of immunization programmes, *Math. Med. Biol.* **1**(3), 233–266.
4. Andreasen, V. (1995) Instability in an SIR-model with age dependent susceptibility, *Mathematical Population Dynamics: Analysis of Heterogeneity*, Vol. 1, Theory of Epidemics (O. Arino, D. Axelrod, M. Kimmel, M. Langlais, eds.), Wuerz, Winnipeg: 3–14.
5. Blythe, S.P. and C. Castillo-Chavez (1989) Like-with-like preference and sexual mixing models, *Math. Biosci.* **96**: 221–238.
6. Blythe, S.P., C. Castillo-Chavez, J. Palmer & M. Cheng (1991) Towards a unified theory of mixing and pair formation, *Math. Biosc.* **107**: 379–405.

7. Busenberg, S. and C. Castillo-Chavez (1989) Interaction, pair formation and force of infection terms in sexually transmitted diseases, in: Mathematical and Statistical Approaches to AIDS Epidemiology, (C. Castillo-Chavez, ed.), Lecture Notes in Biomathematics, Vol. **83**, Springer-Verlag, New York, (1989), pp. 289–300.
8. Busenberg, S. and C. Castillo-Chavez (1991) A general solution of the problem of mixing of subpopulations and its application to risk- and age-structured epidemic models for the spread of AIDS, IMA J. Math. Appl. Med. Biol. **8**: 1–29.
9. Capasso, V. (1993) Mathematical Structures of Epidemic Systems, Lect. Notes in Biomath. **83**, Springer-Verlag, Berlin-Heidelberg-New York.
10. Castillo-Chavez, C. and Z. Feng (1998) Global stability of an age-structure model for TB and its applications to optimal vaccination strategies, Math. Biosci. **151**(2), 135–154.
11. Castillo-Chavez, C., H.W. Hethcote, V. Andreasen, S.A. Levin, and W.M. Liu (1989) Epidemiological models with age structure, proportionate mixing, and cross-immunity, J. Math. Biol. **27**(3), 233–258.
12. Castillo-Chavez, C. and S. P. Blythe (1989) Mixing Framework for Social/Sexual Behavior, in: Mathematical and Statistical Approaches to AIDS Epidemiology, (C. Castillo-Chavez, ed.), Lecture Notes in Biomathematics **83**, Springer-Verlag, New York, pp. 275–288.
13. Cha, Y., M. Iannelli, & F. Milner (1998) Existence and uniqueness of endemic states for the age-structured S-I-R epidemic model, Math. Biosc. **150**: 177–190.
14. Chitnis, N., J.M. Hyman, and J.M. Cushing (2008) Determining important parameters in the spread of malaria through the sensitivity analysis of a mathematical model. Bull. Math. Biol. **70**:1272–1296.
15. Dietz, K. (1975) Transmission and control of arbovirus diseases, Epidemiology (D. Ludwig, K. Cooke, eds.), SIAM, Philadelphia: 104–121.
16. Dietz, K., and D. Schenzle (1985) Proportionate mixing models for age-dependent infection transmission, J. Math. Biol. **22**(1), 117–120.
17. Garret-Jones, C. and G.R. Shidrawi (1969) Malaria vectorial capacity of a population of Anopheles gambiae. Bull. Wld Hlth Org. **40**:531–545.
18. Gurtin, M.L. and R.C. MacCamy (1974) Nonlinear age dependent population dynamics, Arch. Rat. Mech. Analysis, **54**: 281–300.
19. Hadeler, K.P., R. Waldstatter, and A. Worz-Busekros (1988) Models for pair-formation in bisexual populations, J. Math. Biol. **26**: 635–649.
20. Hethcote, H.W. (2000) The mathematics of infectious diseases, SIAM Review **42**: 599–653.
21. Hoppensteadt, F.C. (1975) Mathematical Theories of Populations: Demographics, Genetics and Epidemics, SIAM, Philadelphia.
22. Iannelli, M. (1995) Mathematical Theory of Age-Structured Population Dynamics, Appl. Mathematical Monographs, C.N.R..
23. Kot, M. (2001) Elements of Mathematical Ecology. Cambridge University Press.
24. May, R.M. (1986) Population biology of macroparasitic infections, in Mathematical Ecology; An Introduction, (Hallam, T.G., Levin, S.A. eds.), Biomathematics **18**, Springer-Verlag, Berlin-Heidelberg-New York: 405–442.
25. May, R.M. & Anderson, R.M. (1979) Population biology of infectious diseases II, Nature **280**: 455–461.
26. McKendrick, A.G. (1926) Applications of mathematics to medical problems, Proc. Edinburgh Math. Soc. **44**: 98–130.
27. Müller, J. (1998) Optimal vaccination patterns in age structured populations, SIAM J. Appl. Math. **59**: 222–241.
28. Murray, J.D. (2002) Mathematical Biology, Vol. I, Springer-Verlag, Berlin-Heidelberg-New York.
29. Paaijmans, K. Weather, water and malaria mosquito larvae, 2008.
30. Park, T. (1984) Age-Dependence in Epidemic Models of Vector-Borne Infections, Ph.D. Dissertation, University of Alabama in Huntsville, 2004.
31. Sharpe, F.R., and A.J. Lotka (1911) A problem in age-distribution. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, **21**(124): 435–438.

32. Thieme, H.R. (2003) Mathematics in Population Biology, Princeton University Press, Princeton, N.J.
33. United nations, department of economic and social affairs, population division, population estimates and projections sections. Retrieved January, 2013 from <http://esa.un.org/wpp/index.htm>.
34. Waltman, P. (1974) Deterministic Threshold Models in the Theory of Epidemics, Lect. Notes in Biomath. **1**, Springer-Verlag, Berlin-Heidelberg-New York.
35. Webb, G.F. (1985) Theory of Nonlinear Age-dependent Population Dynamics, Marcel Dekker, New York.

Chapter 14

Spatial Structure in Disease Transmission Models



14.1 Spatial Structure I: Patch Models

With the advent of substantial intercontinental air travel, it is possible for diseases to move from one location to a completely separate location very rapidly. This was an essential aspect of modeling SARS during the epidemic of 2002–2003, and has become a very important part of the study of the spread of epidemics. Mathematically, it has led to the study of metapopulation models or models with patchy environments and movement between patches [4–7, 33, 38].

These models, which are the focus of this section, are called metapopulation models. They usually consist of a system (often a large system) of ordinary differential equations with some coupling between patches. A patch can be a city, community, or some other geographical region. In real life, a metapopulation model for the spread of a communicable disease should consider all the locations for which there are interactions. An example would be two distant cities with some air travel between them but no contact otherwise. It is possible that not all patches are linked directly. For example, we might think of a system consisting of a city and two suburbs, with contact between occupants of each suburb and the city but not between occupants of the two suburbs. Thus a model must keep track of both the patches and the links between them, and should be described in terms of a graph.

In the interest of simplicity, we will confine our attention to models consisting of only two patches, but it is important to be aware of the complications of scale. A more thorough description may be found in [3].

14.1.1 Spatial Heterogeneity

Consider a basic *SIR* compartmental model. We divide the population into two connected sub-populations. Let S_i , I_i , R_i denote, respectively, the number of

susceptible, infective, and recovered individuals in Patch i for $i = 1, 2$. The total population of Patch i is $N_i = S_i + I_i + R_i$. The birth and natural death rate constant μ is assumed to be the same in each patch, so that the total population of each patch remains constant. The average infective period $1/\gamma$ is assumed to be the same in each patch. This spatial model can be written for $i = 1, 2$ as in [24]

$$\begin{aligned} S'_i &= \mu N_i - \mu S_i - \lambda_i S_i \\ I'_i &= \lambda_i S_i - (\mu + \gamma) I_i \\ R'_i &= \gamma I_i - \mu R_i, \end{aligned} \tag{14.1}$$

with the force of infection in Patch i given by a mass action type of incidence

$$\lambda_i = \beta_{i1} I_1 + \beta_{i2} I_2.$$

Thus, infective individuals in one patch can infect susceptible individuals in another patch, but there is no explicit movement of individuals in this model.

For the SIR model (14.1) the disease-free equilibrium is $S_i = N_i$, $I_i = R_i = 0$. Using the next generation matrix [39] \mathcal{R}_0 can be calculated from (14.1) as $\mathcal{R}_0 = \rho(FV^{-1})$, where the (i, j) of FV^{-1} is $\beta_{ij} N_j / (d + \gamma)$.

For the case that each patch has the same population (i.e., $N_i = N$) and β_{ij} are such that the endemic equilibrium values of S_i , I_i , and λ_i are independent of i , then the endemic equilibrium is given explicitly for $\mathcal{R}_0 > 1$ by

$$S_{i\infty} = S_\infty = \frac{N}{\mathcal{R}_0}, \quad I_{i\infty} = I_\infty = \frac{\mu N}{\mu + \gamma} \left(1 - \frac{1}{\mathcal{R}_0}\right), \quad R_{i\infty} = R_\infty = \frac{\gamma I_\infty}{d}, \tag{14.2}$$

with $\lambda_{i\infty} = \lambda_\infty = \mu(\mathcal{R}_0 - 1)$.

As an example of a symmetric situation that satisfies the above requirements, assume that $\beta_{ij} = \beta$ if $i = j$ and $\beta_{ij} = p\beta$ with $p < 1$ if $i \neq j$. Then the contact rate is the same within each patch and has a smaller value between patches. We may calculate that

$$\mathcal{R}_0 = \frac{\beta N(p+1)}{(\mu + \gamma)},$$

which depends on the coupling strength p .

We should note that the assumption in the above example that the two patches have the same total population size is quite unrealistic in practice; it is given only as a simple example.

If we linearize about the endemic equilibrium and solve the linear approximation, we find that the solutions are damped oscillations about the equilibrium which are phase-locked except for very small values of p . Simulations suggest that, as has been found in other metapopulation models, with larger p values (i.e., stronger between patch coupling) the system effectively behaves much like a single patch.

14.1.2 Patch Models with Travel

Sattenspiel and Dietz [33] introduced a metapopulation epidemic model in which individuals are labeled with their city of residence as well as the city in which they are present at a given time.

To formulate a demographic model with travel for two patches, let $N_{ij}(t)$ be the number of residents of Patch i who are present in Patch j at a time t . Residents of Patch i leave this patch at a per-capita rate $g_i \geq 0$ per unit time. For a model with more than two patches we would also need to count the fraction of these travelers going to each patch. Residents of Patch i who are in Patch j return home to Patch i with a per-capita rate of $r_{ij} \geq 0$ with $r_{11} = r_{22} = 0$. It is natural to assume that $g_i > 0$ if and only if $r_{ij} > 0$. These travel rates determine a directed graph with patches as vertices and edges connecting vertices if the travel rates between them are positive. It is assumed that the travel rates are such that this directed graph is strongly connected.

Assume that births occur in the home patch at a per-capita rate $\mu > 0$, and that natural deaths occur in each patch with this same rate. Then the population numbers satisfy the equations

$$\begin{aligned} N'_{ii} &= \sum_{k=1}^2 r_{ik} N_{ik} - g_i N_{ii} + \mu \left(\sum_{k=1}^2 N_{ik} - N_{ii} \right) \\ N'_{ij} &= g_i N_{ii} - r_{ij} N_{ij} - \mu N_{ij}, \quad i, j = 1, 2, \quad i \neq j. \end{aligned} \tag{14.3}$$

These equations describe the evolution of the number of residents in Patch i who are currently in Patch i and those who are currently in Patch $j \neq i$. In the first equation of (14.3) the term μN_{ik} represents births in Patch i to residents of Patch i currently in Patch k . The number of residents of Patch i , namely $N_i^r = N_{i1} + N_{i2}$ is constant, as is the total population of the two-patch system. With initial conditions $N_{ij}(0) > 0$, the system (14.3) has an asymptotically stable equilibrium \hat{N}_{ij} .

We now formulate an epidemic model in each of the patches, with $S_{ij}(t)$ and $I_{ij}(t)$ denoting the number of susceptible and infective individuals resident in Patch i who are present in Patch j at time t . The equations for the evolution of the number of susceptibles and infectives residents in Patch i (with $i = 1, 2$) are

$$\begin{aligned} S'_{ii} &= \sum_{k=1}^2 r_{ik} S_{ik} - g_i S_{ii} - \sum_{k=1}^2 \kappa_i \beta_{iki} \frac{S_{ii} I_{ki}}{N_i^p} + \mu \left(\sum_{k=1}^2 N_{ik} - S_{ii} \right) \\ I'_{ii} &= \sum_{k=1}^2 r_{ik} I_{ik} - g_i I_{ii} + \sum_{k=1}^2 \kappa_i \beta_{iki} \frac{S_{ii} I_{ki}}{N_i^p} - (\gamma + \mu) I_{ii}, \end{aligned} \tag{14.4}$$

and for $j \neq i$

$$\begin{aligned} S'_{ij} &= g_i S_{ii} - r_{ij} S_{ij} - \sum_{k=1}^2 \kappa_j \beta_{ikj} \frac{S_{ij} I_{kj}}{N_j^p} - \mu S_{ij} \\ I'_{ij} &= g_i I_{ii} - r_{ij} I_{ij} + \sum_{k=1}^2 \kappa_j \beta_{ikj} \frac{S_{ij} I_{kj}}{N_j^p} - (\gamma + \mu) I_{ij}, \quad i, j = 1, 2 \end{aligned} \quad (14.5)$$

with $N_i^p = N_{1i} + N_{2i}$ denoting the number present in Patch i . Here $\beta_{ikj} > 0$ is the proportion of adequate contacts in Patch j between a susceptible from Patch i and an infective from Patch k that results in disease transmission, $\beta_{ikj} > 0$ is the average number of such contacts in Patch j per unit time, and $\gamma > 0$ is the recovery rate of infectives (assumed the same in each patch). Note that the disease is assumed to be sufficiently mild so that it does not cause death and does not inhibit travel, and it is assumed that individuals do not change disease status during travel. Equations (14.4) and (14.5) together with non-negative initial conditions constitute the *SIR* metapopulation model.

The disease-free equilibrium is given by $S_{ij} = \hat{N}_{ij}$, $I_{ij} = 0$ for $i, j = 1, 2$. If the system is at an equilibrium and one patch is at the disease-free equilibrium, then all patches are at the disease-free equilibrium; whereas if one patch is at an endemic disease level, then all patches are at an endemic level. These results hold based on the assumption that the directed graph determined by the travel rates is strongly connected. If this is not the case, then the results apply to patches within a strongly connected component.

We may calculate the basic reproduction number \mathcal{R}_0 for the model (14.4), (14.5) using the next generation matrix approach [14, 39]. As the result is somewhat complicated, we do not give it explicitly here, but we note that \mathcal{R}_0 depends on the travel rates as well as the epidemic parameters. If $\mathcal{R}_0 < 1$, then the disease-free equilibrium is locally asymptotically stable; whereas if $\mathcal{R}_0 > 1$, then it is unstable.

If the disease transmission coefficients are equal for all populations present in a patch, i.e., $\beta_{ijk} = \beta_k$ for $i, j = 1, 2$ it is possible to obtain the bounds

$$\min_{i=1,2} \mathcal{R}_{oi} \leq \mathcal{R}_0 \leq \max_{i=1,2} \mathcal{R}_{oi}, \quad (14.6)$$

where $\mathcal{R}_{oi} = \kappa_i \beta_i / (d + \gamma)$ is the basic reproduction number of Patch i in isolation. Thus if $\mathcal{R}_{oi} < 1$ for all i , the disease dies out; whereas if $\mathcal{R}_{oi} > 1$ for all i , then the disease-free equilibrium is unstable.

A change in travel rates g_1, g_2 can induce a bifurcation from $\mathcal{R}_0 < 1$ to $\mathcal{R}_0 > 1$ or vice versa, see [5, Fig. 3a]. Thus travel can stabilize (small travel rates) or destabilize (larger travel rates) the disease-free equilibrium. Numerical simulations support the claim that for $\mathcal{R}_0 > 1$, the endemic equilibrium is unique and that \mathcal{R}_0 acts as a sharp threshold between extinction and invasion of the disease.

Sattenspiel and Dietz [33] have suggested an application of their metapopulation SIR model to the spread of measles in the 1984 epidemic in Dominica. Travel rates of infants, school-age children, and adults are assumed to be different, thus making the model system highly complex and requiring knowledge of much data for simulation. Sattenspiel and coworkers, see [32] and the references therein, have since used this modeling approach for studying other infectious diseases in the historical archives.

The SARS epidemic of 2002–2003 spread rapidly through airline transportation from Asia to North America, and if there is an influenza pandemic in the near future it is likely that it will spread in a similar way. Metapopulation models, perhaps with small numbers of traveling infectives, may be a useful approach to modeling such a spread. Because the airline network is complex and because passenger travel data are difficult to acquire, there are substantial technical problems in the formulation of accurate models. However, it is possible that the qualitative insights that can be obtained from simple metapopulation models may be useful.

14.1.3 *Patch Models with Residence Times*

In this chapter we have been examining patch models with travel rates between patches included explicitly. Another possible perspective, which may be more appropriate in some situations, would be to describe patches with residents who spend a fraction of their time in different patches. For example, the spread of an infectious disease from one village to another through people who visit other patches may be a realistic description. Another interpretation could be to assume that individuals spend some of their time in environments more likely to allow disease transmission.

We consider an *SIR* epidemic model in two patches, one of which has a significantly larger contact rate, with short-term travel between the two patches. The total population resident in each patch is constant. We follow a Lagrangian perspective, that is, we keep track of each individual's place of residence at all times [9, 12, 16]. This is in contrast to an Eulerian perspective, which describes migration between patches.

Thus we consider two patches, with total resident population sizes N_1 and N_2 , respectively, each population being divided into susceptibles, infectives, and removed members. S_i and I_i denote the number of susceptibles and infectives, respectively, who are residents in Patch i , regardless of the patch in which they are present.

Residents of Patch i spend a fraction p_{ij} of their time in Patch j , with

$$\sum_{i=1}^2 p_{i1} + p_{i2} = 1, \quad i = 1, 2.$$

β_i is the risk of infection in Patch i , and we assume $\beta_1 > \beta_2$.

Each of the $p_{11}S_1$ susceptibles from Patch 1 who are present in Patch 1 can be infected by infectives from Patch 1, and infectives from Patch 2 who are present in Patch 1. Similarly, each of the $p_{12}S_1$ susceptibles present in Patch 2 can be infected by infectives from Patch 1, and by infectives from Patch 2 who are present in Patch 2. The number of infectives from both patches who are present in Patch 1 is

$$p_{11}I_1(t) + p_{21}I_2(t)$$

and the total number of individuals present in Patch 1 is

$$p_{11}N_1 + p_{21}N_2.$$

Thus the density of infected individuals in Patch 1 at time t who can infect only individuals currently in Patch 1 at time t , that is, the *effective* infective proportion in Patch 1 is given by

$$\frac{p_{11}I_1(t) + p_{21}I_2(t)}{p_{11}N_1 + p_{21}N_2}.$$

Thus the rate of new infections of members of Patch 1 in Patch 1 is

$$\beta_1 p_{11} S_1 \frac{p_{11}I_1(t) + p_{21}I_2(t)}{p_{11}N_1 + p_{21}N_2}.$$

The rate of new infections of members of Patch 1 in Patch 2 is

$$\beta_2 p_{12} S_1 \frac{p_{12}I_1(t) + p_{22}I_2(t)}{p_{12}N_1 + p_{22}N_2}.$$

Then the differential equations for S_1 and I_1 for an *SIR* infection are given by

$$\begin{aligned} S'_1 &= -\beta_1 p_{11} S_1 \left[\frac{p_{11}I_1(t) + p_{21}I_2(t)}{p_{11}N_1 + p_{21}N_2} \right] - \beta_2 p_{12} S_1 \left[\frac{p_{12}I_1(t) + p_{22}I_2(t)}{p_{12}N_1 + p_{22}N_2} \right] \\ I'_1 &= \beta_1 p_{11} S_1 \left[\frac{p_{11}I_1(t) + p_{21}I_2(t)}{p_{11}N_1 + p_{21}N_2} \right] + \beta_2 p_{12} S_1 \left[\frac{p_{12}I_1(t) + p_{22}I_2(t)}{p_{12}N_1 + p_{22}N_2} \right] - \gamma I_1. \end{aligned} \quad (14.7)$$

There is a corresponding calculation for the rate of new infections of members of Patch 2 in each patch, and the differential equations for S_2 and I_2 are given by

$$\begin{aligned} S'_2 &= -\beta_1 p_{21} S_2 \left[\frac{p_{11}I_1(t) + p_{21}I_2(t)}{p_{11}N_1 + p_{21}N_2} \right] - \beta_2 p_{22} S_2 \left[\frac{p_{12}I_1(t) + p_{22}I_2(t)}{p_{12}N_1 + p_{22}N_2} \right] \\ I'_2 &= \beta_1 p_{21} S_2 \left[\frac{p_{11}I_1(t) + p_{21}I_2(t)}{p_{11}N_1 + p_{21}N_2} \right] + \beta_2 p_{22} S_2 \left[\frac{p_{12}I_1(t) + p_{22}I_2(t)}{p_{12}N_1 + p_{22}N_2} \right] - \gamma I_2. \end{aligned} \quad (14.8)$$

Using the next generation approach to compute the basic reproduction number [39] we define

$$\mathcal{F} = \begin{pmatrix} p_{11}\beta_1 S_1 \frac{p_{11}I_1 + p_{21}I_2}{p_{11}N_1 + p_{21}N_2} + p_{12}\beta_2 S_1 \frac{p_{12}I_1 + p_{22}I_2}{p_{12}N_1 + p_{22}N_2} \\ p_{21}\beta_1 S_2 \frac{p_{11}I_1 + p_{21}I_2}{p_{11}N_1 + p_{21}N_2} + p_{22}\beta_2 S_2 \frac{p_{12}I_1 + p_{22}I_2}{p_{12}N_1 + p_{22}N_2} \end{pmatrix} \quad \text{and} \quad \mathcal{V} = \begin{pmatrix} \gamma I_1 \\ \gamma I_2 \end{pmatrix},$$

then

$$F = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \quad \text{and} \quad V = \begin{pmatrix} \gamma & 0 \\ 0 & \gamma \end{pmatrix},$$

where

$$\begin{aligned} B_{11} &= N_1 \left(\frac{p_{11}^2\beta_1}{p_{11}N_1 + p_{21}N_2} + \frac{p_{12}^2\beta_2}{p_{12}N_1 + p_{22}N_2} \right), \\ B_{12} &= N_1 \left(\frac{p_{11}p_{21}\beta_1}{p_{11}N_1 + p_{21}N_2} + \frac{p_{12}p_{22}\beta_2}{p_{12}N_1 + p_{22}N_2} \right), \\ B_{21} &= N_2 \left(\frac{p_{11}p_{21}\beta_1}{p_{11}N_1 + p_{21}N_2} + \frac{p_{12}p_{22}\beta_2}{p_{12}N_1 + p_{22}N_2} \right), \\ B_{22} &= N_2 \left(\frac{p_{21}^2\beta_1}{p_{11}N_1 + p_{21}N_2} + \frac{p_{22}^2\beta_2}{p_{12}N_1 + p_{22}N_2} \right). \end{aligned}$$

The B matrix explicitly captures the secondary infections produced by Patch 1 and Patch 2 individuals in each environment. For example, B_{12} collects the Patch 1 residents infected by Patch 2 inhabitants in both environments. Finally, the reproduction number is the largest eigenvalue of the matrix FV^{-1} , this is

$$\mathcal{R}_0(\mathbb{P}) = \frac{B_{11} + B_{22} - \sqrt{B_{11}^2 + 4B_{12}B_{21} - 2B_{11}B_{22} + B_{22}^2}}{2\gamma}.$$

Figure 14.1 shows the effect of mobility on $\mathcal{R}_0(\mathbb{P})$ as residence times vary. In the next chapter we show the applications of this approach in the context of Ebola, tuberculosis, and Zika.

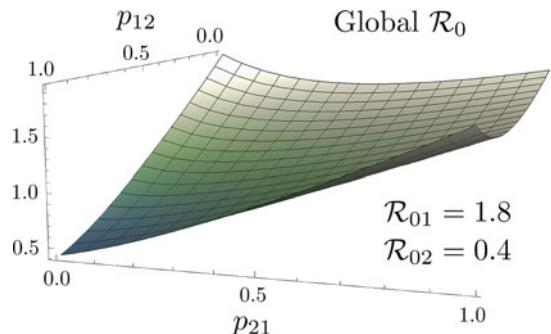
In the special case of no movement between patches

$$p_{11} = p_{22} = 1, \quad p_{12} = p_{21} = 0,$$

we obtain

$$\mathcal{R}_0 = \max \left(\frac{\beta_1}{\gamma}, \frac{\beta_2}{\gamma} \right).$$

Fig. 14.1 Effect of mobility in the global \mathcal{R}_0 . Parameters $\beta_1 = 0.09$, $\beta_2 = 0.02$, and $\gamma = 0.05$



Integration of the equations for $(S_i + I_i)$ ($i = 1, 2$) gives

$$\gamma \int_0^\infty I_i(t) dt = N_i - S_i(\infty), \quad i = 1, 2$$

and these relations combined with the result of integrating the equations for S'_i/S_i ($i = 1, 2$) give the final size relations, whose form is quite complicated.

Choosing different values for p_{ij} gives a way to estimate the effect on the epidemic size of imposing travel restrictions between patches.

14.2 Spatial Structure II: Continuously Distributed Models

In the preceding section, we have discussed the spread of a communicable disease from one patch to another. In this section we will discuss the spatial spread of a disease in a single patch because of the (continuous) motion of individuals. The mathematical analysis is based on partial differential equations of reaction–diffusion type. It is technically complicated and requires substantial mathematical background. As references for some of the mathematical details, we suggest [10, Chapter 5], [15, Chapters 9–11], [23, Chapters 15–18], [27, Chapters 11 and 13], [28, Chapters 1 and 2].

An introduction to models for the spatial spread of epidemics may be found in other references such as [2, 11, 14, 19, 37]. One characteristic feature of such models is the appearance of traveling waves, which have been observed frequently in the spread of epidemics through Europe from medieval times to the more recent studies of fox rabies [1, 22, 25, 29]. The asymptotic speed of spread of disease is the minimum wave speed [8, 13, 26, 31, 35, 41]. Models describing spatial spread and including age of infection are analyzed in [17, 18, 20, 28].

14.2.1 The Diffusion Equation

Let us begin by considering the motion of particles. Here, by a particle we might mean an individual cell, a member of a population, or any object of a set in whose spatial distribution as a function of time we are interested.

Our approach is to take a small region of space and to form a balance equation which says that the rate of change of the number of particles in the region is equal to the rate at which particles flow out of the region minus the rate at which particles flow into the region plus the rate of creation of particles in the region.

We shall confine ourselves mainly to the case in which the dependence is with respect to a single space coordinate. Let us think of a tube of constant cross section area A and let x denote the distance along the tube from some arbitrary starting point $x = 0$. We assume that the tube is a bounded region described by the inequalities $0 \leq x \leq L$.

Let $u(x, t)$ be the concentration of particles (number per unit volume) at location x at time t , meaning that in the portion of the tube between x and $x + h$, with volume Ah the number of particles is approximately $Ahu(x, t)$. By “approximately” we mean that if h is small, the error in this approximation $Ahu(x, t)$ is smaller than a constant multiple of h^2 .

We let $J(x, t)$ be the flux of particles at location x at time t , by which we mean the time rate of the number of particles crossing a unit area in the positive direction. For every x_0 the net rate of flow into the region between x_0 and $x_0 + h$ is $AJ(x_0, t) - AJ(x_0 + h, t)$. We let $Q(x_0, t, u)$ be the net growth rate per unit length at location x_0 at time t , representing births and deaths.

We have a balance relation on the interval $x_0 \leq x \leq x_0 + h$, expressing the fact that the rate of change of population size at time t in this interval is equal to the growth rate of population in this interval plus the net flux, and this leads to the conservation law

$$u_t(x, t) = Q(x, t, u) - \frac{\partial J}{\partial x}(x, t). \quad (14.9)$$

In order to obtain a model which describes the population density $u(x, t)$ we must make some assumption which relates the rate of change of flux density $\frac{\partial J}{\partial x}$ and the population density $u(x, t)$. If the motion is random, then *Fick's law* says that the flux due to random motion is approximately proportional to the rate of change of particle concentration, that is, that J is proportional to u_x . If population density decreases as x increases ($u_x < 0$) we would expect $J > 0$, so that J and u_x have opposite sign and thus that

$$J = -Du_x$$

with D a constant called the diffusivity or diffusion coefficient. More generally, D could be a function of the location x but we shall confine our attention to constant diffusivity. Equation (14.9) then becomes a second-order partial differential equation

$$u_t(x, t) = Q(x, t, u) + Du_{xx}(x, t). \quad (14.10)$$

Equation (14.10) is a *reaction-diffusion* equation. If $Q = 0$, it is called the *heat* or *diffusion* equation. It is possible to solve the heat equation explicitly; the solution for

$$-\infty < x < \infty, \quad 0 \leq t < \infty,$$

with $u(x, 0) = f(x)$ is

$$u(x, t) = \frac{1}{\sqrt{4\pi Dt}} \int_{-\infty}^{\infty} e^{-\frac{(x-\xi)^2}{4Dt}} f(\xi) d\xi.$$

In population ecology, we can translate Fick's law of diffusion into the statement that the individuals move from a region of high concentration to a region of low concentration in search for limited resources. We must, however, use this law with caution when modeling spatial spread of infectious diseases since the individual movement behaviors may be altered during the course of outbreaks of diseases.

For Eq. (14.10) to have a unique solution, we need to impose additional conditions. It is possible to establish the following result.

The diffusion equation (14.10) with a specified initial condition $u(x, 0) = f(x)$ for $0 \leq x \leq L$ and boundary conditions for $x = 0$ and $x = L$ has a unique solution for $0 \leq x \leq L$, $0 \leq t < \infty$. The boundary conditions may specify the value of u or the value of u_x for $x = 0$ and $x = L$.

Such a problem is called an *initial boundary value problem*.

We could also consider problems in an infinite tube defined by $-\infty < x < \infty$ for which no boundary conditions are required, or a semi-infinite tube $0 \leq x < \infty$ for which a boundary condition is required only at $x = 0$. In each case there is a unique bounded solution for any specified initial condition $u(x, 0) = f(x)$ for $-\infty \leq x < \infty$ or $u(x, 0) = f(x)$ for $0 \leq x < \infty$.

A boundary condition specifying that the solution must vanish at a boundary (called an absorbing boundary) may be taken to say that an individual leaving the region must die immediately. This is an idealization, but we may think of a large region with $u = 0$ far enough away. A boundary condition specifying that u_x must vanish at the boundary may be taken to say that the population is confined to the region and there is no flow across the boundary.

There are several types of initial condition which may arise. One possibility is that particles are absent initially, $u(x, 0) = 0$ and enter through the boundary. A second possibility is that particles are inserted at a single point x_0 , $u(x, 0) = u_0\delta(x - x_0)$. Here, $\delta(x)$ denotes the delta "function," which is zero except for $x = 0$ and

$$\int_{-\infty}^{\infty} \delta(x) dx = 1 \quad (14.11)$$

and if f is continuous, then

$$\int_{-\infty}^{\infty} f(x)\delta(x-a) dx = f(a). \quad (14.12)$$

A third kind of initial condition would be to specify a constant initial concentration, $u(x, 0) = u_0$ for $0 \leq x \leq L$.

14.2.2 Nonlinear Reaction–Diffusion Equations

In this section we consider reaction–diffusion equations containing a nonlinear growth rate $g(u)$ with $g(0) = g(K) = 0$, $g'(u) > 0$ for $0 \leq u < K$ and $g'(K) < 0$. Thus, we shall consider the equation

$$u_t(x, t) = Du_{xx}(x, t) + g(u). \quad (14.13)$$

This equation could describe a population with diffusion in space, and births and deaths given by the function $g(u)$.

We begin by looking for solutions $u(t)$ which are independent of x . Then $u_{xx} = 0$ and Eq. (14.13) reduces to the ordinary differential equation

$$u' = g(u). \quad (14.14)$$

Note that K is the carrying capacity, every bounded solution of (14.14) approaches the equilibrium K as $t \rightarrow \infty$.

For solution patterns in space, we can consider time-independent solutions, i.e., $u_t = 0$. In this case, Eq. (14.13) reduces to the following second-order ordinary differential equation:

$$Du'' + g(u) = 0. \quad (14.15)$$

Let $v = u'$, then $v' = -g(u)/D$, Eq. (14.15) is equivalent to the following first-order system:

$$\begin{pmatrix} u \\ v \end{pmatrix}' = \begin{pmatrix} v \\ -g(u)/D \end{pmatrix}. \quad (14.16)$$

From $g(0) = g(K) = 0$ and $g'(u) > 0$ for $0 \leq u < K$, we know that system (14.16) has equilibria $(u_\infty, v_\infty) = (0, 0)$ and $(K, 0)$. The Jacobian matrix is

$$\begin{bmatrix} 0 & 1 \\ -\frac{g'(u_\infty)}{D} & 0 \end{bmatrix}. \quad (14.17)$$

Since $g'(0) > 0$, both eigenvalues at the equilibrium $(0, 0)$ are positive and this equilibrium is asymptotically stable. Since $g'(K) < 0$, there is one positive and one negative eigenvalue and the equilibrium $(K, 0)$ is a saddle point.

Nonlinear reaction-diffusion equations may have traveling wave solutions. A traveling wave solution has the form $u(x, t) = U(x - ct)$ for some constant c . In this case, $u_x(x, t) = U'(x - ct)$, $u_{xx}(x, t) = U''(x - ct)$, and $u_t(x, t) = (-c)U'(x - ct)$. Thus, from Eq. (14.13), the following equation holds:

$$-cU'(x - ct) = g[U(x - ct)] + DU''(x - ct).$$

Let $z = x - ct$, then the function $U(z)$ must satisfy the second-order ordinary differential equation in z

$$DU'' + cU' + g(U) = 0. \quad (14.18)$$

The first-order system equivalent to (14.18) is

$$U' = V, \quad V' = -\frac{g(U)}{D} - c\frac{V}{D}, \quad (14.19)$$

whose equilibria are $(U_\infty, V_\infty) = (0, 0)$ and $(K, 0)$. The Jacobian matrix at $(U_\infty, 0)$ is

$$\begin{bmatrix} 0 & 1 \\ -\frac{g'(U_\infty)}{D} & -\frac{c}{D} \end{bmatrix}.$$

From $g'(K) < 0$ we know that $(K, 0)$ is a saddle point. The equilibrium $(0, 0)$ is an asymptotically stable node if $c^2 > 4Dg'(0)$ and a stable point if $c^2 < 4Dg'(0)$. By studying the phase portrait of the system it is possible to show that, if $(0, 0)$ is a node, then there is an orbit connecting the saddle point as $z \rightarrow -\infty$ and the equilibrium at $(0, 0)$ as $z \rightarrow \infty$. This orbit corresponds to a wave solution $u(x, t)$ traveling to the right, as shown in Fig. 14.2.

14.2.3 Disease Spread Models with Diffusion

If we take the simple endemic *SIR* model (3.1) considered in chapter 3 (with Λ being a constant and $d = 0$) and add diffusion, we obtain the reaction-diffusion model:

$$\begin{aligned} S_t &= \Lambda - \mu S - \beta SI + DS_{xx} \\ I_t &= \beta SI - (\alpha + d + \mu)I + DI_{xx}. \end{aligned} \quad (14.20)$$

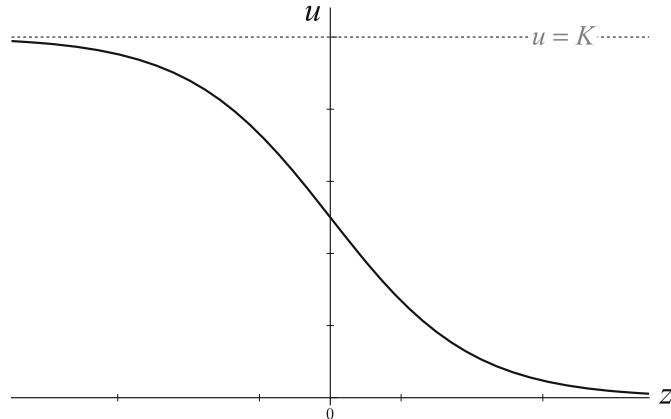


Fig. 14.2 A connecting orbit of Eq. (14.13)

A search for traveling wave solutions would lead to a four-dimensional system of ordinary differential equations. This approach can be carried out but is technically complicated, and we will not pursue it. Instead, we will consider a case study in which it is biologically reasonable to assume that susceptible members of the population do not diffuse, such as the spread of rabies in continental Europe during the period 1945–1985. This will permit a search for traveling wave solutions that requires the analysis of only a two-dimensional system.

The epidemic began on the edge of the German/Polish border, and its front moved westward at an average speed of about 30–60 km per year. The spread of the epidemic was essentially determined by the ecology of the fox population as foxes are the main carrier of the rabies under consideration.

A model was formulated in [22] to describe the front of the wave, its speed, and the total number of foxes infected after the front passes, and the connection of the wave speed to the so-called propagation speed of the disease.

We formulate a model describing susceptible (S) and infective (I) foxes. Assume that susceptible foxes are territorial and do not diffuse, but the rabies virus induces a loss of sense of territory. Consider the case when the population size has been scaled to 1, i.e., $0 < S(x, t) \leq 1$ and $0 \leq I(x, t) < 1$. Assume also a uniform initial density for the susceptibles with $S_0 = 1$. The simplest epidemic model under these assumptions is

$$\begin{aligned} S_t(x, t) &= -\beta S(x, t)I(x, t) \\ I_t(x, t) &= DI_{xx}(x, t) + \beta S(x, t)I(x, t) - \alpha I(x, t), \end{aligned} \tag{14.21}$$

where β is the transmission coefficient, α is the disease death rate of infective foxes, and D is the diffusion coefficient.

Consider the traveling wave solution with speed c :

$$u(z) = S(x - ct), \quad v(z) = I(x - ct),$$

where $z = x - ct$, and u and v are the *waveforms* (or *wave profiles*).

Substituting the above special form into the system (14.21), we obtain the system of ordinary differential equations:

$$Dv'' + cv' + \beta uv - \alpha v = 0, \quad cu' - \beta uv = 0, \quad (14.22)$$

where primes denote differentiation with respect to z . Assume the boundary conditions:

$$u(-\infty) = a, \quad u(\infty) = 1, \quad v(-\infty) = v(\infty) = 0, \quad (14.23)$$

where a is a constant to be determined. Substituting the second equation in (14.22) into the first equation and using the boundary conditions we obtain the system

$$\begin{aligned} u' &= \frac{\beta}{c} uv, \\ v' &= \frac{c}{D} \left[1 - u - v + \frac{\alpha}{\beta} \ln u \right]. \end{aligned} \quad (14.24)$$

Let (u_∞, v_∞) denote an equilibrium of system (14.24). Then $v_\infty = 0$. For u_∞ , it is either 0 or a solution of the equation

$$u - 1 = \frac{\alpha}{\beta} \ln u. \quad (14.25)$$

For $0 < u_\infty < 1$, a solution of (14.25), denoted by a , exists if and only if

$$a < \frac{\alpha}{\beta} < 1. \quad (14.26)$$

Thus, the two equilibrium points are $E_1 = (a, 0)$ and $E_2 = (1, 0)$. Observe that β/α is actually the basic reproduction number \mathcal{R}_0 of the corresponding ODE model (14.24).

The Jacobian matrix at $E = (u_\infty, 0)$ is

$$J(E) = \begin{pmatrix} 0 & \frac{\beta}{c} u_\infty \\ \frac{c}{D} \left(\frac{\alpha}{\beta} \frac{1}{u_\infty} - 1 \right) & -\frac{c}{D} \end{pmatrix}.$$

Using the condition in (14.26) and $\beta/\alpha > 1$, we know that $J(E_1)$ has negative trace and negative determinant, and hence, $E_1 = (a, 0)$ is a saddle point. Let

$$c^* = 2\sqrt{\beta D(1 - \alpha/\beta)},$$

then $E_2 = (1, 0)$ is a stable node if $c > c^*$ and a stable focus if $c < c^*$. Hence, a traveling wave solution satisfies $c > c^*$, in which case there is a connecting orbit from E_1 to E_2 .

In the model considered here we have neglected many important factors, including births and natural deaths and the long latent period for rabies. More accurate models can predict not only the observed wave pattern but also give a close approximation to the shape of the epidemic wave. Some additional sources of information about rabies modeling are [21, 25, 29, 30, 42].

With diffusion in both S and I , other difficult questions arise. One question is *diffusive instability*, meaning that an equilibrium is asymptotically stable for the system of ordinary differential equations obtained by a search for solutions that are constant in time but unstable for the system with diffusion. In general, diffusion tends to have a stabilizing effect and diffusive instability requires very specific conditions on the coefficients.

We have looked only at diffusion in one-dimensional space. In the extension to higher space dimensions, the term u_{xx} can be replaced by the Laplacian of the function u . In many problems for two-dimensional space there is radial symmetry, which can be incorporated by describing the Laplacian in polar coordinates and assuming u to be independent of the angular coordinate. If the radial variable is denoted by r , the term Du_{xx} would be replaced by $u_{rr} + u_r/r$.

Diffusion problems may be mathematically very complicated, and they require a considerable amount of mathematical background. One important possibility is the formation of spatial patterns, first suggested by A. M. Turing in 1952 [36]. These require more knowledge of partial differential equations than wish to assume. Some examples of pattern formation in diffusive epidemic models may be found in [34, 40] and further information may be found in [27].

14.3 Project: A Model with Three Patches

The epidemic patch model, (14.4) and (14.5), is for the case of two patches. We can consider an extension of the model to include three patches. In this case, individuals leaving Patch i can travel to either one of the two other patches. Let $m_{ji} \geq 0$ represent the fractions of individuals moving into Patch j from Patch i . Then $m_{ii} = 0$, $r_{ii} = 0$, $\sum_{j=1}^3 m_{ji} = 1$, and $g_i m_{ji}$ denotes the travel rate entering Patch j from Patch i . Consider the case in which the transmission coefficient β_{ikj} depends only on the

Patch j where the transmission occurs, i.e., $\beta_{ikj} = \beta_j$. Then the system on resident Patch i (with $i = 1, 2, 3$) reads

$$\begin{aligned} S'_{ii} &= \sum_{k=1}^3 r_{ik} S_{ik} - g_i S_{ii} - \sum_{k=1}^3 \kappa_i \beta_i \frac{S_{ii} I_{ki}}{N_i^p} + \mu \left(\sum_{k=1}^3 N_{ik} - S_{ii} \right) \\ I'_{ii} &= \sum_{k=1}^3 r_{ik} I_{ik} - g_i I_{ii} + \sum_{k=1}^3 \kappa_i \beta_i \frac{S_{ii} I_{ki}}{N_i^p} - (\gamma + \mu) I_{ii}, \end{aligned} \quad (14.27)$$

and for $j \neq i$

$$\begin{aligned} S'_{ij} &= g_i m_{ji} S_{ii} - r_{ij} S_{ij} - \sum_{k=1}^3 \kappa_j \beta_j \frac{S_{ij} I_{kj}}{N_j^p} - \mu S_{ij} \\ I'_{ij} &= g_i m_{ji} I_{ii} - r_{ij} I_{ij} + \sum_{k=1}^3 \kappa_j \beta_j \frac{S_{ij} I_{kj}}{N_j^p} - (\gamma + \mu) I_{ij}, \end{aligned} \quad (14.28)$$

with $N_i^p = N_{1i} + N_{2i} + N_{3i}$, the number present in Patch i .

Question 1 The total population sizes satisfy the following equations:

$$\begin{aligned} N'_{ii} &= \sum_{k=1}^3 r_{ik} N_{ik} - g_i N_{ii} + \mu \left(\sum_{k=1}^3 N_{ik} - N_{ii} \right), \quad i = 1, 2, 3, \\ N'_{ij} &= g_i m_{ji} N_{ii} - r_{ij} N_{ij} - \mu N_{ij}, \quad i \neq j. \end{aligned} \quad (14.29)$$

- (a) Show that the total resident population in Patch i , $\sum_{k=1}^3 N_{ik}$, remains constant at all time. Let $N_{i0} = \sum_{k=1}^3 N_{ik}$ for $i = 1, 2, 3$.
- (b) Show that the system (14.29) has the asymptotically stable equilibrium

$$\hat{N}_{ii} = \left(\frac{1}{1 + g_i \sum_{k=1}^3 \frac{m_{ki}}{\mu + r_{ik}}} \right) N_{i0}, \quad (14.30)$$

and for $j \neq i$

$$\hat{N}_{ij} = \frac{g_i m_{ji}}{\mu + r_{ij}} \hat{N}_{ii}. \quad (14.31)$$

Question 2 Let \hat{N}_{iq} be as given in (14.30) and (14.31), and let $\hat{N}_q^p = \sum_{i=1}^3 \hat{N}_{iq}$. It is easy to show that $\mathcal{R}_{0i} = \kappa_i \beta_i / (\mu + \gamma)$ is the isolated basic reproduction number of Patch i (i.e., when there is no travel between patches). Consider the order of infective variables

$$(I_{11}, I_{12}, I_{13}, I_{21}, I_{22}, I_{23}, I_{31}, I_{32}, I_{33}).$$

- (a) Show that the basic reproduction number \mathcal{R}_0 for the model (14.27)–(14.28) is given by the dominant eigenvalue of the matrix FV^{-1} , where F is a block matrix with nine blocks, and each block F_{ij} is a 3×3 matrix with the form $F_{ij} = \text{diag}(f_{ijq})$ with

$$f_{ijq} = \kappa_q \beta_q \frac{\hat{N}_{iq}}{\hat{N}_q^p}, \quad q = 1, 2, 3$$

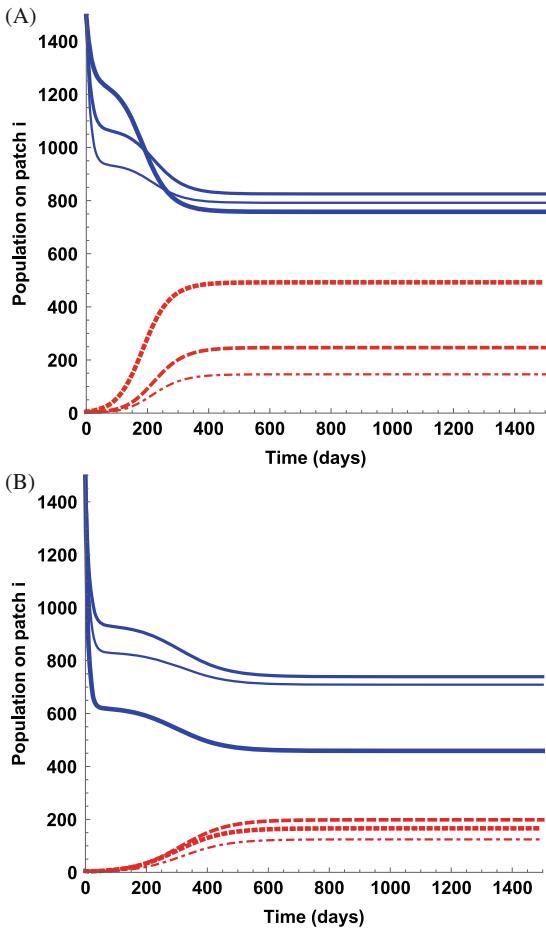
and

$$V = \begin{pmatrix} a_1 & -r_{12} & -r_{13} & 0 & 0 & 0 & 0 & 0 & 0 \\ -g_1 m_{21} & b_{12} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -g_1 m_{31} & 0 & b_{13} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & b_{21} & -g_2 m_{12} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -r_{21} & a_2 & -r_{23} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -g_2 m_{32} & b_{23} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & b_{31} & 0 & -g_3 m_{13} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & b_{32} & -g_3 m_{23} \\ 0 & 0 & 0 & 0 & 0 & 0 & -r_{31} & -r_{32} & a_3 \end{pmatrix},$$

where $a_i = g_i + \gamma + \mu$ and $b_{ik} = r_{ik} + \gamma + \mu$.

- (b) Consider the special case when $\beta_i = \beta$ for $i = 1, 2, 3$. Fix all parameters except β . Then $\mathcal{R}_0 = \mathcal{R}_0(\beta)$ is a function of β . Consider the set of parameters: $\gamma = 1/25$, $\mu = 1/(75 \times 365)$, $g_1 = 0.01$, $g_2 = 0.02$, $g_3 = 0.03$, $m_{ij} = 0.5$ for $i \neq j$, $r_{ij} = 0.05$ for $i \neq j$, $\kappa_i = \kappa = 1$ and $N_{0i} = N_0 = 1500$ for $i = 1, 2, 3$.
- (i) Plot \mathcal{R}_0 as a function of β . What is the threshold value β_c such that $\mathcal{R}_0(\beta) < 1$ for all $\beta < \beta_c$?
 - (ii) Figure 14.3 shows the number of susceptible and infective individuals of three resident populations for the case of $\beta = 0.025$ with different values of κ_i and g_i . Experiment with other set of parameters to observe how the prevalence within these patches will change.
- (c) Consider the same set of parameter values as given in (a) except that the parameters κ_1 and N_{01} for Patch 1 can vary. Numerically plot the solutions for different values of these parameters and describe your observations.
- (d) We can also explore the effect of travel rates g_i on the prevalence. Let $\beta = 0.025$ and fix all parameters as in part (b) except g_1 and κ_1 . Determine a couple of sets of parameter values of g_1 and κ_1 that can determine whether or not the infection on Patch 1 can go extinct.

Fig. 14.3 Time plots of the susceptibles (solid) and infectives (dashed) individuals for the resident populations 1 (thicker curve), 2 (intermediate), and 3 (thinner curve). The parameter values used are $\beta = 0.025$, $\kappa_1 = 3$, $\kappa_2 = 2$, $\kappa_3 = 1$, $g_1 = 0.01$, $g_2 = 0.02$, $g_3 = 0.03$ in (A) and $g_1 = 0.07$, $g_2 = 0.03$, $g_3 = 0.04$ in (B). All other parameters have the same values as in part (A)



14.4 Project: A Patch Model with Residence Time

Consider a model consisting of (14.7) and (14.8) with parameters

$$N_1 = N_2 = 1,000,000, \quad \beta_1 = 0.3056, \quad \beta_2 = 0.1, \quad \gamma = 1/6.5.$$

Question 1 Begin with an assumption that mixing is symmetric, $p_{12} = p_{21}$. Calculate the final epidemic size with several choices of $p_{11} = p_{22}$.

Question 2 Calculate the effect on epidemic size of assuming no travel to the patch with a higher contact rate by assuming $p_{22} = 1$, $p_{21} = 0$ for several choices of p_{12} .

Question 3 Calculate the effect on epidemic size of banning all travel, $p_{12} = p_{21} = 0$.

References

1. Anderson, R.M. H. C. Jackson, R. M. May & A. M. Smith (1981) Population dynamics of fox rabies in Europe, *Nature*, **289**: 765–771.
2. Anderson, R.M. & May, R.M. (1979) Population biology of infectious diseases I, *Nature* **280**: 361–367.
3. Arino, J. (2009) Diseases in metapopulations, in *Modeling and Dynamics of Infectious Diseases*, Z. Ma, Y. Zhou, J. Wu (eds.), Series in Contemporary Applied Mathematics, World Scientific Press, Vol. **11**; 64–122.
4. Arino, J., R. Jordan,& P. van den Driessche (2007) Quarantine in a multispecies epidemic model with spatial dynamics, *Math. Biosoc.* **206**: 46–60.
5. Arino, J. & P. van den Driessche (2003) The basic reproduction number in a multi-city compartmental epidemic model, *Lecture Notes in Control and Information Science* **294**: 135–142.
6. Arino, J. & P. van den Driessche (2003) A multi-city epidemic model, *Mathematical Population Studies* **10**: 175–93.
7. Arino, J. & P. van den Driessche (2006) Metapopulation epidemic models, *Fields Institute Communications* **48**: 1–13.
8. Aronson, D.G. (1977) The asymptotic spread of propagation of a simple epidemic, in *Nonlinear Diffusion* (W.G. Fitzgibbon & H.F. Walker, eds.), *Research Notes in Mathematics* **14**, Pitman, London.
9. Bichara, D., Y. Kang, C. Castillo-Chavez, R. Horan & C. Perrings (2015) *SIS* and *SIR* epidemic models under virtual dispersal, *Bull. Math. Biol.* **77**: 2004–2034.
10. Britton, N.F. (2003) *Essentials of Mathematical Biology*, Springer Verlag, Berlin- Heidelberg.
11. Capasso, V. (1993) *Mathematical Structures of Epidemic Systems*, Lect. Notes in Biomath. **83**, Springer-Verlag, Berlin-Heidelberg-New York.
12. Castillo-Chavez, C., D. Bichara, & B. Morin (2016) Perspectives on the role of mobility, behavior, and time scales on the spread of diseases, *Proc. Nat. Acad. Sci.* **113**: 14582–14588.
13. Diekmann, O. (1978) Run for your life. a note on the asymptotic speed of propagation of an epidemic, *J. Diff. Eqns.* **33**: 58–73.
14. Diekmann, O. & J.A.P. Heesterbeek (2000) *Model building, analysis and interpretation*, John Wiley & Sons, New York.
15. Edelstein-Keshet, L. (2005) *Mathematical Models in Biology*, Classics in Applied Mathematics, SIAM, Philadelphia.
16. Espinoza, B., V. Moreno, D. Bichara, & C. Castillo-Chavez (2016) Assessing the Efficiency of Movement Restriction as a Control Strategy of Ebola. In *Mathematical and Statistical Modeling for Emerging and Re-emerging Infectious Diseases*, pp. 123–145, Springer International Publishing.
17. Fitzgibbon, W.E., M.E. Parrott, & G.F. Webb (1995) Diffusive epidemic models with spatial and age-dependent heterogeneity, *Discrete Contin. Dyn. Syst.* **1**: 35–57.
18. Fitzgibbon, W.E., M.E. Parrott, & G.F. Webb (1996) A diffusive age-structured SEIRS epidemic model, *Methods Appl. Anal.* **3**: 358–369.
19. Grenfell, B.T., & A. Dobson, (eds.) (1995) *Ecology of Infectious Diseases in Natural Populations*, Cambridge University Press, Cambridge, UK.
20. Hadeler, K.P. (2003) The role of migration and contact distributions in epidemic spread, in *Bioterrorism: Mathematical Modeling Applications in Homeland Security* (H. T. Banks and C. Castillo-Chavez eds.), SIAM, Philadelphia, pp. 199–210.
21. Källén, A. (1984) Thresholds and travelling waves in an epidemic model for rabies, *Nonlinear Anal.* **8**: 651–856.
22. Källén, A., P. Arcuri & J. D. Murray (1985) A simple model for the spatial spread and control of rabies, *J. Theor. Biol.* **116**: 377–393.
23. Kot, M. (2001) *Elements of Mathematical Ecology*, Cambridge University Press.

24. Lloyd, A. & R.M. May (1996) Spatial heterogeneity in epidemic models. *J. Theor. Biol.* **179**: 1–11.
25. MacDonald, D.W. (1980) Rabies and Wildlife. A Biologist's Perspective, Oxford University Press, Oxford.
26. Mollison, D. (1977) Spatial contact models for ecological and epidemic spread, *J. Roy. Stat. Soc. Ser. B* **39**: 283–326.
27. Murray, J.D. (2002) Mathematical Biology, Vol. I, Springer-Verlag, Berlin-Heidelberg-New York.
28. Murray, J.D. (2002) Mathematical Biology, Vol. II, Springer-Verlag, Berlin-Heidelberg-New York.
29. Murray, J.D., E. A. Stanley & D. L. Brown (1986) On the spatial spread of rabies among foxes, *Proc. Roy. Soc. Lond. B* **229**: 111–150.
30. Ou, C & J. Wu (2006) Spatial spread of rabies revisited: role of age-dependent different rate of foxes and non-local interaction, *SIAM J. App. Math.* **67**: 138–163.
31. Radcliffe, J. & L. Rass (1986) The asymptotic spread of propagation of the deterministic non-reducible n -type epidemic, *J. Math. Biol.* **23**: 341–359.
32. Sattenspiel, L. (2003) Infectious diseases in the historical archives: a modeling approach. In: D.A. Herring & A.C. Swedlund, (eds) *Human Biologists in the Archives*, Cambridge University Press: 234–265.
33. Sattenspiel, L. & K. Dietz (1995) A structured epidemic model incorporating geographic mobility among regions *Math. Biosc.* **128**: 71–91.
34. Sun, G.-Q. (2012) Pattern formation of an epidemic model with diffusion, *Nonlin. Dyn.* **69**: 1097–1104.
35. Thieme, H.R. (1979) Asymptotic estimates of the solutions of nonlinear integral equations and asymptotic speeds for the spread of populations, *J. Reine Angew. Math.* **306**: 94–121.
36. Turing, A.M. (1952) The chemical basis of morphogenesis, *Phil. Trans. Roy. Soc. London B* **237**: 37–72.
37. van den Bosch, F., J.A.J. Metz, & O. Diekmann (1990) The velocity of spatial population expansion, *J. Math. Biol.* **28**: 529–565.
38. van den Driessche, P. (2008) Spatial structure: Patch models, in *Mathematical Epidemiology* (F. Brauer, P. van den Driessche, J. Wu, eds.), Lecture Notes in Mathematics, Mathematical Biosciences subseries 1945, Springer, pp. 179–189.
39. van den Driessche P. & J. Watmough (2002) Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission, *Math. Biosc.* **180**: 29–48.
40. Wang, W.-M,H.-Y. Liu, Y.-L. Cai, & Z.-Q. Li (2011) Turing pattern selection in a reaction-diffusion epidemic model, *Chinese Phys.* **20**: 07472.
41. Weinberger, H.F. (1981) Some deterministic models for the spread of genetic and other alterations, *Biological Growth and Spread* (W. Jaeger, H. Rost, P. Tautu, eds.), Lecture Notes in Biomathematics **38**, Springer Verlag, Berlin-Heidelberg-New York, pp. 320–333.
42. Wu, J. (2008) Spatial structure: Partial differential equations models, in *Mathematical Epidemiology* (F. Brauer, P. van den Driessche, J. Wu, eds.), Lecture Notes in Mathematics, Mathematical Biosciences subseries 1945, Springer, pp. 191–203.

Chapter 15

Epidemiological Models Incorporating Mobility, Behavior, and Time Scales



15.1 Introduction

The work of Eubank et al. [24], Sara del Valle et al. [44], Chowell et al. [7, 18], and Castillo-Chavez and Song [13] have highlighted the impact of modified modeling approaches that incorporate heterogeneous modes of mobility within variable environments in order to study their impact on the dynamics of infectious diseases. Castillo-Chavez and Song [13], for example, proceeded to highlight a Lagrangian perspective, that is, the use of models that keep track at all times of the identity of each individual. This approach was used to study the consequences of deliberate efforts to transmit smallpox in a highly populated city, involving transient sub-populations and the availability of massive modes of public transportation.

Here, a multi-group epidemic Lagrangian framework where mobility and the risk of infection are functions of patch residence time and local environmental risk is introduced. This Lagrangian approach has been used within classical contact epidemiological (that is, transmission is due to “contacts” between individuals) formulations in the context of a possible deliberate release of biological agents [2, 13]. The Lagrangian approach is introduced here as a modeling approach that explicitly avoids the assignment of heterogeneous contact rates to individuals. The use of contacts or activity levels and the view that transmission is due to collisions between individuals has a long history and it is conceptually consistent with the way we envision disease transmission between susceptible and infectious individuals. However, contacts are hard to define and consequently, at least in the context of communicable diseases, impossible to measure in various settings. Is it possible to capture interactions of individual mathematically in a way different from the notion of contacts? The approach that is proposed focuses on the use of modeling frameworks that involve patches/environments defined or characterized by risks of infection that are functions of the time spent in each environment/patch. These patches/environments may or may not have permanent hosts and they may be used to account for places of “transitory” residence like mass transportation systems or

hospitals or forests, to name but a few possibilities. Each environment or patch is characterized by the expected risk of infection of visitors as a function of time spent in such an environment. For example, a population near a forest may have some of its individuals spend time in the forest. Those who like the outdoors may be exposed longer to vectors than those who do not visit the forest. Consequently, the possibility of acquiring a vector-borne disease is a function of, among other factors, how long an individual spends in the forest each day. Similarly, individuals that use mass transportation routinely (during rush hour) are at a higher risk of acquiring a communicable disease including common colds, and it makes sense to assume that the risk may be a function of how long each individual spends each day commuting to work or to school. In other words, the average time spent in a community defined as a collection of environments that determined *a priori* the risk of acquiring an infection is at the heart of the Lagrangian approach.

What is the Lagrangian approach and what does the theory tell us about the dynamics of such models in epidemic settings? We revisit this framework in possibly the simplest general setting that of a susceptible–infected–susceptible (SIS) epidemic multi-group model. We collect some of the mathematical formulae and results in the context of this general SIS multi-group model as reported in the literature [4, 6, 7, 11]. We proceed to identify basic reproduction numbers \mathcal{R}_0 as a function of the associated multi-patch residence-time matrix \mathbb{P} ($p_{i,j} : i, j = 1, 2, 3 \dots n$), which determines the proportion of time that a resident of Patch i spends in environment j . The analysis shows that the n -patch SIS model (as long as it is a strongly connected system) has a unique globally stable endemic equilibrium when $\mathcal{R}_0 > 1$, and a globally stable disease-free equilibrium when $\mathcal{R}_0 \leq 1$. We have used simulations to generate insights on the impact that the residence matrix \mathbb{P} has on infection levels within each patch. Model results [4, 6, 7, 11] show that the infection risk vector, which characterizes environments by risk to a pre-specified disease (measured by \mathbb{B}), and the residence-time matrix \mathbb{P} both play an important role in determining, for example, whether or not endemicity is reached at the patch level. Further, it is shown that the right combinations of environmental risks (\mathbb{B}) and mobility behavior (\mathbb{P}) are capable of promoting or suppressing infection within particular patches. The theoretical results [4, 6, 7, 11] are used to characterize patch-specific disease dynamics as a function of the time spent by residents and visitors in patches of interest. These results have helped classify patches as sources or sinks of infection, depending, of course, on the risk (\mathbb{B}) and mobility (\mathbb{P}) matrices. In general a residence-time matrix \mathbb{P} cannot be made of constant entries in realistic settings. In fact the entries of \mathbb{P} may depend on disease prevalence levels. We have explored some simple situations, via simulations, when the entries of the matrix \mathbb{P} are state-dependent [4, 6, 7, 11]. The analysis and simulations for specific diseases are illustrated later in this chapter. They are used to highlight some of the possible differences that arise from having a state-dependent residence-time matrix \mathbb{P} .

15.2 General Lagrangian Epidemic Model in an SIS Setting

The following SIS model involving n -patches (environments) is introduced in [7]:

$$\begin{aligned} S'_i &= b_i - d_i S_i + \gamma_i I_i - S_i \lambda_i(t) \\ I'_i &= S_i \lambda_i(t) - \gamma_i I_i - d_i I_i \\ N'_i &= b_i - d_i N_i, \end{aligned} \quad (15.1)$$

where b_i , d_i , and γ_i denote the per-capita birth, natural death, and recovery rates, respectively, for $i = 1, 2, 3 \dots n$. The infection rates $\lambda_i(t)$ have the form:

$$\lambda_i(t) = \sum_{j=1}^n \beta_j p_{ij} \frac{\sum_{k=1}^n p_{kj} I_k}{\sum_{k=1}^n p_{kj} N_k}, \quad i = 1, 2, \dots, n, \quad (15.2)$$

where p_{ij} denotes the proportion of susceptibles from Patch i who are currently in Patch j , β_j is the risk of infection in Patch j , and the last fraction represents the proportion of infected in Patch j . Using the approach of the next generation matrix, the basic reproduction number \mathcal{R}_0 can be derived using the following system:

$$\dot{I}_i = \left(\frac{b_i}{d_i} - I_i \right) \lambda_i(t) - (\gamma_i + d_i) I_i, \quad i = 1, 2, \dots, n.$$

As shown in the next section, \mathcal{R}_0 is a function of the risk vector $\mathcal{B} = (\beta_1, \beta_2, \dots, \beta_n)^t$ and the residence times matrix $\mathbb{P} = (p_{ij})$, $i, j = 1, \dots, n$, and it is shown in [7] that whenever \mathbb{P} is irreducible (patches are strongly connected), the disease-free steady state is globally asymptotically stable if $\mathcal{R}_0 \leq 1$ and a unique interior equilibrium exists and is globally asymptotically stable if $\mathcal{R}_0 > 1$.

While a specific formula for the multi-patch basic reproduction number cannot be computed explicitly, it is possible in this case to find expressions for the patch-specific basic reproduction number. In fact, we have

$$\mathcal{R}_{0i}(\mathbb{P}) = \mathcal{R}_{0i} \times \sum_{j=1}^n p_{ji},$$

where \mathcal{R}_{0i} are the local basic reproduction numbers ($i = 1, 2, 3, \dots, n$) computed when the patches are isolated from each other. From the \mathcal{R}_{0i} ($i = 1, 2, 3, \dots, n$), the role that the relative risk that each environment (patch) plays, namely $\frac{\beta_j}{\beta_i}$, can be assessed. Further, the role that residence times play in keeping track of the appropriate fraction of the population involved in a given patch is given by $\frac{(p_{ij} b_i / d_i)}{\sum_{k=1}^n p_{kj} b_k / d_k}$. In other words, this patch specific \mathcal{R}_{0i} ($i = 1, 2, 3, \dots$) captures the impact of the \mathbb{P} and \mathcal{B} matrices.

In short, if $\mathcal{R}_{0i}(\mathbb{P}) > 1$ the disease persists in Patch i and furthermore, if $p_{kj} = 0$ for all $k = 1, 2, \dots, n$ and $k \neq i$ whenever $p_{ij} > 1$, then it is shown that the disease dies in Patch i if $\mathcal{R}_{0i}(\mathbb{P}) < 1$, that is, patch-specific basic reproduction numbers help characterize disease dynamics at the patch level [7].

We can look first at the following example of a multi-patch SIR model for a single outbreak:

$$\begin{aligned} S'_i &= -S_i \lambda_i(t), \\ I'_i &= S_i \lambda_i(t) - \alpha_i I_i, \\ R'_i &= \alpha_i I_i, \quad i = 1, 2, \end{aligned} \tag{15.3}$$

where S_i , I_i , and R_i denote the population of susceptible, infected, and recovered immune individuals, respectively, in Patch i , and $N_i = S_i + I_i + R_i$. This model is the same as the model (14.7–14.8) in the preceding chapter. The parameter α_i denotes the per-capita recovery rate in Patch i and $\lambda_i(t)$ are given in (15.2).

In the rest of this chapter we make use of this Lagrangian modeling perspective (disease-specific versions) to carry out preliminary studies, in rather simple set ups, of the role of mobility in reducing or enhancing the transmission of specific diseases in regions of variable risk for the case of two patches. Numerical results are used to illustrate the power and limitations of this approach. Lagrangian models are used to explore the role that mobility plays in disease transmission for the cases of Ebola, tuberculosis, and Zika in simplified settings. Figure 15.1 represents a schematic representation of the Lagrangian dispersal between two patches.

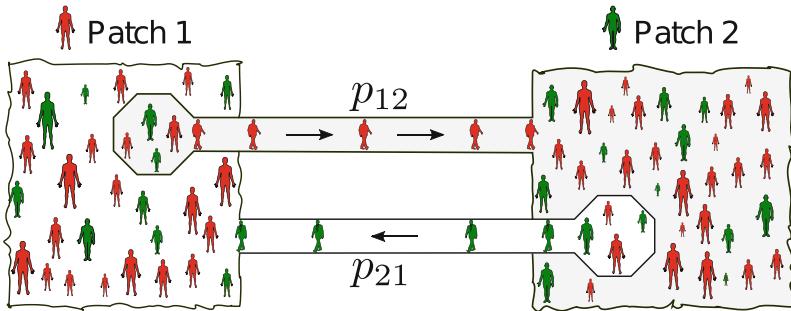


Fig. 15.1 Dispersal of individuals via a Lagrangian approach

15.3 Assessing the Efficiency of Cordon Sanitaire as a Control Strategy of Ebola

During the 2014 Ebola Epidemic in West Africa, it was observed [10, 46] that the rate of growth of the Ebola epidemic seemed to be increasing rather than decreasing as is standard in the study of epidemics. In other words, the reproduction number tends to decrease in time rather than increase. The evidence provided by the data and our analysis indicated that something was not right. We learned that troops were being used to prevent individuals from moving out of the most devastated communities facing Ebola. The use of *cordons sanitaires* seemed to be implemented even though past experiences have shown them to have a deleterious effect. Here, we formulate a two-patch mathematical model for Ebola virus disease (EVD) dynamics to highlight the potential lack of effectiveness or the deleterious impact of impeding mobility (*cordons sanitaires*). Via simulations, we look at the role of mandatory mobility restrictions and their impact on disease dynamics and epidemic final size. It is shown that mobility restrictions between high and low risk areas of closely linked communities are likely to have a deleterious impact on *overall* levels of infection in the total population involved.

15.3.1 Formulation of the Model

The community of interest is assumed to be composed of two adjacent regions facing highly distinct levels of EVD infection and having access to a highly differentiated public health system (the haves and have-nots). There are differences in population density, availability of medical services, and isolation facilities. The need of those in the high-risk area to travel to the low-risk area is high as the jobs are in the well-off community. For Ebola, it may be unrealistic to assume susceptibles and infectives travel at the same rate. We let N_1 denote the population in Patch 1 (high risk) and N_2 the population in Patch 2 (low risk). The classes S_i , E_i , I_i , R_i represent the susceptible, exposed, infective, and recovered sub-populations in Patch i ($i = 1, 2$). The class D_i represents the number of disease induced deaths in Patch i . The dispersal of individuals is modeled via the Lagrangian approach defined in terms of residence times [4, 7].

The numbers of new infections per unit of time are based on the following assumptions:

- The density of infected individuals mingling in Patch 1 at time t , who are only capable of infecting susceptible individuals currently in Patch 1 at time t , that is, the *effective* infectious proportion in Patch 1 is given by

$$\frac{p_{11}I_1 + p_{21}I_2}{p_{11}N_1 + p_{21}N_2},$$

where p_{11} denotes the proportion of time that residents from Patch 1 spend in Patch 1 and p_{21} the proportion of time that residents from Patch 2 spend in Patch 1.

- The number of newly infected Patch 1 residents while sojourning in Patch 1 is therefore given by

$$\beta_1 p_{11} S_1 \left(\frac{p_{11} I_1 + p_{21} I_2}{p_{11} N_1 + p_{21} N_2} \right).$$

- The number of new infections within members of Patch 1, in Patch 2 per unit of time is therefore

$$\beta_2 p_{12} S_1 \left(\frac{p_{12} I_1 + p_{22} I_2}{p_{12} N_1 + p_{22} N_2} \right),$$

where p_{12} denotes the proportion of time that residents from Patch 1 spend in Patch 2 and p_{22} the proportion of time that residents from Patch 2 spend in Patch 2. Hence, the effective density of infected individuals in Patch j is given by

$$p_{1j} N_1 + p_{2j} N_2, \quad j = 1, 2.$$

If we further assume that infection by dead bodies occurs only at the local level (bodies are not moved), then, by following the same rationale as in model (15.3), we arrive at the following model:

$$\begin{aligned} S'_i &= -S_i \lambda_i(t) - \varepsilon_i \beta_i p_{ii} S_i \frac{D_i}{N_i}, \\ E'_i &= S_i \lambda_i(t) + \varepsilon_i \beta_i p_{ii} S_i \frac{D_i}{N_i} - \kappa E_i, \\ I'_i &= \kappa E_i - \gamma I_i, \\ D'_i &= f_d \gamma I_i - v D_i, \\ R'_i &= (1 - f_d) \gamma I_i + v D_i, \\ N_i &= S_i + E_i + I_i + D_i + R_i, \quad i = 1, 2, \end{aligned} \tag{15.4}$$

where $\lambda_i(t)$ are given in (15.2).

15.3.2 Simulations

Simulations show that if only individuals from the high-risk region (Patch 1) were allowed to travel, then the epidemic final size can go under the *cordon sanitaire* level. Figure 15.2 captures the patch-specific prevalence levels for mobility values of $p_{12} = 0, 0.2, 0.4, 0.6$ with $p_{21} = 0$ (no movement). Disease dispersal, if the

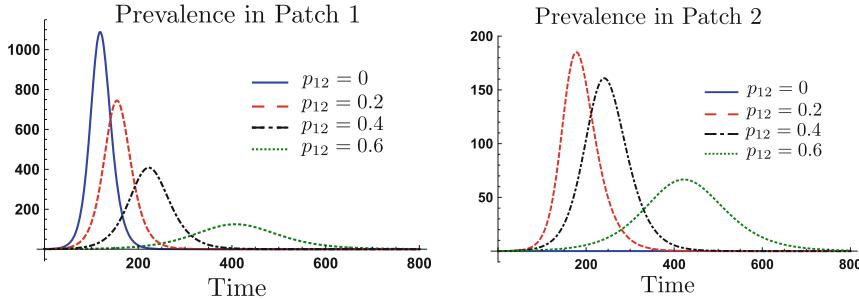


Fig. 15.2 Dynamics of prevalence in each patch for values of mobility $p_{12} = 0\%$, 20% , 40% , 60% and $p_{21} = 0$, with parameters: $\varepsilon_{1,2} = 1.1$, $\mathcal{R}_{01} = 2.45$, $\mathcal{R}_{02} = 0.9$, $f_d = 0.7$, $k = 1/7$, $v = 1/2$, $\gamma = 1/7$

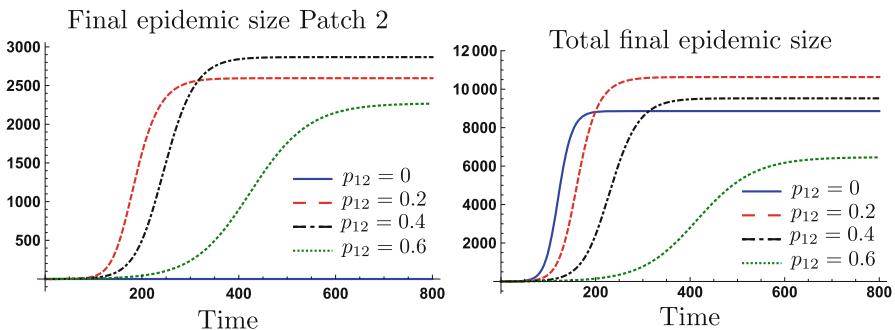


Fig. 15.3 Dynamics of patch specific and total final epidemic size, for mobility values $p_{12} = 0\%$, 20% , 40% , 60% and $p_{21} = 0$, with parameters: $\varepsilon_{1,2} = 1.1$, $\mathcal{R}_{01} = 2.45$, $\mathcal{R}_{02} = 0.9$, $f_d = 0.7$, $k = 1/7$, $v = 1/2$, $\gamma = 1/7$

disease spreads to a totally susceptible region, means that the secondary infections produced in the low-risk region reduce the overall two-patch prevalence, due to its access to better sanitary conditions and resources. However, there is a cost to the low-risk patch not only for the services provided but also for the generation of a larger number of secondary cases than if the “borders” were closed. Figure 15.3 shows that different mobility regimes can increase or decrease the total epidemic final size. In the presence of “low mobility” levels ($p_{12} = 0.2, 0.4$), the total final size curve may turn out to be greater than the *cordon sanitaire* case. We observe that the nonlinear impact of mobility on the total epidemic final size can bring it below the cordoned case even under relative “high mobility” regimes. This result highlights the trade-off that comes from reducing individuals’ time spent in a high-risk region versus exposing a totally susceptible population living in a safer region. Under certain mobility conditions, the results of such a trade-off are beneficial for the Global Commons.

Fig. 15.4 Dynamics of maximum final size and maximum prevalence in Patch 1 with parameters: $\varepsilon_{1,2} = 1.1$, $\mathcal{R}_{01} = 2.45$, $\mathcal{R}_{02} = 0.9$, $f_d = 0.7$, $k = 1/7$, $v = 1/2$, $\gamma = 1/7$

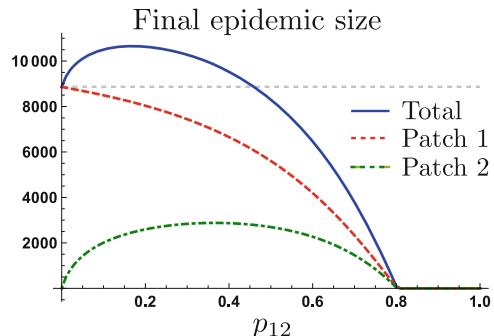
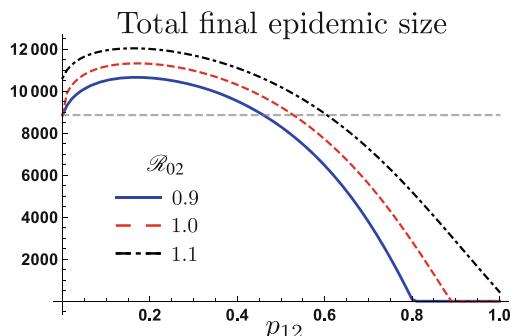


Fig. 15.5 Dynamics of final epidemic size in the one way case with parameters: $\varepsilon_{1,2} = 1.1$, $\mathcal{R}_{01} = 2.45$, $\mathcal{R}_{02} = 0.9$, 1.0 , 1.1 , $f_d = 0.7$, $k = 1/7$, $v = 1/2$, $\gamma = 1/7$



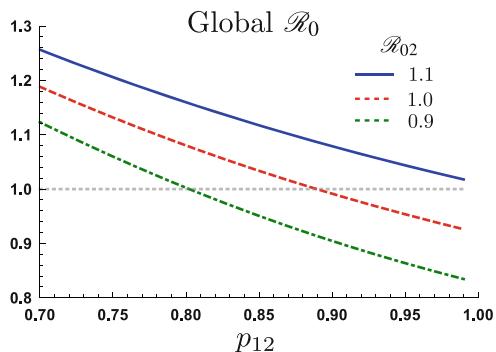
Further, in order to clarify the effects of residence times on total final epidemic size, we proceeded to analyze its behavior under one way mobility. Figure 15.4 shows the *cordon sanitaire* (dashed gray line), patch specific, and the total epidemic final size for various possible mobility scenarios, $p_{12} \in [0, 1]$. We see that one way mobility reduces Patch 1 epidemic final size while increasing the Patch 2 final number of infections. We see that the total epidemic final size under low mobility ($p_{12} < 0.5$) is above the cordoned case. We also observe that Patch 2 sanitary conditions play an important role under high mobility regime bringing the total epidemic final size below the *cordon sanitaire* scenario ($p_{12} > 0.5$).

Moreover, results suggest that for $\mathcal{R}_{02} < 1$ extremely high mobility levels might eradicate an Ebola outbreak. It is important to stress that mobility reducing the total epidemic final size is dependent not only on the residence times and mobility type, but also on the patch-specific prevailing infection rates. Figure 15.5 shows that if $\mathcal{R}_{02} > 1$ mobility is not capable of leading the total epidemic final size towards zero. Figure 15.6 shows that the global basic reproduction number decreases monotonically as one way mobility increases. However, it is not capable of capturing the harmful effect of low mobility levels, increasing the total epidemic final size. Mobility on its own is not always enough to reduce \mathcal{R}_0 below the critical

Fig. 15.6 Dynamics of \mathcal{R}_0

with parameters:

$$\begin{aligned}\varepsilon_{1,2} &= 1, \mathcal{R}_{01} = 2.45, \mathcal{R}_{02} = \\&0.9, 1.0, 1.1, f_d = 0.708, k = \\&1/7, \alpha = 0, v = 1/2, \gamma = \\&1/6.5\end{aligned}$$



threshold. Instead, bringing the global \mathcal{R}_0 less than one requires reducing local risk, that is, getting a lower \mathcal{R}_{02} .

15.4 *Mobility and Health Disparities on the Transmission Dynamics of Tuberculosis

TB dynamics is the result of complex epidemiological and socio-economical interactions between and among individuals living in highly heterogeneous regional conditions. Many factors impact TB transmission and progression. A model is introduced to enhance the understanding of TB dynamics in the presence of diametrically distinct rates of infection and mobility. The dynamics are studied in a simplified world consisting of two patches, that is, two risk-defined environments, where the impact of short-term mobility and variations in reinfection and infection rates are assessed. The modeling framework captures “daily dynamics” of individuals within and between places of residency, work, or business. Activities are modeled by the proportion of time spent in environments (patches) having different TB infection risk. Mobility affects the effective population size of each Patch i (home of i -residents) at time t and they must also account for visitors and residents of Patch i , at time t . The impact that effective population size and the distribution of individuals’ residence times in different patches have on TB transmission and control is explored using selected scenarios where risk is defined by the estimated or perceived first time infection and/or exogenous reinfection rates. Model simulation results suggest that, under certain conditions, allowing infected individuals to move from high to low TB prevalence areas (for example, via the sharing of treatment and isolation facilities) may lead to a reduction in the total TB prevalence in an overall, two-patch, population.

15.4.1 *A Two-Patch TB Model with Heterogeneity in Population Through Residence Times in the Patches

Using a similar approach to model formulation, we consider the following model for the dynamics of TB in two patches:

$$\begin{aligned}\dot{S}_i &= \mu_i N_i - S_i \lambda_i(t) - \mu_i S_i, \\ \dot{L}_i &= q S_i \lambda_i(t) - L_i \hat{\lambda}_i(t) - (\gamma_i + \mu_i) L_i + \rho_i I_i, \\ \dot{I}_i &= (1 - q) S_i \lambda_i(t) + L_i \hat{\lambda}_i(t) + \gamma L_i - (\mu_i + \rho_i) I_i, \quad i = 1, 2,\end{aligned}\tag{15.5}$$

where $\lambda_i(t)$ is the same as in (15.2) and

$$\hat{\lambda}_i(t) = \sum_{j=1}^2 \delta_j p_{ij} \frac{\sum_{k=1}^2 p_{kj} I_k}{\sum_{k=1}^2 p_{kj} N_k}, \quad i = 1, 2.$$

15.4.2 *Results: The Role of Risk and Mobility on TB Prevalence

We highlighted the dynamics of tuberculosis within a two-patch system, described by (15.5), under various residence times schemes via numerical experiments. The simulations were carried out using the two-patch Lagrangian modeling framework on pre-constructed scenarios. We assume that one of the two regions (say, Patch 1) has high TB prevalence. We do not model specific cities or regions. Nomenclature of some terms and scenarios are defined in Table 15.1.

The interconnection of the two idealized patches demands that individuals from Patch 1 travel to the “safer” Patch 2 to work, to school, or for other social activities. It is assumed that the proportion of time that Patch 2 residents spend in Patch 1 is negligible.

Here we define “high risk” based on the value of the probability of developing active TB using two distinct definitions: (i) patch having high direct first time transmission potential but no difference in exogenous reinfection potential between patches ($\beta_1 > \beta_2$ and $\delta_1 = \delta_2$) and (ii) the patch with high exogenous reinfection potential ($\delta_1 > \delta_2$ and $\beta_1 = \beta_2$). In addition, we assume a fixed population size for Patch 1 and vary the population size of Patch 2. Particularly, we assume that Patch 1 is the denser patch, while Patch 2 is assumed to have $\frac{1}{2}N_1$ and $\frac{1}{4}N_1$. That is, contact rates are higher in the Patch 1 population as compared to corresponding rates in Patch 2.

Table 15.1 Definitions and scenarios

Nomenclature	
Risk	Interpreted based on levels of infection rate, prevalence, or average contacts (via population size)
High-risk patch	Defined either by high direct first time infection rate (i.e., high β) which leads to high corresponding \mathcal{R}_0 or by high exogenous reinfection rate (i.e., high δ)
Enhanced socio-economic conditions (reducing health disparity)	Defined by better healthcare infrastructure which is incorporated by high prevalence of a disease (i.e., high $I(0)/N$) in a large population (i.e., large N)
Mobility	Captured by average residence times of an individual in different patches (i.e., by using \mathbb{P} matrix)
<i>Scenarios (assume high-risk and diminished socio-economic conditions in Patch 1 as compared to Patch 2)</i>	
Scenario 1	$\beta_1 > \beta_2$, $\delta_1 = \delta_2$; $\frac{I_1(0)}{N_1} > \frac{I_2(0)}{N_2}$, $N_1 > N_2$; vary p_{12} & $p_{21} \approx 0$
Scenario 2	$\beta_1 = \beta_2$, $\delta_1 > \delta_2$; $\frac{I_1(0)}{N_1} > \frac{I_2(0)}{N_2}$, $N_1 > N_2$; vary p_{12} & $p_{21} \approx 0$

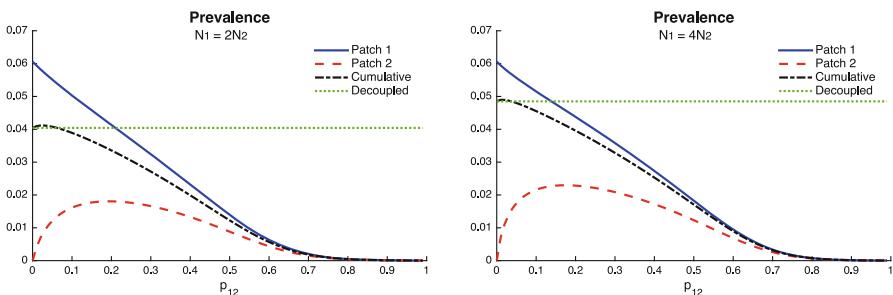


Fig. 15.7 Effect of mobility in the case of different transmission rates $0.13 = \beta_1 > \beta_2 = 0.07$ (which gives $\mathcal{R}_{01} = 1.5$, $\mathcal{R}_{02} = 0.8$) and $\delta_1 = \delta_2 = 0.0026$, on the endemic prevalence. The cumulative prevalence and prevalence for each patch using the following population size proportions $N_2 = \frac{1}{2}N_1$ (left figure) and $N_2 = \frac{1}{4}N_1$ (right figure) are shown here. The green horizontal dotted line represents the decoupled case (i.e., the case when there is no movement between patches)

15.4.3 The Role of Risk as Defined by Direct First Time Transmission Rates

In this subsection, we explore the impact of differences in transmission rates between patches. Patch 1 is high risk ($\mathcal{R}_{01} > 1$; obtained by assuming $\beta_1 > \beta_2$), while Patch 2 in the absence of visitors would be unable to sustain an epidemic ($\mathcal{R}_{02} < 1$). In addition the effect of different population ratios N_1/N_2 is explored.

Figure 15.7 uses mobility values p_{12} as it looks at their impact on increases in cumulative two-patch prevalence. At the individual patch level, increase in mobility values reduces the prevalence in Patch 1 but increases the prevalence in Patch 2 initially and then decreases past a threshold value of p_{12} (see red and black curves in Fig. 15.7). That is, completely cordoning off infected regions may not be a good idea to control disease. However, the movement rate of individuals between high-risk infection region and low-risk region must be maintained above a critical value to control an outbreak. Thus, it is possible that when Patch 1 (riskier patch) has a bigger population size, then mobility may turn out to be beneficial; the higher the ratio in population sizes, the higher the range of beneficial “traveling” times.

15.4.4 The Impact of Risk as Defined by Exogenous Reinfection Rates

Here, we focus our attention on the impact of exogenous reinfection on TB’s transmission dynamics when transmission rates are the same in both patches, $\beta_1 = \beta_2$. In this scenario, we assume the disease in both patches have reached an endemic state, that is, $\mathcal{R}_{01} > 1$ and $\mathcal{R}_{02} > 1$. However, Patch 1 remains the riskier, due to the assumption that exogenous reactivation of TB in Patch 1 is higher than in Patch 2, $\delta_1 > \delta_2$.

Figure 15.8 shows the combined role of exogenous reinfection and mobility values when the population of Patch 1 is twice or four times the population of Patch 2.

It is possible to see a small reduction in the overall prevalence, given for all mobility values from Patch 1 to Patch 2. Within this framework, parameters, and scenarios, our model suggests that direct first time transmission plays a central role

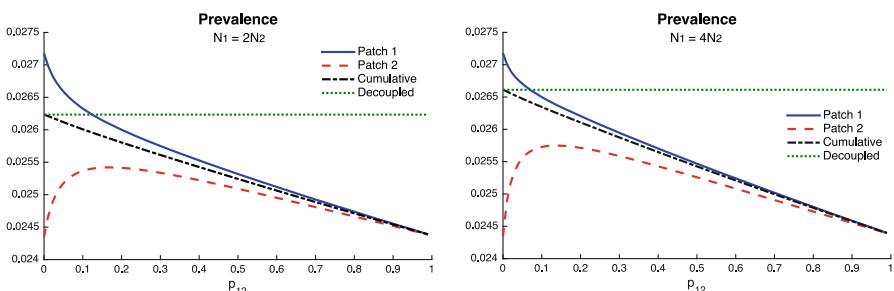
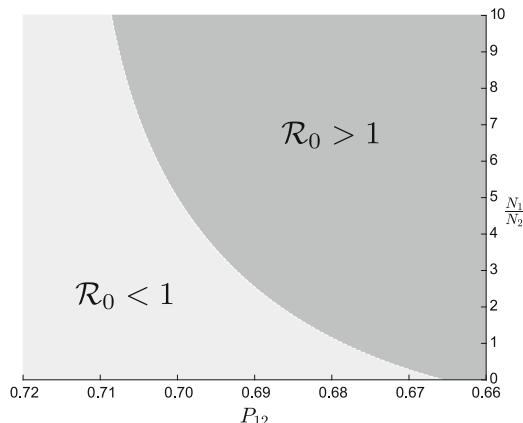


Fig. 15.8 Effect of mobility when risk is defined by the exogenous reinfection rates $0.0053 = \delta_1 > \delta_2 = 0.0026$ and $\beta_1 = \beta_2 = 0.1$ (which gives $\mathcal{R}_{01} = \mathcal{R}_{02} = 1.155$), on the endemic prevalence. The cumulative prevalence and prevalence for each patch using the following population size proportions $N_2 = \frac{1}{2}N_1$ (left figure) and $N_2 = \frac{1}{4}N_1$ (right figure) are shown here. The green dotted line represents the decoupled case (i.e., the case when there is no movement between patches)

Fig. 15.9 Effect of mobility and population size proportions on the global basic reproduction number \mathcal{R}_0 when $0.13 = \beta_1 > \beta_2 = 0.07$ and $\delta_1 = \delta_2 = 0.0026$



in TB dynamics when mobility is considered. Although mobility also reduces the overall prevalence when exogenous reinfection differs between patches, its impact is small compared to direct first time transmission results.

Finally, Fig. 15.9 shows the relationship between population densities and mobility (p_{12}) with respect to the basic reproduction number \mathcal{R}_0 . In this case we only explore the first case: direct first time transmission heterogeneity and found out that in this case mobility could indeed eliminate a TB outbreak.

According to the World Health Organization (WHO) [48], in 2014, 80% of the reported TB cases occurred in 22 countries, all developing countries. Efforts to control TB have been successful in many regions of the globe and yet we still see 1.5 million people die each year. And so, TB, faithful to its history [19], still poses one of the greatest challenges to global health. Recent reports suggest that established control measures for TB have not been adequately implemented, particularly in sub-Saharan countries [1, 15]. In Brazil rates have decreased with relapse being more important than reinfection [20, 33]. Finally, in Cape Town, South Africa, a study [47] showed that in high incidence areas, individuals who have received TB treatment and are no longer infectious are at the highest risk of developing TB instead of being the most protected. Hence, policies that do not account for population specific factors are unlikely to be effective. Without a complete description of the attributes of the community in question, it is almost impossible to implement successful intervention programs capable of generating low reinfection rates through multiple pathways and low number of drug resistant cases. Intervention must account for the risks that are inherent with high levels of migration as well as with local and regional mobility patterns between areas defined by high differences in TB risk. This discussion of TB dynamics within a simplified framework of a two-patch system has captured in a rather dramatic way the dynamics in two worlds: the world of the haves and the world of the have-nots. Simulations of simplified extreme scenarios highlight the impact of disparities.

TB dynamics depend on the basic reproduction number (\mathcal{R}_0), a function of model parameters that includes direct first transmission and exogenous (reinfection)

transmission rates. The simulations of specific extreme scenarios suggest that short-term mobility between heterogeneous patches does not always contribute to overall increases in TB prevalence. The results show that when risk is considered only in terms of exogenous reinfection, global TB prevalence remains almost unchanged when compared to the effect of direct new infection transmission. In the case of a high-risk direct first time transmission, it is observed that mobile populations may contribute to prevalence levels in both environments (patches). The simulations show that when the individuals from the risky population spend 25% of their time or less in the safer patch this is bad for the overall prevalence. However, if they spend more, the overall prevalence decreases. Further, in the absence of exogenous reinfections, the model is robust, that is, the disease dies out or persists based on whether or not the basic \mathcal{R}_0 is below or above unity, respectively. Although, the role of exogenous reinfection seems not that relevant to overall prevalence, the fact remains that such mode of transmission increases the risk that comes from a large displacement of individuals into a particular TB-free areas, due to catastrophes or conflict. As noted in [25], ignoring exogenous reinfections, that is, establishing policies that focus exclusively on the reproduction number \mathcal{R}_0 , would amount to ignoring the role of dramatic changes in initial conditions, now more common than before, due to the displacement of large groups of individuals, the result of catastrophes, and/or conflict.

15.5 *ZIKA

In November 2015, El Salvador reported their first case of Zika virus (ZIKV), an event followed by an explosive outbreak that generated over 6000 suspected cases in a period of 2 months. National agencies began implementing control measures that included vector control and recommending an increased use of repellents. In addition in response to the alarming and growing number of microcephaly cases in Brazil, the importance of avoiding pregnancies for 2 years was stressed. The role of mobility within communities characterized by extreme poverty, crime, and violence where public health services are not functioning is the set up for this example. We use a Lagrangian modeling approach within a two-patch setting in order to highlight the possible effects that short-term mobility, within two highly distinct environments, may have on the dynamics of ZIKV when the overall goal is to reduce the number of cases in both patches. The results of simulations in highly polarized and simplified scenarios are used to highlight the role of mobility on ZIKV dynamics. We found that matching observed patterns of ZIKV outbreaks was not possible without incorporating increasing levels of heterogeneity (more patches). A lack of attention to the threats posed by the weakest links in the global spread of diseases poses a serious threat to global health policies (see [12, 16, 23, 34, 40–42, 50]). Our results highlight the importance of focusing on key nodes of global transmission networks, which in the case of many regions correspond to places where the level of violence is highest. Latin America and the Caribbean, which

house 9% of the global population are a particular hot spot because this region accounts for 33% of the world's homicides [29]. Hence, it is essential to assess how much public safety conditions may affect mobility and the level of local risk, which may affect the dynamics of ZIKV.

15.5.1 *Single Patch Model

Assume that individuals while in Patch 1 will be experiencing high risk of infection, while those in Patch 2 will be experiencing low risk. Movement (daily activities) will alter the amount of time that each individual spends on each patch, the longer that an individual is found in Patch 1, the more likely that he/she will become infected. The level of patch-specific risk to infection is captured via the use of a single parameter $\hat{\beta}_i$, $i = 1, 2$ with $\hat{\beta}_1 \gg \hat{\beta}_2$. This assumption pretends to capture health disparities in a rather simplistic way. The case of Johannesburg and Soweto in South Africa, or North and South Bogota in Colombia, or Rio de Janeiro and adjacent favelas in Brazil, or gang-controlled and gang-free areas within San Salvador are but some of the unfortunately large number of pockets dominated by conflict, high crime or highly differentiated health structures within urban centers around world. The short time scale dynamics of individuals (going to work or attending schools or universities) are incorporated within this model. The dynamics of transmission is carried out via simulations over the duration of a single outbreak.

The ZIKV dynamics single patch model involves host and vector populations of size N_h and N_v , respectively. Both populations are subdivided by epidemiological states; the transmission process is modeled as the result of the interactions of these populations. On that account, we let S_h , E_h , $I_{h,a}$, $I_{h,s}$, and R_h denote the susceptible, latent, infectious asymptomatic, infectious symptomatic, and recovered host sub-populations. Similarly, S_v , E_v , and I_v are used to denote the susceptible, latent, and infectious mosquito sub-populations. Since the focus is on the study of disease dynamics over a single outbreak, we neglect the host demographics while assuming that the vector demographics do not change, meaning that it is assumed the birth and death per-capita mosquito rates cancel each other out. New reports [14, 21] point out the presence of large numbers of asymptomatic ZIKV infectious individuals. Consequently, we consider two classes of infectious $I_{h,a}$ and $I_{h,s}$, that is, asymptomatic and symptomatic infectious individuals. Further, since there is no full knowledge of the dynamics of ZIKV transmission, it is assumed that $I_{h,a}$ and $I_{h,s}$ individuals are equally infectious with their periods of infectiousness roughly the same. Our assumptions could be used to reduce the model to one that considers a single infectious class $I_h = I_{h,a} + I_{h,s}$. We keep both infectious classes as it may be desirable to keep track of each type. These assumptions may not be too bad given our current knowledge of ZIKV epidemiology and the fact that ZIKV infections, in general, are not severe. Furthermore, given that the infectious process of ZIKV is somewhat similar to that of dengue, we use some of the parameters estimated in dengue transmission studies within El Salvador. ZIKV basic reproduction number

estimates are taken from those that we just estimated using outbreak data from Barranquilla Colombia [45]. Furthermore, the selection of model parameters ranges used also benefited from prior estimates conducted with data from the 2013–2014 French Polynesia outbreak [31], some of the best available. The dynamics of the prototypic single patch system, single epidemic outbreak, can therefore be modeled using the following standard nonlinear system of differential equations [9]:

$$\begin{aligned} S'_h &= -b\beta_{vh}S_h \frac{I_v}{N_h} \\ E'_h &= b\beta_{vh}S_h \frac{I_v}{N_h} - \nu_h E_h \\ I'_{h,s} &= (1-q)\nu_h E_h - \gamma_h I_{h,s} \\ I'_{h,a} &= q\nu_h E_h - \gamma_h I_{h,a} \\ R'_h &= \gamma_h(I_{h,s} + I_{h,a}) \\ S'_v &= \mu_v N_v - b\beta_{hv}S_v \frac{I_{h,s} + I_{h,a}}{N_h} - \mu_v S_v \\ E'_v &= b\beta_{hv}S_v \frac{I_{h,s} + I_{h,a}}{N_h} - (\mu_v + \nu_v)E_v \\ I'_v &= \nu_v E_v - \mu_v I_v. \end{aligned} \tag{15.6}$$

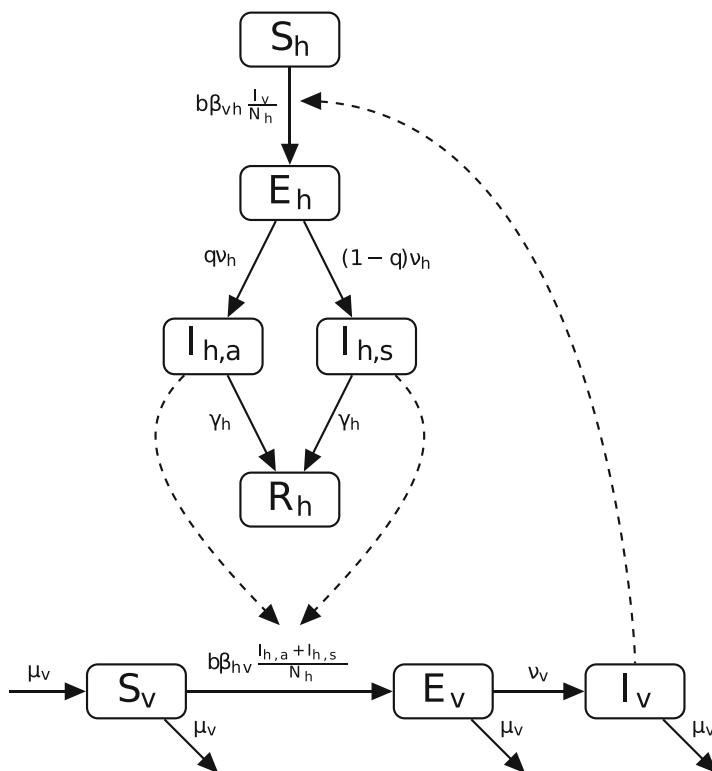
15.5.2 *Residence Times and Two-Patch Models

The role of mobility between two communities, within the same city, living under dramatically distinct health, economic, social, and security settings is explored using a model as simple as possible, that is, a model that only considers two patches (prior modeling efforts that didn't account for the effective population size but that incorporated specific controls include [32]). Patch 2 has access to working health facilities, low crime rate, adequate human and financial resources, and adequate public health policies, in place. Patch 1 lacks nearly everything and crime is high. The differences in risk are captured by postulating very different transmission rates. We study the dynamics of host mobility in highly distinct environments, with risk being captured by the transmission rate, $\hat{\beta}$. Hence, $\hat{\beta}_1 \gg \hat{\beta}_2$, where $\hat{\beta}_i$ defines the risk in Patch i , $i = 1, 2$ [Patch 1 (high risk) and Patch 2 (low risk)].

The host populations are stratified by epidemiological classes indexed by the patch of residence. In particular, $S_{h,i}$, $E_{h,i}$, $I_{h,a,i}$, $I_{h,s,i}$, and $R_{h,i}$ denote the susceptible, latent, infectious asymptomatic, infectious symptomatic, and recovered host populations in Patch i , $i = 1, 2$ with $S_{v,i}$, $E_{v,i}$, and $I_{v,i}$ denoting the susceptible, latent, and infectious mosquito populations in Patch i , $i = 1, 2$. As before, $N_{h,i}$ denotes the host patch population size ($i, i = 1, 2$) and $N_{v,i}$ the total vector population in Patch i , $i = 1, 2$. The vector is assumed to be incapable of moving between patches, a reasonable assumption in the case of *Aedes aegypti* under the appropriate spatial scale. The patch model parameters are presented in Table 15.2 with the flow diagram, single patch dynamics model, capturing the situation when residents and visitors do not move; that is, when the 2×2 residence times matrix \mathbb{P} is such that $p_{11} = p_{22} = 1$ (Fig 15.10).

Table 15.2 Description of the parameters used in system (15.6)

Parameters	Description	Value
β_{vh}	Infectiousness of human to mosquitoes	0.41
β_{hv}	Infectiousness of mosquitoes to humans	0.5
b_i	Biting rate in Patch i	0.8
v_h	Humans' incubation rate	$\frac{1}{7}$
q	Fraction of latent that become asymptomatic and infectious	0.1218
γ_i	Recovery rate in Patch i	$\frac{1}{5}$
p_{ij}	Proportion of time residents of Patch i spend in Patch j	$[0, 1]$
μ_v	Vectors' natural mortality rate	$\frac{1}{13}$
v_v	Vectors' incubation rate	$\frac{1}{9.5}$

**Fig. 15.10** Flow diagram of model (15.6)

Since individuals experience a higher risk of ZIKV infection while in Patch 1, then it is assumed that mobility from Patch 2 to Patch 1 is unappealing with typical Patch 2 residents spending (on the average) a reduced amount of time each unit of

time in Patch 1. Parameters are chosen so that the dynamics of ZIKV within Patch 2 cannot be supported in the absence of mobility between Patch 1 and Patch 2. Thus, the Patch 2 local basic reproduction number is taken to be less than one, namely $\mathcal{R}_{02} = 0.9$. Mobility is modeled under the residence times matrix \mathbb{P} with entries given *initially* by $p_{21} = 0.10$ and $p_{12} = 0$.

Two cases are explored: A “worst case” scenario where control measures are hardly implemented due to crime, conflict, or other factors on Patch 1, that is, Patch 1 is a place where the risk of acquiring a ZIKV infection is high since $\mathcal{R}_{01} = 2$. The “best case” scenario corresponds to the case when Patch 1 can implement some control measures with some degree of effectiveness, and consequently Patch 1 has an $\mathcal{R}_{01} = 1.52$. The \mathcal{R}_{0i} values used are in line with those previously estimated for ZIKV outbreaks [31, 45]. Simulations are seeded by introducing an asymptomatic infected individual in Patch 1 under the assumption that the host and vector populations are fully susceptible in both patches.

Figure 15.11 (top) shows the incidence and final ZIKV epidemic size when Patch 1 is under the “worst case scenario,” defined by a basic reproduction number of

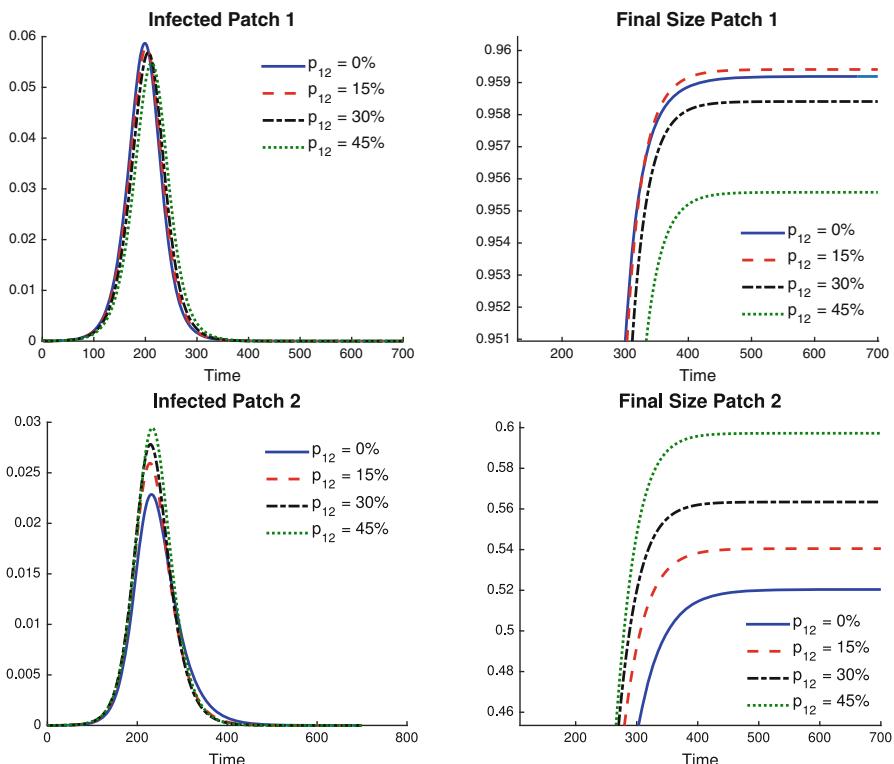


Fig. 15.11 Per patch incidence and final size proportions for $p_{21} = 0.10$, $p_{12} = 0, 0.15, 0.30$, and 0.45 . Mobility shifts the behavior of the Patch 1 final size in the “worst case” scenario: $\mathcal{R}_{01} = 2$ and $\mathcal{R}_{02} = 0.9$

$\mathcal{R}_{01} = 2$ [45]. Figure 15.11 shows that at $p_{12} = 0.15$ the final number of infected residents in Patch 1 is larger to the number in the baseline scenario ($p_{12} = 0$). In fact, it reaches almost 96% of the population, an unrealistic value. Additional simulated p_{12} values show that final Patch 1 size would go below the baseline case; a benefit of mobility. Figure 15.11 highlights the case when the Patch 2 epidemic final size grows with increases in mobility when compared with the baseline case (no mobility from Patch 1). We see reductions in the Patch 1 epidemic final size for some mobility values accompanied by increments in the Patch 2 epidemic final size when compared to the baseline scenario (no mobility from Patch 1). Specifically, reductions in Patch 1 epidemic final size are around 1×10^{-3} , while increments in Patch 2 are around 1×10^{-2} , under the assumption that the population in Patch 1 is the same as that in Patch 2. Thus while mobility may provide benefits within Patch 1 (under the above assumptions) the fact remains that it does it at a cost. In short, it is also observed that the epidemic final size per patch does not respond linearly to changes in mobility even when only the mobility p_{12} is increased (see Figs. 15.11 and 15.12).

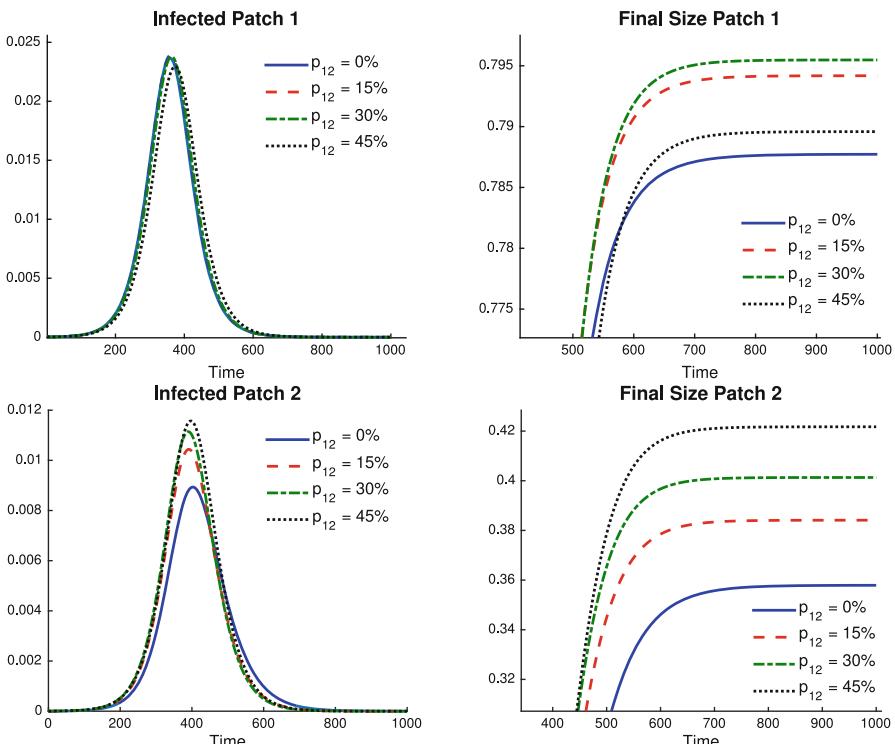


Fig. 15.12 Per patch incidence and final size proportions for $p_{21} = 0.10$, $p_{12} = 0, 0.15, 0.30$, and 0.45 . Mobility significantly shapes the per patch final sizes in the “worst case” scenario $\mathcal{R}_{01} = 2$ and $\mathcal{R}_{02} = 0.9$

Consider now the “best case” scenario, a basic reproduction number $\mathcal{R}_{01} = 1.52$, under the assumption that the population in Patch 1 is the same as that in Patch 2. The results of simulations collected in Fig. 15.12 show a final size epidemic curve similar to that generated in the “worst case” scenario for Patch 1. Some mobility values can increase the Patch 1 epidemic final size, reaching almost 80% of the population when $p_{12} = 0.30$, an unrealistic level, albeit, as expected lower than in the “worst” case scenario. The existence of a *mobility threshold* after which the final epidemic sizes in Patch 1 start to decrease is also observed. The results in Fig. 15.12 suggest that under all p_{12} mobility levels, Patch 2 ZIKV epidemic final size supports monotone growth in the total number of infected individuals. The changes in the epidemic final size in each patch in Fig. 15.12 are roughly equivalent (the same order, 1×10^{-2}) given that the population in Patch 1 is the same as that in Patch 2.

The simulation results presented so far provide only partial information on the impact that short-term mobility may have on the transmission dynamics of ZIKV. Now, by fixing the mobility from Patch 2 to Patch 1, we are just focusing only on the impact of changes in mobility from Patch 1 to Patch 2. Further, the potential changes in mobility patterns that host populations may have in response to ZIKV dynamics are ignored by our use of a mobility matrix \mathbb{P} with constant entries p_{ij} . We found that epidemic final size within Patch 1 is qualitatively similar in the worst and best case scenarios: increasing at first, decreasing after a certain threshold, and crossing down the baseline case under some mobility regimes. Further, it has been observed that the qualitative behavior of the epidemic final size in Patch 2 grows monotonically as mobility increases. Patch 1 and Patch 2 responses are of different orders of magnitude in the “worst case” scenario but roughly of the same order of magnitude in the “best case” scenario, which means, under our restrictive conditions and assumptions, that reductions in risk in Patch 1 do help significantly.

15.5.2.1 *The Role of Risk Heterogeneity in the Dynamics of ZIKV Transmission

The impact of risk heterogeneity on ZIKV dynamics within the overall two-patch system is explored, an analysis that requires the numerical estimation of the global reproduction number as a function of the mobility matrix \mathbb{P} . Using the previous scenarios ($\mathcal{R}_{01} = 1.52$, 2) simulations are carried out first assuming equal population sizes ($N_1 = N_2$). However, when looking at the impact of changes in risk on Patch 2 ($\mathcal{R}_{02} = 0.8, 0.9, 1, 1.1$), our simulations identify a growing epidemic in Patch 2 as risk increases with the overall community experiencing nonlinear changes in risk as residency times change from the baseline scenario given by $p_{12} = 0$. Specifically, Fig. 15.13 captures overall reductions on the global reproduction number for all residence times while identifying the existence of a residence time interval for which mobility decreases the total size of the outbreak in the two-patch community, when compared to the corresponding baseline case ($p_{12} = 0$). In the absence of mobility from Patch 1 ($p_{12} = 0$), increases in the epidemic final size as \mathcal{R}_{0i} increases are observed. These simulations show that

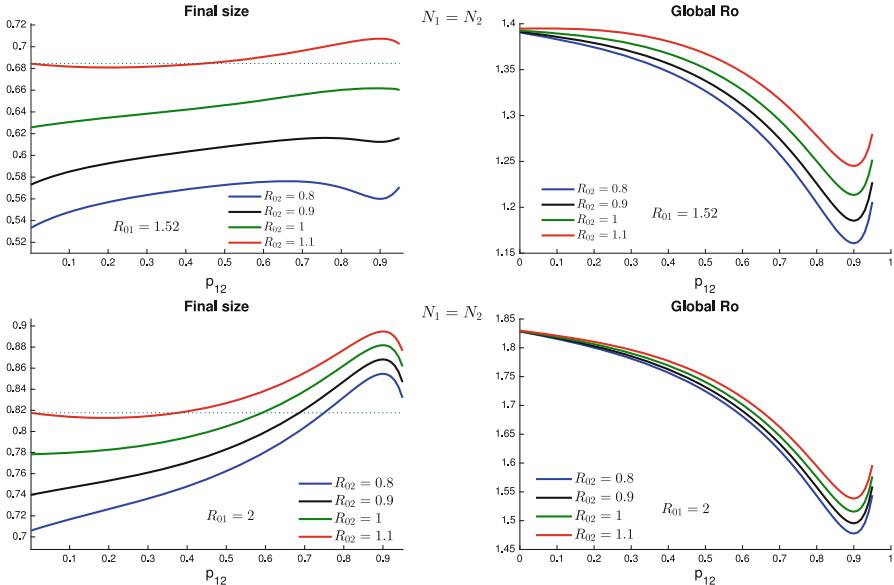


Fig. 15.13 Local and global final sizes through mobility values when $p_{21} = 0.10$. Although mobility reduces the global \mathcal{R}_0 , allowing mobility in the case of El Salvador ($\mathcal{R}_0 = 2$) might lead to a detrimental effect in the global final size

mobility can slow down the speed of the outbreak (smaller global \mathcal{R}_0). Of course, the simulation results also re-affirm the obvious, that is, that the existence of a high risk, mobile, and well-connected patch can serve as an outbreak magnifier; a situation that has been explored within an n -patch system under various connective schemes [7, 10]. This is because, in the two-patch case, it is observed that the global reproduction number \mathcal{R}_0 experiences reductions for almost all mobility values. For the scenarios selected \mathcal{R}_0 never drops below 1. Hence, under our assumptions and scenarios, it is seen that the use of fixed mobility patterns makes the elimination of ZIKV extremely difficult if not impossible under our two scenarios. Figure 15.13 provides an example that highlights the relationship between the global reproduction number and corresponding epidemic final size.

15.5.2.2 *The Role of Population Size Heterogeneity in the Dynamics of ZIKV Transmission

The role of population density in the total epidemic final size and global basic reproduction number is explored under our two scenarios, now under the changed assumption that the densities (population sizes) of Patch 1 and Patch 2 differ. Specifically, we take $N_1 = 2N_2$, $3N_2$, $5N_2$, and $10N_2$.

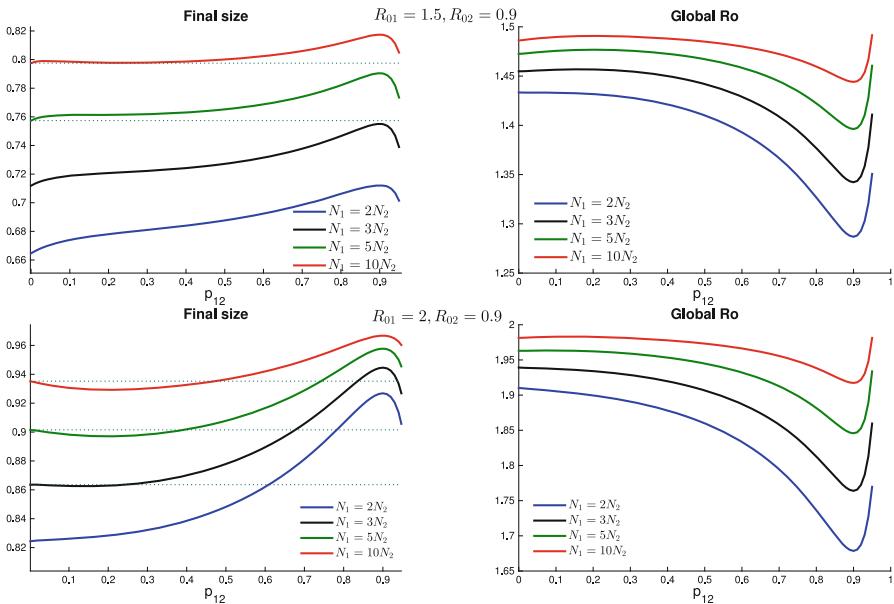


Fig. 15.14 Total final size and global basic reproduction number through mobility values when $p_{21} = 0.10$. Local risk values are set up to $\mathcal{R}_{02} = 0.9$ and $\mathcal{R}_{01} = 1.52, 2$

It is observed that difference in population sizes do matter. Specifically, it is observed that (under our selections) a big difference in density indicate that a higher epidemic final size is reached. The value of 90% for the “worst case” is possible with changes in the global reproduction number exhibiting different patterns (see Fig. 15.14). We observe that despite increases in the total epidemic final size as mobility changes the global \mathcal{R}_0 actually decreases monotonically for most residence times, never falling below one. A sensible degree of magnification on the spread of the disease as residence times change is observed whenever the differences between N_1 and N_2 are not too extreme. In fact, it is possible for mobility to be beneficial in the control of ZIKV under the above simplistic extreme scenarios. Simulations continue to show that under the prescribed conditions and assumptions, model generated ZIKV outbreaks remain unrealistically high. The simulations show, for example, that the global reproduction number reaches its minimum at around $p_{12} = 0.90$ with Fig. 15.14 showing that the larger the high risk population gets ($N_1 >> N_2$), the greater the total epidemic final size becomes as individuals from Patch 1 spend more than half of their day in Patch 2. Using a low p_{12} value, a small benefit is observed, namely the total epidemic final size is reduced, when the differences between \mathcal{R}_{0i} are high.

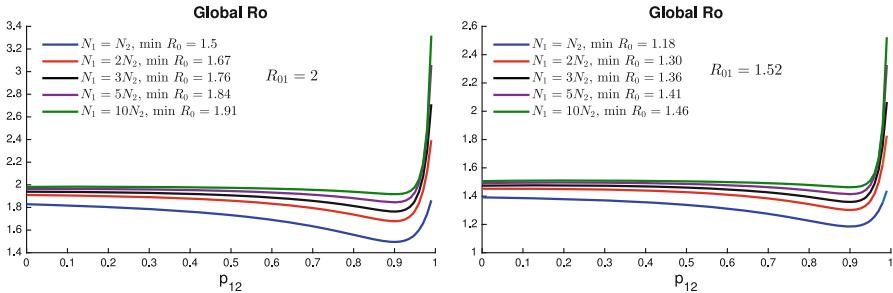
For the two epidemiological scenarios $\mathcal{R}_{01} = 2$ and $\mathcal{R}_{01} = 1.52$, Tables 15.3 and 15.4 provide a summary of the average proportion of infected for low

Table 15.3 Final size (Patch 1, Patch 2) $N_1 = 10,000$, $\mathcal{R}_{01} = 2$, $\mathcal{R}_{02} = 0.9$, and $p_{21} = 0.10$

N_2	Low mobility	Intermediate mobility	High mobility	Min \mathcal{R}_0
$N_1 = N_2$	(0.9594, 0.5333)	(0.9583, 0.5633)	(0.9539, 0.6122)	1.4954
$N_1 = 2N_2$	(0.9683, 0.5418)	(0.9685, 0.5599)	(0.9667, 0.6116)	1.6786
$N_1 = 3N_2$	(0.9709, 0.5390)	(0.9713, 0.5478)	(0.9701, 0.6018)	1.7640
$N_1 = 5N_2$	(0.9729, 0.5283)	(0.9732, 0.5255)	(0.9725, 0.5852)	1.8457
$N_1 = 10N_2$	(0.9741, 0.5030)	(0.9743, 0.4908)	(0.9739, 0.5624)	1,9173

Table 15.4 Final size (Patch 1, Patch 2) $N_1 = 10,000$, $\mathcal{R}_{01} = 1.52$, $\mathcal{R}_{02} = 0.9$, and $p_{21} = 0.10$

N_2	Low mobility	Intermediate mobility	High mobility	Min \mathcal{R}_0
$N_1 = N_2$	(0.7920, 0.3756)	(0.7950, 0.4010)	(0.7849, 0.4304)	1.1853
$N_1 = 2N_2$	(0.8287, 0.3938)	(0.8340, 0.4061)	(0.8300, 0.4356)	1.3023
$N_1 = 3N_2$	(0.8398, 0.3948)	(0.8448, 0.3956)	(0.8422, 0.4248)	1.3590
$N_1 = 5N_2$	(0.8480, 0.3877)	(0.8520, 0.3731)	(0.8500, 0.4046)	1.4141
$N_1 = 10N_2$	(0.8533, 0.3652)	(0.8556, 0.3352)	(0.8542, 0.3756)	1.4630

**Fig. 15.15** Global \mathcal{R}_0 dynamics through mobility when $p_{21} = 0.10$. Patch 2 populations vary from $N_1 = N_2$, $2N_2$, $3N_2$, $5N_2$ up to $N_1 = 10N_2$. The global \mathcal{R}_0 hits its minimum always at an unrealistic 91% of mobility. As N_1 approaches N_2 , this minimum value decreases

($p_{12} = 0\text{--}0.2$), intermediate ($p_{12} = 0.2\text{--}0.4$), and high mobility ($p_{12} = 0.4\text{--}0.6$) when $p_{21} = 0.10$. The role of population scaling $N_1 = 2N_2$, $3N_2$, $5N_2$, and $10N_2$ is also explored. Figure 15.15 shows the global \mathcal{R}_0 over all mobility values for different population weights in the two epidemic scenarios. The minimum \mathcal{R}_0 value is reached for all cases when mobility is at an unrealistic 91% and when $N_1 \approx N_2$. The results collected in Fig. 15.15 show that short-term mobility plays an important role in ZIKV dynamics, again, under a system involving two highly differentiated patches. Simulations also suggest that, even though mobility can reduce the global reproduction number, mobility by itself is not enough to eliminate an outbreak or make a real difference under our two scenarios.

15.5.3 What Did We Learn from These Single Outbreak Simulations?

The study of the role of mobility at large spatial scales may be best captured using question-specific related models that account for the possibility of long-term mobility (see, for example, [2, 3, 18, 22, 27, 28, 30, 43]). Here, we made use of two patches, as distinct as they can be would be able to shed some light on the transmission dynamics of ZIKV, whenever extreme health disparities within neighboring communities or within urban centers were the norm. Although the goal is not to fit specific outbreaks, we decided to make use of recently published parameter ranges, including some reported by us [45]. The impact of ZIKV can be assessed locally (each patch) or globally, that is, over the two-patch system. Here, system risk assessment was carried out by computing \mathcal{R}_0 , via the numerical solution of a system of nonlinear equations. Changes in the system \mathcal{R}_0 were computed (as residence times were varied) in relationship to the local \mathcal{R}_{0i} , that is, local basic reproduction numbers (in the absence of mobility). Further, the mobility-dependent system epidemic final sizes were computed via simulations that assessed the impact of mobility (and risk) locally and on the overall system. The metrics used in our assessment included the overall epidemic final size (a measure of the overall impact of an outbreak), a function of mobility within the two selected scenarios ($\mathcal{R}_{01} = 1.52$ and $\mathcal{R}_{01} = 2$).

The challenges posed by policies that may be beneficial to the system but detrimental to each patch were explored within our two-patch system. Situations where the total final epidemic size increased with increments in \mathcal{R}_{02} , and situations where the total final epidemic size decreased under low mobility values for $\mathcal{R}_{(02)}$ were documented. Population density does make a difference and examples when $\mathcal{R}_{02} < 1$ with mobility incapable of reducing the total epidemic final size under no differences in patch density (here measured by total population size in each patch, both assumed to have roughly the same area) were also identified. Differences in population density were also shown to be capable of generating reductions on the total final epidemic size within some mobility regimes.

The highly simplified two-patch model used seemed to have shed some light on the role of mobility on the spread of ZIKV in areas where huge differences in the availability of public health programs and services—the result of endemic crime, generalized violence, and neglect—exist. Model simulations seemed to have shed some light on the potential relevance of factors that we failed to account for. The value of the use of single patch-specific risk parameters ($\hat{\beta}$) has strengths and limitations. The model used did not account explicitly for changes in the levels of infection within the vector population nor did it account for the impact of substantial differences in patch vector population sizes. The simplified model failed to account for the responses to outbreaks by patch residents as individuals may alter mobility patterns, use more protective clothing while responding individually and independently to official control programs in the face of dramatic increases on the vector population or due to a surge in the number of cases. The use of two patches

and severe assumptions limits the outcomes that such an oversimplified system can support. Communities can't in generally be modeled under a highly differentiated two-tier system and in the case of ZIKV, the possibility of vertical transmission in humans and vectors as well as sexually transmitted ZIKV cannot be completely neglected [8, 39]. The introduction of changes in behavior in response to individuals' assessment of the levels of risk infection over time needs to be addressed [10]; a challenge that has yet to be met to the satisfaction of the scientific community involved in the study of epidemiological processes as complex adaptive systems (see, for example, [26, 36, 42]).

The limitations of the role of technology in the absence of the public health infrastructure—there is no silver bullet—have been addressed in the context of Ebola [16, 49]. It would be interesting to see the impact of technology in settings where health disparities are pervasive, using a two-patch Lagrangian epidemic model in the context of communicable and vector-borne diseases, including dengue, tuberculosis , and Ebola [5, 23, 34, 35]. Further, its often the case that the use of simplified models quite often overestimates the impact of an outbreak (see [37, 38]) and so find the right level of model heterogeneity (number of patches) becomes a pressing and challenging question. What is the right level of aggregation to address these questions?

Certainly, we have seen the use of dramatic measures to limit the spread of diseases like SARS, influenza , or Ebola [16, 17, 27], as well as the rise of vector-borne diseases like dengue and Zika, and the dramatic implications that some measures have had on local and global economies. The question remains, what can we do to mitigate or limit the spread of disease, particularly emergent diseases without disrupting central components? Discussions on these issues are recurrent [26, 36], most intensely in the context of SARS, influenza, Ebola , and Zika, in the last decade or so. The vulnerability of world societies is directly linked to the lack of action in addressing the challenges faced by the weakest links in the system. This must be accepted and acted on by the world community. We need global investments in communities and nations where health disparities and lack of resources are the norm. We must invest in research and surveillance within clearly identified world hot spots, where the emergence of new diseases is most likely to occur. We must do so with involvement at all levels of the affected communities [12, 44].

References

1. Andrews, J.R., C. Morrow, and R. Wood (2013) Modeling the role of public transportation in sustaining tuberculosis transmission in South Africa. *Am J Epidemiol* **177**: 556–561.
2. Banks, H.T. and C. Castillo-Chavez (2003)*Bioterrorism: Mathematical Modeling Applications in Homeland Security*. SIAM, 2003.
3. Baroyan, O., L. Rvachev, U. Basilevsky, V. Ermakov, K. Frank, M. Rvachev, and V. Shashkov (1971) Computer modelling of influenza epidemics for the whole country (USSR), *Advances in Applied Probability* **3**: 224–226.

4. Bichara, D. and C. Castillo-Chavez (2016) Vector-borne diseases models with residence times—a Lagrangian perspective, *Math. Biosc.* **281**: 128–138.
5. Bichara, D., S. A. Holechek, J. Velázquez-Castro, A. L. Murillo, and C. Castillo-Chavez (2016) On the dynamics of dengue virus type 2 with residence times and vertical transmission. *Letters in Biomathematics* **3**:140–160.
6. Bichara, D. and A. Iggidr (2017) Multi-patch and multi-group epidemic models: a new framework. *J. Math. Biol.* <https://doi.org/10.1007/s285-017-1191-9>.
7. Bichara, D., Y. Kang, C. Castillo-Chavez, R. Horan, and C. Perrings (2015) SIS and SIR epidemic models under virtual dispersal, *Bull. Math. Biol.* **77**: 2004–2034.
8. Brauer, F., C. Castillo-Chavez, A. Mubayi, and S. Towers (2016) Some models for epidemics of vector-transmitted diseases. *Infectious Disease Modelling* **1**: 79–87.
9. Brauer, F. and C. Castillo-Chavez (2012) Mathematical Models for Communicable Diseases **84** SIAM.
10. Castillo-Chavez, C., K. Barley, D. Bichara, D. Chowell, E.D. Herrera, B. Espinoza, V. Moreno, S. Towers, and K. Yong (2016) Modeling Ebola at the mathematical and theoretical biology institute (MTBI). *Notices of the AMS* **63**.
11. Castillo-Chavez, C., D. Bichara, and B.R. Morin (2016) Perspectives on the role of mobility, behavior, and time scales in the spread of diseases, *Proc. Nat. Acad. Sci.* **113**: 14582–14588.
12. Castillo-Chavez, C., R. Curtiss, P. Daszak, S.A. Levin, O. Patterson-Lomba, C. Perrings, G. Poste, and S. Towers (2015) Beyond Ebola: Lessons to mitigate future pandemics. *The Lancet Global Health* **3**: e354–e355.
13. Castillo-Chavez, C., B. Song, and J. Zhangi (2003) An epidemic model with virtual mass transportation: The case of smallpox. *Bioterrorism: Mathematical Modeling Applications in Homeland Security* **28**: 173.
14. CDC(a) (2016) Zika virus. Online, February 01, 2016.
15. Chatterjee, D. and A. K. Pramanik (2015) Tuberculosis in the African continent: A comprehensive review. *Pathophysiology* **22**: 73–83.
16. Chowell, D., C. Castillo-Chavez, S. Krishna, X. Qiu, and K.S. Anderson (2015) Modelling the effect of early detection of Ebola. *The Lancet Infectious Diseases*
17. Chowell, G., P.W. Fenimore, M.A. Castillo-Garsow, and C. Castillo-Chavez (2003) SARS outbreaks in Ontario, Hong Kong and Singapore: the role of diagnosis and isolation as a control mechanism, *J. Theor. Biol.* **224**: 1–8.
18. Chowell, G., J.M. Hyman, S. Eubank, and C. Castillo-Chavez (2003) Scaling laws for the movement of people between locations in a large city. *Phys. Rev.* **68**: 066102.
19. Daniel, T.M. (2006) The history of tuberculosis. *Respiratory Medicine* **100**: 1862–1870.
20. de Oliveira, G.P., A.W. Torrens, P. Bartholomay, and D. Barreira (2013) Tuberculosis in Brazil: last ten years analysis – 2001–2010 in *Journal of Infectious Diseases* **17**: 218–233.
21. Duffy, M.R., T.-H. Chen, W.T. Hancock, A.M. Powers, J.L. Kool, R.S. Lanciotti, M. Pretrick, M. Marfel, S. Holzbauer, C. Dubray, et al (2009) Zika virus outbreak on yap island, federated states of Micronesia. *New England J. Medicine* **360**: 2536–2543.
22. Elveback, L.R., J.P. Fox, E. Ackerman, A. Langworthy, M. Boyd, and L. Gatewood (1976) An influenza simulation model for immunization studies. *Am. J. Epidemiology* **103**: 152–165.
23. Espinoza, B., V. Moreno, D. Bichara, and C. Castillo-Chavez (2016) Assessing the efficiency of movement restriction as a control strategy of Ebola. In *Mathematical and Statistical Modeling for Emerging and Re-emerging Infectious Diseases*, pages 123–145. Springer.
24. Eubank, S., H. Guclu, V. A. Kumar, M.V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang (2004) Modelling disease outbreaks in realistic urban social networks. *Nature* **429**: 180–184.
25. Feng, Z., C. Castillo-Chavez, and A.F. Capurro (2000) A model for tuberculosis with exogenous reinfection. *Theor. Pop. Biol.* **57**: 235–247.
26. Fenichel, E.P., C. Castillo-Chavez, M.G. Ceddia, G. Chowell, P.A.G. Parra, G.J. Hickling, G. Holloway, R. Horan, B. Morin, C. Perrings, et al (2011) Adaptive human behavior in epidemiological models. *Proc. Nat. Acad. Sci.* **108**: 6306–6311.

27. Herrera-Valdez, M.A., M. Cruz-Aponte, and C. Castillo-Chavez (2011) Multiple outbreaks for the same pandemic: Local transportation and social distancing explain the different “waves” of A-H1N1pdm cases observed in México during 2009. *Math. Biosc. Eng.* **8**: 21–48.
28. Hyman, J.M. and T. LaForce (2003) Modeling the spread of influenza among cities. In *Bioterrorism: Mathematical Modeling Applications in Homeland Security*, pages 211–236. SIAM.
29. Jaitman, L. (2015) Los costos del crimen y la violencia en el bienestar en America Latina y el Caribe, L. Jaitman Ed. Banco Interamericano del Desarrollo.
30. Khan, K., J. Arino, W. Hu, P. Raposo, J. Sears, F. Calderon, C. Heidebrecht, M. Macdonald, J. Liauw, A. Chan, et al (2009) Spread of a novel influenza a (H1N1) virus via global airline transportation. *New England J. Medicine* **361**: 212–214.
31. Kucharski, A.J., S. Funk, R.M. Eggo, H.-P. Mallet, J. Edmunds, and E.J. Nilles (2016) Transmission dynamics of Zika virus in island populations: a modelling analysis of the 2013–14 French Polynesia outbreak. *bioRxiv*, page 038588.
32. Lee, S. and C. Castillo-Chavez (2015) The role of residence times in two-patch dengue transmission dynamics and optimal strategies, *J. Theor. Biol.* **374**:152–164.
33. Luzzé, H., D.F. Johnson, K. Dickman, H. Mayanja-Kizza, A. Okwera, K. Eisenach, M. D. Cave, C.C. Whalen, J.L. Johnson, W.H. Boom, M. Joloba, and Tuberculosis Research Unit (2013) Relapse more common than reinfection in recurrent tuberculosis 1–2 years post treatment in urban Uganda. *Int. J. of Tuberculosis and Lung Disease* **17**: 361–367.
34. Moreno, V., B. Espinoza, K. Barley, M. Paredes, D. Bichara, A. Mubayi, and C. Castillo-Chavez (2017) The role of mobility and health disparities on the transmission dynamics of tuberculosis. *Theoretical Biology and Medical Modelling* **14** :3, 2017.
35. Moreno, V.M., B. Espinoza, D. Bichara, S. A. Holechek, and C. Castillo-Chavez (2017) Role of short-term dispersal on the dynamics of Zika virus in an extreme idealized environment. *Infectious Disease Modelling* **2**: 21–34.
36. Morin, B.R., E.P. Fenichel, and C. Castillo-Chavez (2013) SIR dynamics with economically driven contact rates, *Natural resource modeling* **26**: 505–525.
37. Nishiura, H., C. Castillo-Chavez, M. Safan, and G. Chowell (2009) Transmission potential of the new influenza A (H1N1) virus and its age-specificity in Japan, *Eurosurveillance* **14**: 19227.
38. Nishiura, H., G. Chowell, and C. Castillo-Chavez (2011) Did modeling overestimate the transmission potential of pandemic (H1N1-2009)? sample size estimation for post-epidemic seroepidemiological studies. *PLoS One* **6**: e17908.
39. Padmanabhan, P., P. Seshaiyer, and C. Castillo-Chavez (2017) Mathematical modeling, analysis and simulation of the spread of Zika with influence of sexual transmission and preventive measures. *Letters in Biomathematics* **4**: 148–166.
40. Patterson-Lomba, O., E. Goldstein, A. Gómez-Liévano, C. Castillo-Chavez, and S. Towers (2015) Infections increases systematically with urban population size: a cross-sectional study. *Sex Transm Infect* **91**: 610–614.
41. Patterson-Lomba, O., M. Safan, S. Towers, and J. Taylor (2016) Modeling the role of healthcare access inequalities in epidemic outcomes. *Math. Biosc. Eng.* **13**: 1011–1041.
42. Perrings, C. Castillo-Chavez, G. Chowell, P. Daszak, E.P. Fenichel, D. Finnoff, R.D. Horan, A.M. Kilpatrick, A.P. Kinzig, N.V. Kuminoff, et al (2014) Merging economics and epidemiology to improve the prediction and management of infectious disease. *EcoHealth* **11**: 464–475.
43. Rvachev, L.A. and I.M. Longini Jr. (1985) A mathematical model for the global spread of influenza. *Math. Biosc.* **75**: 3–22.
44. Stroud, P., S. Del Valle, S. Sydoriak, J. Riese, and S. Mniszewski (2007) Spatial dynamics of pandemic influenza in a massive artificial society. *Journal of Artificial Societies and Social Simulation* **10**: 9.
45. Towers, S., F. Brauer, C. Castillo-Chavez, A.K. Falconar, A. Mubayi, and C.M. Romero-Vivas (2016) Estimation of the reproduction number of the 2015 Zika virus outbreak in Barranquilla, Colombia, and estimation of the relative role of sexual transmission. *Epidemics* **17**: 50–55.

46. Towers, S., O. Patterson-Lomba, and C. Castillo-Chavez (2014) Temporal variations in the effective reproduction number of the 2014 West Africa Ebola outbreak. *PLoS currents* **6**.
47. Verver, S., R.M. Warren, N. Beyers, M. Richardson, G.D. van der Spuy, M.W. Borgdorff, D. A. Enarson, M.A. Behr, and P.D. van Helden (2005) Rate of reinfection tuberculosis after successful treatment is higher than rate of new tuberculosis, *Am. J. Respiratory and Critical Care Medicine* **171**: 1430–1435.
48. W. H. O. (WHO) (2015) Tuberculosis, fact sheet no. 104. Online, October 2015.
49. Yong, K., E.D. Herrera, and C. Castillo-Chavez (2016) From bee species aggregation to models of disease avoidance: The Ben-Hur effect. In *Mathematical and Statistical Modeling for Emerging and Re-emerging Infectious Diseases*, pages 169–185. Springer.
50. Zhao, H., Z. Feng, and C. Castillo-Chavez (2014) The dynamics of poverty and crime. *J. Shanghai Normal University (Natural Sciences· Mathematics)*, pages 225–235.

Part IV

The Future

Chapter 16

Challenges, Opportunities and Theoretical Epidemiology



Lessons learned from the HIV pandemic, SARS in 2003, the 2009 H1N1 influenza pandemic, the 2014 Ebola outbreak in West Africa, and the ongoing Zika outbreaks in the Americas can be framed under a public health policy model that responds after the fact. Responses often come through reallocation of resources from one disease control effort to a new pressing need. The operating models of preparedness and response are ill-equipped to prevent or ameliorate disease emergence or reemergence at global scales [27]. Epidemiological challenges that are a threat to the economic stability of many regions of the world, particularly those depending on travel and trade [132], remain at the forefront of the Global Commons. Consequently, efforts to quantify the impact of mobility and trade on disease dynamics have dominated the interests of theoreticians for some time [14, 143]. Our experience includes an H1N1 influenza pandemic crisscrossing the world during 2009 and 2010, the 2014 Ebola outbreaks, limited to regions of West Africa lacking appropriate medical facilities, health infrastructure, and sufficient levels of preparedness and education, and the expanding Zika outbreaks, moving expeditiously across habitats suitable for *Aedes aegypti*. These provide opportunities to quantify the impact of disease emergence or reemergence on the decisions that individuals take in response to real or perceived disease risks [11, 62, 93]. The case of SARS in 2003 [40], the efforts to reduce the burden of H1N1 influenza cases in 2009 [33, 62, 80, 93] and the challenges faced in reducing the number of Ebola cases in 2014 [24, 27] are but three recent scenarios that required a timely global response. Studies addressing the impact of centralized sources of information [150], the impact of information along social connections [33, 37, 42], or the role of past disease outbreak experiences [105, 130] on the risk-aversion decisions that individuals undertake may help identify and quantify the role of human responses to disease dynamics while recognizing the importance of assessing the timing of disease emergence and reemergence. The co-evolving human responses to disease dynamics are prototypical of the feedbacks that define

complex adaptive systems. In short, we live in a socioepisphere being reshaped by ecoepidemiology in the “Era of Information”.

What are the questions and modes of thinking that should be driving ongoing research on the dynamics, evolution, and control of epidemic diseases at the population level? The challenges of SARS, Ebola, Influenza, Zika, and other diseases are immense. While we may guess which emerging or re-emerging disease may lead to the next possible catastrophe, we cannot know. The contemporary philosopher Yogi Berra is rumored to have said, “Making predictions is hard, especially about the future”. There are some epidemiological topics that have already received some attention but are not yet fully developed. In the rest of this chapter we highlight some challenges, opportunities, and promising approaches in the study of disease dynamics at the population level.

16.1 Disease and the Global Commons

As has been noted, “The identification of a theoretical explanatory framework that accounts for the pattern regularity exhibited by a large number of host–parasite systems, including those sustained by host–vector epidemiological dynamics, is but one of the challenges facing the co-evolving fields of computational, evolutionary, and theoretical epidemiology” [25]. Furthermore, “The emergence of new diseases, the persistence of recurring diseases and the re-emergence of old foes, the result of genetic changes or shifts in demographic, and environmental shifts have increased due to mobility, global connectedness, trade, bird migration, poverty and long-lasting violent conflicts. These diseases often present modeling challenges which may yield to existing analytic techniques but sometimes require new mathematics” [25].

The Global Commons are continuously reshaped by the ability of an increasing proportion of the human population to live, move, or trade nearly anywhere. Therefore, understanding the patterns of interactions between humans, or between humans and vectors, as well as their relationships to patterns of individual movement, particularly those of the highly mobile, is critical to public health responses that effectively ameliorate the ability of a disease to spread. In today’s world, hosts’ risk knowledge (good or bad information) when combined with the ability of public health officials to measure and properly communicate, in a timely manner, real or perceived information on disease risks, limit our ability to derail the spread of emergent and re-emergent diseases, at time scales that make a difference.

16.1.1 Contagion and Tipping Points

Contagion is believed to be the direct or indirect result of interactions between individuals experiencing radically different epidemiological, or immunological, or

social states. Contagion tends to succeed within environments or communities that “facilitate” modes of infection among its members. Contagion is an “understood” or “believed” mode of disease transmission or of “socially transmitted” behaviors, popularized by Malcolm Gladwell, a journalist who made use of his general understanding of the concept of “contagion.” In his construction of reasonable or plausible explanations for the observed and documented dramatic reductions in car thefts and violent crimes in New York City in the 1990s, [71] Gladwell expanded the use of the concept of contagion and tipping point in his development of a framework that captures—as the result of contagion—the spread of a multitude of social ills or virtues [72]. Specifically, contagion is seen in [71] as a force capable of starting and sustaining growth in criminal activity as long as a “critical mass” of individuals capable and willing to commit crimes is available. The growth in criminal activity in New York City is, according to Gladwell, the result of the “interactions” between a large enough pool of criminally active (infected) individuals and individuals susceptible to criminal contagion [71]. Gladwell extends the perspective pioneered by Sir Ronald Ross [141] and his “students” [90–92] to the field of social dynamics.

Gladwell concludes as Ross did in 1911 that implementing control measures (crime contact-reduction measures) that bring the size of the population of criminals (the core) below a critical threshold (tipping point), are sufficient to explain the drastic reductions in criminal activity in NYC. Gladwell concludes, “There is probably no other place [NYC] in the country where violent crime has declined so far, so fast” [71].

16.1.2 Geographic and Spatial Disease Spread

The SARS epidemic of 2002–2003 emphasized the possibility of disease transmission over long distances through air travel, and this has led to metapopulation studies that account for long-distance transmission [5–8]. A metapopulation, in this context, is a population of populations linked by travel. A metapopulation model would have an associated, independent of travel, reproduction number as well as reproduction numbers that account for travel between patches, either temporary travel or permanent migration . This is an Eulerian perspective, describing migration between patches.

An alternative approach to the modeling of the spatial spread of diseases is based on a Lagrangian perspective, which can be formulated, for example, in terms of residence times [18, 25] . This approach has been introduced in Chap. 15. In this structure, actual travel between patches is not described explicitly, and this makes the analysis less complicated. Calculation of the reproduction number and the final size relations is possible.

Another aspect of the study of the spread of diseases is the spatial spread of diseases through diffusion. This has been introduced in Chap. 14 of this book and has been examined in considerable detail in [136], with particular emphasis on epidemic waves.

16.2 Heterogeneity of Mixing, Cross-immunity, and Coinfection

In epidemics, as in the rest of biology, the role of heterogeneity plays a fundamental role and a critical question arises: what is the level of heterogeneity that must be included to address a specific question properly? For example, first-order estimates of the fraction that must be vaccinated to eliminate a communicable disease can be handled with homogeneous mixing models while the elaboration of optimal vaccination strategies in real-life situations often require an age-structured model [28, 81]. The study of nosocomial (in-hospital) infections provides an additional example of the role of heterogeneity in transmission or degree of susceptibility or resistance [38, 39, 106]. The SARS epidemic provided a timely example of the criticality of heterogeneous mixing, in nosocomial transmission [85, 152]. Since there was no treatment available during the SARS epidemic, the main management approach rested on the effectiveness of isolation of diagnosed infectives, quarantine of suspected infectives, and early diagnosis. Quarantine was decided by tracing of contacts made by infectives but in fact few quarantined individuals developed SARS symptoms. The role of early diagnosis and the effectiveness of isolation seemed to have been the key to SARS control with improvements in contact tracing also playing an important role in epidemic control.

Another set of questions arises when one considers the immunological history of individuals or populations. There are many instances in which more than one strain of a disease is circulating within a population and the possibility of cross-immunity between strains becomes important [3, 4]. Mathematically, co-strain co-circulation may lead to models that support a disease-free equilibrium (or non-uniform age distribution), equilibria in which only one strain persists, and an equilibrium in which two strains coexist. The role of cross-immunity in destabilizing disease dynamics (periodic solutions) has been studied extensively in the case of influenza models without age structure [57, 123, 124, 151] and also in age-structured models [29, 30]. Coinfections of more than one disease are also possible and their analysis requires more elaborate models. This is a real possibility with HIV and tuberculosis [96, 119, 133, 135, 140, 144, 154].

16.3 Antibiotic Resistance

In short-term disease outbreaks, antiviral treatment is one of the methods used to treat illness and also to decrease the basic reproduction number \mathcal{R}_0 and thus to lessen the number of cases of disease. However, many infectious pathogens can evolve and generate successor strains that confer drug resistance [55]. The evolution of resistance is generally associated with impaired transmission fitness compared to the sensitive strains of the infectious pathogen [112]. In the absence of treatment, resistant strains may be competitively disadvantaged compared to the

sensitive strains and may go extinct. However, treatment prevents the growth and spread of sensitive strains, and therefore induces a selective pressure that favors the resistant strain to replicate and restores its fitness to a level suitable for successful transmission [2]. This phenomenon has been observed in several infectious diseases, in particular for management of influenza infection using antiviral drugs [138].

Previous models of influenza epidemics and pandemics have investigated strategies for antiviral treatment in order to reduce the epidemic final size (the total number of infections throughout the epidemic), while preventing widespread drug resistance in the population [77, 107, 111, 113, 114]. Through computer simulations, these studies have shown that, when resistance is highly transmissible, there may be situations in which increasing the treatment rate may do more harm than good by causing a larger number of resistant cases than the decrease in cases produced by treatment of sensitive infections. A recent epidemic model [156] has exhibited such behavior and suggested that there may be an optimal treatment rate for minimizing the final size [107, 111, 114].

In diseases such as tuberculosis, which operate on a very long time scale, the same problems arise but the modeling scenario is quite different. It is necessary to include demographic effects such as births and natural deaths in a model. This means that there may be an endemic equilibrium, and that the disease is always present in the population. Instead of studying the final size of an epidemic to measure the severity of a disease outbreak, it is more appropriate to consider the degree of prevalence of the disease in the population at endemic equilibrium as a measure of severity. For diseases such as tuberculosis, in which there are additional aspects such as reinfection, there may be additional difficulties caused by the possibility of backward bifurcations. The importance of understanding the dynamics of tuberculosis treatment suggests that this is a topic that should be pursued [60].

Antibiotic-resistant bacterial infections in hospitals are considered one of the biggest threat to public health. The British Chief Medical Officer, Dame Sally Davies, noted that “the problem of microbes becoming increasingly resistant to the most powerful drugs should be ranked alongside terrorism and climate change on the list of critical risks to the nation.” Yet while antibiotic use is rising, not least in agriculture for farmed animals and fish, resistance is steadily growing.

The challenges posed by the persistence, evolution, and expansion of resistance to antimicrobials are critically important because the number of drugs is limited and no new ones have been created for three decades [2, 16, 38, 65, 99]. We are facing a global crisis in antibiotics, the result of rapidly evolving resistance among microbes responsible for common infections that threaten to turn them into untreatable diseases. Every antibiotic ever developed is at risk of becoming useless. Antimicrobial resistance is on the rise in Europe, and elsewhere in the world.

Dr. Margaret Chan, Director General of the World Health Organization, while addressing a meeting of infectious disease experts in Copenhagen, noted that “A post-antibiotic era means, in effect, an end to modern medicine as we know it. Things as common as strep throat or a child’s scratched knee could once again kill.

For patients infected with some drug-resistant pathogens, mortality has been shown to increase by around 50 per cent.”¹

Strategies suggested to curb the development of resistant hospital-acquired infections include antimicrobial cycling and mixing, that is, models of antibiotic use that make use of two distinct classes of antibiotics that may be distributed over different schedules with the goal of slowing down the evolution of resistance. Cycling alternates both classes of drugs over pre-specified periods of time while mixing distributes both drugs simultaneously at random, that is, roughly half of the physicians would prescribe the first drug class while the other half would prescribe the second class. If the goal is to slow down single class drug resistance then “mixing” is the answer [16] while if the goal is to minimize dual resistance (if such a possibility exists) then the best option is cycling [38]. Of course, there are other factors that may accelerate resistance (physicians’ compliance) or slow down resistance (quarantine and isolation). All the above questions may be addressed via the use of contagion models [38].

16.4 Mobility

The Global Commons are continuously reshaped by the ability of an increasing proportion of the human population to live, move, or trade nearly anywhere. Therefore, understanding the patterns of interactions between humans, between humans and vectors, and the patterns of individuals’ movement, particularly those who are highly mobile, is critical in guiding public health responses to disease spread. In today’s world, hosts’ knowledge of information about risk, combined with the ability of public health officials to measure and properly communicate, in a timely manner, real or perceived information on disease risks, affects our ability to derail the spread of emergent and re-emergent diseases, at scales that make a difference.

Simon Levin showed that understanding scale-dependent phenomena is intimately tied in to our understanding on how information at particular scales impact other scales. His four decade old seminal paper establishing the relationships between processes operating at different scales that highlighted how macroscopic features arise from microscopic processes open the door to the theoretical advances that have dominated the study of ecological and epidemiological systems [101]. Specifically, the theory of metapopulations, common to the study of the models in this book [104, 155], was used to establish the role that localized disturbances have had in maintaining biodiversity [103, 127]. Kareiva et al. observe that there is a multitude of frameworks to study the role of disturbance, noting that, “Models that deal with dispersal and spatially distributed populations are extraordinarily varied, partly because they employ three distinct characterizations of space: as

¹The Independent, Friday 16 March 2012.

‘islands’ (or ‘metapopulations’), as ‘stepping-stones’, or as a continuum” [88]. We choose to deal with mobility in Chap. 15 and this chapter uses a metapopulation approach [80, 93, 104], with populations that exist on discrete “patches” defined by some characteristic(s) (i.e., location, disease risk, water availability, etc.). As is customary, patches are connected by their ability to transfer relevant information among themselves, which, in the context of disease dynamics, is modeled by the ability of individuals to move between patches. Patches may be constructed (defined) by species (human and mosquito) with movement explicitly modeled via patch-specific residence times and under a framework that sees disease dynamics as the result of location-dependent interactions and location characteristic average risks of infection [17, 18].

We observed that “It is therefore important to identify and quantify the processes responsible for observed epidemiological macroscopic patterns: the result of individual interactions in changing social and ecological landscapes” [25]. In the rest of this chapter, we touch on some of the issues calling for the identification of an encompassing theoretical explanatory framework or frameworks. We do this by identifying some of the limitations of existing theory, in the context of particular epidemiological systems. The goal is fostering and re-energizing research that aims at disentangling the role of epidemiological and socioeconomic forces on disease dynamics. In short, epidemic models on social landscapes are better formulated as complex adaptive systems. Now the question becomes, “How does such a perspective help our understanding of epidemics and our ability to make informed adaptive decisions?” These are huge complex questions whose answers have engaged a large number of interdisciplinary and trans-disciplinary teams of researchers. What may be promising directions? In what follows, we discuss some of the modeling used to address some of the challenges and opportunities that we believe must be considered in the field of theoretical epidemiology.

16.4.1 A Lagrangian Approach to Modeling Mobility and Infectious Disease Dynamics

The deleterious impact of the use of cordons sanitaires [58, 100] to limit the spread of Ebola in West Africa points to the importance of developing and implementing novel approaches that may ameliorate the impact of disease outbreaks in areas where timely response to the emergence of novel pathogens is not possible at this time.

Disease risk is a function of the scale and the level of heterogeneity considered. Risk varies by countries and within a country by areas of localized poverty, or as a function of the availability and quality of sanitary/phytosanitary conditions, or as a result of access and the quality of health care, or variability on the levels of individual education, or as a result of engrained cultural practices and norms. Travel and trade, easily bypassing in today’s world the natural or cultural boundaries defined by many of factors just outlined, are now seen as engines that

drive the spread of pests and pathogens across regional and global scales. Hence, the identification of explanatory frameworks that help to disentangle the role of epidemiological, socioeconomic, and cultural perspectives on disease dynamics becomes evident and necessary in the Global Commons. Further, since the work of Sir Ronald Ross over a century ago [141], efforts to develop a mathematical framework that allow us to tease out the role of various mechanisms on disease spread while enhancing our understanding of what may be the most effective measures to manage or eliminate a disease, the fields of mathematical and theoretical epidemiology have developed into rich and useful fields of their own. Their role in the development of public health policy and the study of disease evolution within hosts (immunology) and between populations and its relationship to the study of host-pathogen interactions within ecology or community ecology are now integral components of the education and training of theoreticians and practitioners alike [1, 12, 19–23, 26, 41, 53, 54, 59, 74, 76, 82, 102, 122, 157].

The use of (per capita) contact or activity rates in modeling the interactions between individuals, that is, who mixes with whom or who interacts with whom, has been the natural social dynamics currency used to model human-to-human or vector-to-human interactions in the context of the transmission dynamics of communicable diseases. The “physics or chemistry traditions” are used to model disease transmission as the result of the “collisions” between individuals (with different energy or activity levels) in different epidemiological states. Further, movement, typically modeled using a metapopulation approach, is seen as the relocation between patches of non-identifiable individuals. The scholarly and extensive review in [83] addresses this perspective within homogeneous and heterogeneous mixing (age-structured) population models (see also [30]). Weakening the assumption of homogeneous mixing via contacts in epidemiology has been addressed using network-based analyses that identify host contact patterns and clusters [13, 120, 121, 128] (and references therein with [121] offering an extensive review). Focusing on how each individual is connected within the population has been used to address the effects of host behavioral response on disease prevalence (see [67, 68, 110] for a review). Other approaches have included the effects of behavioral changes triggered by “fear” and/or awareness of disease [56, 66, 131, 134]. Although this stress-induced behavior may benefit public health efforts in some cases, it can also cause somewhat unpredictable outcomes [75].

However, the fact remains that our ability to determine (and hence define) what an effective contact is in the context of communicable diseases, that is, our ability to measure the average number of contacts that a typical patch resident has per unit of time and where, has been hampered by high levels of uncertainty. Therefore, when we ask, what is the average rate of contacts that an individual has while riding a packed subway in Japan or Mexico City, or what is the average rate of contacts that an individual has at a religious event involving hundreds of thousands of people, including pilgrimages, one quickly arrives at the conclusion that different observers are extremely likely to arrive at very distinct understandings and quantifications of the frequency, intensity, and levels of heterogeneity involved. In short, this perspective puts emphasis on the use of a different currency (residence

times) because measuring contacts at the places where the risk of infection is the highest, pilgrimages, massive religious ceremonies, “Woodstock time events”, packed subways, and other forms of mass gathering or transportation have not been done to the satisfaction of most researchers. The risk of acquiring an infectious disease within a flight can be measured at least in principle as a function of the time that each individual of x -type spends flying, the number of passengers, and the likelihood that an infectious individual is on board. For example, measuring the risk of acquiring tuberculosis, an airborne disease that may spread by air circulation in a flight, may be more a function of the duration of the flight and the seating arrangement than the average rate of contacts per passenger within the flight (see [31] and references therein). Furthermore, replication studies that measure risk of infection in a given environment may indeed be possible under a residence time model. In short, the risks of acquiring an infection can be quantified as a function of the time spent (residence time) within each particular environment. The Lagrangian modeling approach builds (epidemiological) models by tracking individuals’ patch-residence times and estimating their contacts according to the time spent in each environment [32]. The value of these models increases when we have the ability to assess risk as a patch-specific characteristic. In short, the use of a Lagrangian modeling perspective rather than the use of contacts is tied to the difficulties that must be faced when the goal is to measure the average rate of contacts per type- x individual in the environments that facilitate transmission the most.

The Lagrangian approach is highlighted here via the formulation of a disease model involving the joint dynamics of an n -patch geographically structured population with individuals moving back and forth from their place of residence to other patches. Each of these patches (or environments) is defined by its associated risk of residence-time infection. Patch risk measurements account for environmental, health, and socioeconomic conditions. The Lagrangian approach [73, 125, 126] keeps track of the identity of hosts regardless of their geographical/spatial position. The use of Lagrangian modeling in living systems was, to the best of our knowledge, pioneered and popularized by Okubo and Levin [125, 126] in the context of animal aggregation. Recently, Lagrangian approaches have also been used to model human crowd movement and behavior [15, 49, 78, 79] and in the context of bioterrorism [31].

Here, host-residence status and mobility across patches are monitored with the help of a residence times matrix $\mathbb{P} = (p_{ij})_{1 \leq i, j \leq n}$, where p_{ij} is the proportion of time residents of Patch i spend in Patch j . Letting N_i denote the population of Patch i predispersal, that is, when patches are isolated, we conclude that effective population size in Patch i , at time t , is given by $\sum_{j=1}^n p_{ji} N_j$. That is, the effective population within each patch must account for the residents and visitors to Patch i at time t . A typical SIS model captures this Lagrangian approach in an n - patch setting via the system of nonlinear differential equations:

$$\begin{aligned}\dot{S}_i &= b_i - d_i S_i + \gamma_i I_i - \sum_{j=1}^n (S_i \text{ infected in Patch } j) \\ \dot{I}_i &= \sum_{j=1}^n (S_i \text{ infected in Patch } j) - \gamma_i I_i - d_i I_i,\end{aligned}\tag{16.1}$$

where b_i , d_i , and γ_i denote the constant recruitment, the per capita natural death, and recovery rates, respectively, in Patch i . The effective population $\sum_{j=1}^n p_{ij} N_j$ in each Patch i , $i = 1, \dots, n$ includes $\sum_{j=1}^n p_{ij} I_j$ infected individuals. Therefore, the infection term is modeled as follows:

$$S_i \text{ infected in patch } j = \beta_j \times p_{ij} S_i \times \frac{\sum_{k=1}^n p_{kj} I_k}{\sum_{k=1}^n p_{kj} N_k}.$$

The likelihood of infection in each patch is tied to the environmental risks, defined by the “transmission/risk” vector $\mathcal{B} = (\beta_1, \beta_2, \dots, \beta_n)^t$ and the proportion of time spent in particular area. Letting $I = (I_1, I_2, \dots, I_n)^t$, $\bar{N} = (\frac{b_1}{d_1}, \frac{b_2}{d_2}, \dots, \frac{b_n}{d_n})^t$, $\tilde{N} = \mathbb{P}t\bar{N}$, $d = (d_1, d_2, \dots, d_n)^t$, and $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_n)^t$ allows to rewrite System 16.1 in the following single vectorial form

$$\dot{I} = \text{diag}(\tilde{N} - I)\mathbb{P}\text{diag}(\mathcal{B})\text{diag}(\tilde{N})^{-1}\mathbb{P}^t I - \text{diag}(d + \gamma)I.\tag{16.2}$$

The dynamics of the disease in all of the patches depends on the patch connectivity structure. Therefore, if the residence-time matrix \mathbb{P} is irreducible, patches are strongly connected, then system 2 supports a sharp threshold property. That is, the disease dies out or persists (in all patches) whenever the basic reproduction number \mathcal{R}_0 is less than or greater than unity [18]. \mathcal{R}_0 is given by

$$\mathcal{R}_0 = \rho(\text{diag}(\tilde{N})\mathbb{P}\text{diag}(\mathcal{B})\text{diag}(\tilde{N})^{-1}\mathbb{P}^t V^{-1}),$$

where ρ denotes the spectral radius and $V = -\text{diag}(d + \gamma)$. The dynamics of the system when the matrix \mathbb{P} is not irreducible can be characterized using the following patch-specific basic reproduction numbers:

$$\mathcal{R}_0^i(\mathbb{P}) = \frac{\beta_i}{\gamma_i + d_i} \times \sum_{j=1}^n \left(\frac{\beta_j}{\beta_i} \right) p_{ij} \left(\frac{p_{ij} \left(\frac{b_i}{d_i} \right)}{\sum_{k=1}^n p_{kj} b_k d_k} \right).$$

The disease persists in Patch i whenever $\mathcal{R}_0^i(\mathbb{P}) > 1$, whereas the disease dies out in Patch i if $p_{kj} = 0$ for all $k = 1, \dots, n$, and $k \neq i$, provided $p_{ij} > 0$ and $\mathcal{R}_0^i(\mathbb{P}) < 1$.

Patch-specific disease persistence can be established using the average Lyapunov theorem [86] (see [18] for more details).

In Model 16.2, human behavior is crudely incorporated through the use of a constant mobility matrix \mathbb{P} . The role that adaptive human behavior may play in response to disease dynamics is captured, also rather crudely, via a phenomenological approach that assumes that individuals avoid or spend less time in areas of high prevalence. This effect is captured by placing natural restrictions on the entries of \mathbb{P} . The inequalities $\frac{p_{ij}(I_i, I_j)}{\partial I_j} \leq 0$ and $\frac{p_{ij}(I_i, I_j)}{\partial I_i} \geq 0$, for $(i, j) \in 1, 2$, guarantee the expected behavioral response. An example of such dependency could be captured by the following functions: $p_{ii}(I_i, I_j) = \frac{\sigma_{ii} + \sigma_{ii} I_i + I_j}{1 + I_i + I_j}$ and $p_{ij}(I_1, I_2) = \sigma_{ij} \frac{1 + I_1}{1 + I_1 + I_2}$, for $(i, j) \in 1, 2$ and $\sigma_{ij} = p_{ij}(0, 0)$, are such that $\sum_{j=1}^2 \sigma_{ij} = 1$. The simulation below shows how a crude, density-dependent modeling mobility approach can alter the expected disease dynamics from those generated under constant \mathbb{P} (Figs. 16.1 and 16.2). In the special case, where there is no movement between patches ($p_{12} = p_{21} = \sigma_{12} = \sigma_{21} = 0$), that is, there is no behavioral change, the two populations support, as expected, the same dynamics (see the blue curves in Figs. 16.1 and 16.2).

The speed at which the vector-borne Zika virus disease spread throughout Latin America, Central America, and the Caribbean (then hitting Mexico and the United States) was strongly linked to human mobility patterns. Travelers transport the disease and infect native mosquitoes. Here, it is assumed that vector mobility is negligible and the assumptions proceed to incorporate the life history and epidemiology of mosquitoes [10, 84, 98, 108, 109, 141], which can be effectively captured

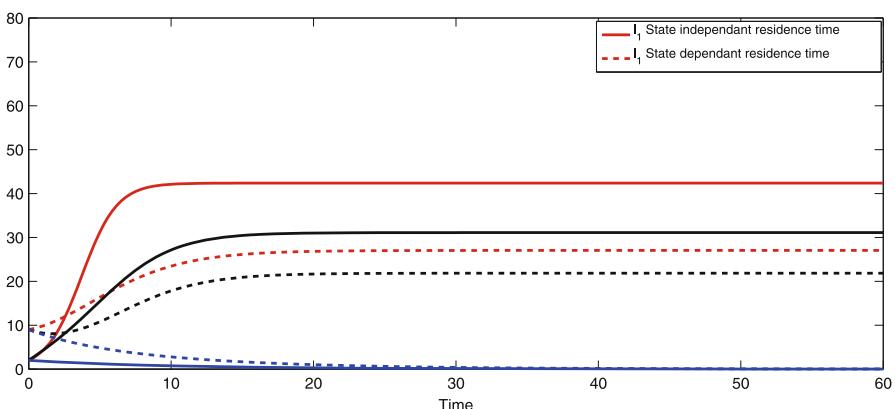


Fig. 16.1 Dynamics of the disease in Patch 1 for three special cases. The symmetric residence times ($p_{12} = p_{21} = \sigma_{12} = \sigma_{21} = 0.5$) are described by the solid and dashed black curves. The blue curves represent the case where there is no movement between patches, that is, $p_{12} = p_{21} = \sigma_{12} = \sigma_{21} = 0$. The red curves represent the high-mobility case for which $p_{12} = p_{21} = \sigma_{12} = \sigma_{21} = 1$. If there is no movement between the patches (blue curves), the disease dies out in the low risk Patch 1 in both approaches with $\mathcal{R}_0^1 = \frac{\beta_1}{d_1 + \gamma_1} = 0.7636$. The vertical axis represents the prevalence of the disease in Patch 1. Figure courtesy of Ref. [18]

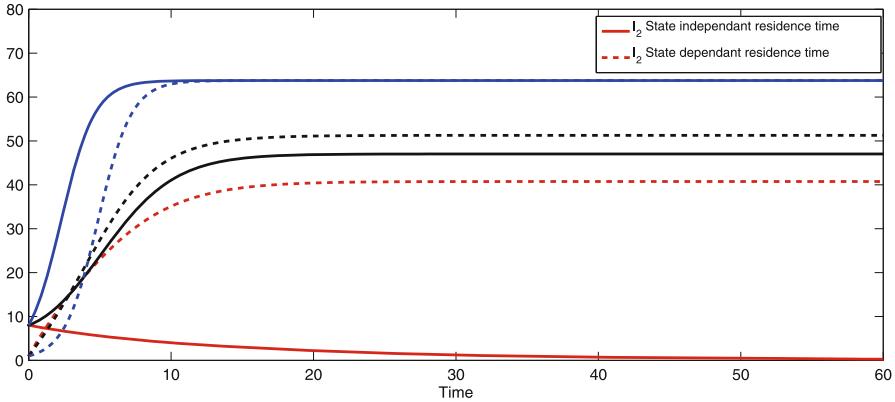


Fig. 16.2 Dynamics of the disease in Patch 2. In the high-mobility case $p_{12} = p_{21} = \sigma_{12} = \sigma_{21} = 1$ (and then $p_{11} = p_{22} = \sigma_{11} = \sigma_{22} = 0$), the disease dies out (solid red curve) for \mathbb{P} constant, with $\tilde{\mathcal{R}}_0^2 = \frac{\beta_1}{\gamma_2 + d_2} = 0.8571$. For the constant residence-time matrix, the system is strangely decoupled because individuals of Patch 1 spend all their time in Patch 2, whereas individuals of Patch 2 spend all their time in Patch 1. Hence, Patch 2 individuals (d_2 and μ_2) are subject exclusively to the environmental conditions that define Patch 1 (β_1), and so the basic reproduction of the “isolated” Patch 1 is $\tilde{\mathcal{R}}_0^2 = \frac{\beta_1}{\gamma_2 + d_2}$ and the disease dies out because $\tilde{\mathcal{R}}_0^2 = 0.8571$. The disease persists if \mathbb{P} state-dependent (dashed red curve) as $p_{12}(I_1, I_2) = \frac{1+I_1}{1+I_1+I_2}$, $p_{21}(I_1, I_2) = \frac{1+I_2}{1+I_1+I_2}$, $p_{11}(I_1, I_2) = \frac{I_2}{1+I_1+I_2}$, and $p_{22}(I_1, I_2) = \frac{I_1}{1+I_1+I_2}$. Figure courtesy of Ref. [18]

by decoupling host and vector mobility [98, 145]. Figure 16.3 and System 16.3 illustrate the approach. A Lagrangian model based on residence times has been proposed recently for vector-borne diseases like dengue, malaria, and Zika [17]. The appropriateness of the Lagrangian approach for the study of the dynamics of vector-borne diseases lies also in its assessment of the life-history specifics of the vector involved [145].

$$\begin{aligned}\dot{I}_h &= \beta_{vh} \text{diag}(N_h - I_h) \mathbb{P} \text{diag}(a) \text{diag}(\mathbb{P}^t N_h)^{-1} I_v - \text{diag}(\mu + \gamma) I_h \\ \dot{I}_v &= \beta_{hv} \text{diag}(a) \text{diag}(N_v - I_v) \text{diag}(\mathbb{P}^t N_h)^{-1} \mathbb{P}^t I_h - \text{diag}(\mu_v + \delta) I_v.\end{aligned}\quad (16.3)$$

Lagrangian approaches have been used to model vector-borne diseases (see [48, 87, 139, 142, 153] and other references contained therein), although these researchers have not considered the impact that the residence-time matrix \mathbb{P} may have on patch effective population size. Specifically, in [48, 142], the effects of movement on patch population size at time t are ignored, namely, the population size in each Patch j is fixed at N_j . In [139], it is assumed that human mobility across patches does not produce any “net” change on the patch population size. On the other hand, in Model 16.3 the relationship between each patch population and mobility is dynamic and explicitly formulated. Moreover, the limited (vector

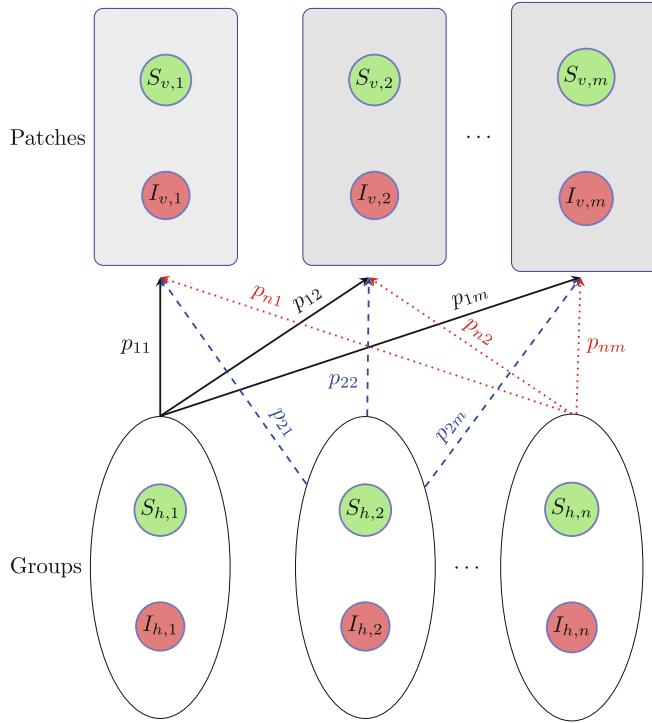


Fig. 16.3 Flow diagram of a Lagrangian model in which the host structure is decoupled from the vectors' structure. Figure courtesy of Ref. [17]

mobility is ignored) Lagrangian approach used here to model the dynamics of vector-borne diseases captures some unique features because the “spatial” structure of mosquitoes is not the same as that of humans. Mosquitoes are stratified into m patches (that may represent, for example, oviposition or breeding sites or forests) with infection taking place still within each Patch j , characterized by its particular risk $\beta_{vh}a_j$ for $j = 1, \dots, m$. Here, β_{vh} represents the infectiousness of human to mosquitoes bite with a_j denoting the per capita biting rate in Patch j . Hosts, on the other hand, are structured by social groups or age classes (n). This residence habitat division can be particularly useful in the study of the impact of target control strategies.

The model in [17] describes the interactions of n host groups in m patches via System 16.3, where

$$I_h = [I_{h,1}, I_{h,2}, \dots, I_{h,n}]^t, I_v = [I_{v,1}, I_{v,2}, \dots, I_{v,m}]^t$$

$$N_h = [N_{h,1}, N_{h,2}, \dots, N_{h,n}]^t, \bar{N}_v = [\bar{N}_{v,1}, \bar{N}_{v,2}, \dots, \bar{N}_{v,m}]^t$$

$$\delta = [\delta_1, \delta_2, \dots, \delta_m]^t, a = [a_1, a_2, \dots, a_m]^t, \text{ and } \mu = [\mu_1, \mu_2, \dots, \mu_n]^t.$$

The infected host population is denoted by the vector I_h and the host population by N_h . The infected vector population is denoted by I_v and the mosquito population by N_v . The parameters a_i , δ_i , and μ_v denote the biting, death rate of control, and natural death rate of mosquitoes in Patch j , for $j = 1, \dots, m$. The infectiousness of human to mosquitoes is β_{vh} , whereas the infectiousness of mosquitoes to humans is given by β_{hv} . The host recovery and natural mortality rates are given, respectively, by γ and μ . Finally, the matrix \mathbb{P} represents the proportion of time host of group i , $i = 1, \dots, n$, spend in Patch j , $j = 1, \dots, m$. The basic reproduction number of Model 3, with m patches and n groups, is given by $\mathcal{R}_0^2(m, n) = \rho(M_{vh}M_{hv})$, where

$$M_{hv} = \beta_{hv} \text{diag}(a) \text{diag}(\mathbb{P}^t N_h)^{-1} \text{diag}(N_v) \mathbb{P}^t \text{diag}(\mu + \gamma)^{-1}$$

and

$$M_{vh} = \beta_{vh} \text{diag}(N_h) \mathbb{P} \text{diag}(\mathbb{P}^t N_h)^{-1} \text{diag}(a) \text{diag}(\mu_v + \delta)^{-1}.$$

If the host–vector network configuration is irreducible, then System 16.3 is cooperative and strongly concave with an irreducible Jacobian, hence the theory of monotone systems, particularly Smith’s results [146], guarantee the existence of a sharp threshold. That is, the disease-free equilibrium is globally asymptotically stable if $\mathcal{R}_0^2(m, n)$ is less than unity and a unique globally asymptotic stable interior endemic equilibrium exists otherwise. The effects of various forms of heterogeneity on the basic reproduction number have been explored in [17], and we have found, for example, that the irreducibility of the residence-time matrix \mathbb{P} is no longer sufficient to ensure a sharp threshold property, although the irreducibility of the host–vector network configuration is necessary for such property [17].

The Lagrangian approach to disease modeling can use contacts [32] or times or both as its currency. Here, we choose time-spatial-dependent risk, that is, we choose to handle social heterogeneity by keeping track of individuals’ social or geographical membership. In this context, it is possible to include adaptive responses, for example, via the inclusion of prevalence-dependent dispersal coefficients. In this setting, the underlying hypothesis is that host behavioral responses to disease are automatic: either constant or following a predefined function. The average residence time \mathbb{P} incorporates the average behavior of all hosts in each patch. This assumption is rather crude because it implicitly assumes that hosts have accurate information on health status and patch prevalence and respond to risk of infection accordingly. The incorporation of the role that human decisions, as a function of what individuals value and the cost that individuals place on these choices and trade-offs, within systems that account for the overall population disease dynamics has been addressed recently [61, 132] and discussed in economic epidemiology.

16.5 Behavior, Economic Epidemiology, and Mobility

The movement/behavior of individuals within and between patches may be driven by real or perceived personal economic risk and associated social dynamics. Embedding behavioral-driven decisions in epidemiological models has shed new perspectives on the modeling of disease dynamics [61], expanding available option to manage infectious diseases [44, 62]. Economic epidemiological modeling (EEM) has a history of addressing the role of individuals' behavior when facing the risk of disease. However, it has often failed to incorporate within host-pathogen feedback mechanisms [34–36, 52, 70, 97, 137]. EEMs that account for host-pathogen feedback mechanisms has propelled the study of the ways that contact decisions impact disease emergence or alter infectious disease-transmission dynamics. Decisions involved may include the determination to engage in trade on particular routes [89, 94, 95, 129], or to travel to specific places [62, 147–149], or to make contact with or to avoid particular types of people [61, 63, 116]. EEMs advance the view that the emergence of novel zoonotic diseases, such as SARS or the Nipah virus, depend on the choices that bring people into contact with other species [50, 51]. EEMs are usually built under the assumption that associated disease risks are among the factors that individuals must consider when making decisions. Individual decision-making processes, within epidemic outbreaks, must incorporate the humans' cost-benefit-driven adaptive responses to risk.

16.5.1 Economic Epidemiology

Simple EEMs are, by mathematical necessity, initially built on classical compartmental epidemiological models that account for the orderly transition of individuals facing a communicable disease, through the susceptible, infected, and recovered disease stages: the result of social and environmental interactions. EEMs assume that the amount of activity one participates in, with whom, and where may all be envisioned as the solutions to an individual decision problem. It is assumed that individual decision problems are generated by rational-value formulations based on (driven by) personal, real or perceived, cost of disease, and disease avoidance: decisions constrained by underlying population-level disease dynamics. Thus, finding effective ways of modeling rational value connections to individualized cost-benefit analyses of disease risk is fundamental to the building of useful EEMs. It is a quite challenging enterprise.

EEM approaches have precursors in the epidemiological literature [9, 64, 69]. EEM construction has been strongly influenced by past and ongoing work on the exploitation of species [45–47], a literature that addresses optimal harvesting questions in the context of wild species, or the control of invasive pests, or the management of forestry system. The methodology for modeling behavior within an EEM rests on a proper specification of behavioral costs and a description of

the payoffs linked to such behaviors; the stipulation of an appropriate objective function, congruent with the decision-makers' goals; the coupling to the dynamics of the natural resource and/or infectious human capital; and the mechanisms available for a decision-maker to alter his or her behavior and the behaviors of those around him or her. Although not all motivations for mitigation against infection are monetary in nature, we choose to refer to them as economic.

Modeling whether or not an individual undertakes infection-causing behavior provides a natural starting point since it is connected to the rate of generation of secondary cases of infection per unit of time, the so-called incidence rate. A simple incidence function that captures the instantaneous expectation of the rate of new infections at a given time is therefore given by

$$S(t)cP_{SI}(t)\rho,$$

where $S(t)$ is the number of individuals susceptible to the disease, c is the average amount of activity they engage in, $P_{SI}(t)$ is the probability that a unit of such activity takes the susceptible individual in contact with infectious individuals/material, and ρ is the probability that such contact successfully infects.

A decision to reduce the volume of activity one engages in (lowering c) has been shown in many cases to be phenomenologically identical to reducing one's chances of coming in contact with infection (lowering $P_{SI}(t)$) by altering where the activity takes place and with whom one engages or by substituting a particular behavior for a riskier one [62, 118]. The modeling assumes that individuals derive benefits from making contacts but may incur costs associated with an infection. Hence, the modeling assumes that activity volume or contacts are chosen to maximize expected utility (rudimentarily, benefit less cost), balancing the marginal value of a contact against the increased risk of infection. The utility function is assumed to depend on the health status of the individual and the contacts that they make, that is, the utility of a representative individual of health status h is given, for example, by the function

$$U_h = U(h, C^h). \quad (16.4)$$

The utility function is assumed to be concave, decreasing in illness and increasing in contacts. If the probability of transitioning from susceptible to infected health status depends on the rate of contacts, the optimal choice of contacts is the solution to a dynamic programming problem:

$$V_t(h) = \max_{C_s} \left\{ U_t(h_t, C_t^h) + r \sum_j \rho^{hj} V_{t+1}(j) \right\}, \quad (16.5)$$

where r is the discount rate and ρ^{hj} is the probability of transition from health state h to health state j . This probability depends on the current state of the system, $\{S(t), I(t), R(t)\}$, the behavior of individuals in other health classes, C^{-h} , and

the behavior of individuals in the decision-makers' own health class, \bar{C}^h . In short, we have a complex adaptive system where individuals within the model, in this example, impact disease outcomes (through changes in the incidence). Eqs. (16.4) and (16.5) are both optimized from an individual perspective. Within this individual context, EEMs have shown that individual distancing, conditional on health status, plays an important role in the spread of infectious disease. However, it has also been shown that the provision of incentives for infectious individuals to self-quarantine is likely to be welfare-enhancing [43, 44, 61, 115]. Thus, understanding how the individual responds to relative costs of disease and disease prevention is critical to the design of public policy that affects those costs. Indeed, the role of recovered individuals in protecting susceptible individuals has been generally overlooked in public health interventions, and yet it is known that their behavior is, in fact, critical to disease management due to the positive externality the individuals' contacts generate once in an immune, non-disease-transmitting state [63]. The benefits of herd immunity include the positive externality associated with acquired immunity but may, in turn, be nullified by nontargeted social-distancing policies that induce such immune individuals to reduce contacts. By incentivizing the maintenance of contacts by recovered individuals policy may lower the probability of susceptible individuals contacting infected individuals and/or allow susceptible and infected individuals to individually increase contacts without changing the probability of infection.

16.5.2 Lagrangian and Economic Epidemiology Models

Theoretical epidemiology aims to disentangle the role of epidemiological and socioeconomic forces on disease dynamics. However, the role of behavior and individual decisions in response to a changing epidemic landscape has not been tackled systematically. In this chapter, we highlight alternative ways for modeling disease transmission that can use contacts as its currency or residence times or both. It seems evident that the use of contacts, in the context of influenza, Ebola, tuberculosis, or other communicable diseases (as opposed to sexually transmitted diseases), while intellectually satisfying, fails to recognize the fact that contacts cannot be measured effectively in settings where the risk of acquiring such infections is the highest. In fact, when contact-based models are fitted to data, it has become clear that contact rates play primarily the role of fitting parameters; in other words, if the goal is connecting models to data that include transmission mechanisms, then the use of contacts has serious shortcomings. Therefore, in order to advance the role of theory, we need models that are informed by data. Hence, the need to invest on efforts that bring forth alternative modes of modeling. While Lagrangian approaches are not a panacea, their use extends the possibilities because they depend on parameters like residence times and average time to infection for a given environment (risk), that is, parameters that can be measured. Frameworks should be explored and compared

and their analyses contrasted. We have revisited recent work that equates behavior with cost–benefit decisions, which, in turn, are linked, within our framework, to health status and population-level dynamics, the components of a complex adaptive system. Connecting the Lagrangian movement-modeling approach with EEMs seems promising, albeit computationally and mathematically challenging. However, as discussed in [117], the perception that the benefits of disease control are limited by the capacity of the weakest link in the chain to respond effectively is not a basic result of EEM models, which actually show that it may not be in within the ability of an individual in a poor community/country to do more risk mitigation. The need for richer communities or nations to find ways to incentivize greater levels of disease-risk mitigation in poor countries may be the best approach.

Simon Levin, in his address as the 2004 recipient of the Heineken award, placed our narrow perspective in a broader powerful context:

A great challenge before us is thus to understand the dynamics of social norms, how they arise, how they spread, how they are sustained and how they change. Models of these dynamics have many of the same features as models of epidemic spread, no great surprise, since many aspects of culture have the characteristics of being social diseases. 1998 Heineken award winner Paul Ehrlich and I have been directing our collective energies to this problem, convinced that it is as important to understand the dynamics of the social systems in which we live as it is to understand the ecological systems themselves. Understanding the links between individual behavior and societal consequences, and characterizing the networks of interaction and influence, create the potential to change the reward structures so that the social costs of individual actions are brought down to the level of individual payoffs. It is a daunting task, both because of the amount we still must learn, and because of the ethical dilemmas that are implicit in any form of social engineering. But it is a task from which we cannot shrink, lest we squander the last of our diminishing resources.

References

1. Anderson, R.M., R. M. May, and B. Anderson (1992) *Infectious Diseases of Humans: Dynamics and Control*, volume 28. Wiley Online Library.
2. Andersson, D.I. and B. R. Levin (1999) The biological cost of antibiotic resistance. *Current opinion in microbiology* **2**: 489–493.
3. Andreasen, V. (2003) Dynamics of annual influenza a epidemics with immuno-selection, *J. Math. Biol.* **46**: 504–536.
4. Andreasen, V., J. Lin, and S. A. Levin (1997) The dynamics of cocirculating influenza strains conferring partial cross-immunity, *J. Math. Biol.* **35**: 825–842.
5. J. Arino, J., R. Jordan, and P. Van den Driessche (2007) Quarantine in a multi-species epidemic model with spatial dynamics, *Math. Biosc.* **206**: 46–60.
6. Arino, J. and P. Van Den Driessche (2003) The basic reproduction number in a multi-city compartmental epidemic model, In *Positive Systems*, pages 135–142. Springer
7. Arino, J. and P. Van den Driessche (2003) A multi-city epidemic model, *Mathematical Population Studies* **10**: 175–193.
8. Arino, J. and P. van den Driessche (2006) Metapopulation epidemic models, a survey, *Fields Institute Communications* **48**: 1–13.
9. Auld, M.C. (2003) Choices, beliefs, and infectious disease dynamics, *J. Health Economics* **22**:361–377.

10. Bailey, N.T., et al, (1982) The biomathematics of malaria: the biomathematics of diseases: 1, The biomathematics of malaria. *The Biomathematics of Diseases* **1**.
11. Bajardi, P. C. Poletto, J. J. Ramasco, M. Tizzoni, V. Colizza, and A. Vespignani (2011) Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic, *PloS One* **6**: e16591.
12. Banks, H.T. and C. Castillo-Chavez (2003) Bioterrorism: Mathematical Modeling Applications in Homeland Security. SIAM.
13. Bansal, S., B. T. Grenfell, and L. A. Meyers (2007) When individual behaviour matters: homogeneous and network models in epidemiology, *J. Roy. Soc. Interface* **4**: 879–891.
14. Baroyan, O., L. Rvachev, U. Basilevsky, V. Ermakov, K. Frank, M. Rvachev, and V. Shashkov (1971) Computer modelling of influenza epidemics for the whole country (USSR), *Adv. in Applied Probability* **3**: 224–226.
15. Bellomo, N., B. Piccoli, and A. Tosin (2012) Modeling crowd dynamics from a complex system viewpoint, *Math. models and methods in applied sciences* **22**,(supp02): 1230004.
16. Bergstrom, C. T., Lo, M., & Lipsitch, M. (2004) Ecological theory suggests that antimicrobial cycling will not reduce antimicrobial resistance, *Proc Natl Acad Sci.* **101**: 13285–13290.
17. Bichara, D. and C. Castillo-Chavez (2016) Vector-borne diseases models with residence times—a Lagrangian perspective, *Math. Biosc.* **281**: 128–138.
18. Bichara, D., Y. Kang, C. Castillo-Chavez, R. Horan, and C. Perrings (2015) SIS and SIR epidemic models under virtual dispersal, *Bull. Math. Biol.* **77**: 2004–2034.
19. Brauer F., and C. Castillo-Chávez (2013) Mathematical models for communicable diseases. No. 84 in CBMS-NSF regional conference series in applied mathematics Philadelphia: Society for Industrial and Applied Mathematics.
20. Brauer, F., P. van den Driessche and J. Wu, eds. (2008) Mathematical Epidemiology, Lecture Notes in Mathematics, Mathematical Biosciences subseries 1945, Springer, Berlin - Heidelberg - New York.
21. Castillo-Chavez, C. (2002) Mathematical Approaches for Emerging and Reemerging Infectious Diseases: an Introduction, Volume 1. Springer Science & Business Media.
22. Castillo Chavez, C. (2002) Mathematical Approaches for Emerging and Reemerging Infectious Diseases: Models, Methods, and Theory, volume 126. Springer, 2002.
23. Castillo-Chavez, C. (2013) Mathematical and Statistical Approaches to AIDS Epidemiology, volume 83. Springer Science & Business Media.
24. Castillo-Chavez, C., K. Barley, D. Bichara, D. Chowell, E. D. Herrera, B. Espinoza, V. Moreno, S. Towers, and K. Yong (2016) Modeling Ebola at the mathematical and theoretical biology institute (MTBI). *Notices of the AMS* **63**.
25. Castillo-Chavez, C., D. Bichara, and B. R. Morin (2016) Perspectives on the role of mobility, behavior, and time scales in the spread of diseases, *Proc. Nat. Acad. Sci.* **113**: 14582–14588.
26. Castillo-Chavez, C. and G. Chowell (2011) Preface: Mathematical models, challenges, and lessons learned Special volume on influenza dynamics, volume 8. *Math. Biosc. Eng.* **8**:1–6.
27. Castillo-Chavez, C., R. Curtiss, P. Daszak, S. A. Levin, O. Patterson-Lomba, C. Perrings, G. Poste, and S. Towers (2015) Beyond Ebola: Lessons to mitigate future pandemics, *The Lancet Global Health* **3**: e354–e355.
28. Castillo-Chavez,C., Z. Feng, et al (1996) Optimal vaccination strategies for TB in age-structure populations, *J. Math. Biol.* **35**: 629–656.
29. Castillo-Chavez, C., H. W. Hethcote, V. Andreasen, S. A. Levin, and W. M. Liu (1988) Cross-immunity in the dynamics of homogeneous and heterogeneous populations, *Proceedings of the Autumn Course Research Seminars Mathematical Ecology*, pages 303–316.
30. Castillo-Chavez, C., H. W. Hethcote, V. Andreasen, S. A. Levin, and W. M. Liu (1989) Epidemiological models with age structure, proportionate mixing, and cross-immunity, *J. Math. Biol.* **27**: 233–258.
31. Castillo-Chavez, C. and B. Song (2004) Dynamical models of tuberculosis and their applications, *Math. Biosc. Eng.* **1**: 361–404.
32. Castillo-Chavez, C., B. Song, and J. Zhang (2003) An epidemic model with virtual mass transportation: the case of smallpox in a large city, In *Bioterrorism: Mathematical Modeling Applications in Homeland Security*, pages 173–197. SIAM.

33. Cauchemez, S., A. Bhattacharai, T. L. Marchbanks, R. P. Fagan, S. Ostroff, N. M. Ferguson, D. Swerdlow, S. V. Sodha, M. E. Moll, F. J. Angulo, et al (2011) Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza, *Proc. Nat. Acad. Sci.* **108**: 2825–2830.
34. Chen, F., M. Jiang, S. Rabidoux, and S. Robinson (2011) Public avoidance and epidemics: insights from an economic model, *J. Theor. Biol.* **278**: 107–119.
35. Chen F. (2004) Rational behavioral response and the transmission of stds. *Theor. Pop. Biol.* **66**: 307–316.
36. Chen, F.H. (2009) Modeling the effect of information quality on risk behavior change and the transmission of infectious diseases, *Math. Biosc.* **217**: 125–133.
37. Chew, C. and G. Eysenbach (2010) Pandemics in the age of twitter: content analysis of tweets during the 2009 h1n1 outbreak, *PloS One* **5**: e14118, 2010.
38. Chow, K., X. Wang, R. Curtiss III, and C. Castillo-Chavez (2011) Evaluating the efficacy of antimicrobial cycling programmes and patient isolation on dual resistance in hospitals, *J. Biol. Dyn.* **5**: 27–43, 2011.
39. Chow, K.C., X. Wang, and C. Castillo-Chávez (2007) A mathematical model of nosocomial infection and antibiotic resistance: evaluating the efficacy of antimicrobial cycling programs and patient isolation on dual resistance Mathematical and Theoretical Biology Institute archive, 2007.
40. Chowell, G., P. W. Fenimore, M. A. Castillo-Garsow, and C. Castillo-Chavez (2003) SARS outbreaks in Ontario, Hong Kong and Singapore: the role of diagnosis and isolation as a control mechanism, *J. Theor. Biol.* **224**: 1–8, 2003.
41. Chowell, G., J. M. Hyman, L. M. Bettencourt, and C. Castillo-Chavez (2009) Mathematical and Statistical Estimation Approaches in Epidemiology, Springer.
42. Chowell, G., J. M. Hyman, S. Eubank, and C. Castillo-Chavez (2003) Scaling laws for the movement of people between locations in a large city, *Phys. Rev. E* **68**: 066102, 2003.
43. Chowell, G., H. Nishiura, and L. M. Bettencourt (2007) Comparative estimation of the reproduction number for pandemic influenza from daily case notification data, *J. Roy. Soc. Interface* **4**: 155–166.
44. Chowell, G., C. Viboud, X. Wang, S. M. Bertozzi, and M. A. Miller (2009) Adaptive vaccination strategies to mitigate pandemic influenza: Mexico as a case study, *PLoS One* **4**: e8164.
45. Clark, C.W. (1973) Profit maximization and the extinction of animal species, *J. Political Economy* **81**: 950–961.
46. Clark, C.W. (1976) A delayed-recruitment model of population dynamics, with an application to baleen whale populations, *J. Math. Biol.* **3**: 381–391.
47. Clark, C.W. (1979) Mathematical models in the economics of renewable resources, *SIAM Review* **21**: 81–99.
48. Cosner, C., J. C. Beier, R. Cantrell, D. Impoinvil, L. Kapitanski, M. Potts, A. Troyo, and S. Ruan (2009) The effects of human movement on the persistence of vector-borne diseases, *J. Theor. Biol.* **258**: 550–560.
49. Cristiani, E., B. Piccoli, and A. Tosin (2014) Multiscale Modeling of Pedestrian Dynamics, volume 12. Springer.
50. Daszak, P., R. Plowright, J. Epstein, J. Pulliam, S. Abdul Rahman, H. Field, C. Smith, K. Olival, S. Luby, K. Halpin, et al (2006) The emergence of Nipah and Hendra virus: pathogen dynamics across a wildlife-livestock-human continuum, Oxford University Press: Oxford, UK.
51. Daszak, P., G. M. Tabor, A. Kilpatrick, J. Epstein, and R. Plowright (2004) Conservation medicine and a new agenda for emerging diseases. *Ann. New York Academy of Sciences* **1026**: 1–11.
52. Del Valle,S., H. Hethcote, J. M. Hyman, and C. Castillo-Chavez (2005) Effects of behavioral changes in a smallpox attack model. *Math. Biosc.* **195**: 228–251.
53. Diekmann, O., H. Heesterbeek, and T. Britton (2012) Mathematical Tools for Understanding Infectious Disease Dynamics. Princeton University Press.

54. Diekmann, O. and J. A. P. Heesterbeek (2000) Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation, Volume 5. John Wiley & Sons.
55. Domingo, E. and J. Holland (1997) RNA virus mutations and fitness for survival, *Annual Reviews in Microbiology* **51**: 151–178.
56. Epstein, J.M., J. Parker, D. Cummings, and R. A. Hammond (2008) Coupled contagion dynamics of fear and disease: mathematical and computational explorations, *PLoS One* **3**: e3955.
57. Erdem, M., M. Safan, and C. Castillo-Chavez (2017) Mathematical analysis of an SIQR influenza model with imperfect quarantine, *Bull. Math. Biol.* **79**: 1612–1636.
58. Espinoza, B., V. Moreno, D. Bichara, and C. Castillo-Chavez (2016) Assessing the efficiency of movement restriction as a control strategy of Ebola, In *Mathematical and Statistical Modeling for Emerging and Re-emerging Infectious Diseases*, pages 123–145. Springer.
59. Feng, Z. (2014) *Applications of Epidemiological Models to Public Health Policymaking: the Role of Heterogeneity in Model Predictions*. World Scientific.
60. Feng, Z., C. Castillo-Chavez, and A. F. Capurro (2000) A model for tuberculosis with exogenous reinfection, *Theor. Pop. Biol.* **57**: 235–247.
61. Fenichel, E.P., C. Castillo-Chavez, M. G. Coddia, G. Chowell, P. A. G. Parra, G. J. Hickling, G. Holloway, R. Horan, B. Morin, C. Perrings, et al. (2011) Adaptive human behavior in epidemiological models, *Proc. Nat. Acad. Sci.* **108**: 6306–6311.
62. Fenichel, E.P., N. V. Kuminoff, and G. Chowell(2013) Skip the trip: Air travelers' behavioral responses to pandemic influenza, *PLoS One* **8**: e58249.
63. Fenichel, E.P. and X. Wang (2013) The mechanism and phenomena of adaptive human behavior during an epidemic and the role of information, In *Modeling the Interplay Between Human Behavior and the Spread of Infectious Diseases*, pages 153–168. Springer, 2013.
64. Ferguson, N. (2007) Capturing human behaviour. *Nature* **446**:733.
65. Fraser, C., S. Riley, R.M. Anderson, & N.M. Ferguson (2004) Factors that make an infectious disease outbreak controllable, *Proc. Nat. Acad. Sci.* **101**: 6146–6151.
66. Funk, S., E. Gilad, and V. Jansen (2010) Endemic disease, awareness, and local behavioural response, *J. Theor. Biol.* **264**: 501–509.
67. Funk, S., E. Gilad, C. Watkins, and V. A. Jansen (2009) The spread of awareness and its impact on epidemic outbreaks, *Proc. Nat. Acad. Sci.* **106**: 6872–6877.
68. Funk, S., M. Salathé, and V. A. Jansen (2010) Modelling the influence of human behaviour on the spread of infectious diseases: a review, *J. Roy. Soc. Interface* page rsif20100142.
69. Geffard, P.Y. and T. Philipson (1996) Rational epidemics and their public control, *International economic review*: 603–624.
70. Gersovitz, M. (2011) The economics of infection control, *Ann. Rev. Resour. Econ.* **3**: 277–296.
71. Gladwell, M. (1996) *The Tipping Point*. New Yorker, June 3, 1996.
72. Gladwell, M. (2002) *The Tipping Point: How Little Things Can Make a Big Difference*. Back Bay Books/LittleLittle, Brown and Company, Time Warner Book Group.
73. Grünbaum, D. (1994) Translating stochastic density-dependent individual behavior with sensory constraints to an Eulerian model of animal swarming, *J. Math. Biol.* **33**: 139–161.
74. Gumel, A.B., C. Castillo-Chávez, R. E. Mickens, and D. P. Clemence (2006) Mathematical Studies on Human Disease Dynamics: Emerging Paradigms and Challenges: AMS-IMS-SIAM Joint Summer Research Conference on Modeling the Dynamics of Human Diseases: Emerging Paradigms and Challenges, July 17–21, 2005, Snowbird, Utah, volume 410. American Mathematical Soc., 2006.
75. Hadeler, K.P. and C. Castillo-Chávez (1995) A core group model for disease transmission, *Math. Biosc.* **128**: 41–55.
76. Hadeler, K.P. and J. Müller (2017) *Cellular Automata: Analysis and Applications*. Springer, 2017.
77. Hansen, E. and T. Day (2011) Optimal antiviral treatment strategies and the effects of resistance, *Proc. Roy. Soc. London B: Biological Sciences* **278**: 1082–1089.

78. Helbing, D., I. Farkas, and T. Vicsek (2000) Simulating dynamical features of escape panic, *Nature* **407**: 487.
79. Helbing, D., J. Keltsch, and P. Molnar (1997) Modelling the evolution of human trail systems, *Nature* **388**: 47.
80. Herrera-Valdez, M.A., M. Cruz-Aponte, and C. Castillo-Chavez (2011) Multiple outbreaks for the same pandemic: Local transportation and social distancing explain the different “waves” of A-H1N1pdm cases observed in México during 2009, *Math. Biosc. Eng.* **8**: 21–48.
81. Hethcote, H.W. (2000) The mathematics of infectious diseases, *SIAM Review* **42**: 599–653.
82. Hethcote, H.W. and J. W. Van Ark (2013) Modeling HIV Transmission and AIDS in the United States, volume 95. Springer Science & Business Media.
83. Hethcote, H.W., L. Yi, and J. Zhujun (1999) Hopf bifurcation in models for pertussis epidemiology, *Mathematical and Computer Modelling* **30**: 29–45.
84. Honório, N.W., W. d. C. Silva, P. J. Leite, J. M. Gonçalves, L. P. Lourenço-de Oliveira (2003) Dispersal of Aedes aegypti and Aedes albopictus (diptera: Culicidae) in an urban endemic dengue area in the state of Rio de Janeiro, Brazil, *Memórias do Instituto Oswaldo Cruz* **98**: 191–198.
85. Hsieh, Y.-H., J. Liu, Y.-H. Tzeng, and J. Wu (2014) Impact of visitors and hospital staff on nosocomial transmission and spread to community, *J. Theor. Biol.* **356**: 20–29.
86. Hutson, V. (1984) A theorem on average Liapunov functions, *Monatshefte für Mathematik*, **98**: 267–275.
87. Iggidr, A., G. Sallet, and M. O. Souza (2016) On the dynamics of a class of multi-group models for vector-borne diseases, *J. Math. Anal. Appl.* **441**: 723–743.
88. Kareiva, P. (1990) Population dynamics in spatially complex environments: theory and data, *Phil. Trans. Roy. Soc. Lond. B* **330**: 175–190.
89. Karesh, W.B., R. A. Cook, E. L. Bennett, and J. Newcomb (2005) Wildlife trade and global disease emergence. *Emerging Inf. Diseases* **11**: 1000.
90. Kermack, W.O. & A.G. McKendrick (1927) A contribution to the mathematical theory of epidemics, *Proc. Royal Soc. London*, **115**: 700–721.
91. Kermack, W.O. & A.G. McKendrick (1932) Contributions to the mathematical theory of epidemics, part. II, *Proc. Roy. Soc. London*, **138**: 55–83.
92. Kermack, W.O. & A.G. McKendrick (1933) Contributions to the mathematical theory of epidemics, part. III, *Proc. Roy. Soc. London*, **141**: 94–112.
93. Khan, K., J. Arino, W. Hu, P. Raposo, J. Sears, F. Calderon, C. Heidebrecht, M. Macdonald, J. Liauw, A. Chan, et al. (2009) Spread of a novel influenza A (H1N1) virus via global airline transportation, *New England J. Medicine* **361**: 212–214.
94. Kilpatrick, A.M., A. A. Chmura, D. W. Gibbons, R. C. Fleischer, P. P. Marra, and P. Daszak (2006) Predicting the global spread of H5N1 avian influenza, *Proc. Nat. Acad. Sci* **103**: 19368–19373.
95. Kimball A.N. (2016) Risky Trade: Infectious Disease in the Era of Global Trade, Routledge.
96. Kirschner, D. (1999) Dynamics of co-infection with tuberculosis and HIV-1, *Theor. Pop. Biol.* **55**: 94–109.
97. Klein, E., R. Laxminarayan, D. L. Smith, and C. A. Gilligan (2007) Economic incentives and mathematical models of disease, *Environment and Development Economics* **12**: 707–732.
98. Koram, K., S. Bennett, J. Adiamah, and B. Greenwood (1995) Socio-economic risk factors for malaria in a peri-urban area of the Gambia, *Trans. Roy. Soc. Tropical Medicine and Hygiene* **89**: 146–150.
99. Laxminarayan, R. (2001) Bacterial resistance and optimal use of antibiotics, *Discussion Papers dp-01-23 Resources for the Future*.
100. Legrand, J., R. F. Grais, P.-Y. Boelle, A.-J. Valleron, and A. Flahault (2007) Understanding the dynamics of Ebola epidemics, *Epidemiology & Infection* **135**: 610–621.
101. Levin, S.A. (1992) The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture, *Ecology* **7**: 1943–1967.
102. Levin, S.A. (2001) Fragile Dominion: Complexity and the Commons, volume 18. Springer.

103. Levin, S.A. and R. T. Paine (1974) Disturbance, patch formation, and community structure, *Proc. Nat. Acad. Sci.* **71**: 2744–2747.
104. Levins, R. (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control, *Am. Entomologist* **15**: 237–240.
105. Levinthal, D.A. and J. G. March (1993) The myopia of learning, *Strategic Management J.* **14(S2)**: 95–112.
106. Lipsitch, M., C. T. Bergstrom, and B. R. Levin (2000) The epidemiology of antibiotic resistance in hospitals: paradoxes and prescriptions, *Proc. Nat. Acad. Sci.* **97**: 1938–1943.
107. Lipsitch, M., T. Cohen, M. Murray, and B. R. Levin (2007) Antiviral resistance and the control of pandemic influenza, *PLoS Medicine* **4**: e15.
108. Macdonald, G. (1956) Epidemiological basis of malaria control, *Bull.* **15**: 613.
109. Macdonald, G. (1956) Theory of the eradication of malaria, *Bull. WHO* **15**: 369.
110. Meloni, S., N. Perra, A. Arenas, S. Gómez, Y. Moreno, and A. Vespignani (2011) Modeling human mobility responses to the large-scale spreading of infectious diseases, *Scientific reports* **1**: 62.
111. Moghadas, S.M. (2008) Management of drug resistance in the population: influenza as a case study, *Proc. Roy. Soc. London B: Biological Sciences* **275**: 1163–1169.
112. Moghadas, S.M. (2011) Emergence of resistance in influenza with compensatory mutations, *Math. Pop. Studies* **18**: 106–121.
113. Moghadas, S.M., C. S. Bowman, G. Röst, D. N. Fisman, and J. Wu (2008) Post-exposure prophylaxis during pandemic outbreaks, *BMC Medicine* **7**: 73.
114. Moghadas, S.M., C. S. Bowman, G. Röst, and J. Wu (2008) Population-wide emergence of antiviral resistance during pandemic influenza, *PLoS One* **3**: e1839.
115. Morin, B.R., A. Kinzig, S. Levin, and C. Perrings (2017) Economic incentives in the socially optimal management of infectious disease: When $r_{\{0\}}$ is not enough, *EcoHealth*, pages 1–16, 2017.
116. Morin, B.R., E. P. Fenichel, and C. Castillo-Chavez (2013) SIR dynamics with economically driven contact rates, *Nat. Res. Modeling* **26**: 505–525.
117. Morin, B.R., C. Perrings, A. Kinzig, and S. Levin (2015) The social benefits of private infectious disease-risk mitigation, *Theor. Ecology* **8**: 467–479.
118. Morin, B.R., C. Perrings, S. Levin, and A. Kinzig (2014) Disease risk mitigation: The equivalence of two selective mixing strategies on aggregate contact patterns and resulting epidemic spread, *J. Theor. Biol.* **363**: 262–270.
119. Naresh, R. and A. Tripathi (2005) Modelling and analysis of HIV-TB co-infection in a variable size population, *Math. Modelling and Analysis* **10**: 275–286.
120. Newman, M.E. (2002) Spread of epidemic disease on networks, *Phys. Rev.* **66**: 016128.
121. Newman, M.E. (2003) The structure and function of complex networks, *SIAM Review* **45**: 167–256.
122. Nowak, M. and R. M. May (2000) *Virus Dynamics: Mathematical Principles of Immunology and Virology: Mathematical Principles of Immunology and Virology*, Oxford University Press, UK.
123. Nuno, M., C. Castillo-Chavez, Z. Feng, and M. Martcheva (2008) Mathematical models of influenza: the role of cross-immunity, quarantine and age-structure, In *Mathematical Epidemiology*, pages 349–364. Springer, 2008.
124. Nuño, M., T. A. Reichert, G. Chowell, and A. B. Gumel (2008) Protecting residential care facilities from pandemic influenza, *Proc. Nat. Acad. Sci.* **105**: 10625–10630.
125. Okubo, A. (1980) Diffusion and ecological problems:(mathematical models), *Biomathematics*.
126. Okubo, A. and S. A. Levin (2013) Diffusion and ecological problems: modern perspectives, volume 14. Springer Science & Business Media.
127. Paine, R.T. and S. A. Levin (1981) Intertidal landscapes: disturbance and the dynamics of pattern, *Ecological Monographs* **51**: 145–178.
128. Pastor-Satorras, R. and A. Vespignani (2001) Epidemic dynamics and endemic states in complex networks, *Phys. Rev. E* **63**: 066117.

129. Pavlin, B.I., L. M. Schloegel, and P. Daszak (2009) Risk of importing zoonotic diseases through wildlife trade, united states, *Emerging Infectious Diseases* **15**: 1721.
130. Penning, J.M., B. Wansink, and M. T. Meulenber (2002) A note on modeling consumer reactions to a crisis: The case of the mad cow disease, *Int. J. Research in Marketing* **19**: 91–100.
131. Perra, N., D. Balcan, B. Gonçalves, and A. Vespignani (2011) Towards a characterization of behavior-disease models, *PloS One* **6**: e23084.
132. Perrings, C., C. Castillo-Chavez, G. Chowell, P. Daszak, E. P. Fenichel, D. Finnoff, R. D. Horan, A. M. Kilpatrick, A. P. Kinzig, N. V. Kuminoff, et al (2014) Merging economics and epidemiology to improve the prediction and management of infectious disease, *EcoHealth* **11**: 464–475.
133. Porco, T.C., P. M. Small, S. M. Blower, et al. (2001) Amplification dynamics: predicting the effect of HIV on tuberculosis outbreaks, *J. AIDS-Hagerstown Md.* **28**: 437–444.
134. Preisser, E.L. and D. I. Bolnick (2008) The many faces of fear: comparing the pathways and impacts of nonconsumptive predator effects on prey populations, *PloS One* **3**: e2465.
135. Raimundo, S.M., A. B. Engel, H. M. Yang, and R. C. Bassanezi (2003) An approach to estimating the transmission coefficients for AIDS and for tuberculosis using mathematical models, *Systems Analysis Modelling Simulation* **43**: 423–442.
136. Rass, L. and J. Radcliffe (2003) *Spatial Deterministic Epidemics*, volume 102. Am. Math. Soc.
137. Reluga T.C. (2010) Game theory of social distancing in response to an epidemic, *PLoS Computational Biology* **6**: e1000793.
138. Rimmelzwaan, G., E. Berkhoff, N. Nieuwkoop, D. J. Smith, R. Fouchier, and A. Osterhaus (2005) Full restoration of viral fitness by multiple compensatory co-mutations in the nucleoprotein of influenza a virus cytotoxic t-lymphocyte escape mutants, *J. General Virology* **86**: 1801–1805.
139. Rodríguez, D.J. and L. Torres-Sorando (2001) Models of infectious diseases in spatially heterogeneous environments, *Bull. Math. Biol.* **63**: 547–571.
140. Roeger, L.I.W., Z. Feng, C. Castillo-Chavez, et al. (2009) Modeling TB and HIV coinfections, *Math. Biosc. Eng.* **6**: 815–837.
141. Ross, R. (1911) *The Prevention of Malaria*, John Murray, London.
142. Ruktanonchai, N.W., D. L. Smith, and P. De Leenheer (2016) Parasite sources and sinks in a patched Ross–MacDonald malaria model with human and mosquito movement: Implications for control, *Math. Biosc.* **279**: 90–101.
143. Rvachev, L.A. and I. M. Longini Jr. (1985) A mathematical model for the global spread of influenza, *Math. Biosc.* **75**: 3–22.
144. Schulzer, M., M. Radhamani, S. Grzybowski, E. Mak, and J. M. Fitzgerald (1994) A mathematical model for the prediction of the impact of HIV infection on tuberculosis, *Int. J. Epidemiology* **23**: 400–407.
145. Smith, D.L., J. Dushoff, and F. E. McKenzie (2004) The risk of a mosquito-borne infection in a heterogeneous environment: Supplementary material, *PLoS Biology* <https://doi.org/10.1371/journal.pbio.0020368>.
146. Smith, H. (1986) Cooperative systems of differential equations with concave nonlinearities, *Nonlinear Analysis: Theory, Methods & Applications* **10**: 1037–1052.
147. Tatem, A.J. (2009) The worldwide airline network and the dispersal of exotic species: 2007–2010, *Ecography* **32**: 94–102.
148. Tatem, A.J., S. I. Hay, and D. J. Rogers (2006) Global traffic and disease vector dispersal, *Proc. Nat. Acad. Sci.* **103**: 6242–6247.
149. Tatem, A.J., D. J. Rogers, and S. Hay (2006) Global transport networks and infectious disease spread, *Advances in Parasitology* **62**: 293–343.
150. Towers, S., S. Afzal, G. Bernal, N. Bliss, S. Brown, B. Espinoza, J. Jackson, J. Judson-Garcia, M. Khan, M. Lin, et al. (2015) Mass media and the contagion of fear: the case of Ebola in America. *PloS One* **10**: e0129179.

151. Vivas-Barber, A.L., C. Castillo-Chavez, and E. Barany (2014) Dynamics of an SAIQR influenza model. *Biomath* **3**: 1–13.
152. Webb, G.F., M. J. Blaser, H. Zhu, S. Ardal, and J. Wu (2004) Critical role of nosocomial transmission in the Toronto SARS outbreak, *Math. Biosc. Eng.* **1**: 1–13.
153. Wesolowski, A., N. Eagle, A. J. Tatem, D. L. Smith, A. M. Noor, R. W. Snow, and C. O. Buckee (2012) Quantifying the impact of human mobility on malaria, *Science* **338**: 267–270.
154. West, R.W. and J. R. Thompson (1997) Modeling the impact of HIV on the spread of tuberculosis in the United States, *Math. Biosc.* **143**: 35–60.
155. Wilson, E.O. (1973) Group selection and its significance for ecology, *Bioscience* **23**: 631–638.
156. Y. Xiao, Y., F. Brauer, and S. M. Moghadas (2016) Can treatment increase the epidemic size?, *J. Math. Biol.* **72**: 343–361.
157. Yorke, J.A., H. W. Hethcote, and A. Nold (1978) Dynamics and control of the transmission of gonorrhea, *Sexually Transmitted Diseases* **5**: 51–56.

Epilogue

This book, inspired by our textbook *Mathematical Models in Population Biology and Epidemiology* [4], attempts to draw attention to the fields of computational, mathematical, and theoretical epidemiology (CMTE) with the goal of reaching researchers and practitioners whose interests and questions overlap with those posed by individuals or groups working in the fields of epidemiology, ecology, population biology, public health, and applied mathematics. Different variants of contagion models are used to study the transmission dynamics of communicable and vector-borne diseases. Contagion models are now being used extensively in the study of the dynamics of socio-economic-epidemiological processes [3, 11, 16].

One of the objectives of this book is to highlight the commonality of modeling approaches within the fields of epidemiology and the population dynamics of host-pathogen or host-parasite systems and consequently the intimate mathematical connections between ecology, epidemiology, and population biology. The mathematical focus is primarily on the applications of CMTE to the study of disease dynamics in large populations and, consequently, the book is dominated by the use of deterministic models. We tacitly recognize the challenges posed by the socio-behavioral dynamics that underline the world where disease appears, expands, persists, and evolves. Researchers and practitioners willing to immerse themselves in the topics addressed will value the importance and relevance of building research programs at the interface of CMTE and public health, guided by adopting a holistic perspective. The challenges faced by a world shaped and re-shaped by mobility, trade, changing demographics, war, violence, and health and economic disparities highlight the importance of training researchers at the intersection of CMTE and public health. The contents of this volume have been shaped by the understanding gained from the study of models for specific diseases and their analyses as well as by the challenges of connecting and simplifying processes that impact disease dynamics across scales and levels of organization, as is the case, for example, of immunology and epidemiology.

Our systems level thinking was instigated by the 1992 NSF workshop *Mathematics and Biology: the Interface Challenges and Opportunities* [13]. The workshop and the meeting’s report focused on the interface between mathematics and biology, which “...has long been a rich area of research, with mutual benefit to each supporting discipline.” Classical areas of investigation “such as population genetics, ecology, neurobiology, and 3-D reconstructions, have flourished [over] the past twenty years.” Interactions between mathematicians and biologists have changed dramatically, reaching out to encompass areas of both biology and mathematics that previously had not benefited from the closer integration of theory and experiment due to the increased reliance on high-speed computation. The report documents the participants’ enthusiastic conclusion that mathematical and computational approaches are essential to the future of biology and that biological applications will continue to contribute to the vitality of mathematics, as they have since the days of Vito Volterra. The ideas and vision expressed in this nearly three-decades old report have been expanded, reformulated, and enriched to the point that three decades later it has become routine to hold specialized workshops in mathematical epidemiology, mathematical immunology, mathematical physiology, or mathematical ecology, to name but a few.

In 2009, most of the NSF directorates supported the workshop *Towards a Science of Sustainability*, an effort to bring to the forefront of NSF the importance of convergence research. The report *Toward a Science of Sustainability* [9] states that “Building a science of sustainability... requires a truly multi-disciplinary approach that integrates practical experience with knowledge and know-how drawn from across the natural and social sciences, medicine and engineering, and mathematics and computation.” This National Science Foundation (NSF) report, carried out with the full support of most Directorates at the Foundation, places in context the importance of the challenges and opportunities posed by sustainability science. The importance of sustainability thinking in epidemiology and public health has been central to their goal of reducing or ameliorating the impact of disease from the individual to the population level and beyond. The need to manage antibiotic resistance, or to formulate and implement sustainable health solutions, or the effective management of vector-borne diseases, or working on identifying ways of assessing the role of travel and trade and their impact on disease dynamics, or the challenges posed by the ecology of infectious disease, or the efforts to reduce the prevalence or stop the expansion of health disparities would all benefit from what is being learned within the science of sustainability [14].

We believe that the chapters of this book provide a good start for those interested not only on epidemiological applications but also on processes that can be modeled as contagion processes as the surge in applications in the context of the behavioral, economic, and social sciences demonstrate [3, 11].

The possible deliberate release of biological agents and their consequences has moved to the forefront of the concerns of the Global Commons. Its potential impact is being addressed in part by the models and methods developed over the past century within the field of computational and mathematical epidemiology.

Specific models have been used to address and assess the role of disease and interventions on human or animal populations or agricultural systems. Studies on the potential impact of the deliberate release of biological agents have instigated the development of mathematical frameworks capable of addressing the impact of foot and mouth disease, or the role of mobility and trade, and particularly situations when a released agent is aimed at unsuspecting populations.

In the Report generated at a 2002 DIMACS working group meeting, *Mathematical Sciences Methods for the Study of Deliberate Releases of Biological Agents and their Consequences* [5] these issues were addressed soon after the events of September 11 of 2001.

The Centers for Disease Control (CDC) “...in the early 1950s established and developed intelligence epidemiological services. This decision, driven in part by national concerns about the potential use of biological agents as a source of terror, was one of the first systematic responses to bioterrorism ... The impact of deliberate releases of biological agents (Foot and Mouth, Mediterranean Fruit Flies, Citrus rust, etc.) on agricultural systems and/or our food supply needs to be addressed and evaluated. For example, foot and mouth disease was most likely accidentally introduced in Britain nearly simultaneously at multiple sites via the cattle food supply and agricultural personnel movement. Hence, it was difficult to contain this outbreak despite Britain’s effective post-detection response (stamping-out). The costs associated with its containment have been estimated to be over 15 billion dollars. The use of agents like anthrax highlights the need to look at existing models for the dispersion of pathogens in buildings (models of air-flow in buildings) and in water systems (e.g. dispersion while flowing through pipes) ... new paradigms are needed for the study of releases of these agents in rather unconventional ways.”

The challenges posed by the Global Commons have reached critical stages as travel, trade, immigration, population growth, conflicts, disease and changing economic and demographic factors continuously exhibit our inability to reduce the levels of uncertainty when assessing the impact of disease within the socio-epi sphere that we live in. The opportunities given to modelers, epidemiologists, and public health experts to improve health outcomes using mathematical modeling have re-energized the intersecting fields of computational, mathematical, and theoretical epidemiology or CMTE just as the world’s population surpasses the seven billion mark. Further, the fields of data science, demography, social dynamics, public health policy, behavioral sciences, modeling, computer science and applied mathematics continue to enhance our understanding of problems that can be better understood and managed from what can be learned from the study of specific questions that can be often addressed through synergistic approaches involving the mathematical, health, and behavioral sciences. Efforts to advance relevant theoretical work while searching for solutions to pressing epidemiological questions is becoming the new standard, some would say the norm. Hence, grounding ourselves within the context of human managed or operated systems is not only critical but also highlights the importance of the following question: How can decision-makers build and implement robust policies anchored on the best information and knowledge available given that social landscapes evolve and change as a function of the

decisions that individuals make every single moment in response to real or perceived information?

The challenges of the twenty-first century have brought the role of individual and group decisions to the forefront of most efforts aimed at ameliorating the impact and consequences of emergent, re-emergent, and endemic diseases. The technical advances of the twenty-first century have put the power of the quantitative sciences in the hands of policy and decision makers. Progress depends on the researchers' ability to address disease dynamics, evolution and epidemiological management challenges tied in to current and changing understanding of disease processes across various levels of organization within systems often operating simultaneously over fast, intermediate, and slow time scales.

The need to model and quantify uncertainty within multi-scale, multi-state human-disease systems increases the role that the impact of human behavior and decisions play in altering the social and biological landscapes on which diseases operate. Disentangling the complications of disease dynamics, control, and individuals' decisions demand the use of a holistic perspective since, in fact, most often we are dealing with complex adaptive systems [12].

Simon Levin put it best when he observed (in his address as the 2004 recipient of the Heineken award), "A great challenge before us is thus to understand the dynamics of social norms, how they arise, how they spread, how they are sustained and how they change. Models of these dynamics have many of the same features as models of epidemic spread, no great surprise, since many aspects of culture have the characteristics of being social diseases. The 1998 Heineken award winner Paul Ehrlich and I have been directing our collective energies to this problem, convinced that it is as important to understand the dynamics of the social systems in which we live as it is to understand the ecological systems themselves. Understanding the links between individual behavior and societal consequences, and characterizing the networks of interaction and influence, create the potential to change the reward structures so that the social costs of individual actions are brought down to the level of individual payoffs. It is a daunting task, both because of the amount we still must learn, and because of the ethical dilemmas that are implicit in any form of social engineering. But it is a task from which we cannot shrink, lest we squander the last of our diminishing resources."

The challenges posed by the explosive growth of antibiotic resistance are immense. The head of the WHO while addressing a meeting of infectious disease experts in Copenhagen highlighted recently the global crisis in antibiotics, the result of "rapidly evolving resistance among microbes responsible for common infections that threaten to turn them into untreatable diseases every antibiotic ever developed was at risk of becoming useless."

No new classes of antibiotics have been developed since 1987. Professor Nigel Brown, president of the Society for General Microbiology remarks that immediate action by scientists is required if we are going to identify and mass produce new antibiotics; the kind of effort needed to tackle the problem of antimicrobial resistance and its transmission, particularly in the context of nosocomial (in hospital) infections.

The challenges and opportunities facing the fields of epidemiology and public health are timely, pressing, and complex. The contributions of the fields of computational, mathematical, and theoretical epidemiology continue to expand often responding immediately epidemiological emergencies [2, 6, 7].

Bill Gates, in *Business Insider* has recently warned the world (April 27, 2018, and May 12, 2018) about the consequences of its lack of preparedness for the sure-to-be coming influenza pandemic. This book, through the study of communicable and vector-borne diseases, some re-emergent, and some emergent diseases, provides a solid training on modeling approaches and the mathematical methods that may be useful in the formulation and analyzes of population-level modeling efforts aimed at containing or ameliorating the impact of childhood diseases, sexually-transmitted diseases, HIV, tuberculosis, influenza, vector-borne diseases, and more.

We have included applications that account for the impact of host or pathogen heterogeneity on disease dynamics and control. We have made efforts to enhance the modeling tool kit by outlining and comparing the use of “Lagrangian” and “Eulerian” modeling perspectives.

The issue of time scales has been addressed in various forms throughout including via the clear differentiating that we have made of the differences that come from the study and modeling of short and long-term dynamics as well as on the consequences of including population structure [15].

Some of the limitations found in this volume are tied in to our deliberate omission of stochastic models [1] or at least some of the specifics of parameter estimation [8]. Further, we have not included the role of agent based modeling [10] or many of the advances and perspectives that network theory has brought into epidemiology.

This book captures our biased perspectives on the role of computational, mathematical, and theoretical epidemiology in the study of disease dynamics and public health. However, it does not provide a comprehensive view of CMTE, just the basics. We believe nevertheless that its content would be useful and relevant to those interested in addressing the challenges posed by disease dynamics and control.

References

1. Allen, L. (2010) An Introduction to Stochastic Processes with Applications to Biology, CRC Press.
2. Arino, J., F. Brauer, P. van den Driessche, J. Watmough, J. Wu (2006) Simple models for containment of a pandemic, *J. Roy. Soc. Interface*, <https://doi.org/10.1098/rsif.2006.0112>.
3. Bettencourt, L.M., A. Cintrón-Arias, D. J. Kaiser, and C. Castillo-Chavez (2006) The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models, *Physica A: Statistical Mechanics and its Applications*. **364**: 513-536.
4. Brauer F. and C. Castillo-Chavez (2012) Mathematical models in population biology and epidemiology. No. 40 in Texts in applied mathematics, New York: Springer, 2nd ed. 2012.
5. Castillo-Chavez, C. and F. Roberts (2002) Report on DIMACS working group meeting: Mathematical sciences methods for the study of deliberate releases of biological agents and their consequences. Biometrics Unit Technical Report (BU-1600-M) Cornell University

6. Chowell, D., C. Castillo-Chavez, S. Krishna, X. Qiu, and K.S. Anderson (2015) Modelling the effect of early detection of Ebola. *The Lancet Infectious Diseases*, **15**: 148–149.
7. Chowell, G., P.W. Fenimore, M.A. Castillo-Garsow, and C. Castillo-Chavez (2003) SARS outbreaks in Ontario, Hong Kong and Singapore: the role of diagnosis and isolation as a control mechanism. *J. Theor. Biol.* **224**: 1–8.
8. Chowell, G., J. M. Hayman, L. M. A. Bettencourt, C. Castillo-Chavez (2014) Mathematical and Statistical Estimation Approaches in Epidemiology, Springer.
9. Clark W.C., S.A. Levin (2010) Toward a science of sustainability: Executive, Levin SA, Clark WC (Eds.), Report from Toward a Science of Sustainability Conference, Airlie Center, Warrenton, Virginia.
10. Epstein, J. and S. Epstein (2007) Generative Social Science: Studies in Agent-Based Computational Modeling, Princeton University Press.
11. Fenichel, E. P., Castillo-Chavez, C., Cuddia, M. G., Chowell, G., Parra, P. A. G., Hickling, G. J., . . . & Springborn, M. (2011) Adaptive human behavior in epidemiological models. *Proc. Nat. Acad. Sci.* textbf{108}: 6306–6311.
12. Levin, S. (1999) Fragile Dominion: Complexity and the Commons. Perseus Publishing, Cambridge, Massachusetts.
13. Levin, S. (1992) Mathematics and Biology: The Interface, Challenges and Opportunities (Ed. Simon A Levin).
14. Muneepeerakul, R., & Castillo-Chavez, C. (2015) Towards a Quantitative Science of Sustainability.
15. Song, B., C. Castillo-Chavez, J. P. Aparicio (2002) Tuberculosis models with fast and slow dynamics: the role of close and casual contacts, **180**, 187–205.
16. Towers, S., A. Gomez-Lievano, M. Khan, A. Mubayi, C. Castillo-Chavez (2015) Contagion in Mass Killings and School Shootings, PLoS One, <https://doi.org/10.1371/journal.pone.0117259>.

Appendix A

Some Properties of Vectors and Matrices

A.1 Introduction

A single linear algebraic equation in one unknown has the form

$$Ax = b.$$

A system of m linear equations in n variables is a system of equations of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \end{aligned}$$

.....

$$a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n$$

The language of vectors and matrices makes it possible to write this system in the simple form $Ax = b$, where A is an $m \times n$ matrix and b is a column vector. We will develop the elementary theory of vectors and matrices to show how to use this language to simplify the analysis of linear systems. This will be useful in the study of systems of differential equations that arise in epidemic models. In models consisting of a system of two differential equations in two unknown functions the matrices and vectors involved will have $m = n = 2$, and readers should interpret the results in this chapter with $m = n = 2$. We may think of matrix algebra as machinery that will allow us to simplify the language of epidemiological models.

A.2 Vectors and Matrices

A real n -vector is an ordered n -tuple of real numbers of the form

$$v = [a_1, a_2, \dots, a_n].$$

We also write v in the form (*column vector*)

$$v = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}. \quad (\text{A.1})$$

In these notes, we use the form (A.1), since this will simplify calculations later. The real numbers will be called *scalars*.

We have the following two vector operations:

1. *Addition*, which is given by the formula

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{bmatrix};$$

2. *Scalar multiplication*, defined by

$$\lambda \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \lambda a_1 \\ \vdots \\ \lambda a_n \end{bmatrix}.$$

We define the *scalar product* (also called *inner* or *dot* product) of two vectors by

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = a_1 b_1 + \dots + a_n b_n.$$

Example 1 Consider the vectors

$$v_1 = \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}, v_2 = \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}.$$

Then

$$v_1 + v_2 = \begin{bmatrix} -1 \\ 4 \\ 0 \end{bmatrix}, \quad 3v_1 = \begin{bmatrix} 3 \\ 9 \\ -6 \end{bmatrix},$$

$$v_1 \cdot v_2 = (1)(-2) + (3)(1) + (-2)(2) = -3.$$

It is not hard to see that the operations defined above have the following properties:

$$\begin{aligned} u + v &= v + u \\ (u + v) + w &= u + (v + w) \\ \lambda(u + v) &= \lambda u + \lambda v \\ u \cdot v &= v \cdot u \\ u \cdot (v + w) &= u \cdot v + u \cdot w \\ u \cdot (\lambda v) &= \lambda u \cdot v \end{aligned}$$

An $m \times n$ matrix is an array of mn real numbers with m rows and n columns:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}. \quad (\text{A.2})$$

The matrix A is also written as $A = (a_{ij})$. The size of A is $m \times n$. Matrices of the same size can be added, and matrices can be multiplied by scalars, according to the following rules:

$$\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & \dots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \dots & b_{mn} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & \dots & a_{1n} + b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & \dots & a_{mn} + b_{mn} \end{bmatrix},$$

$$\lambda \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} \lambda a_{11} & \dots & \lambda a_{1n} \\ \vdots & \ddots & \vdots \\ \lambda a_{m1} & \dots & \lambda a_{mn} \end{bmatrix}.$$

Example 2 Consider the matrices

$$A = \begin{bmatrix} 2 & -1 & 3 \\ -2 & 5 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & -1 & 3 \\ 4 & 4 & -2 \end{bmatrix}.$$

Then we have

$$A + B = \begin{bmatrix} 2 & -2 & 6 \\ 2 & 9 & -2 \end{bmatrix}, 5A = \begin{bmatrix} 10 & -5 & 15 \\ -10 & 25 & 0 \end{bmatrix}, -B = \begin{bmatrix} 0 & 1 & -3 \\ -4 & -4 & 2 \end{bmatrix}.$$

The matrix operations satisfy the following properties:

$$\begin{aligned} A + B &= B + A \\ A + (B + C) &= (A + B) + C \\ \lambda(A + B) &= \lambda A + \lambda B. \end{aligned}$$

If $A = (a_{ij})$ is an $m \times n$ matrix, and $B = (b_{jk})$ is an $n \times p$ matrix, then the product AB is defined as the $m \times p$ matrix given by $C = (c_{ik})$ where

$$c_{ik} = \sum_{j=1}^n a_{ij}b_{jk}.$$

That is, the (i, k) -element of C is the scalar product of the i th row of A and the k th column of B . Note that the product of two matrices A and B is defined only when the number of columns of A is equal to the number of rows of B .

Example 3 Let

$$A = \begin{bmatrix} 2 & -1 & 3 \\ -2 & 5 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & -1 & 4 \\ 0 & 1 & -1 \\ -2 & 5 & 0 \end{bmatrix},$$

then

$$AB = \begin{bmatrix} -2 & 12 & 9 \\ -4 & 7 & -13 \end{bmatrix}.$$

We can check that the matrix product satisfies the following properties:

$$\begin{aligned} A(BC) &= (AB)C \\ A(B + C) &= AB + AC \\ A(\lambda B) &= \lambda AB. \end{aligned}$$

It is important to note that the matrix product **is not commutative**, that is, in general $AB \neq BA$. Moreover, if the product AB is defined, in general BA is not defined.

A matrix is called a *square matrix* if the number of its rows is equal to the number of its columns. The *main diagonal* of an $n \times n$ square matrix $A = (a_{ij})$ is the n -tuple $(a_{11}, a_{22}, \dots, a_{nn})$. A square matrix D is called a *diagonal matrix* if all its elements are zero with the exception of its diagonal:

$$D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}.$$

We also write D as $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$.

Example 4 Let

$$A = \begin{bmatrix} 2 & 1 & 4 \\ 0 & 1 & -1 \\ -2 & 5 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 5 \end{bmatrix} = \text{diag}(2, -1, 5).$$

Then

$$AD = \begin{bmatrix} 4 & 1 & 20 \\ 0 & -1 & -5 \\ -4 & -5 & 0 \end{bmatrix}, \quad DA = \begin{bmatrix} 4 & -2 & 8 \\ 0 & -1 & 1 \\ -10 & 25 & 0 \end{bmatrix}.$$

The example above suggests the following fact, which is indeed true: If $A = (v_1 \ v_2 \ \dots \ v_n)$ is a square matrix whose columns are the n -vectors v_1, v_2, \dots, v_n ,

and $B = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ is a square matrix whose rows are the n -vectors u_1, u_2, \dots, u_n ,

then

$$A \ \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) = (\lambda_1 v_1 \ \lambda_2 v_2 \ \dots \ \lambda_n v_n) \quad (\text{A.3a})$$

$$\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) B = \begin{bmatrix} \lambda_1 u_1 \\ \lambda_2 u_2 \\ \vdots \\ \lambda_n u_n \end{bmatrix}. \quad (\text{A.3b})$$

The $n \times n$ *identity matrix* is the matrix $I \operatorname{diag}(1, 1, \dots, 1)$, and satisfies

$$AI = IA = A \quad (\text{A.4})$$

for all $n \times n$ matrices A .

A.3 Systems of Linear Equations

A system of m linear equations in n variables is a system of equations of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdots &\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned}$$

We can write this system in the form $Ax = b$, where

$$A = (a_{ij}), \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

The information contained in the above system is also contained in the *augmented matrix* $A|b$ with $n + 1$ rows and m columns. The solutions of the system are not changed by adding a multiple of one equation to another equation, multiplication of an equation by a non-zero constant, or interchanging two equations. This statement is equivalent to the statement that the set of solutions is not changed if the augmented matrix is subjected to a sequence of *elementary row operations*. There are three types of elementary row operations, namely

1. addition of a multiple of one row to another row,
2. multiplication of a row by a non-zero constant,
3. interchange of two rows.

The system can be solved by reducing the matrix A to a triangular matrix using elementary row operations, a method called the Gauss–Jordan method. We demonstrate the method by an example.

Example 1 Solve the system

$$\begin{aligned} x - 3y - 5z &= -8 \\ -x + 2y + 4z &= 5 \\ 2x - 5y - 11z &= -9. \end{aligned}$$

We add the first row to the second row, subtract double the first row from the third row, and multiply the new second row by -1 . Next we add 3 times the second row to the first row and subtract the second row from the third row. Finally we subtract the third row from the first row, multiply the third row by $-1/2$, and subtract the third row from the second row. These operations give the sequence of augmented matrices

$$\begin{aligned} \left(\begin{bmatrix} 1 & -3 & -5 \\ -1 & 2 & 4 \\ 2 & -5 & -11 \end{bmatrix} \middle| \begin{bmatrix} -8 \\ 5 \\ -9 \end{bmatrix} \right) &\sim \left(\begin{bmatrix} 1 & -3 & -5 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix} \middle| \begin{bmatrix} -8 \\ 3 \\ 7 \end{bmatrix} \right) \\ &\sim \left(\begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & -2 \end{bmatrix} \middle| \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} \right) \\ &\sim \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \middle| \begin{bmatrix} -3 \\ 5 \\ -2 \end{bmatrix} \right). \end{aligned}$$

Then we may read off the solution from the final form as $x = -3$, $y = 5$, $z = -2$.

A.4 The Inverse Matrix

Let A be a square $n \times n$ matrix. The inverse of A , if it exists, is a matrix A^{-1} such that

$$AA^{-1} = A^{-1}A = I.$$

Not every matrix A has an inverse. If A has an inverse, then it is called *invertible*, and its inverse is unique.

Example 1 Find the inverse of the matrix

$$A = \begin{bmatrix} 1 & -3 & -5 \\ -1 & 2 & 4 \\ 2 & -5 & -11 \end{bmatrix}.$$

Solution The inverse of A must be of the form (A.2), so we have to solve the system of 9 linear equations

$$A \cdot (x_{ij}) = I.$$

As in the previous example, we use the Gauss–Jordan method to solve this system of equations. The first and last steps of the algorithm are

$$\left(\begin{bmatrix} 1 & -3 & -5 \\ -1 & 2 & 4 \\ 2 & -5 & -11 \end{bmatrix} \middle| \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \sim \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \middle| \begin{bmatrix} -1 & -4 & -1 \\ -\frac{3}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix} \right).$$

Therefore the inverse of A is the matrix

$$A^{-1} = \begin{bmatrix} -1 & -4 & -1 \\ -\frac{3}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix}.$$

A.5 Determinants

The *determinant* of a square matrix is defined inductively by

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

$$\begin{aligned} \det \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} &= a_{11} \det \begin{bmatrix} a_{22} & \dots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n2} & \dots & a_{nn} \end{bmatrix} \\ &\quad - a_{12} \det \begin{bmatrix} a_{21} & \dots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} + \dots \\ &\quad + (-1)^{n+1} a_{1n} \det \begin{bmatrix} a_{21} & a_{22} & \dots & a_{2(n-1)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{n(n-1)} \end{bmatrix}. \end{aligned}$$

Example 1 Calculate the determinant of the matrix

$$A = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 4 & 1 \\ 0 & -2 & -1 \end{bmatrix}.$$

Solution We have

$$\det A = 2 \det \begin{bmatrix} 4 & 1 \\ -2 & -1 \end{bmatrix} - 0 + \det \begin{bmatrix} 1 & 4 \\ 0 & -2 \end{bmatrix} = 2(-4 + 2) + (-2) = -6.$$

Theorem A.1

$$\det(v_1 \dots \alpha u + \beta v \dots v_n) = \alpha \det(v_1 \dots u \dots v_n) + \beta \det(v_1 \dots v \dots v_n)$$

$$\det(v_1 \dots v_i \dots v_j \dots v_n) = -\det(v_1 \dots v_j \dots v_i \dots v_n)$$

From (A.5), we can deduce the following properties of the determinant.

$$\det(v_1 \dots 0 \dots v_n) = 0$$

$$\det(v_1 \dots u \dots u \dots v_n) = 0 \quad (\text{A.5})$$

$$\det(v_1 \dots u \dots v + \alpha u \dots v_n) = \det(v_1 \dots u \dots v \dots v_n).$$

The transpose of a matrix is the matrix whose columns are the rows of A , and whose rows are the columns of A . If A is an $m \times n$ matrix, then A^T is an $n \times m$ matrix.

One can show inductively that

$$\det A^T = \det A. \quad (\text{A.6})$$

From (A.6) we conclude that the alternating multilinear properties (A.5-A.5) also hold for the rows of a matrix.

Example 2 Calculate the determinant of the matrix

$$A = \begin{bmatrix} 1 & 0 & 4 & 3 & -1 \\ 0 & 4 & 2 & -2 & 0 \\ -2 & 1 & -1 & 3 & 2 \\ 10 & 4 & -2 & 0 & 1 \\ 4 & 6 & -1 & 0 & 3 \end{bmatrix}.$$

Solution We use properties (A.5-A.5) applied to the rows of A to calculate this determinant.

$$\det A = \det \begin{bmatrix} 1 & 0 & 4 & 3 & -1 \\ 0 & 4 & 2 & -2 & 0 \\ 0 & 1 & 7 & 9 & 0 \\ 0 & 4 & -42 & -30 & 11 \\ 0 & 6 & -17 & -12 & 7 \end{bmatrix} = -\det \begin{bmatrix} 1 & 0 & 4 & 3 & -1 \\ 0 & 1 & 7 & 9 & 0 \\ 0 & 4 & 2 & -2 & 0 \\ 0 & 4 & -42 & -30 & 11 \\ 0 & 6 & -17 & -12 & 7 \end{bmatrix}$$

$$\begin{aligned}
&= -\det \begin{bmatrix} 1 & 0 & 4 & 3 & -1 \\ 0 & 1 & 7 & 9 & 0 \\ 0 & 0 & -26 & -38 & 0 \\ 0 & 0 & -70 & -66 & 11 \\ 0 & 0 & -59 & -66 & 7 \end{bmatrix} = -\det \begin{bmatrix} -26 & -38 & 0 \\ -70 & -66 & 11 \\ -59 & -66 & 7 \end{bmatrix} \\
&= 2 \det \begin{bmatrix} 13 & 19 & 0 \\ -70 & -66 & 11 \\ -59 & -66 & 7 \end{bmatrix} = 2 \left(13 \det \begin{bmatrix} -66 & 11 \\ -66 & 7 \end{bmatrix} \right) - \left(-19 \det \begin{bmatrix} -70 & 11 \\ -59 & 7 \end{bmatrix} \right) \\
&= 2(13(-462 + 726) - 19(-490 + 649)) = 822.
\end{aligned}$$

Theorem A.2 If A and B are square matrices of the same size, then

$$\det(AB) = (\det A)(\det B). \quad (\text{A.6a})$$

If the matrix A has an inverse A^{-1} , then Eq. (A.6a) implies

$$(\det A)(\det A^{-1}) = 1. \quad (\text{A.7})$$

Thus, from Eq. (A.7), we can conclude that if A has an inverse, then $\det A \neq 0$. The converse is true and is contained in the following theorem.

Theorem A.3 Let A be an $n \times n$ matrix. Then the following are equivalent:

1. The system $Ax = b$ has a unique solution for each n -vector b .
2. The matrix A is invertible.
3. $\det A \neq 0$.

Note that if $\det A = 0$, then the system $Ax = 0$ has non-zero solutions x . We use this fact in the following section.

A.6 Eigenvalues and Eigenvectors

Let A be a square matrix. We say that $v \neq 0$ is an *eigenvector* of A if λ is a scalar such that

$$Av = \lambda v \quad (\text{A.8})$$

for some possibly complex scalar λ . The scalar λ is called an *eigenvalue* of A with respect to v . Note that each eigenvector of a matrix A corresponds to a unique eigenvalue; however, several eigenvectors may correspond to a single eigenvalue.

Example 1 Consider

$$A = \begin{bmatrix} 5 & 8 & -3 \\ 0 & -3 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad v = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Then v is an eigenvector of A with respect to the eigenvalue $\lambda = 2$:

$$Av = \begin{bmatrix} 5 & 8 & -3 \\ 0 & -3 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

We now describe an algorithm to find the eigenvalues of a matrix. First, note that if v is an eigenvector of A with respect to the eigenvalue λ , then λ and v are solutions to Eq. (A.8) with $v \neq 0$. Equation (A.8) can be written in the form

$$(A - \lambda I)v = 0, \tag{A.9}$$

where I is the identity matrix of the same size as A . By Theorem A.3, Eq. (A.9) has a non-zero solution in v if and only if

$$\det(A - \lambda I) = 0. \tag{A.10}$$

Equation (A.10) is called the characteristic equation of the matrix A . Observe that if A is an $n \times n$ matrix, then the characteristic equation (A.10) is a polynomial equation in λ of degree n .

Example 2 Calculate the eigenvalues and eigenvectors of the matrix

$$A = \begin{bmatrix} 5 & 8 & -3 \\ 0 & -3 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Solution The characteristic equation of A is given by

$$0 = \det(A - \lambda I) = \det \begin{bmatrix} 5 - \lambda & 8 & -3 \\ 0 & -3 - \lambda & 0 \\ 0 & 0 & 2 - \lambda \end{bmatrix} = (5 - \lambda)(-3 - \lambda)(2 - \lambda).$$

Thus the eigenvalues of A are $\lambda_1 = 5$, $\lambda_2 = -3$, and $\lambda_3 = 2$. To calculate the eigenvectors, we solve the Eq. (A.9) for each λ_i . Using the Gauss–Jordan method we obtain that

$$v_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad v_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

are the eigenvectors of A with respect to $\lambda_1 = 5$, $\lambda_2 = -3$, and $\lambda_3 = 2$.

Example 3 Calculate the eigenvalues and eigenvectors of the matrix

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Solution The characteristic equation of A is given by

$$\begin{aligned} 0 &= \det \begin{bmatrix} 1-\lambda & 1 & 0 \\ 2 & -\lambda & 0 \\ 0 & 0 & 3-\lambda \end{bmatrix} = (1-\lambda) \det \begin{bmatrix} -\lambda & 0 \\ 0 & 3-\lambda \end{bmatrix} - \det \begin{bmatrix} 2 & 0 \\ 0 & 3-\lambda \end{bmatrix} \\ &= (1-\lambda)(-\lambda)(3-\lambda) - 2(3-\lambda) = -(\lambda-1)(\lambda+1)(\lambda-2). \end{aligned}$$

Thus, the eigenvalues of A are then $\lambda_1 = 3$, $\lambda_2 = -1$, and $\lambda_3 = 2$. The eigenvectors of A with respect to these eigenvalues are, respectively,

$$v_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \quad v_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Example 4 Find the eigenvalues and eigenvectors of the matrix

$$A = \begin{bmatrix} 1 & 3 \\ -4 & 2 \end{bmatrix}.$$

Solution The characteristic equation of A is the equation

$$0 = \det \begin{bmatrix} 1-\lambda & 3 \\ -4 & 2-\lambda \end{bmatrix} = \lambda^2 - 3\lambda + 14.$$

Then the eigenvalues of A are

$$\lambda_1 = \frac{3}{2} + \frac{\sqrt{47}}{2}i \text{ and } \lambda_2 = \frac{3}{2} - \frac{\sqrt{47}}{2}i.$$

The eigenvectors of A corresponding to λ_1 and λ_2 are, respectively,

$$v_1 = \begin{bmatrix} 3 \\ \frac{1}{2} + \frac{\sqrt{47}}{2}i \end{bmatrix}, \quad v_2 = \begin{bmatrix} 3 \\ \frac{1}{2} - \frac{\sqrt{47}}{2}i \end{bmatrix}.$$

Let A be an $n \times n$ matrix, and suppose that the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ are all different. If v_1, v_2, \dots, v_n are the eigenvectors of A with respect to the λ_i , let

$$\begin{aligned} D &= \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n), \\ S &= (v_1 \ v_2 \ \dots \ v_n). \end{aligned}$$

The matrix D is the diagonal matrix whose diagonal elements are the eigenvalues of A , and S is the matrix whose columns are the eigenvectors of A . Hence

$$\begin{aligned} AS &= A(v_1 \ v_2 \ \dots \ v_n) = (Av_1 \ Av_2 \ \dots \ Av_n) = (\lambda_1 v_1 \ \lambda_2 v_2 \ \dots \ \lambda_n v_n) \\ &= (v_1 \ v_2 \ \dots \ v_n) \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) = SD, \end{aligned}$$

where we have used (A.3a). One can show that the matrix S has an inverse. Then, we factor A in the form

$$A = SDS^{-1}. \quad (\text{A.11})$$

The expression (A.11) is called the *diagonalization* of A . Not every matrix has a diagonalization. If the matrix A has a diagonalization, then we say that A is *diagonalizable*. We have seen, in the case where all the eigenvalues of A are distinct, that A is diagonalizable.

Example 5 Diagonalize the matrix

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Solution The eigenvalues of A are $\lambda_1 = 3, \lambda_2 = 2$, and $\lambda_3 = -1$, and the eigenvectors with respect to these eigenvalues are

$$v_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad v_3 = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}.$$

Let

$$S = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & -2 \\ 1 & 0 & 0 \end{bmatrix}, D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Then we have $A = SDS^{-1}$. Thus, the diagonalization of A is

$$\begin{bmatrix} 1 & 1 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & -2 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 2/3 & 1/3 & 0 \\ 1/3 & -1/3 & 0 \end{bmatrix}.$$

From (A.3) one can check that

$$(\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n))^k = \text{diag}(\lambda_1^k, \lambda_2^k, \dots, \lambda_n^k).$$

Thus, if $A = SDS^{-1}$,

$$\begin{aligned} A^k &= (SDS^{-1})(SDS^{-1}) \cdots (SDS^{-1}) \quad (k \text{ factors}) \\ &= SD^k S^{-1}. \end{aligned}$$

If $p(x) = a_k x^k + \dots + a_1 x + a_0$ is a polynomial and A is an $n \times n$ matrix, we define

$$p(A) = a_k A^k + \dots + a_1 A + a_0 I.$$

By the above observations, we see that if A is diagonalizable, then

$$p(A) = p(SDS^{-1}) = S \cdot p(D) \cdot S^{-1} = S \cdot \text{diag}(p(\lambda_1), p(\lambda_2), \dots, p(\lambda_n)) \cdot S^{-1}.$$

Example 6 Calculate $p(A)$ where $p(x) = x^2 - 3x + 14$, and

$$A = \begin{bmatrix} 1 & 3 \\ -4 & 2 \end{bmatrix}.$$

Solution The characteristic equation of A is $\lambda^2 - 3\lambda + 14 = 0$. Therefore

$$p(A) = S \cdot p(D) \cdot S^{-1} = S \cdot \text{diag}(p(\lambda_1), p(\lambda_2)) \cdot S^{-1} = S \cdot \text{diag}(0, 0) \cdot S^{-1} = 0.$$

The above example suggests the following fact, which is indeed true: Every diagonalizable matrix satisfies its characteristic equation. In fact, one can show that every matrix satisfies its characteristic equation.

Appendix B

First Order Ordinary Differential Equations

B.1 Exponential Growth and Decay

The rate of change of some quantity is often proportional to the amount of the quantity present. This may be true, for example, of the size of a population with enough resources that its growth is unrestricted, and depends only on an inherent per capita reproductive rate. It can also apply to a decaying population—for example, the mass of a piece of a radioactive substance. In such a case, if $y(t)$ is the quantity at time t , then $y(t)$ satisfies the *differential equation*

$$\frac{dy}{dt} = ay \quad (\text{B.1})$$

where a is a constant representing the proportional growth or decay rate with a positive if the quantity is increasing and negative if the quantity is decreasing. It is easy to verify that $y = ce^{at}$ is a solution of the differential equation (B.1) for every choice of the constant c . By this we mean that if we substitute the function $y = ce^{at}$ into the differential equation (B.1) it becomes an identity. If $y = ce^{at}$, then $y' = ace^{at} = ay$, and this is the necessary verification. Thus the differential equation (B.1) has an infinite family of solutions (one for every choice of the constant c , including $c = 0$), namely

$$y = ce^{at}. \quad (\text{B.2})$$

In order for a mathematical problem to be a plausible description of a scientific situation, the mathematical problem must have only one solution; if there were multiple solutions, we would not know which solution represents the situation. This suggests that the differential equation (B.1) by itself is not enough to specify a description of a physical situation. We must also specify the value of the function y for some initial time when we may measure the quantity y and then allow the

system to start running. For example, we might impose the additional requirement, called an *initial condition*, that

$$y(0) = y_0. \quad (\text{B.3})$$

A problem consisting of a differential equation together with an initial condition is called an *initial value problem*. We may determine the value of c for which the solution (B.2) of the differential equation (B.1) also satisfies the initial condition (B.3) by substituting $t = 0$, $y = y_0$ into the form (B.2). This gives the equation

$$y_0 = c e^0 = c$$

and thus $c = y_0$. We now use this value of c to give the solution of the differential equation (B.1) which also satisfies the initial condition (B.3), namely $y = y_0 e^{at}$. In order to show that this is the only solution of the initial value problem (B.1), (B.3), we must show that every solution of the differential equation (B.1) is of the form (B.2).

To prove this, suppose that $y(t)$ is a solution of the differential equation (B.1), that is, that $y'(t) = ay(t)$ for every value of t . If $y(t) \neq 0$, division of this equation by $y(t)$ gives

$$\frac{y'(t)}{y(t)} = \frac{d}{dt} \log |y(t)| = a. \quad (\text{B.4})$$

Integration of both sides of (B.4) gives $\log |y(t)| = at + k$ for some constant of integration k . Then

$$|y(t)| = e^{at+k} = e^k e^{at}.$$

Because e^{at} and e^k are positive for every value of t , $|y(t)|$ cannot be zero, and thus $y(t)$ cannot change sign. We may remove the absolute value and conclude that $y(t)$ is a constant multiple of e^{at} , $y = ce^{at}$. We note also that if $y(t)$ is different from zero for one value of t then $y(t)$ is different from zero for every value of t . Thus the division by $y(t)$ at the beginning of the proof is legitimate unless the solution $y(t)$ is identically zero. The identically zero function is a solution of the differential equation, as is easily verified by substitution, and it is contained in the family of solutions $y = ce^{at}$ with $c = 0$.

The absolute value which appears in the integration produces some complications which may be avoided if we know that the solution must be non-negative, so that $|y(t)| = y(t)$, as is the case in many applications. If we know that a solution $y(t)$ of the differential equation $y' = ay$ is positive for all t , we could replace (B.4) by

$$\frac{d}{dt} \log y(t) = a$$

and then integrate to obtain $\log y(t) = at + k$, $y(t) = e^{at+k} = e^{at} e^k = ce^{at}$.

The logical argument in the above proof is that if the differential equation has a solution, then that solution must have a certain form. However, it also derives the form and thus serves as a method of determining the solution.

We now have a family of solutions of the differential equation (B.1). In order to determine the member of this family which satisfies a given initial condition, that is, in order to determine the value of the constant c , we merely substitute the initial condition into the family of solutions. This procedure may be followed in any situation which is described by an initial value problem, including population growth models and radioactive decay.

Example 1 Suppose that a given population of protozoa develops according to a simple growth law with a growth rate of 0.6 per member per day, that there are no deaths, and that on day zero the population consists of two members. Find the population size after 10 days.

Solution The population size satisfies the differential equation (B.1) with $a = 0.6$, and is therefore given by $y(t) = ce^{0.6t}$. Since $y(0) = 2$, we substitute $t = 0$, $y = 2$, and we obtain $2 = c$. Thus the solution which satisfies the initial condition is $y(t) = 2e^{0.6t}$, and the population size after 10 days is $y(10) = 2e^{(0.6)(10)} = 403$ (with population size rounded off to the nearest integer). \square

If we know that a population grows exponentially according to an exponential growth law but do not know the rate of growth we view the solution $y = ce^{at}$ as containing two parameters which must be determined. This requires knowledge of the population size at two different times to provide two equations which may be solved for these two parameters.

Example 2 Suppose that a population which follows an exponential growth law has 50 members at a starting time and 100 members at the end of 10 days. Find the population at the end of 20 days.

Solution The population size at time t satisfies $y(t) = ce^{at}$ and $y(0) = 50$, $y(10) = 100$. Thus $y(0) = 50 = ce^0$, $y(10) = ce^{10a} = 100$. It follows that $c = 50$ and $100 = 50e^{10a}$. We obtain $e^{10a} = 2$, $a = \frac{\log 2}{10} = 0.0693$. Finally, we obtain

$$y(20) = 50e^{(20)(\log 2)/10} = 50e^{2\log 2} = 50 \cdot 2^2 = 200. \quad \square$$

B.2 Radioactive Decay

Radioactive materials decay because a fraction of their atoms decompose into other substances. If $y(t)$ represents the mass of a sample of a radioactive substance at time t , and a fraction k of its atoms decompose in unit time, then $y(t + h) - y(t)$ is approximately $-ky(t)$ and we are led to the differential equation (B.1) with a replaced by $-k$. If it is clear from the nature of the problem that the constant of

proportionality must be negative, we will use $-k$ for the constant of proportionality, giving a differential equation

$$y' = -ky \quad (\text{B.5})$$

with $k > 0$.

Example 3 The radioactive element strontium 90 has a decay constant 2.48×10^{-2} years $^{-1}$. How long will it take for a quantity of strontium 90 to decrease to half of its original mass?

Solution The mass $y(t)$ of strontium 90 at time t satisfies the differential equation (B.7) with $k = 2.48 \times 10^{-2}$. If we denote the mass at time $t = 0$ by y_0 , then $y(t) = y_0 e^{-(2.48 \times 10^{-2})t}$. The value of t for which $y(t) = y_0/2$ is the solution of

$$\frac{y_0}{2} = y_0 e^{-(2.48 \times 10^{-2})t}.$$

If we divide both sides of this equation by y_0 and then take natural logarithms, we have

$$-(2.48 \times 10^{-2})t = \log \frac{1}{2} = -\log 2$$

so that $t = (\log 2) / (2.48 \times 10^{-2}) = 27.9$ years. \square

The time required for the mass of a radioactive substance to decrease to half of its starting value is called the *half-life* of the substance. The half-life T is related to the decay constant k by the equation

$$T = \frac{\log 2}{k}$$

because if $y(t) = y_0 e^{-kt}$ and (by definition) $y(T) = \frac{y_0}{2}$, then $e^{-kT} = \frac{1}{2}$, so that $-kT = \log \frac{1}{2} = -\log 2$. For radioactive substances it is common to give the half-life rather than the decay constant.

Example 4 Radium 226 is known to have a half-life of 1620 years. Find the length of time required for a sample of radium 226 to be reduced to one fourth of its original size.

Solution The decay constant for radium 226 is $k = \frac{\log 2}{1620} = 4.28 \times 10^{-4}$ years $^{-1}$. In terms of k , the mass of a sample at time t is $y_0 e^{-kt}$ if the starting mass is y_0 . The time τ at which the mass is $y_0/4$ is obtained by solving the equation $y_0/4 = y_0 e^{-k\tau}$ or $e^{-k\tau} = 1/4$. Taking natural logarithms we obtain $-k\tau = \log 4$, which gives

$$\tau = -\frac{\log \frac{3}{4}}{k} = \frac{1620(\log \frac{3}{4})}{\log 2} = 672 \text{ years. } \square$$

The radioactive element carbon 14 decays to ordinary carbon (carbon 12) with a decay constant 1.244×10^{-4} years $^{-1}$, and thus the half-life of carbon 14 is 5570 years. This has an important application, called carbon dating, for determining the approximate age of fossil materials. The carbon in living matter contains a small proportion of carbon 14 absorbed from the atmosphere. When a plant or animal dies, it no longer absorbs carbon 14 and the proportion of carbon 14 decreases because of radioactive decay. By comparing the proportion of carbon 14 in a fossil with the proportion assumed to have been present before death, it is possible to calculate the time since absorption of carbon 14 ceased.

Example 5 Living tissue contains approximately 6×10^{10} atoms of carbon 14 per gram of carbon. A wooden beam in an ancient Egyptian tomb from the First Dynasty contained approximately 3.33×10^{10} atoms of carbon 14 per gram of carbon. How old is the tomb?

Solution The number of atoms of carbon 14 per gram of carbon, $y(t)$, is given by $y(t) = y_0 e^{-kt}$, with $y_0 = 6 \times 10^{10}$, $k = 1.244 \times 10^{-4}$, and $y(t) = 3.33 \times 10^{10}$ for this particular t value. Thus the age of the tomb is given by the solution of the equation

$$e^{-(1.244 \times 10^{-4})t} = \frac{3.33 \times 10^{10}}{6 \times 10^{10}} = \frac{3.33}{6},$$

and if we take natural logarithms this reduces to

$$t = -\frac{\log 3.33 - \log 6}{1.244 \times 10^{-4}} = 4733 \text{ years. } \square$$

B.3 Solutions and Direction Fields

By a differential equation we will mean simply a relation between an unknown function and its derivatives. We will confine ourselves to *ordinary differential equations*, which are differential equations whose unknown function is a function of one variable so that its derivatives are ordinary derivatives. A *partial differential equation* is a differential equation whose unknown function is a function of more than one variable, so that the derivatives involved are partial derivatives. The *order* of a differential equation is the order of the highest-order derivative appearing in the differential equation. In this chapter we shall consider first-order differential equations, relations involving an unknown function $y(t)$ and its first derivative $y'(t) = \frac{dy}{dt}$. The general form of a first-order differential equation is

$$y' = \frac{dy}{dt} = f(t, y), \quad (\text{B.6})$$

with f a given function of the two variables t and y .

By a solution of the differential equation (B.6) we mean a differentiable function y of t on some t -interval I such that, for every t in the interval I ,

$$y'(t) = f(t, y(t)).$$

In other words, differentiating the function $y(t)$ results in the function $f(t, y(t))$. For example, we have seen in the preceding section that, whatever the value of the constant c , the function $y = ce^{at}$ is a solution of the differential equation $y' = ay$ on every t -interval. We see this by differentiating ce^{at} to get $a ce^{at}$, which we can rewrite as ay .

To verify whether a given function is a solution of a given differential equation, we need only substitute into the differential equation and check whether it then reduces to an identity.

Example 1 Show that the function $y = \frac{1}{t+1}$ is a solution of the differential equation $y' = -y^2$.

Solution For the given function,

$$\frac{dy}{dt} = -\frac{1}{(t+1)^2} = -y^2$$

and this shows that it is indeed a solution. \square

In the same way we can verify that a family of functions satisfies a given differential equation. By *a family of functions* we will mean a function which includes an arbitrary constant, so that each value of the constant defines a distinct function. The family ce^{5t} , for instance, includes the functions e^{5t} , $-4e^{5t}$, $12e^{5t}$, and $\sqrt{3}e^{5t}$, among others. When we say that a family of functions satisfies a differential equation, we mean that substitution of the family (i.e., the general form) into the differential equation gives an identity satisfied for every choice of the constant.

Example 2 Show that for every c the function $y = \frac{1}{t+c}$ is a solution of the differential equation $y' = -y^2$.

Solution For the given function,

$$\frac{dy}{dt} = -\frac{1}{(t+c)^2} = -y^2,$$

and thus each member of the given family of functions is a solution. \square

In applications we are usually interested in finding not a family of solutions of a differential equation but a solution which satisfies some additional requirement. In the various examples in Sect. B.1 the additional requirement was that the solution should have a specified value for a specified value of the independent variable t . Such a requirement, of the form

$$y(t_0) = y_0 \quad (\text{B.7})$$

is called an *initial condition*, and t_0 is called the *initial time* while y_0 is called the *initial value*. A problem consisting of a differential equation (B.1) together with an initial condition (B.2) is called an *initial value problem*. Geometrically, an initial condition picks out the solution from a family of solutions which passes through the point (t_0, y_0) in the t - y plane. Physically, this corresponds to measuring the state of a system at the time t_0 and using the solution of the initial value problem to predict the future behavior of the system.

Example 4 Find the solution of the differential equation $y' = -y^2$ of the form $y = \frac{1}{t+c}$ which satisfies the initial condition $y(0) = 1$.

Solution We substitute the values $t = 0$, $y = 1$ into the equation $y = \frac{1}{t+c}$, and we obtain a condition on c , namely $1 = \frac{1}{c}$, whose solution is $c = 1$. The required solution is the function in the given family with $c = 1$, namely $y = \frac{1}{t+1}$. \square

Example 5 Find the solution of the differential equation $y' = -y^2$ which satisfies the general initial condition $y(0) = y_0$, where y_0 is arbitrary.

Solution We substitute the values $t = 0$, $y = y_0$ into the equation $y = \frac{1}{t+c}$ and solve the resulting equation $y_0 = \frac{1}{c}$ for c , obtaining $c = \frac{1}{y_0}$ provided $y_0 \neq 0$. Thus the solution of the initial value problem is

$$y = \frac{1}{(t + \frac{1}{y_0})^2}$$

except if $y_0 = 0$. If $y_0 = 0$, there is no solution of the initial value problem of the given form; in this case the identically zero function, $y = 0$, is a solution. We have now obtained a solution of the initial value problem with arbitrary initial value at $y = 0$ for the differential equation $y' = -y^2$. \square

A family of solutions may arise if we are considering a differential equation with no initial condition imposed, and we will then also be concerned with the question of whether the given family contains all solutions of the differential equation. To answer this question, we will need to make use of a theorem which guarantees that each initial value problem for the given differential equation has exactly one solution. More specifically, if an initial value problem is to be a usable mathematical description of a scientific problem, it must have a solution, for otherwise it would be of no use in predicting behavior. Furthermore, it should have only one solution, for otherwise we would not know which solution describes the system. Thus for applications it is vital that there be a mathematical theory telling us that an initial value problem has exactly one solution. Fortunately, there is a very general theorem which tells us that this is true for the initial value problem (B.6), (B.3) provided

the function f is reasonably smooth. We will state this result and ask the reader to accept it without proof because the proof requires more advanced mathematical knowledge than we have at present.

Existence and Uniqueness Theorem *If the function $f(t, y)$ is differentiable with respect to y in some region of the plane which contains the point (t_0, y_0) , then the initial value problem consisting of the differential equation $y' = f(t, y)$ and the initial condition $y(t_0) = y_0$ has a unique solution which is defined on some t -interval containing t_0 in its interior.*

Even though the function $f(t, y)$ may be well-behaved in the whole t - y plane, there is no assurance that a solution will be defined for all t . As we have seen in Example 1, the solution $y = \frac{1}{t+1}$ of $y' = -y^2$, $y(0) = 1$ exists only for $-1 < t < \infty$. As we have seen in Example 2, each solution of the family of solutions $y = \frac{1}{t+c}$ has a different interval of existence. In Example 5, we have shown how to rewrite a family of solutions for a differential equation in terms of an arbitrary initial condition—that is, as a solution of an initial value problem for that differential equation. We have also seen how to identify those initial conditions which cannot be satisfied by a member of the given family. Often there are constant functions which are not members of the given family but which are solutions and satisfy initial conditions that cannot be satisfied by a member of the family. The existence and uniqueness theorem tells us that if we can find a family of solutions, possibly supplemented by some additional solutions, so that we can find this collection contains a solution corresponding to each possible initial condition, then we have found the set of all solutions of the differential equation.

A differential equation which arises in various applications, including models for population growth and spread of rumors, is the *logistic differential equation*,

$$y' = ry \left(1 - \frac{y}{K}\right) \quad (\text{B.8})$$

containing two parameters r and K . The basic idea behind this form is that instead of a constant per capita growth rate r as in the exponential growth equation it is more realistic to assume that the per capita growth rate decreases as the population size increases. The form $1 - \frac{y}{K}$ used in the logistic equation is the simplest form for a decreasing per capita growth rate. We may verify that for every constant c the function

$$y = \frac{K}{1 + ce^{-rt}} \quad (\text{B.9})$$

is a solution of this differential equation. To see this, note that for the given function y ,

$$y' = \frac{Kcr e^{-rt}}{(1 + ce^{-rt})^2}$$

and

$$1 - \frac{y}{K} = \frac{K - y}{K} = \frac{ce^{-rt}}{(1 + ce^{-rt})}.$$

Thus

$$ry \left(1 - \frac{y}{K} \right) = \frac{Krc e^{-rt}}{(1 + ce^{-rt})^2} = y',$$

and the given function satisfies the logistic equation for every choice of c .

To find the solution which obeys the initial condition $y(0) = y_0$, we substitute $t = 0$, $y = y_0$ into the form (B.9), obtaining $\frac{K}{1+c} = y_0$ which implies $c = \frac{K-y_0}{y_0}$ as long as $y_0 \neq 0$ and gives the solution

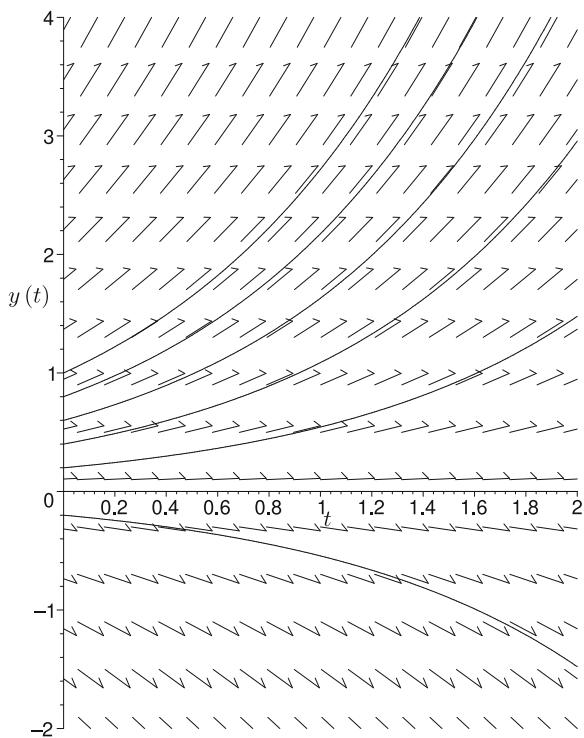
$$y = \frac{K}{1 + \left(\frac{K - y_0}{y_0} \right) e^{-rt}} = \frac{Ky_0}{y_0 + (K - y_0) e^{-rt}} \quad (\text{B.10})$$

to the initial value problem with $y_0 \neq 0$. Note that the denominator begins at K (for $t = 0$) and moves toward y_0 as $t \rightarrow \infty$. Now suppose that K represents some physical quantity such that $K > 0$. It is easy to see from the form (B.10) that if $y_0 > 0$, then the solution $y(t)$ exists for all $t > 0$, and $\lim_{t \rightarrow \infty} y(t) = K$. If $y_0 < 0$, then this solution does not exist for all $t > 0$, because $y(t) \rightarrow -\infty$ where the denominator changes sign: as $y_0 + (K - y_0)e^{-rt} \rightarrow 0$, or $t \rightarrow -\log(\frac{-y_0}{K-y_0})$. If $y_0 = 0$, the solution of the initial value problem is not given by (B.10), but is the identically zero function $y = 0$. We observe that the family of solutions (B.9) of the logistic differential equation (B.8) includes the constant solution $y = K$ (with $c = 0$) but not the constant solution $y = 0$. The existence and uniqueness theorem shows that, since we have now obtained a solution corresponding to each possible initial condition, we have obtained all solutions of the logistic differential equation.

The geometric interpretation of a solution $y(t)$ to a differential equation (B.6) is that the curve $y = y(t)$ has slope $f(t, y)$ at each point (t, y) along its length. Thus we might think of approximating the solution curve by piecing together short line segments whose slope at each point (t, y) is $f(t, y)$. To realize this idea, we construct at each point (t, y) in some region of the plane a short line segment with slope $f(t, y)$. The collection of line segments is called the *direction field* of the differential equation (B.6). The direction field can help us to visualize solutions of the differential equation since at each point on its graph a solution curve is tangent to the line segment at that point. We may sketch the solutions of a differential equation by connecting these line segments by smooth curves.

A direction field and some solutions of the differential equation $y' = y$ are shown in Fig. B.1. The direction field suggests exponential solutions, which we know from Sect. B.1 to be correct.

Fig. B.1 Direction field and solutions for $y' = y$



A direction field and some solutions of the differential equation $y' = y(1-y)$ are shown in Fig. B.2. The direction field indicates that solutions below the t -axis are unbounded below while solutions above the t -axis tend to 1 as $t \rightarrow \infty$, consistent with what we have established for this logistic differential equation.

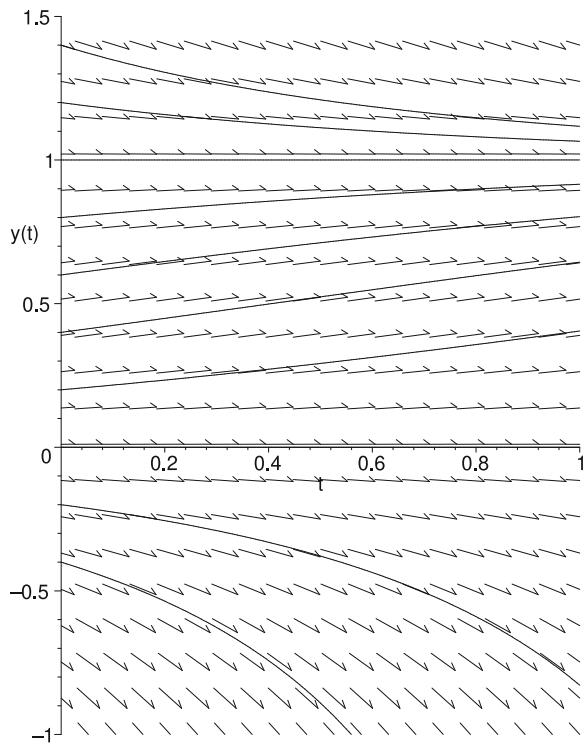
Drawing direction fields by hand is a difficult and time-consuming task. There are computer programs, both self-contained and portions of more elaborate computational systems such as Maple, Matlab, and Mathematica, which can generate direction fields for a differential equation and can also sketch solution curves corresponding to these direction fields. The examples here have been produced by Maple.

The geometric view of differential equations presented by the direction field will appear again when we examine some qualitative properties in Sect. B.5.

B.4 Equations with Variables Separable

In this section, we shall learn a method for finding solutions of a class of differential equations. The method is based on the method we used in Sect. B.1 for solving the differential equation of exponential growth or decay, and is applicable to differential equations with *variables separable*. A differential equation $y' = f(t, y)$ is called *separable* or is said to have *variables separable* if the function f can be expressed

Fig. B.2 Direction field and solutions for $y' = y(1 - y)$



in the form

$$f(t, y) = p(t)q(y). \quad (\text{B.11})$$

In particular, if f is independent of t and is a function of y only, the differential equation (B.6) is separable; such an equation is said to be *autonomous*. The reason for the name separable is that the differential equation can then be written as

$$\frac{y'}{q(y)} = p(t)$$

with all the dependence on y on the left and all the dependence on t on the right-hand side of the equation, provided that $q(y) \neq 0$. For example, $y' = \frac{y}{1+t^2}$ and $y' = y^2$ are both separable, whereas $y' = \sin t - 2ty$ is not separable. The examples discussed in Sect. B.1 are also separable and the method of solution described for Eq. (B.1) of Sect. B.1 is a special case of the general method of solution to be developed for separable equations. Before explaining this general technique, let us work out some more examples.

Example 1 Find all solutions of the differential equation

$$y' = -y^2.$$

Solution We divide the equation by y^2 , permissible if $y \neq 0$, to give

$$\frac{1}{y^2} \frac{dy}{dt} = -1.$$

We then integrate both sides of this equation with respect to t . In order to integrate the left side of the equation we must make the substitution $y = y(t)$, where $y(t)$ is the (as yet unknown) solution. This substitution gives

$$\int \frac{1}{y^2} \frac{dy}{dt} dt = \int \frac{dy}{y^2} = -\frac{1}{y} + c$$

and we obtain

$$-\frac{1}{y} = -t - c \quad (\text{B.12})$$

or $y = \frac{1}{t+c}$, with c a constant of integration. Observe that no matter what value of c is chosen this solution is never equal to zero. We began by dividing the equation by y^2 , which was legitimate provided $y \neq 0$. If $y = 0$, we cannot divide, but the constant function $y = 0$ is a solution. We now have all the solutions of the differential equation, namely the family (B.12) together with the zero solution. \square

Example 2 Solve the initial value problem

$$y' = -y^2, \quad y(0) = 1.$$

Solution We begin by using a more colloquial form of separation of variables than the procedure of Example 1. We write the differential equation in the form $\frac{dy}{dt} = -y^2$ and separate variables by dividing the equation through by y^2 to give

$$-\frac{dy}{y^2} = dt.$$

This form of the equation is meaningless without an interpretation of differentials, but the integrated form

$$-\int \frac{dy}{y^2} = \int dt$$

is meaningful. Carrying out the integration, we obtain $\frac{1}{y} = t + c$, where c is a constant of integration. Since $y = 1$ for $t = 0$ we may substitute these values into the solution and obtain $1 = c$. Substituting this value of c we now obtain

$$\frac{1}{y} = t + 1.$$

Solving for y , we obtain the solution $y = \frac{1}{t+1}$, and we may easily verify that this is a solution of the initial value problem, defined for all $t \geq 0$. \square

The version of separation of variables used in Example 2 is easier to apply in practice than the more precise version used in Example 1. When we treat the general case we will justify this approach.

Example 3 Solve the initial value problem

$$y' = -y^2, \quad y(0) = 0.$$

Solution The procedure used in Example 1 leads to the family of solutions $y = \frac{1}{t+c}$ for the differential equation $y' = -y^2$. When we substitute $t = 0$, $y = 0$ and attempt to solve for the constant c , we find that there is no solution. When we divided the differential equation by y^2 we had to assume $y \neq 0$, but the constant function $y \equiv 0$ is also a solution of the differential equation, as may easily be verified. Since this function also satisfies the initial condition, it is the solution of the given initial value problem. \square

Let us now apply the technique of Example 2 to show how to obtain the solution in the general case of a differential equation with variables separable. We will make use of the idea of the definition of the indefinite integral of a function as a new function whose derivative is the derivative of the given function. We first solve the differential equation

$$y' = p(t)q(y) \tag{B.13}$$

where p is continuous on some interval $a < t < b$ and q is continuous on some interval $c < y < d$. We solve the differential equation (B.13) by separating variables as in Example 1 dividing by $q(y)$ and integrating to give

$$\int \frac{dy}{q(y)} = \int p(t)dt + c \tag{B.14}$$

where c is a constant of integration. We now define

$$Q(y) = \int \frac{dy}{q(y)}, \quad P(t) = \int p(t)dt,$$

by which we mean that $Q(y)$ is a function of y whose derivative with respect to y is $\frac{1}{q(y)}$ and $P(t)$ is a function of t whose derivative with respect to t is $p(t)$. Then (B.14) becomes

$$Q(y) = P(t) + c \quad (\text{B.15})$$

and this equation defines a family of solutions of the differential equation (B.13). To verify that each function $y(t)$ defined implicitly by (B.15) is indeed a solution of the differential equation (B.13), we differentiate (B.15) implicitly with respect to t , obtaining

$$\frac{1}{q(y(t))} \frac{dy}{dt} = p(t)$$

on any interval on which $q(y(t)) \neq 0$, and this shows that $y(t)$ satisfies the differential equation (B.13). There may be constant solutions of (B.13) which are not included in the family (B.15). A constant solution to the original equation (B.13) corresponds to a solution of the equation $q(y) = 0$.

In order to find the one solution of the differential equation (B.13) which satisfies the initial condition $y(t_0) = y_0$, it will help to be more specific in the choice of the indefinite integrals $Q(y)$ and $P(t)$. We let $Q(y)$ be the indefinite integral of $\frac{1}{q(y)}$ such that $Q(y_0) = 0$, and we let $P(t)$ be the indefinite integral of $p(t)$ such that $P(t_0) = 0$. Then, because of the fundamental theorem of calculus, we have

$$Q(y) = \int_{y_0}^y \frac{du}{q(u)}, \quad P(t) = \int_{t_0}^t p(s) ds$$

and we may write the solution (B.15) in the form

$$\int_{y_0}^y \frac{du}{q(u)} = \int_{t_0}^t p(s) ds.$$

Now substitution of the initial conditions $t = t_0$, $y = y_0$ gives $c = 0$. Thus the solution of the initial value problem is given implicitly by

$$\int_{y_0}^y \frac{du}{q(u)} = \int_{t_0}^t p(s) ds \quad (\text{B.16})$$

We have now solved the initial value problem in the case $q(y_0) \neq 0$. Since $y(t_0) = y_0$ and the function q is continuous at y_0 , $q(y(t))$ is continuous, and therefore $q(y(t)) \neq 0$ on some interval containing t_0 (possibly smaller than the original interval $a < t < b$). On this interval, $y(t)$ is the unique solution of the initial value problem. If $q(y_0) = 0$, we have the constant function $y = y_0$ as a solution of the initial value problem in place of the solution given by (B.16).

We shall see in Sect. B.5 that the constant solutions of a separable differentiable equation (B.13) play an important role in describing the behavior of all solutions as $t \rightarrow \infty$.

The differential equation

$$y' = -ay + b \quad (\text{B.17})$$

where a and b are given constants appears in some of the examples in Sect. B.4. In order to solve it, we separate variables, obtaining

$$\int \frac{dy}{-ay + b} = \int dt.$$

Integration gives

$$-\frac{1}{a} \log(-ay + b) = t + c$$

with c an arbitrary constant of integration. Algebraic solution now gives

$$\begin{aligned} \log(-ay + b) &= -a(t + c) \\ -ay + b &= e^{-a(t+c)} = e^{-ac} e^{-at} \\ ay &= b - e^{-ac} e^{-at} \\ y &= \frac{b}{a} - \frac{e^{-ac}}{a} e^{-at}. \end{aligned}$$

We now rename the arbitrary constant $-\frac{e^{-ac}}{a}$ as a new arbitrary constant, which we again call c , and obtain the family of solutions

$$y = \frac{b}{a} + ce^{-at}. \quad (\text{B.18})$$

In our separation of variables, we overlooked the constant solution $y = \frac{b}{a}$, but this solution is contained in the family (B.18).

In order to find the solution of (B.17) which satisfies the initial condition

$$y(0) = y_0 \quad (\text{B.19})$$

we substitute $t = 0$, $y = y_0$ into (B.18), obtaining

$$y_0 = \frac{b}{a} + c$$

or $c = y_0 - \frac{b}{a}$. This value of c gives the solution

$$y = \frac{b}{a} + \left(y_0 - \frac{b}{a}\right)e^{-at} = \frac{b}{a}(1 - e^{-at}) + y_0 e^{-at} \quad (\text{B.20})$$

of the initial value problem.

The solutions (B.18) and (B.20) have been obtained without regard for the sign of the coefficients b and a . We will think of a in (B.18) and (B.20) as positive, and will rewrite the solutions for a negative by making the replacement $r = -a$, thinking of r as positive. The result of this is that the solutions of the differential equation

$$y' = ry + b \quad (\text{B.21})$$

are given by the family of functions

$$y = -\frac{b}{r} + ce^{rt} \quad (\text{B.22})$$

and the solution of the differential equation (B.21) which satisfies the initial condition (B.19) is

$$y = -\frac{b}{r}(1 - e^{rt}) + y_0 e^{rt} = \frac{b}{r}(e^{rt} - 1) + y_0 e^{rt}. \quad (\text{B.23})$$

In applications, the specific form of the solution of a differential equation is often less important than the behavior of the solution for large values of t . Because $e^{-at} \rightarrow 0$ as $t \rightarrow \infty$ when $a > 0$, we see from (B.18) that every solution of (B.17) tends to the limit $\frac{b}{a}$ as $t \rightarrow \infty$ if $a > 0$. This is an example of *qualitative* information about the behavior of solutions which will be useful in applications. In Sect. B.5, we shall examine other qualitative questions—information about the behavior of solutions of a differential equation which may be obtained indirectly rather than by an explicit solution.

To conclude this section, we return to the logistic differential equation (B.8) whose solutions were described in Sect. B.3. In Sect. B.3, we verified that the solutions were given by Eq. (B.9) but did not show how to obtain these solutions. The differential equation (B.8) is separable, and separation of variables and integration gives

$$\int \frac{K dy}{y(K - y)} = \int r dt$$

provided $y \neq 0, y \neq K$. In order to evaluate the integral on the left-hand side, we use the algebraic relation (which may be obtained by partial fractions)

$$\frac{K}{y(K - y)} = \frac{1}{y} + \frac{1}{K - y}$$

and then rewrite

$$\int \frac{K dy}{y(K - y)} = \int \frac{dy}{y} + \int \frac{dy}{K - y} = \log |y| - \log |K - y| = rt + c,$$

so

$$\log \left| \frac{y}{K-y} \right| = rt + c.$$

We can now exponentiate both sides of the equation to obtain

$$\left| \frac{y}{K-y} \right| = e^{rt+c} = e^{rt} e^c.$$

If we remove the absolute value bars from the left-hand side, we can define a new constant C on the right-hand side, equal to e^{-c} if $0 < y < K$, and equal to $-e^{-c}$ otherwise, thus rewriting the right-hand side as $\frac{1}{C} e^{rt}$. Finally, we solve for y , and obtain the family of solutions

$$y = \frac{K}{1 + Ce^{-rt}} \quad (\text{B.24})$$

as in Sect. B.3. This family contains all solutions of (B.8) except for the constant solution $y = 0$. A qualitative observation is that if $r > 0$ every positive solution approaches the limit K as $t \rightarrow \infty$. (The only other non-negative solution is the constant solution $y \equiv 0$).

The explicit solution of the logistic differential equation is complicated, requiring some knowledge of techniques of integration (separation into partial fractions) as well as some algebraic manipulations to convert the implicit function given by integration into an explicit expression. In Sect. B.5, we will see how to obtain the qualitative observation that if $r > 0$, every positive solution approaches the limit K as $t \rightarrow \infty$ without the need to go through these calculations.

B.5 Qualitative Properties of Differential Equations

We have seen some instances in which all solutions of a differential equation, or at least all solutions with initial values in some interval, tend to the same limit as $t \rightarrow \infty$. Two examples are the (separable) linear differential equation with constant coefficients

$$y' = -ay + b \quad (\text{B.25})$$

for which every solution approaches the limit $\frac{b}{a}$ as $t \rightarrow \infty$ provided $a > 0$ (Sect. B.4), and the logistic differential equation (B.8) for which every solution with positive initial value approaches the limit K as $t \rightarrow \infty$ (Sect. B.4) if $r > 0$. In applications we are often particularly interested in the long-term behavior of solutions, especially since many of the models we develop will be complex enough to make finding an explicit solution impractical. In this section, we describe some

information of this nature which may be obtained indirectly, without explicit solution of a differential equation. Properties of solutions obtained without actually finding an expression for the solutions are called *qualitative* properties.

We shall study only differential equations which do not depend explicitly on the independent variable t , of the general form

$$y' = g(y). \quad (\text{B.26})$$

Such differential equations are called autonomous. We will always assume that the function $g(y)$ is sufficiently smooth that the existence and uniqueness theorem of Sect. B.3 is valid, and there is a unique solution of the differential equation (B.26) through each initial point. An autonomous differential equation is always separable, but if the integral $\int \frac{dy}{g(y)}$ is difficult or impossible to evaluate, solution by separation of variables is impractical. In Sect. B.4, when we established the method of separation of variables, we pointed out the possibility of constant solutions which do not come from separation of variables. It will turn out that examination of these constant solutions is of central importance in the qualitative analysis. We begin with an example, the logistic equation already solved explicitly in Sect. B.3.

Example 1 Determine the behavior as $t \rightarrow \infty$ of solutions of the logistic equation (B.8) with $y(0) > 0$.

Solution We write

$$g(y) = ry \left(1 - \frac{y}{K}\right),$$

and note that $g(y) > 0$ if $0 < y < K$, $g(y) < 0$ if $y > K$. Since a solution $y(t)$ of (B.8) has the property that $y(t)$ is increasing if $g(y(t)) > 0$, a solution with $0 < y(0) < K$ increases for all t such that $y(t) < K$. However, this solution must remain less than K for all t , because if there were a value t^* such that $y(t^*) = K$, there would be two solutions with $y(t^*) = K$, namely the solution $y(t)$ and the constant solution $y = K$, and this would violate the uniqueness theorem. Thus $y(t)$ is an increasing function for all t bounded above by K , and therefore tends to a limit y_∞ as $t \rightarrow \infty$. If $y_\infty < K$, then

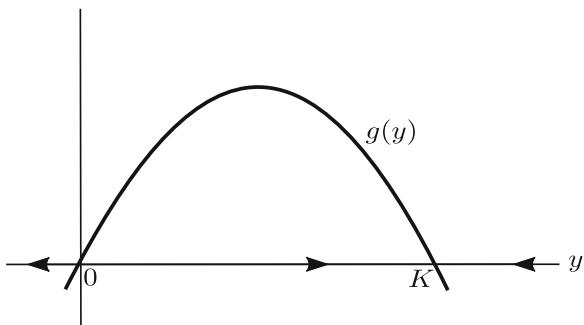
$$\lim_{t \rightarrow \infty} g(y(t)) = g(y_\infty) > 0,$$

and this is a contradiction because if $y(t)$ is a monotone increasing differentiable function that tends to a limit, its derivative must approach zero. Therefore, we have shown that

$$\lim_{t \rightarrow \infty} y(t) = K.$$

A similar argument if $y(0) > K$ shows that the solution $y(t)$ decreases monotonically to the limit K , and we have shown that every solution of (B.8) with $y(0) > 0$

Fig. B.3 Phase line for the logistic equation



approaches the limit K as $t \rightarrow \infty$ using only information about the sign of the function $g(y)$. We may describe this pictorially by drawing the *phase line*. This is the graph of the function $g(y)$ together with arrows drawn along the y -axis to the right on intervals in which the function $g(y)$ is positive (corresponding to intervals in which a solution $y(t)$ is increasing), and arrows drawn to the left on intervals on which the function $g(y)$ is negative. Points where $g(y) = 0$ correspond to constant solutions of the differential equation, and the arrows indicate which of these *equilibrium points* are approached by other solutions (the equilibrium $y = K$ for the logistic equation), and which equilibrium points repel solutions beginning near them (the equilibrium $y = 0$ for the logistic equation (Fig. B.3). \square

The information given by the phase line is reflected in the graphs of solutions of the logistic equation (Fig. B.4).

The analysis in general of the behavior as $t \rightarrow \infty$ of the solutions of the differential equation (B.26) follows a similar pattern, determined from the nature of the constant solutions. The theory depends on the following properties of autonomous differential equations. For simplicity, we shall consider only non-negative values of t , and we will think of solutions as determined by their initial values for $t = 0$.

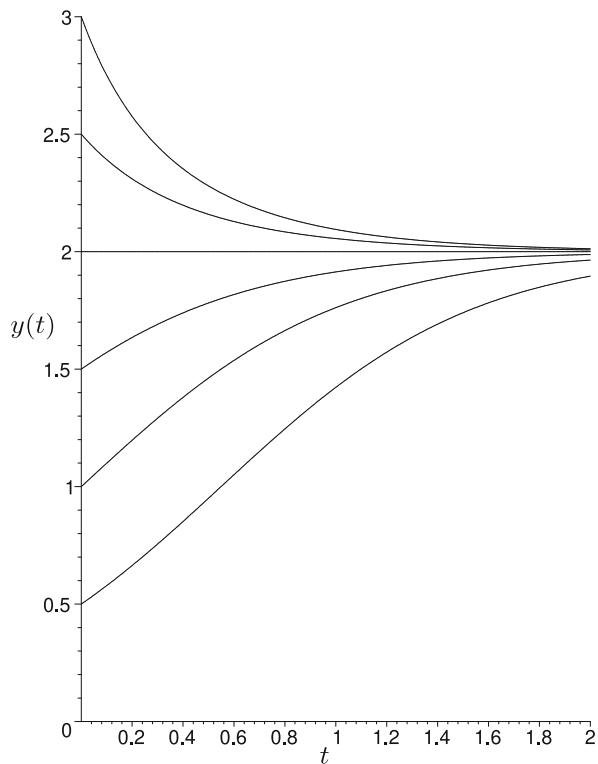
Property 1 If \hat{y} is a solution of the equation $g(y) = 0$, then the constant function $y = \hat{y}$ is a solution of the differential equation $y' = g(y)$. Conversely, if $y = \hat{y}$ is a constant solution of the differential equation $y' = g(y)$, then \hat{y} is a solution of the equation $g(y) = 0$.

To establish this property, we need only observe that because the derivative of a constant function is the zero function, a constant function $y = \hat{y}$ is a solution of (B.26) if and only if $g(\hat{y}) = 0$. A solution \hat{y} of $g(y) = 0$ is called an *equilibrium* or *critical point* of the differential equation (B.26), and the corresponding constant solution of (B.26) is called an *equilibrium solution*.

The graphs of equilibrium solutions of an autonomous differential equation (B.26) are horizontal lines which separate the $t - y$ plane into horizontal bands of the form

$$\{(t, y) \mid t \geq 0, y_1 < y < y_2\}$$

Fig. B.4 Some solutions of the logistic equation (B.8) with $r = 2$, $K = 2$



where y_1 and y_2 are consecutive equilibria of (B.26), with no equilibrium of (B.26) between y_1 and y_2 .

Property 2 The graph of every solution curve of the differential equation $y' = g(y)$ remains in the same band and is either monotone increasing [$y'(t) > 0$] or monotone decreasing [$y'(t) < 0$] for all $t \geq 0$, depending on whether $g(y) > 0$ or $g(y) < 0$, respectively, in the band.

Suppose that y_1 and y_2 are consecutive equilibria of (B.26) and that $y_1 < y(0) < y_2$. If $g(y)$ is indeed smooth enough to apply the existence and uniqueness theorem of Sect. B.3, then the graph of a solution cannot cross either of the constant solutions which form the boundaries of the band; otherwise at the crossing there would be a point in the $t - y$ plane with two solutions passing through it, violating the uniqueness. Therefore the solution must remain in this band for all $t \geq 0$. If, for example, $g(y) > 0$ for $y_1 < y < y_2$, then $y'(t) = g\{y(t)\} > 0$ for $t \geq 0$, and the solution $y(t)$ is monotone increasing. A similar argument shows that if $g(y) < 0$ for $y_1 < y < y_2$, the solution $y(t)$ is monotone decreasing.

If $y(0)$ is above the largest equilibrium of (B.26), the band containing the solution is unbounded, and if $g(y) > 0$ in this band, then the solution $y(t)$ may be (positively) unbounded and does not necessarily exist for all $t \geq 0$. Likewise, if $y(0)$ is below the

smallest equilibrium of (B.26) and if $g(y) < 0$ in the band containing the solution, then the solution may be unbounded (negatively) and may fail to exist for all $t \geq 0$.

Property 3 Every solution of the differential equation $y' = g(y)$ which remains bounded for $0 \leq t < \infty$ approaches a limit as $t \rightarrow \infty$.

This property is an immediate consequence of Property 2 and the fact from calculus that a function which is bounded and monotone (either increasing or decreasing) must approach a limit as $t \rightarrow \infty$. In order to show which solutions of a given differential equation have a limit, it is necessary to show which solutions remain bounded. A solution may become unbounded positively (i.e., $y(t) \rightarrow +\infty$) if it is monotone increasing when y is large. In many applications, $g(y) < 0$ for large y , and in such a case every solution remains bounded, because solutions which become large and positive must be decreasing. If, as is also frequently the case in applications, $y = 0$ is an equilibrium and only non-negative solutions are of interest, solutions cannot cross the line $y = 0$ and become negatively unbounded (i.e., $y(t) \rightarrow -\infty$). In many applications, y stands for a quantity such as the number of members of a population, the mass of a radioactive substance, the quantity of money in an account, or the height of a particle above ground which cannot become negative. In such a situation, only non-negative solutions are significant, and if $y = 0$ is not an equilibrium but a solution reaches the value zero for some finite t , we will consider the population system to have collapsed and the population to be zero for all larger t . If this is the case we need not be concerned with the possibility of solutions becoming negatively unbounded even if $y = 0$ is not an equilibrium.

Property 4 The only possible limits as $t \rightarrow \infty$ of solutions of the differential equation $y' = g(y)$ are the equilibria of the differential equation.

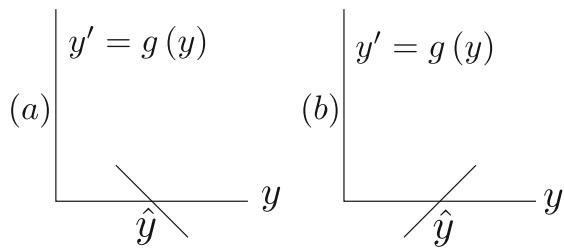
To see why this property is true, we use the fact from calculus that if a differentiable function tends to a limit as $t \rightarrow \infty$, then its derivative must tend to zero. If a solution $y(t)$ of $y' = g(y)$ tends to zero, then by the continuity of the function $g(y)$ we have

$$0 = \lim_{t \rightarrow \infty} y'(t) = \lim_{t \rightarrow \infty} g\{y(t)\} = g\{\lim_{t \rightarrow \infty} y(t)\}.$$

Thus, $\lim_{t \rightarrow \infty} y(t)$ must be a root of the equation $g(y) = 0$, and hence an equilibrium of (B.26).

In applications, the initial condition usually comes from observations and is subject to experimental error. For a model to be a plausible predictor of what will actually occur, it is important that a small change in the initial value does not produce a large change in the solution. An equilibrium \hat{y} of (B.26) such that every solution with initial value sufficiently close to \hat{y} approaches \hat{y} as $t \rightarrow \infty$ is said to be *asymptotically stable*. If there are solutions which start arbitrarily close to an equilibrium but move away from it, then the equilibrium is said to be *unstable*. For the logistic differential equation (B.8) the equilibrium $y = 0$ is unstable, and the equilibrium $y = K$ is asymptotically stable. In applications, unstable equilibria

Fig. B.5 Equilibria \hat{y} with
(a) $g'(\hat{y}) < 0$, (b) $g'(\hat{y}) > 0$



have no significance, because they can be observed only if the initial condition is “just right.”

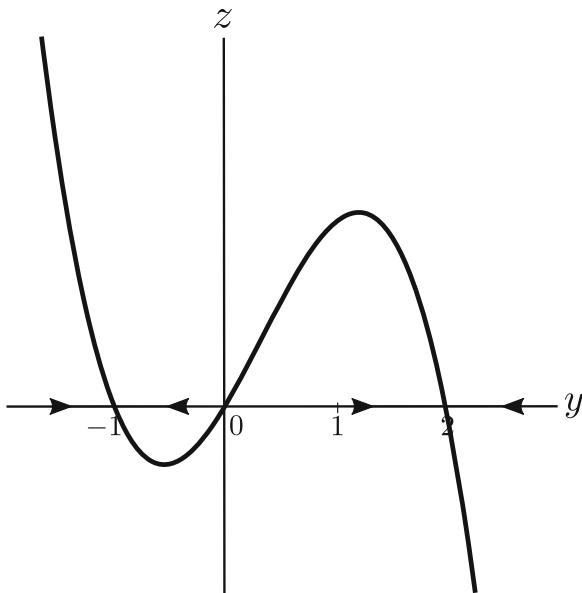
Property 5 An equilibrium \hat{y} of $y' = g(y)$ with $g'(\hat{y}) < 0$ is asymptotically stable; an equilibrium \hat{y} with $g'(\hat{y}) > 0$ is unstable.

To establish this property, we note that if \hat{y} is an equilibrium with $g'(\hat{y}) < 0$ ($g(y)$ has a negative slope at $y = \hat{y}$), then $g(y)$ is positive if $y < \hat{y}$ and negative if $y > \hat{y}$ (Fig. B.5a). In this case solutions above the equilibrium decrease toward the equilibrium, while solutions below the equilibrium increase toward the equilibrium. This shows that an equilibrium \hat{y} with $g'(\hat{y}) < 0$ is asymptotically stable. By a similar argument, if $g'(\hat{y}) > 0$, solutions above the equilibrium increase and solutions below the equilibrium decrease, with both moving away from the equilibrium (Fig. B.5b).

The properties we have developed make it possible for us to make a complete analysis of the asymptotic behavior, that is, the behavior as $t \rightarrow \infty$, of solutions of an autonomous differential equation (B.26) merely by examining the equilibria and the nature of the function $g(y)$ for large values of y . We begin by finding all the equilibria (roots of the equation $g(y) = 0$). An equilibrium \hat{y} with $g'(\hat{y}) < 0$ is asymptotically stable, and all solutions with initial value in the two bands adjoining this equilibrium tend to it. An equilibrium \hat{y} with $g'(\hat{y}) > 0$ is unstable and repels all solutions with initial value in the two bands adjoining it. An equilibrium \hat{y} with $g'(\hat{y}) = 0$ must be analyzed more carefully. If $g(y)$ is negative for values of y above the largest equilibrium, then no solutions become positively unbounded. If $g(y)$ is positive for values of y above the largest equilibrium, then this equilibrium is unstable, and solutions with initial value above this equilibrium become unbounded. If $g(y)$ is negative for values of y below the smallest equilibrium, then this equilibrium is likewise unstable, and solutions with initial value below this equilibrium become negatively unbounded.

A convenient way to display the qualitative behavior of solutions of an autonomous differential equation (B.26) is by drawing the *phase line*. We draw the graph of the function $g(y)$ and on the y -axis we may draw arrows to the right where the graph is above the y -axis and to the left where the graph is below the y -axis. The reason for doing this is that where the graph is above the axis $g(y)$ is positive and therefore the solution y of (B.26) is increasing, while where the graph is below the axis $g(y)$ is negative and therefore the solution y of (B.26) is

Fig. B.6 A phase line superimposed on the corresponding graph of $g(y)$



decreasing. The points where the graph crosses the y -axis are the equilibria, and we can see from the directions of the arrows along the axis which equilibria are asymptotically stable and which are unstable. The phase line is the y -axis viewed as the line on which the solution curve moves, thinking of t as a parameter. Thus, the solution is described by the motion along the line, whose direction is given by the arrows. The graph of Fig. B.6 describes a situation in which there are asymptotically stable equilibria at $y = -1$ and $y = 2$, and an unstable equilibrium at $y = 0$.

Example 8 Describe the asymptotic behavior of solutions of the differential equation (B.25).

Solution Here $g(y) = -ay + b$, $g'(y) = -a$. The only equilibrium is $y = \frac{b}{a}$. If $a > 0$, this equilibrium is asymptotically stable, and $g(y) < 0$ if y is large and positive, $g(y) > 0$ if y is large and negative. This means that every solution is bounded and approaches the limit $\frac{b}{a}$. If $a < 0$, however, the equilibrium is unstable. Further, since $g(y) > 0$ above the equilibrium and $g(y) < 0$ below the equilibrium, every solution is unbounded, either positively or negatively. \square

Example 9 Describe the asymptotic behavior of solutions with $y(0) \geq 0$ of the differential equation

$$y' = y(r e^{-y} - d), \quad (\text{B.27})$$

where r and d are positive constants.

Fig. B.7 A phase line and graph of $g(y)$ for (B.27), with $r < d$

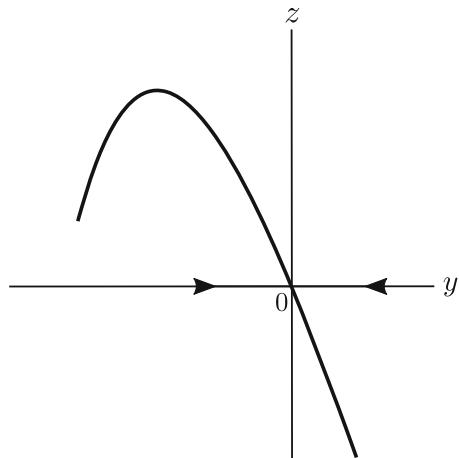
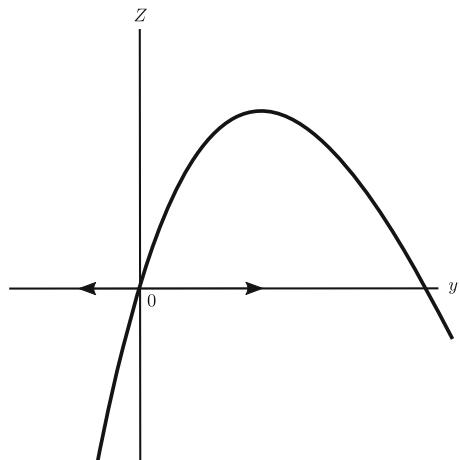


Fig. B.8 A phase line and graph of $g(y)$ for (B.27), with $r > d$



Solution The equilibria of (B.27) are the solutions of $y(re^{-y} - d) = 0$. Thus there are two equilibria, namely $y = 0$ and the solution \hat{y} of $re^{-y} = d$, which is $\hat{y} = \log \frac{r}{d}$. If $r < d$, $\hat{y} < 0$, and only the equilibrium $y = 0$ is of interest. In this case, $g(y) < 0$ for $y > 0$, and solutions with $y(0) > 0$ decrease to 0 (Fig. B.7).

If $r > d$, we define K to be $\log \frac{r}{d}$, so that the positive equilibrium is $y = K$. We may now rewrite the differential equation (B.27) as $y' = ry(e^{-y} - e^{-K})$. For the function $g(y) = ry(e^{-y} - e^{-K})$, we have $g'(y) = r(e^{-y} - e^{-K}) - rye^{-y}$ and $g'(0) = r(1 - e^{-K}) > 0$, implying that the equilibrium $y = 0$ is unstable. The equilibrium $y = K$ is asymptotically stable since $g'(K) = -rKe^{-K} < 0$. All positive solutions are bounded, because $g(y) < 0$ for $y > K$. Thus every solution with $y(0) > 0$ tends to the limit K , while the solution with initial value zero is the zero function and has limit zero (Fig. B.8). \square

We see that the solutions of (B.27), which has been suggested as a population growth model with per capita birth rate re^{-y} and per capita death rate d , behave qualitatively in the same manner as solutions of the logistic equation. Some other differential equations which exhibit the same behavior have been proposed as population models. The original formulation of the logistic equation was based on a per capita growth rate which should be a decreasing function of population size, positive for small y but negative for large y . It is not difficult to show that every population model of the form (B.26) for which the per capita growth rate is a decreasing function of the population size y , and which is positive for $0 < y < K$ and negative for $y > K$, has the property that every solution with $y(0) > 0$ approaches the limit K as $t \rightarrow \infty$.

In many applications, the function $g(y)$ has the form $g(y) = yf(y)$, which has the effect of guaranteeing that $y = 0$ is an equilibrium. Then $g'(y) = f(y) + yf'(y)$. At the equilibrium $y = 0$, $g'(0) = f(0)$ and at a non-zero equilibrium \hat{y} with $f(\hat{y}) = 0$, $g'(\hat{y}) = \hat{y}f'(\hat{y})$. Thus the equilibrium $y = 0$ is asymptotically stable if $f(0) < 0$ and unstable if $f(0) > 0$; a non-zero equilibrium \hat{y} is asymptotically stable if $f'(\hat{y}) < 0$ and unstable if $f'(\hat{y}) > 0$. We shall restate this result formally as a theorem.

Equilibrium Stability Theorem *An equilibrium \hat{y} of $y' = g(y)$ with $g'(\hat{y}) < 0$ is asymptotically stable; an equilibrium \hat{y} with $g'(\hat{y}) > 0$ is unstable. The equilibrium $y = 0$ of $y' = yf(y)$ is asymptotically stable if $f(0) < 0$ and unstable if $f(0) > 0$, while a non-zero equilibrium \hat{y} is asymptotically stable if $f'(\hat{y}) < 0$ and unstable if $f'(\hat{y}) > 0$.*

Appendix C

Systems of Differential Equations

In the previous chapter we saw how to model and analyze continuously changing quantities using differential equations. In many applications of interest there may be two or more interacting quantities—populations of two or more species, for instance, or parts of a whole, which depend upon each other. When the amount or size of one quantity depends in part on the amount of another, and vice versa, they are said to be *coupled*, and it is not possible or appropriate to model each one separately. In these cases we write models which consist of *systems* of differential equations. In this chapter, we will find that the quantitative and qualitative approaches we used to analyze individual differential equations in Sects. B.2 and B.5 extend in a more or less natural way to cover systems of differential equations. Extending them will require some basic multivariable calculus, principally the use of partial derivatives and the idea of linear approximation.

C.1 The Phase Plane

Our purpose is to study *two-dimensional autonomous systems of first-order differential equations* of the general form

$$\begin{aligned}y' &= F(y, z) \\z' &= G(y, z)\end{aligned}\tag{C.1}$$

Usually, it is not possible to solve such a system, by which we mean to find y and z as functions of t so that

$$y'(t) = F(y(t), z(t)), \quad z'(t) = G(y(t), z(t))$$

for all t in some interval. However, we can often obtain information about the relation between the functions $y(t)$ and $z(t)$. Geometrically, such information is displayed as a curve in the y - z plane, called the *phase plane* for this system.

An *equilibrium* of the system (C.1) is a solution (y_∞, z_∞) of the pair of equations

$$F(y, z) = 0, \quad G(y, z) = 0. \quad (\text{C.2})$$

Geometrically, an equilibrium is a point in the phase plane. In terms of the system (C.1), an equilibrium gives a constant solution $y = y_\infty$, $z = z_\infty$ of the system. This definition is completely analogous to the definition of an equilibrium given for a first-order differential equation in Sect. B.5.

The *orbit* of a solution $y = y(t)$, $z = z(t)$ of the system (C.1) is the curve in the y - z phase plane consisting of all points $(y(t), z(t))$ for $0 \leq t < \infty$. A closed orbit corresponds to a periodic solution because the orbit must travel repeatedly around the closed orbit as t increases.

There is a geometric interpretation of orbits which is analogous to the interpretation given for solutions of first-order differential equations in Sect. B.3. Just as the curve $y = y(t)$ has slope $y' = f(t, y)$ at each point (t, y) along its length, an orbit of (C.1) (considering z as an implicit function of y) has slope

$$\frac{dz}{dy} = \frac{z'}{y'} = \frac{G(y, z)}{F(y, z)}$$

at each point of the orbit. The *direction field* for a two-dimensional autonomous system is a collection of line segments in the phase plane with this slope at each point (y, z) , and an orbit must be a curve which is tangent to the direction field at each point of the curve. Computer algebra systems such as Maple and Mathematica may be used to draw the direction field for a given system.

Example 1 Describe the orbits of the system

$$y' = z, \quad z' = -y.$$

Solution If we consider z as a function of y , we have

$$\frac{dz}{dy} = z'/y' = -\frac{y}{z}.$$

Solution by separation of variables gives

$$\int z \, dz = - \int y \, dy,$$

and integration gives $\frac{z^2}{2} = -\frac{y^2}{2} + c$. Thus every orbit is a circle $y^2 + z^2 = 2c$ with center at the origin, and every solution is periodic. \square

To find equilibria of a system (C.1), it is helpful to draw the *nullclines*, namely the curves $F(y, z) = 0$ on which $y' = 0$, and $G(y, z) = 0$ on which $z' = 0$. An equilibrium is an intersection of these two curves.

C.2 Linearization of a System at an Equilibrium

Sometimes it is possible to find the orbits in the phase plane of a system of differential equations, but it is rarely possible to solve a system of differential equations analytically. For this reason, our study of systems will concentrate on qualitative properties. The linearization of a system of differential equations at an equilibrium is a linear system with constant coefficients, whose solutions approximate the solutions of the original system near the equilibrium. In this section, we shall see how to find the linearization of a system. In the next section, we shall see how to solve linear systems with constant coefficients, and this will enable us to understand much of the behavior of solutions of a system near an equilibrium.

Let (y_∞, z_∞) be an equilibrium of a system (C.1) that is, a point in the phase plane such that (C.2). We will assume that the equilibrium is *isolated*, that is, that there is a circle centered around (y_∞, z_∞) which does not contain any other equilibrium. We shift the origin to the equilibrium by letting $y = y_\infty + u$, $z = z_\infty + v$, and then make linear approximations to $F(y_\infty + u, z_\infty + v)$ and $G(y_\infty + u, z_\infty + v)$. The difference here between a one-dimensional system and a two-dimensional one is that the linear approximation in two or more dimensions uses partial derivatives. Our approximations are

$$F(y_\infty + u, z_\infty + v) \approx F(y_\infty, z_\infty) + F_y(y_\infty, z_\infty)u + F_z(y_\infty, z_\infty)v \quad (\text{C.3})$$

$$G(y_\infty + u, z_\infty + v) \approx G(y_\infty, z_\infty) + G_y(y_\infty, z_\infty)u + G_z(y_\infty, z_\infty)v$$

with error terms h_1 and h_2 , respectively, which are negligible relative to the linear terms in (C.3) when u and v are small (i.e., close to the equilibrium).

The linearization of the system (C.1) at the equilibrium (y_∞, z_∞) is defined to be the linear system with constant coefficients

$$\begin{aligned} u' &= F_y(y_\infty, z_\infty)u + F_z(y_\infty, z_\infty)v \\ v' &= G_y(y_\infty, z_\infty)u + G_z(y_\infty, z_\infty)v. \end{aligned} \quad (\text{C.4})$$

To obtain it, we first note that $y' = u'$, $z' = v'$, and then substitute (C.3) into (C.1). By (C.2) the constant terms are zero, and for the linearization we neglect the higher-order terms h_1 and h_2 . The *coefficient matrix* of the linear system (C.4) is the matrix of constants

$$\begin{bmatrix} F_y(y_\infty, z_\infty) & F_z(y_\infty, z_\infty) \\ G_y(y_\infty, z_\infty) & G_z(y_\infty, z_\infty) \end{bmatrix}.$$

The entries of this Jacobian matrix describe the effect of a change in each variable on the growth rates of the two variables.

Example 1 Find the linearization at each equilibrium of the system

$$y' = A - \beta yz - \mu y, \quad z' = \beta yz - (\gamma + \mu)z.$$

(This corresponds to the classical Kermack–McKendrick SIR model for a possibly endemic disease. Here y corresponds to the number of susceptible, uninfected individuals, z to the number of infected, infective individuals, A to the birth rate, μ to the death rate and β and γ to the infection and recovery rates, respectively.)

Solution The equilibria are the solutions of $A = y(\beta z + \mu)$, $z(\beta y) = (\gamma + \mu)z$. To satisfy the second of these equations, we must have either $z = 0$ or $\beta y = \gamma + \mu$. If $z = 0$, the first equation gives $y = A/\mu$. If $z > 0$, the second equation gives $\beta y = \gamma + \mu$. Since

$$\begin{aligned} \frac{\partial}{\partial y} [-\beta yz - \mu y] &= -\beta z - \mu, & \frac{\partial}{\partial z} [-\beta yz] &= -\beta y, \\ \frac{\partial}{\partial y} [\beta yz - (\gamma + \mu)z] &= \beta z, & \frac{\partial}{\partial z} [\beta yz - (\gamma + \mu)z] &= \beta y - (\gamma + \mu), \end{aligned}$$

the linearization at an equilibrium (y_∞, z_∞) is

$$\begin{aligned} u' &= -(\beta z_\infty + \mu)u - \beta y_\infty v, \\ v' &= \beta z_\infty u + (\beta y_\infty - (\gamma + \mu))v. \end{aligned}$$

At the equilibrium $(A/\mu, 0)$ the linearization is

$$\begin{aligned} u' &= -\mu u - \beta \frac{A}{\mu} v \\ v' &= \beta \frac{A}{\mu} u - (\gamma + \mu)v. \end{aligned}$$

At the other equilibrium with $I > 0$ the linearization is

$$\begin{aligned} u' &= -(\beta I_\infty + \mu)u - (\mu + \alpha)v \\ v' &= \beta I_\infty u. \quad \square \end{aligned}$$

An equilibrium of the system (C.1) with the property that every orbit with initial value sufficiently close to the equilibrium remains close to the equilibrium for all $t \geq 0$, and approaches the equilibrium as $t \rightarrow \infty$, is said to be *locally asymptotically stable*. An equilibrium of (C.1) with the property that some solutions starting arbitrarily close to the equilibrium move away from it is said to be *unstable*.

These definitions are completely analogous to those given in Sect. B.5 for first-order differential equations. We speak of local asymptotic stability to distinguish from global asymptotic stability, which is the property that *all* solutions, not merely those with initial value sufficiently close to the equilibrium, approach the equilibrium. If we speak of asymptotic stability of an equilibrium, we will mean local asymptotic stability unless we specify that the asymptotic stability is global.

The fundamental property of the linearization which we will use to study stability of equilibria is the following result, which we state without proof. The proof may be found in any text which covers the qualitative study of nonlinear differential equations. Here we suppose F and G to be twice differentiable, that is, smooth enough for the linearization to give a correct picture.

Linearization Theorem *If (y_∞, z_∞) is an equilibrium of the system*

$$y' = F(y, z), \quad z' = G(y, z)$$

and if every solution of the linearization at this equilibrium approaches zero as $t \rightarrow \infty$, then the equilibrium (y_∞, z_∞) is (locally) asymptotically stable. If the linearization has unbounded solutions, then the equilibrium (y_∞, z_∞) is unstable.

For a first-order differential equation $y' = g(y)$ at an equilibrium y_∞ , the linearization is the first-order linear differential equation $u' = g'(y_\infty)u$. We may solve this differential equation by separation of variables and see that all solutions approach zero if $g'(y_\infty) < 0$, and there are unbounded solutions if $g'(y_\infty) > 0$. We have seen in Sect. B.5, without recourse to the linearization, that the equilibrium is locally asymptotically stable if $g'(y_\infty) < 0$ and unstable if $g'(y_\infty) > 0$. The linearization theorem is valid for systems of any dimension and is the approach needed for the study of stability of equilibria for systems of dimension higher than 1.

Note that there is a case where the theorem above does not draw any conclusions, namely the case where the linearization about the equilibrium is neither asymptotically stable nor unstable. In this case, the equilibrium of the original (nonlinear) system may be asymptotically stable, unstable, or neither.

Example 2 For each equilibrium of the system

$$y' = z, \quad z' = -2(y^2 - 1)z - y$$

determine whether the equilibrium is asymptotically stable or unstable.

Solution The equilibria are the solutions of $z = 0, -2(y^2 - 1)z - y = 0$, and thus the only equilibrium is $(0,0)$. Since $\frac{\partial}{\partial y}[z] = 0$, $\frac{\partial}{\partial z}[z] = 1$, and

$$\frac{\partial}{\partial y}[-2(y^2 - 1)z - y] = -4yz - 1, \quad \frac{\partial}{\partial z}[-2(y^2 - 1)z - y] = -2(y^2 - 1),$$

the linearization at $(0,0)$ is $u' = 0u + 1v = v$, $v' = -u + 2v$. We can actually solve this system of equations outright by using a clever trick to reduce it to a single equation. We subtract the first equation from the second to give $(v - u)' = (v - u)$. This is a first-order differential equation for $(v - u)$, whose solution is $v - u = c_1 e^t$. This partial result is already enough to tell us that the equilibrium is unstable, as since the difference between u and v grows exponentially, at least one of them must therefore grow exponentially as well. As there are unbounded solutions, we conclude that the equilibrium $(0,0)$ is unstable. \square

C.3 Solution of Linear Systems with Constant Coefficients

We have seen in the preceding section that the stability of an equilibrium of a system of differential equations is determined by the behavior of solutions of the system's linearization at the equilibrium. This linearization is a linear system with constant coefficients (recall that a linear system has right-hand sides linear in y and z). Thus, in order to be able to decide whether an equilibrium is asymptotically stable, we need to be able to solve linear systems with constant coefficients. We were able to do this in the examples of the preceding section because the linearizations took a simple form with one of the equations of the system containing only a single variable. In this section we shall develop a more general technique.

The problem we wish to solve is a general two-dimensional linear system with constant coefficients,

$$\begin{aligned} y' &= ay + bz, \\ z' &= cy + dz, \end{aligned} \tag{C.5}$$

where a , b , c , and d are constants. We look for solutions of the form

$$y = Ye^{\lambda t}, \quad z = Ze^{\lambda t}, \tag{C.6}$$

where λ , Y , and Z are constants to be determined, with Y and Z not both zero. When we substitute the form (C.6) into the system (C.5), using $y' = \lambda Ye^{\lambda t}$, $z' = \lambda Ze^{\lambda t}$, we obtain two conditions

$$\begin{aligned} \lambda Ye^{\lambda t} &= aYe^{\lambda t} + bZe^{\lambda t}, \\ \lambda Ze^{\lambda t} &= cYe^{\lambda t} + dZe^{\lambda t}, \end{aligned}$$

which must be satisfied for all t . Because $e^{\lambda t} \neq 0$ for all t , we may divide these equations by $e^{\lambda t}$ to obtain a system of two equations which do not depend on t , namely

$$\lambda Y = aY + bZ,$$

$$\lambda Z = cY + dZ$$

or

$$(a - \lambda)Y + bZ = 0, \quad (C.7)$$

$$cY + (d - \lambda)Z = 0.$$

The pair of equations (C.7) is a system of two homogeneous linear algebraic equations for the unknowns Y and Z . For certain values of the parameter λ this system will have a solution other than the obvious solution $Y = 0, Z = 0$. In order that the system (C.7) have a non-trivial solution for Y and Z , it is necessary that the determinant of the coefficient matrix, which is $(a - \lambda)(d - \lambda) - bc$, be equal to zero (this result comes from linear algebra). This gives a quadratic equation, called the *characteristic equation*, of the system (C.5) for λ . We may rewrite the characteristic equation as

$$\lambda^2 - (a + d)\lambda + (ad - bc) = 0. \quad (C.8)$$

We will assume that $ad - bc \neq 0$, which is equivalent to the assumption that $\lambda = 0$ is not a root of (C.8). Our reason for this assumption is the following: If $\lambda = 0$ is a root (or, equivalently, $ad - bc = 0$), the equilibrium conditions will reduce to a single equation since each equation is a constant multiple of the other and there is effectively only one equilibrium equation. Consequently, there will be a line of non-isolated equilibria. We do not wish to explore this problem in part because the treatment of non-isolated equilibria is complicated and in part because it is not possible to apply the linearization theorem of the previous section when the linearization of a system about an equilibrium is of this form.

If $ad - bc \neq 0$, the characteristic equation has two roots λ_1 and λ_2 , which may be real and distinct, real and equal, or complex conjugates. If λ_1 and λ_2 are the roots of (C.8), then there is a solution (Y_1, Z_1) of (C.7) corresponding to the root λ_1 , and a solution (Y_2, Z_2) of (C.7) corresponding to the root λ_2 . These, in turn, give us two solutions

$$y = Y_1 e^{\lambda_1 t}, \quad z = Z_1 e^{\lambda_1 t}$$

$$y = Y_2 e^{\lambda_2 t}, \quad z = Z_2 e^{\lambda_2 t}$$

of the system (C.5). We note that if λ is a root of (C.8), then equations (C.7) reduce to a single equation. Thus we may make an arbitrary choice for one of Y , Z and the other is then determined by (C.7).

Because the system (C.5) is linear, it is easy to see that every constant multiple of a solution of (C.7) is also a solution and also that the sum of two solutions of (C.7) is also a solution. It is possible to show that if we have two different solutions of the system (C.5), then every solution of (C.5) is a constant multiple of the first solution plus a constant multiple of the second solution. By “different” we mean that neither solution is a constant multiple of the other. Here, if the roots λ_1 and λ_2 of the characteristic equation (C.8) are distinct, we do have two different solutions of (C.5), and then every solution of system (C.5) has the form

$$\begin{aligned}y &= K_1 Y_1 e^{\lambda_1 t} + K_2 Y_2 e^{\lambda_2 t}, \\z &= K_1 Z_1 e^{\lambda_1 t} + K_2 Z_2 e^{\lambda_2 t}\end{aligned}\tag{C.9}$$

for some constants K_1 and K_2 . The form (C.9) with two arbitrary constants K_1 and K_2 is called the *general solution* of the system (C.5). If initial values $y(0)$ and $z(0)$ are specified, these two initial values may be used to determine values for the constants K_1 and K_2 , and thus to obtain a *particular solution* in the family (C.9).

Example 1 Find the general solution of the system

$$y' = -y - 2z, \quad z' = y - 4z$$

and also the solution such that $y(0) = 3$, $z(0) = 1$.

Solution Here $a = -1$, $b = -2$, $c = 1$, $d = -4$, so that $a + d = -5$, $ad - bc = 6$, and the characteristic equation is $\lambda^2 + 5\lambda + 6 = 0$, with roots $\lambda_1 = -2$, $\lambda_2 = -3$. With $\lambda = -2$, both equations of the algebraic system (C.7) are $Y - 2Z = 0$, and we may take $Y = 2$, $Z = 1$. The resulting solution of (C.5) is $y = 2e^{-2t}$, $z = e^{-2t}$. With $\lambda = -3$, both equations of the algebraic system (C.7) are $Y - Z = 0$, and we may take $Y = 1$, $Z = 1$. The resulting solution of (C.5) is $y = e^{-3t}$, $z = e^{-3t}$. Thus the general solution of the system is

$$y = 2K_1 e^{-2t} + K_2 e^{-3t}, \quad z = K_1 e^{-2t} + K_2 e^{-3t}.$$

To satisfy the initial conditions, we substitute $t = 0$, $y = 3$, $z = 1$ into this form, obtaining a pair of equations $2K_1 + K_2 = 3$, $K_1 + K_2 = 1$. We may subtract the second of these from the first to give $K_1 = 2$, and then $K_2 = -1$. This gives, as the solution of the initial value problem,

$$y = 4e^{-2t} - e^{-3t}, \quad z = 2e^{-2t} - e^{-3t}. \quad \square$$

In the language of linear algebra, λ is an *eigenvalue* and (Y, Z) is a corresponding *eigenvector* of the 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Note that $a + d$ is the *trace*, denoted by $\text{tr } A$, and $ad - bc$ is the *determinant*, denoted by $\det A$, of the matrix A . The sum of the eigenvalues of A is the trace, and the product of the eigenvalues is the determinant. This is seen easily by writing (C.8) as

$$\lambda^2 - \text{tr } A \lambda + \det A = 0$$

We will consider vectors to be column vectors, that is, matrices with two rows and one column. Suppose the matrix A has two distinct real eigenvalues λ_1, λ_2 with corresponding eigenvectors

$$[Y_1, Z_1], \quad [Y_2, Z_2]$$

respectively (this approach can also be used if $\lambda_1 = \lambda_2$ provided there are two independent corresponding eigenvectors). We define the matrix

$$P = \begin{bmatrix} Y_1 & Y_2 \\ Z_1 & Z_2 \end{bmatrix}$$

and then if we make the change of variable

$$\begin{bmatrix} y \\ z \end{bmatrix} = P \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} Y_1 & Y_2 \\ Z_1 & Z_2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

the system (C.5) is transformed to the system

$$\begin{aligned} u' &= \lambda_1 u \\ v' &= \lambda_2 v. \end{aligned} \tag{C.10}$$

Since the system (C.10) is uncoupled, it may easily be solved explicitly to give

$$u = K_1 e^{\lambda_1 t}, \quad v = K_2 e^{\lambda_2 t}.$$

We may then transform back to the original variables y, z to give the solution (C.9). The linear algebra approach gives the same result that was obtained by our initial approach of trying to find exponential solutions. Since it is constructive, it does not make use of the result which we stated without proof that if we have two different solutions of the system (C.5), then every solution of (C.5) is a constant multiple of the first solution plus a constant multiple of the second solution.

If the characteristic equation (C.8) has a double root, the method we have used gives only one solution of the system (C.5), and we need to find a second solution in order to form the general solution. In order to see what form the second solution must have we consider the special case $c = 0$ of (C.5). Then (C.5) becomes

$$\begin{aligned} y' &= ay + bz, \\ z' &= dz. \end{aligned} \tag{C.11}$$

Then the eigenvalues of the corresponding matrix A are a and d ; if $a = d$, the characteristic equation has a double root. However, the assumption $c = 0$ means that we can solve the system (C.11) recursively. Every solution of the second equation in (C.11) has the form

$$z = Ze^{dt}$$

and we may substitute this into the first equation to give

$$y' = ay + bZe^{dt}$$

This first order linear differential equation is easily solved. We multiply the equation by e^{-at} to give

$$y'e^{-at} - aye^{-at} = bZe^{(d-a)t}. \quad (\text{C.12})$$

If $a \neq d$, integration of (C.12) gives

$$ye^{-at} = \frac{bZ}{d-a}e^{(d-a)t} + Y$$

where Y is a constant of integration. From this we obtain the solution

$$\begin{aligned} y &= \frac{bZ}{d}e^{dt} + Ye^{at} \\ z &= Ze^{dt} \end{aligned}$$

which is equivalent to the solution (C.9) obtained earlier; note that here $\lambda_1 = a$, $\lambda_2 = d$.

A more interesting case arises when $a = d$, so that the two roots of the characteristic equation are equal. Then (C.12) becomes

$$y'e^{-at} - aye^{-at} = bZ \quad (\text{C.13})$$

and integration gives

$$ye^{-at} = bZt + Y$$

where Y is a constant of integration. From this we obtain the solution

$$\begin{aligned} y &= bZte^{at} + Ye^{at} \\ z &= Ze^{at}. \end{aligned}$$

This suggests that if there is a double root λ of the characteristic equation the needed second solution of (C.5) will contain terms $te^{\lambda t}$ as well as $e^{\lambda t}$. It is possible to show (and the reader can verify) that if λ is a double root of (C.8), we must have $\lambda = \frac{a+d}{2}$

and in addition to the solution $y = Y_1 e^{\lambda t}$, $z = Z_1 e^{\lambda t}$ of (C.5) there is a second solution, of the form

$$y = (Y_2 + Y_1 t) e^{\lambda t}, \quad z = (Z_2 + Z_1 t) e^{\lambda t}$$

where Y_1 , Z_1 are as in (C.7) and Y_2 , Z_2 are given by

$$\begin{aligned} (a - \lambda)Y_2 + bZ_2 &= Y_1, \\ cY_2 + (d - \lambda)Z_2 &= Z_1. \end{aligned}$$

Thus the general solution of the system (C.5) in the case of equal roots is

$$\begin{aligned} y &= (K_1 Y_1 + K_2 Y_2) e^{\lambda t} + K_2 Y_1 t e^{\lambda t}, \\ z &= (K_1 Z_1 + K_2 Z_2) e^{\lambda t} + K_2 Z_1 t e^{\lambda t}. \end{aligned} \quad (\text{C.14})$$

Note that if (C.8) has a single root and $b = 0$, then $\lambda = a = d$, and we have $Y_1 = 0$, $Z_1 = cY_2$, so that the general solution becomes

$$y = K_3 e^{\lambda t}, \quad z = K_4 e^{\lambda t} + cK_3 t e^{\lambda t},$$

where $K_3 \equiv K_2 Y_2$ and $K_4 \equiv cK_1 Y_2 + K_2 Z_2$ are arbitrary constants. Likewise if (C.8) has a single root and $c = 0$, then $\lambda = a = d$, $Z_1 = 0$, $Y_1 = bZ_2$, and the general solution reduces to

$$y = K_5 e^{\lambda t} + bK_6 t e^{\lambda t}, \quad z = K_6 e^{\lambda t}$$

with arbitrary constants $K_5 \equiv bK_1 Z_2 + K_2 Y_2$ and $K_6 \equiv K_2 Z_2$. If b and c are both zero, the system becomes uncoupled

$$y' = ay, \quad z' = dz$$

and is easily solved by integration to give $y = K_1 e^{at}$, $z = K_2 e^{dt}$. This is the solution in all cases, regardless of whether the characteristic equation has distinct roots or equal roots.

Example 2 Find the general solution of the system

$$y' = z, \quad z' = -y + 2z$$

and also the solution such that $y(0) = 2$, $z(0) = 3$.

Solution Since $a = 0$, $b = 1$, $c = -1$, $d = 2$, we have $a + d = 2$, $ad - bc = 1$. The characteristic equation is $\lambda^2 - 2\lambda + 1 = 0$, with a double root $\lambda = 1$. With $\lambda = 1$, both equations of the system (C.7) are $Y - Z = 0$, and we may take $Y = 1$, $Z = 1$ to give the solution $y = e^t$, $z = e^t$ of (C.5). Substituting these values into the

equations for Y_2, Z_2 , we find that they reduce to the single equation $-Y_2 + Z_2 = 1$, so we may take $Y_2 = 0, Z_2 = 1$. Equations (C.14) now give the general solution $y = K_1 e^t + K_2 t e^t, z = (K_1 + K_2)e^t + K_2 t e^t$. To find the solution with $y(0) = 2, z(0) = 3$, we substitute $t = 0, y = 2, z = 3$ into this form, obtaining the pair of equations $K_1 = 2, K_1 + K_2 = 3$. Then $K_2 = 1$, and the solution satisfying the initial conditions is $y = 2e^t + te^t, z = 3e^t + te^t$. \square

From a linear algebra perspective we may make a change of variable of the system (C.5) just as in the case where we had two eigenvectors. We define the matrix

$$P = \begin{bmatrix} Y_1 & Y_2 \\ Z_1 & Z_2 \end{bmatrix}$$

with

$$\begin{bmatrix} Y_1 \\ Z_1 \end{bmatrix}$$

an eigenvector of the matrix A as before, but with

$$\begin{bmatrix} Y_2 \\ Z_2 \end{bmatrix}$$

chosen arbitrarily, except that the determinant of P must be non-zero. Then under the change of variable

$$\begin{bmatrix} y \\ z \end{bmatrix} = P \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} Y_1 & Y_2 \\ Z_1 & Z_2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

the system (C.5) is transformed to a system of the form (C.5) with $c = 0$ which may be solved explicitly, as we have seen.

A particularly clever choice of $[Y_2, Z_2]$ is the solution of the system of linear equations

$$\begin{aligned} aY_2 + bZ_2 &= \lambda Y_2 + Y_1 \\ cY_2 + dZ_2 &= \lambda Z_2 + Y_2. \end{aligned}$$

It is possible to show that this system always has a solution. With this choice of Y_2, Z_2 , the system (C.5) is transformed to

$$\begin{aligned} u' &= \lambda_1 u + v & (C.15) \\ v' &= \lambda_2 v. \end{aligned}$$

If we solve the system (C.15) recursively by the method used to solve the system (C.11) and the solution is then transformed back to the original variables, we obtain the solution (C.14).

Another complication arises if the characteristic equation (C.8) has complex roots. While the general solution of (C.5) is still given by (C.9), in this case the constants λ_1 and λ_2 are complex, and the solution is in terms of complex functions. Complex exponentials, however, can be defined with the aid of trigonometric functions ($e^{i\theta} \equiv \cos \theta + i \sin \theta$ for real θ), so it is still possible to give the solution of (C.5) in terms of real exponential and trigonometric functions. In this case, if the characteristic equation (C.8) has conjugate complex roots $\lambda = \alpha \pm i\beta$, where α and β are real and $\beta > 0$, equations (C.9) become

$$\begin{aligned} y &= (K_1 Y_1 + K_2 Y_2) e^{\alpha t} \cos \beta t + i(K_1 Y_1 - K_2 Y_2) e^{\alpha t} \sin \beta t, \\ z &= (K_1 Z_1 + K_2 Z_2) e^{\alpha t} \cos \beta t + i(K_1 Z_1 - K_2 Z_2) e^{\alpha t} \sin \beta t. \end{aligned}$$

We can eliminate the imaginary coefficients by defining $Q_1 \equiv i(K_1 - K_2)$, $Q_2 \equiv K_1 + K_2$, and taking $(a - \lambda_i)Y_i + bZ_i = 0$ for $i = 1, 2$ from (C.9) to arrive at the form

$$\begin{aligned} y &= Q_1 b e^{\alpha t} \sin \beta t + Q_2 b e^{\alpha t} \cos \beta t, \\ z &= -[Q_1(a - \alpha) - Q_2 \beta] e^{\alpha t} \sin \beta t + [Q_1 \beta - Q_2(a - \alpha)] e^{\alpha t} \cos \beta t. \end{aligned}$$

Example 3 Find the general solution of the system

$$y' = -2z, \quad z' = y + 2z$$

and also the solution with $y(0) = -2$, $z(0) = 0$.

Solution We have $a = 0$, $b = -2$, $c = 1$, $d = 2$, and the characteristic equation is $\lambda^2 - 2\lambda + 2 = 0$, with roots $\lambda = 1 \pm i$. The general solution is then

$$y = -2K_1 e^t \sin t - 2K_2 e^t \cos t, \quad z = (K_1 - K_2) e^t \sin t + (K_1 + K_2) e^t \cos t.$$

To find the solution with $y(0) = -2$, $z(0) = 0$, we substitute $t = 0$, $y = -2$, $z = 0$ into this form, obtaining $-2K_2 = -2$, $K_1 + K_2 = 0$, whose solution is $K_1 = -1$, $K_2 = 1$. This gives the particular solution

$$y = 2e^t \sin t - 2e^t \cos t, \quad z = -2e^t \sin t. \quad \square$$

In many applications, especially in analyzing stability of an equilibrium, the precise form of the solution of a linear system is less important to us than the qualitative behavior of solutions. It will turn out that often the crucial question is whether all solutions of a linear system approach zero as $t \rightarrow \infty$. The nature of the origin as an equilibrium of the linear system (C.5) depends on the roots of the

characteristic equation (C.8). If both roots of (C.8) are real and negative, then the solutions of (C.5) are combinations of negative exponentials. This implies that all orbits approach the origin and the origin is asymptotically stable. If the roots of (C.8) are real and of opposite sign, then there are solutions which are positive exponentials and solutions which are negative exponentials. Thus there are solutions approaching the origin and solutions moving away from the origin, and the origin is unstable. If the roots of (C.8) are complex, the orbits approach the origin if their real parts are negative. We conclude that the origin is an asymptotically stable equilibrium of (C.5) if both roots of (C.8) have negative real part and unstable if at least one root of (C.8) has positive real part.

Solution of systems of linear differential equations with constant coefficients can be carried out more economically with the use of vectors and matrices. A system

$$\begin{aligned}y'_1 &= a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\y'_2 &= a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\&\vdots \quad \vdots \quad \vdots \\y'_n &= a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n\end{aligned}$$

may be written in the form

$$y' = Ay, \quad (\text{C.16})$$

where A is an $n \times n$ matrix and y and y' are column vectors,

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

We attempt to find solutions of the form $y = e^{\lambda t}c$ with c a constant column vector. Substituting this form into (C.16) we obtain the condition

$$\lambda e^{\lambda t}c = Ae^{\lambda t}c,$$

or

$$Ac = \lambda c.$$

Thus λ be an eigenvalue of the matrix A and c must be a corresponding eigenvector. If the eigenvalues of A are distinct, this procedure generates n solutions of (C.16) and it can be shown that every solution is a linear combination of these solutions. This explains the exponential solutions that we obtained. If there are multiple

eigenvalues of A , it may be necessary to obtain additional solutions, and these will be exponential functions multiplied by powers of t . If some eigenvalues are complex, the corresponding complex exponential solutions may be replaced by products of exponential functions and sines or cosines.

C.4 Stability of Equilibria

In order to apply the linearization theorem of Sect. C.2 to questions of stability of an equilibrium, we must determine conditions under which all solutions of a linear system with constant coefficients

$$\begin{aligned}y' &= ay + bz, \\z' &= cy + dz\end{aligned}\tag{C.17}$$

approach zero as $t \rightarrow \infty$. As we saw in Sect. C.3, the nature of the solutions of (C.17) is determined by the roots of the characteristic equation

$$\lambda^2 - (a+d)\lambda + (ad - bc) = 0.\tag{C.18}$$

If the roots λ_1 and λ_2 of (C.18) are real, then the solutions of (C.17) are made up of terms $e^{\lambda_1 t}$ and $e^{\lambda_2 t}$, or $e^{\lambda_1 t}$ and $te^{\lambda_1 t}$ if the roots are equal. In order that all solutions of (C.17) approach zero, we require $\lambda_1 < 0$ and $\lambda_2 < 0$, so that the terms will be negative exponentials. If the roots are complex conjugates, $\lambda = \alpha \pm i\beta$, then in order that all solutions of (C.17) approach zero, we require $\alpha < 0$. Thus if the roots of the characteristic equation have negative real part, all solutions of the system (C.17) approach zero as $t \rightarrow \infty$. In a similar manner, we may see that if a root of the characteristic equation has positive real part, then (C.17) has unbounded solutions.

It turns out, however, that it is not necessary to solve the characteristic equation in order to determine whether all solutions of (C.17) approach zero, as there is a useful criterion in terms of the coefficients of the characteristic equation. The basic result is that the roots of a quadratic equation $\lambda^2 + a_1\lambda + a_2 = 0$ have negative real part if and only if $a_1 > 0$ and $a_2 > 0$. Applying this to the characteristic equation (C.18) and the system (C.17), we obtain the following result for linear systems with constant coefficients:

Stability Theorem for Linear Systems *Every solution of the linear system with constant coefficients (C.17) approaches zero as $t \rightarrow \infty$ if and only if the trace $a+d$ of the coefficient matrix of the system is negative and the determinant $ad - bc$ of the system's coefficient matrix is positive. If either the trace is positive or the determinant is negative, there is at least one unbounded solution.*

Example 1 Determine whether all solutions tend to zero or whether there are unbounded solutions for each of the following systems:

- (i) $y' = -y - 2z, \quad z' = y - 4z$
- (ii) $y' = z, \quad z' = -u - 2v$
- (iii) $y' = -2z, \quad z' = y + 2z$

Solution (i) The characteristic equation is $\lambda^2 + 5\lambda + 6 = 0$, with roots $\lambda = -2, \lambda = -3$. Thus all solutions tend to zero. Alternatively, since the trace of the coefficient matrix is $-5 < 0$ and the determinant is $6 > 0$, the stability theorem gives the same conclusion. For (ii), the characteristic equation is $\lambda^2 + 2\lambda + 1 = 0$ with a double root $\lambda = -1$, and thus all solutions tend to zero. For (iii), the characteristic equation is $\lambda^2 - 2\lambda + 2 = 0$, and since the trace is positive, there are unbounded solutions. As we indicated in the previous section, we could also have drawn this conclusion from a phase portrait. \square

If we apply the stability theorem for linear systems to the linearization (C.4) of a system (C.2) at an equilibrium (y_∞, z_∞) , we obtain the following result.

Equilibrium Stability Theorem *Let (y_∞, z_∞) be an equilibrium of a system $y' = F(y, z)$, $z' = G(y, z)$, with F and G twice differentiable. Then if the trace of the Jacobian matrix at the equilibrium is negative and the determinant of the Jacobian matrix at the equilibrium is positive, the equilibrium (y_∞, z_∞) is locally asymptotically stable. If either the trace is positive or the determinant is negative, then the equilibrium (y_∞, z_∞) is unstable.*

Example 2 Determine whether each equilibrium of the system

$$y' = z, \quad z' = 2(y^2 - 1)z - y$$

is locally asymptotically stable or unstable.

Solution The equilibria are the solutions of $z = 0, 2(y^2 - 1)z - y = 0$, and thus the only equilibrium is $(0,0)$. Here $F(y, z) = z$, with partial derivatives 0 and 1, respectively, and $G(y, z) = 2(y^2 - 1)z - y$, with partial derivatives $4yz - 1$ and $2(y^2 - 1)$, respectively. Therefore the Jacobian matrix at the equilibrium is

$$\begin{bmatrix} 0 & 1 \\ -1 & -2 \end{bmatrix}$$

with trace -2 and determinant 1 , as in Example 1(ii). Thus the equilibrium $(0,0)$ is locally asymptotically stable. \square

Example 3 Determine whether each equilibrium of the system

$$\begin{aligned}y' &= y(1 - 2y - z) \\z' &= z(1 - y - 2z)\end{aligned}$$

is locally asymptotically stable or unstable.

Solution The equilibria are the solutions of $y(1 - 2y - z) = 0, z(1 - y - 2z) = 0$. One solution is $(0,0)$; a second is the solution of $y = 0, 1 - y - 2z = 0$, which is $(0, \frac{1}{2})$; a third is the solution of $z = 0, 1 - 2y - z = 0$, which is $(\frac{1}{2}, 0)$; and a fourth is the solution of $1 - 2y - z = 0, 1 - y - 2z = 0$, which is $(\frac{1}{3}, \frac{1}{3})$. The Jacobian matrix at an equilibrium (y_∞, z_∞) is

$$\begin{bmatrix} 1 - 4y_\infty - z_\infty & -y_\infty \\ -z_\infty & 1 - y_\infty - 4z_\infty \end{bmatrix}.$$

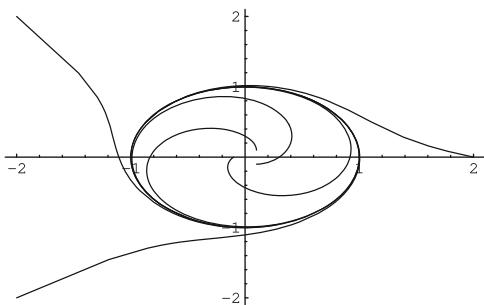
At $(0,0)$, this matrix has trace 1 and determinant 1, and thus the equilibrium is unstable. At $(0, \frac{1}{2})$, this matrix has trace $-\frac{1}{2}$ and determinant $-\frac{1}{2}$, and thus the equilibrium is unstable. At $(\frac{1}{2}, 0)$, this matrix has trace $-\frac{1}{2}$ and determinant $-\frac{1}{2}$, and thus the equilibrium is unstable. At $(\frac{1}{3}, \frac{1}{3})$, this matrix has trace $-\frac{4}{3}$ and determinant $\frac{1}{3}$, and thus this equilibrium is locally asymptotically stable. \square

The careful reader will have noticed that, like the linearization theorem of Sect. C.2, the equilibrium stability theorem given above has a hole of sorts in its result, in that the theorem says nothing about the stability of equilibria for which the trace and determinant lie on the boundary of conditions, in other words, for which the trace or the determinant of the Jacobian matrix at the equilibrium is zero, so that the linearization has solutions which do not approach zero as $t \rightarrow \infty$ but stay bounded. The reason for this “hole” is that in such cases, the linearization does not give enough information to determine stability.

If all orbits beginning near an equilibrium remain near the equilibrium for $t \geq 0$, but some orbits do not approach the equilibrium as $t \rightarrow \infty$, the equilibrium is said to be *stable*, or sometimes *neutrally stable*. If the origin is neutrally stable for the linearization at an equilibrium, then the equilibrium may also be neutrally stable for the nonlinear system. However, it is also possible for the origin to be neutrally stable for the linearization at an equilibrium, while the equilibrium is asymptotically stable or unstable. Thus neutral stability of the origin for a linearization at an equilibrium gives no information about the stability of the equilibrium.

We have seen in Sect. B.5 that a solution of an autonomous first-order differential equation is either unbounded or approaches a limit as $t \rightarrow \infty$. For an autonomous system of two first-order differential equations, these same two possibilities exist. In addition, however, there is the possibility of an orbit which is a closed curve, corresponding to a periodic solution. Such an orbit is called a *periodic orbit* because it is traversed repeatedly.

Fig. C.1 Trajectories approaching a limit cycle



There is a remarkable result which says essentially that these are the only possibilities.

Poincaré-Bendixson Theorem *A bounded orbit of a system of two first-order differential equations which does not approach an equilibrium as $t \rightarrow \infty$ either is a periodic orbit or approaches a periodic orbit as $t \rightarrow \infty$.*

It is possible to show that a periodic orbit must enclose an equilibrium point in its interior. In many examples, there is an unstable equilibrium, and orbits beginning near this equilibrium spiral out towards a periodic orbit. A periodic orbit which is approached by other (non-periodic) orbits is called a *limit cycle*. One example of a limit cycle involves the system

$$\begin{aligned} y' &= y(1 - y^2 - z^2) - z, \\ z' &= z(1 - y^2 - z^2) + y, \end{aligned}$$

which has its only equilibrium at the origin. The equilibrium is unstable, and Fig. C.1 illustrates the fact that all orbits not beginning at the origin spiral counterclockwise [in or out] toward the unit circle $y^2 + z^2 = 1$.

In many applications the functions $y(t)$ and $z(t)$ are restricted by the nature of the problem to non-negative values. For example, this is the case if $y(t)$ and $z(t)$ are population sizes. In such a case, only the first quadrant $y \geq 0, z \geq 0$ of the phase plane is of interest. For a system

$$y' = F(y, z), \quad z' = G(y, z)$$

which has $F(0, z) \geq 0$ for $z \geq 0$ and $G(y, 0) \geq 0$ for $y \geq 0$, then since $y' \geq 0$ along the positive z -axis (where $y = 0$) and $z' \geq 0$ along the positive y -axis (where $z = 0$), no orbit can leave the first quadrant by crossing one of the axes. The Poincaré–Bendixson theorem may then be applied to orbits in the first quadrant. In such a case, the first quadrant is called an *invariant set*, a region with the property that orbits must remain in the region.

If instead F and G are identically zero along the respective [half-]axes, then $y' = 0$ for $\{y = 0, z \geq 0\}$, and $z' = 0$ for $\{y \geq 0, z = 0\}$. In this case, orbits which

begin on an axis must remain on that axis, and orbits beginning in the interior of the first quadrant (with $y(0) > 0, z(0) > 0$) must remain in the interior of the first quadrant (i.e., $y(t) > 0$ and $z(t) > 0$ for $t \geq 0$). If there is no equilibrium in the first quadrant, there cannot be a periodic orbit, because a periodic orbit *must* enclose an equilibrium. Thus, if there is no equilibrium in the first quadrant every orbit must be unbounded.

Example 4 Show that every orbit in the region $y > 0, z > 0$ of the system

$$\begin{aligned}y' &= y(2 - y) - \frac{yz}{y+1}, \\z' &= 4 \frac{yz}{y+1} - z\end{aligned}$$

approaches a periodic orbit as $t \rightarrow \infty$.

Solution We have $y' = 0$ when $y = 0$, and $z' = 0$ when $z = 0$, so orbits starting in the first quadrant remain in the first quadrant. Equilibria are the solutions of either $y = 0$ or $2 - y = \frac{z}{y+1}$, and either $z = 0$ or $\frac{4y}{y+1} = 1$. If $y = 0$, we must also have $z = 0$. If $2 - y = \frac{z}{y+1}$, we could have $z = 0$, which implies $y = 2$, or $y = \frac{1}{3}$, which implies $z = \frac{20}{9}$. Thus there are three equilibria, namely $(0,0)$, $(2,0)$, and $(\frac{1}{3}, \frac{20}{9})$. By checking the values of the trace and determinant of the Jacobian matrix, which is

$$\begin{bmatrix} 2 - 2y_\infty - \frac{z_\infty}{(1+y_\infty)^2} & -\frac{y_\infty}{y_\infty+1} \\ \frac{4z_\infty}{(y_\infty+1)^2} & \frac{4y_\infty}{y_\infty+1} - 1 \end{bmatrix},$$

we may see that each of the three equilibria is unstable. In order to apply the Poincaré–Bendixson theorem, we must show that every orbit starting in the first quadrant of the phase plane is bounded.

To show this, we might like to show that y' and z' are negative when y and/or z are sufficiently large, but a glance at the equations tells us this isn't necessarily so. Therefore we instead consider some positive combination of y and z whose time derivative does become negative far enough from the origin. In particular, consider the function $V(y, z) = 4y + z$. If an orbit is unbounded, then along this orbit the function $V(y, z)$ must also be unbounded. The derivative of $V(y, z)$ along an orbit is

$$\frac{d}{dt} V[y(t), z(t)] = 4y'(t) + z'(t) = 4y(2 - y) - z.$$

This is negative except in the bounded region defined by the inequality $z < 4y(2 - y)$. Therefore the function $V(y, z)$ cannot become unbounded, because it is decreasing ($dV/dt < 0$) whenever it becomes large ($z > 4y(2 - y)$), which is true, for example, whenever $V > 9$). This proves that all orbits of the system are bounded. Now we may apply the Poincaré–Bendixson theorem to see that every orbit approaches a limit cycle. \square

C.5 Qualitative Behavior of Solutions of Linear Systems

In Sect. C.2 we reduced the analysis of the stability of an equilibrium (x_∞, y_∞) of a system of differential equations

$$x' = F(x, y), \quad y' = G(x, y)$$

to the determination of the behavior of solutions of the linearization at the equilibrium

$$\begin{aligned} u' &= F_x(x_\infty, y_\infty)u + F_y(x_\infty, y_\infty)v, \\ v' &= G_x(x_\infty, y_\infty)u + G_y(x_\infty, y_\infty)v. \end{aligned} \tag{C.19}$$

Next we shall analyze the various possibilities for the behavior of solutions of the two-dimensional linear homogeneous system with constant coefficients

$$\begin{aligned} x' &= ax + by, \\ y' &= cx + dy, \end{aligned} \tag{C.20}$$

where a, b, c , and d are constants, in order that we might describe the behavior of solutions of the linearization at an equilibrium. We will then be able to state some refinements of the linearization theorem of Sect. C.3 that give more specific information about the behavior of solutions near an equilibrium as determined by the Jacobian matrix at the equilibrium. We will assume throughout that $ad - bc \neq 0$. This implies that the origin is the only equilibrium of the system (C.20). If this system is the linearization at an equilibrium (x_∞, y_∞) of a nonlinear system (C.19), then this equilibrium is *isolated*, meaning that there is a disc centered at (x_∞, y_∞) containing no other equilibrium of the nonlinear system. It is convenient to use vector-matrix notation: We let \mathbf{x} denote the column vector $\begin{bmatrix} x \\ y \end{bmatrix}$, \mathbf{x}' the column vector $\begin{bmatrix} x' \\ y' \end{bmatrix}$, and A the 2×2 matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Then, using the properties of matrix multiplication, we may rewrite the linear system $x' = ax + by$, $y' = cx + dy$ in the form

$$\mathbf{x}' = A\mathbf{x}.$$

A linear change of variable $\mathbf{x} = P\mathbf{u}$, with P a nonsingular 2×2 matrix (which represents a rotation of the axes and a change of scale along the axes), transforms the system to $P\mathbf{u}' = AP\mathbf{u}$, or

$$\mathbf{u}' = P^{-1}AP\mathbf{u} = B\mathbf{u}.$$

This is of the same type as the original system $\mathbf{x}' = A\mathbf{x}$, and its coefficient matrix $B = P^{-1}AP$ is *similar* to A . If we can solve for \mathbf{u} , we can then reconstruct $\mathbf{x} = P\mathbf{u}$, and in fact the qualitative properties of solutions for \mathbf{u} are preserved in this reconstruction. Thus, we may describe the various possible phase portraits of $\mathbf{x}' = A\mathbf{x}$ by listing the various possible canonical forms of A under similarity and constructing the phase portrait for each possibility.

Canonical Form Theorem *The matrix*

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

with $\det A = ad - bc \neq 0$, is similar under a real transformation to one of

- (i) $\begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix}$, $\lambda > \mu > 0$ or $\lambda < \mu < 0$.
- (ii) $\begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$, $\lambda > 0$ or $\lambda < 0$.
- (iii) $\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$, $\lambda > 0$ or $\lambda < 0$.
- (iv) $\begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix}$, $\lambda > 0 > \mu$
- (v) $\begin{bmatrix} 0 & \beta \\ -\beta & 0 \end{bmatrix}$, $\beta \neq 0$.
- (vi) $\begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}$, $\alpha > 0$, $\beta \neq 0$ or $\alpha < 0$, $\beta \neq 0$.

We now describe the phase portraits for each of these cases in turn.

Case (i) The transformed system is $u' = \lambda u$, $v' = \mu v$, with solution $u = u_0 e^{\lambda t}$, $v = v_0 e^{\mu t}$. If $\lambda < \mu < 0$ then u and v both tend to zero as $t \rightarrow \infty$ and $v/u = v_0 e^{(\mu-\lambda)t}/u_0 \rightarrow +\infty$. Thus, every orbit tends to the origin with infinite slope (except if $v_0 = 0$, in which case the orbit is on the u -axis), and the phase portrait is as shown in Fig. C.2. If $\lambda > \mu > 0$ the portrait is the same, except that the arrows are reversed.

Case (ii) The system is $u' = \lambda u$, $v' = \lambda v$, with solution $u = u_0 e^{\lambda t}$, $v = v_0 e^{\lambda t}$. If $\lambda < 0$ both u and v tend to zero as $t \rightarrow \infty$ and $v/u = v_0/u_0$. Thus every orbit is a straight line going to the origin, and all slopes as the orbit approaches the origin are possible. The phase portrait is as shown in Fig. C.3. If $\lambda > 0$, the portrait is the same except that the arrows are reversed.

Case (iii) The system is $u' = \lambda u + v$, $v' = \lambda v$. The solution of the second equation is $v = v_0 e^{\lambda t}$, and substitution into the first equation gives the first-order linear equation $u' = \lambda u + v_0 e^{\lambda t}$ whose solution is $u = (u_0 + v_0 t) e^{\lambda t}$. If $\lambda < 0$, both u and v tend to zero as $t \rightarrow \infty$, and $v/u = v_0/(u_0 + v_0 t)$ tends to zero unless $v_0 = 0$, in

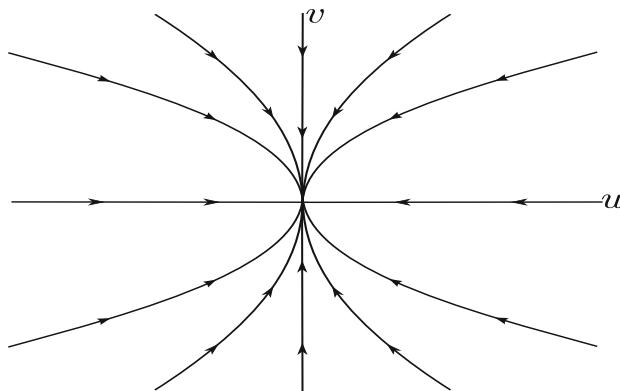


Fig. C.2 Case (i)

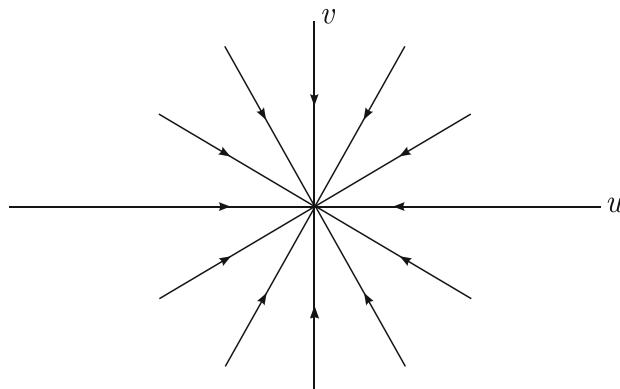


Fig. C.3 Case (ii)

which case the orbit is the u -axis. Because $u = (u_0 + v_0 t)e^{\lambda t}$, we have $du/dt = ((u_0\lambda + v_0) + \lambda v_0 t)e^{\lambda t}$, and therefore $du/dt = 0$ when $t = -(v_0 + u_0\lambda)/\lambda v_0$. Thus except for the orbits on the u -axis, every orbit has a maximum or minimum u -value and then turns back toward the origin. The phase portrait is as shown in Fig. C.4. If $\lambda > 0$, the portrait is the same except that the arrows are reversed.

Case (iv) The solution is $u = u_0 e^{\lambda t}$, $v = v_0 e^{\mu t}$ just as in Case (i), but now u is unbounded and $v \rightarrow 0$ as $t \rightarrow \infty$, unless $u_0 = 0$, in which case the orbit is on the v -axis. The phase portrait is as shown in Fig. C.5.

Case (v) The system is $u' = \beta v$, $v' = -\beta u$. Then $u'' = \beta v' = -\beta^2 u$ and $u = A \cos \beta t + B \sin \beta t$ for some A, B . Thus $v = u'/\beta = -A \sin \beta t + B \cos \beta t$ and $u^2 + v^2 = A^2 + B^2$. Every orbit is a circle, clockwise if $\beta > 0$ and counterclockwise if $\beta < 0$. The phase portrait for $\beta > 0$ is as shown in Fig. C.6.

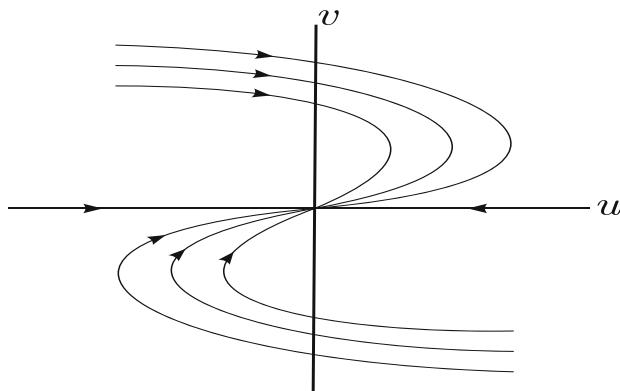


Fig. C.4 Case (iii)

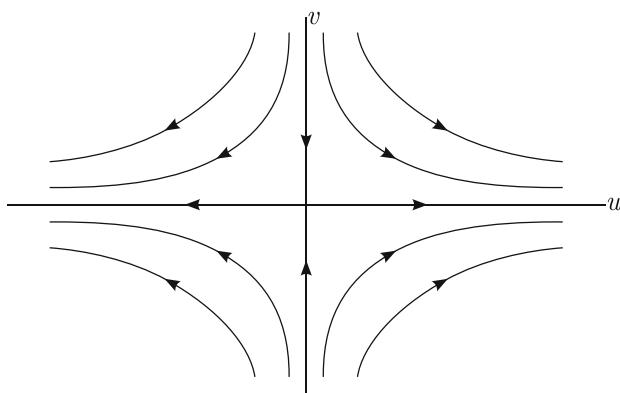


Fig. C.5 Case (iv)

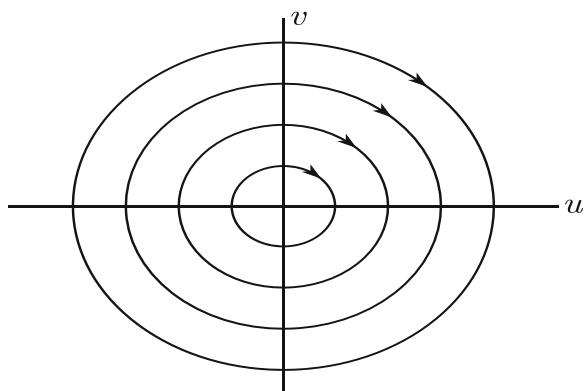


Fig. C.6 Case (v)

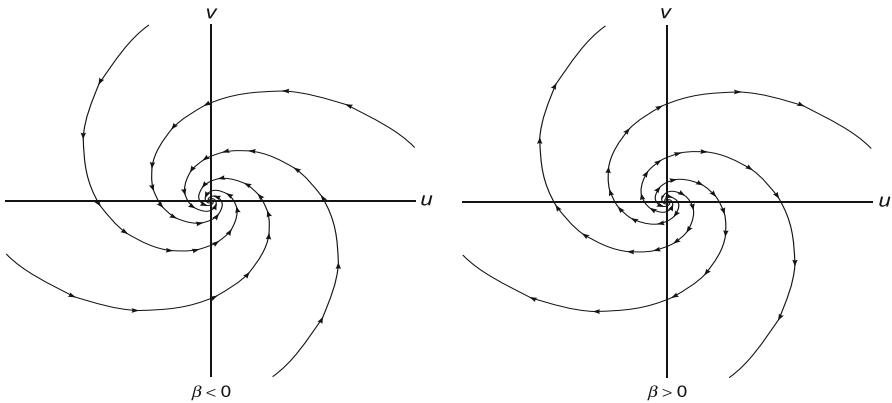


Fig. C.7 Case (vi)

Case (vi) The system is $u' = \alpha u + \beta v$, $v' = -\beta u + \alpha v$. We make the change of variables $u = e^{\alpha t} p$, $v = e^{\alpha t} q$, so that $u' = \alpha e^{\alpha t} p + e^{\alpha t} p'$, $v' = \alpha e^{\alpha t} q + e^{\alpha t} q'$ reduces the system to $p' = \beta q$, $q' = -\beta p$, which has been solved in Case (v). Thus, $u = e^{\alpha t}(A \cos \beta t + B \sin \beta t)$, $v = e^{\alpha t}(-A \sin \beta t + B \cos \beta t)$, and $u^2 + v^2 = e^{2\alpha t}(A^2 + B^2)$. If $\alpha < 0$, $u^2 + v^2$ decreases exponentially, and the orbits are spirals inward to the origin, clockwise if $\beta > 0$ and counterclockwise if $\beta < 0$. The phase portrait is as shown in Fig. C.7. If $\alpha > 0$, the portraits are the same except that the arrows are reversed.

These six cases may be classified as being of four distinct types. In Cases (i), (ii), and (iii) all orbits approach the origin as $t \rightarrow +\infty$ (or as $t \rightarrow -\infty$ depending on the signs of λ and μ) with a limiting direction, and the origin is said to be a *node* of the system. In Case (iv), only two orbits approach the origin as $t \rightarrow +\infty$ or as $t \rightarrow -\infty$, and all other orbits move away from the origin. In this case the origin is said to be a *saddle point*. In Case (vi) every orbit winds around the origin in the sense that its angular argument tends to $+\infty$ or to $-\infty$, and the origin is said to be a *vortex*, *spiral point*, or *focus*. In Case (v), every orbit is periodic; in this case the origin is said to be a *center*. According to the equilibrium stability theorem of Sect. C.4, asymptotic stability of the origin for the linearization implies asymptotic stability of an equilibrium of a nonlinear system. In addition, instability of the origin for the linearization implies instability of an equilibrium of a nonlinear system. The asymptotic stability or instability of the origin for a linear system is determined by the *eigenvalues* of the matrix A , defined to be the roots of the *characteristic equation*

$$\begin{aligned}\det(A - \lambda I) &= \det \begin{bmatrix} a - \lambda & b \\ c & d - \lambda \end{bmatrix} = (a - \lambda)(d - \lambda) - bc \\ &= \lambda^2 - (a + d)\lambda + (ad - bc) = 0.\end{aligned}$$

The sum of the eigenvalues is the *trace* of the matrix A , namely $a + d$, and the product of the eigenvalues is the determinant of the matrix A , namely $ad - bc$. A similarity transformation preserves the trace and determinant and therefore does not change the eigenvalues. Thus, the eigenvalues of A are λ, μ in Cases (i) and (iv); λ (a double eigenvalue) in Cases (ii) and (iii); the complex conjugates $\pm i\beta$ in Case (v); and $\alpha \pm i\beta$ in Case (vi). Examination of the phase portraits in the various cases shows that the origin is asymptotically stable in Case (i) if $\lambda < \mu < 0$; in Cases (ii) and (iii) if $\lambda < 0$; and in Case (vi) if $\alpha < 0$. Similarly the origin is unstable in Case (i) if $\lambda > \mu > 0$; in Cases (ii) and (iii) if $\lambda > 0$; in Case (iv); and in Case (vi) if $\alpha > 0$. The origin is stable but not asymptotically stable in Case (v) (center). A simpler description is that the origin is asymptotically stable if both eigenvalues have negative real part and unstable if at least one eigenvalue has positive real part. If both eigenvalues have real part zero, the origin is stable but not asymptotically stable. Our assumption that $ad - bc = \det A \neq 0$ rules out the possibility that $\lambda = 0$ is an eigenvalue; thus, eigenvalues with real part zero can occur only if the eigenvalues are pure imaginary, as in Case (v).

Combining this analysis with the linearization result, we have the following result (which is the equilibrium stability theorem of Sect. C.4 with some added details):

Stability Theorem *If (x_∞, y_∞) is an equilibrium of the system (C.19) and if all eigenvalues of the coefficient matrix of the linearization at this equilibrium have negative real part, specifically if*

$$\operatorname{tr} A(x_\infty, y_\infty) = F_x(x_\infty, y_\infty) + G_y(x_\infty, y_\infty) < 0,$$

$$\det A(x_\infty, y_\infty) = F_x(x_\infty, y_\infty)G_y(x_\infty, y_\infty) - F_y(x_\infty, y_\infty)G_x(x_\infty, y_\infty) > 0,$$

then the equilibrium (x_∞, y_∞) is asymptotically stable.

We can be more specific about the nature of the orbits near an equilibrium. In terms of the elements of the matrix A , we may characterize the cases as follows, using the remark that the eigenvalues are complex if and only if

$$\Delta = (a + d)^2 - 4(ad - bc) = (a - d)^2 + 4bc < 0.$$

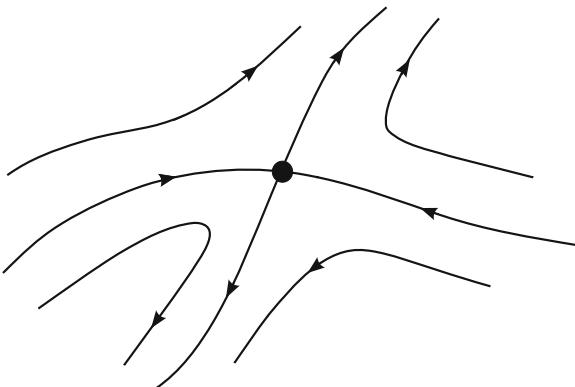
1. If $\det A = ad - bc < 0$, the origin is a saddle point.
2. If $\det A > 0$ and $\operatorname{tr} A = a + d < 0$, the origin is asymptotically stable, a node if $\Delta \geq 0$ and a spiral point if $\Delta < 0$.
3. If $\det A > 0$ and $\operatorname{tr} A > 0$, the origin is unstable, a node if $\Delta \geq 0$ and a spiral point if $\Delta > 0$.
4. If $\det A > 0$ and $\operatorname{tr} A = 0$, the origin is a center.

It is possible to show that in general, the phase portrait of a nonlinear system at an equilibrium is similar to the phase portrait of the linearization at the equilibrium, except possibly if the linearization has a center. This is true under the assumption that the functions $F(x, y)$ and $G(x, y)$ in the system (C.19) are smooth enough that Taylor's theorem is applicable, so that the terms neglected in the linearization

process are of higher order. If the linearization at an equilibrium has a *node*, then the equilibrium of the nonlinear system is also a node, defined to mean that every orbit tends to the equilibrium (either as $t \rightarrow \infty$ or as $t \rightarrow -\infty$) with a limiting direction. If the linearization at an equilibrium has a *spiral point*, then the equilibrium of the nonlinear system is also a spiral point, defined to mean that every orbit tends to the equilibrium (either as $t \rightarrow \infty$ or as $t \rightarrow -\infty$) with its angular variable becoming infinite. If the linearization at an equilibrium has a *saddle point*, then the equilibrium of the nonlinear system is also a saddle point. A saddle point is defined by the characterization that there is a curve through the equilibrium such that orbits starting on this curve tend to the equilibrium but orbits starting off this curve cannot stay near the equilibrium. An equivalent formulation is that there are two orbits tending to the equilibrium as $t \rightarrow +\infty$ and two orbits tending away from the equilibrium, or tending to the equilibrium as $t \rightarrow -\infty$. These orbits are called *separatrices*; the two orbits tending to the saddle point are the stable separatrices, while the two orbits tending away from the saddle point are the unstable separatrices. Other orbits appear like hyperbolas (Fig. C.8).

A *center* is defined to be an equilibrium for which there is an infinite sequence of periodic orbits around the equilibrium with the orbits approaching this equilibrium. If the linearization at an equilibrium has a center, then the equilibrium of the nonlinear system *may* be a center, but is not necessarily a center: It could be an asymptotically stable spiral point or an unstable spiral point.

Fig. C.8 Separatrices at a saddle point



Index

A

- Absorbing boundary, 466
Accelerated progression, 267
Acquired immunity, 394
Acquired resistance, 251
Active dynamics, 274
Activity level, 189, 190, 210, 211, 477
Adjoint equation, 258
Adjoint problem, 258
Adverse effect of control, 161
Aedes aegypti, 409, 410, 414
Age, 14
 cohort, 69
 at infection, 69, 71
 of infection, 6, 10, 37, 139–141, 143, 145, 149, 153, 155, 156, 168–171, 175, 195, 209, 210, 269, 273, 281, 347, 452, 464
 profile, 331
 structure, 15, 265, 344, 429
Age dependent birth rate, 449
Agent-based model, 362, 384
Agents
 bacterial, 23
 helminth, 23
Age progression rate, 447
Age specific disease induced death rate, 447
Age specific natural death rate, 447
Age specific recovery rate, 447
Age-structured models, 15
Age-structured populations, 15
AIDS, 10, 23, 273–276, 279, 282, 283, 286, 302, 303
Airline network, 461
Airline travel, 461
Air travel, 457
Algebraic equation, 539, 585
Algorithm, 549
Amplitude, 34, 71, 72
Antibiotic, 250
 resistance, 256, 511
 resistant strain, 256
Antigenically distinct strain, 347
Antigenic drift, 323
Antigenic shift, 326
Antimalarial therapy, 405
Antimicrobial resistance, 511
Antiviral drug, 159, 312, 319, 322, 511
Antiviral efficacy, 322
Antiviral treatment, 221, 312, 319, 320, 322, 334, 510, 511
Approximate exponential growth, 356
Arbitrary constant, 567
Arbitrary distribution, 376
Arbitrary stage distribution, 367
Arbitrary stage duration distribution, 162, 163
Arbovirus, 414
Aristotle, 5
Assessment, 338
Asymptomatic, 11, 319, 347, 413, 414, 419
 case, 315
 compartment, 313
 infection, 313
 infective, 217
 period, 312, 314
 stage, 133, 195
Asymptotically stable, 573, 582
Asymptotic behavior, 26, 574
Asymptotic spread, 464

- Asymptotic stability, 11, 12, 26, 29, 31–33, 49, 51, 55, 63, 64, 67, 68, 74, 75, 77–82, 87, 88, 101, 109, 113, 130, 182, 183, 185, 186, 223, 234, 238, 243, 254, 255, 264, 265, 269, 278, 284, 288, 304, 339, 340, 344, 345, 393, 400, 402, 404, 406, 436, 444, 445, 449, 450, 459, 460, 468, 471, 472, 520, 573–577, 582–584, 592, 594, 595, 602, 603
- global, 64, 583
 - local, 583
- Attack rate, 37, 314, 315, 318
- symptomatic, 323
- Attack ratio, 195
- Augmented matrix, 544
- Autonomous, 563, 570, 571, 579, 580, 595
- B**
- Backward bifurcation, 263, 265, 511
- Bacteria, 22
- Balance equation, 222, 465
- Balance relation, 234
- Baseline value, 315
- Basic reproduction number, 6, 8, 10–12, 14, 22, 26, 28, 33, 39, 44, 53, 63, 65, 68, 71, 74, 78, 83, 112, 117, 120, 123, 125, 132, 135, 140, 141, 146, 148, 151–153, 156, 168, 169, 171, 180, 182, 184, 187, 191, 192, 196, 211–214, 223, 232, 234, 235, 243, 275, 277, 280, 283, 304, 305, 313–315, 317, 322, 325, 327, 339, 344, 347, 348, 351, 352, 361, 380, 385, 387, 392–396, 400, 406, 411–417, 422, 449, 450, 452, 453, 460, 478, 479, 484, 489, 492, 496–498
- Behavior, 161
- Behavioral change, 14, 384
- Behavioral class, 29
- Behavioral heterogeneity, 429
- Behavioral response, 334
- Bernouilli, 5
- Berra, Yogi, 508
- Between-host system, 108, 109
- Bifurcation, 81, 338, 460
- backward, 12, 81, 82, 85–87
 - curve, 12, 81, 87, 88
 - diagram, 341, 403, 404
 - forward, 12, 81, 82
 - parameter, 275, 339, 403
 - point, 340, 341
 - transcritical, 12
- Bilinear incidence, 43
- Birth defects, 13, 418
- Birth modulus, 437
- Birth rate, 100, 404, 479
- density dependent, 31, 400
 - maximum, 403
- Bi-stability, 404
- Bite, 231
- Black Death, 3, 337
- Blood meal, 229
- Boosting of immunity, 396, 397
- Boundary condition, 448, 466
- Boundary equilibrium, 254, 255, 288
- asymptotically stable, 344
- Boundary layer, 238, 239, 243
- Branching process, 7, 8, 118, 129
- Broad Street pump, 5
- Bubonic plague, 41, 100, 102
- Budd William, 5
- Burial, 360
- C**
- Carbon dating, 557
- Carrier, 406, 407
- Carrying capacity, 28, 72, 82
- Case, 352
- data, 336
 - fatality rate, 385
 - fatality ratio, 45
- Case-finding, 251, 257
- Case-holding, 251, 257, 260
- Cell mortality, 108
- Cell–parasite interaction, 107
- Census data, 102
- Center, 602, 604
- manifold, 341
- Chagas’ disease, 13, 67, 229
- Chain binomial model, 8
- Changing behavior, 274
- Characteristic equation, 33, 75, 76, 130, 133, 137, 233, 352, 417, 436, 549, 585, 587–589, 591–594, 602
- Characteristics, 437
- Chemoprophylaxis, 262
- Chicken pox, 15, 22, 71, 429
- Chikungunya, 13, 229, 414
- Cholera, 3, 5, 11, 21, 107, 167
- Clinical attack rate, 315
- Clinical case, 314, 315
- Coefficient matrix, 32, 33
- Coevolutionary interaction, 346
- Coexistence, 254, 256, 345, 348
- equilibrium, 254

- multiple strain, 347
region, 254, 345
steady state, 269
sub-threshold, 347
- Cohort, 25
Coinfection, 266, 267, 284, 334, 347, 510
Communicable disease, 478
Compartment, 117
Compartmental diagram, 284
Compartmental model, 6–9, 11, 22–24, 26, 35, 81, 137, 141, 182, 189, 195, 312, 331
Compartmental structure, 23, 26, 64, 144, 155, 230
Competing disease strains, 49
Competing risks, 4
Competitive exclusion principle, 347
Compliance, 250, 252
Computer network, 7, 118
Conditional probability, 304
Condom use, 273
Configuration model, 126
Conservation law, 465
Constant coefficients, 569, 570, 581, 584, 592, 593
Constant solution, 566, 567, 570, 571, 580
Constraint, 222, 266, 367, 378, 449
Contact, 7, 44, 118, 131, 139, 161, 188, 189, 196, 201, 216, 219, 276, 281, 286, 304, 313, 316, 319, 404, 460, 477
effective, 24
intervention, 125
matrix, 196
network, 7, 8, 118, 123, 124, 312
pattern, 118
random, 23, 24
tracing, 134, 216, 510
- Contact rate, 15, 31, 39, 69, 72, 151, 156, 189, 201, 204, 208, 210, 212, 216, 221, 242, 274, 322, 332, 333, 385, 386, 458, 486
pandemic, 327
variable, 332
- Contact tracing, 360, 361
Contagion, 509
Contagious agent, 337
Continuous time model, 46
Control, 159, 188, 259, 318, 380, 485, 491
effort, 256
measure, 9, 14, 97, 134–136, 151, 165, 166, 188, 315, 351, 360, 381, 490, 494, 509
parameter, 97, 136, 148
program, 259, 263, 351, 500
- reproduction number, 135, 148, 163, 165, 166, 180, 188, 196, 197, 199, 261, 287, 317, 322, 324, 379, 380, 382
strategy, 97, 216, 386
- Cordon sanitaire, 386, 387, 481, 483, 484
Core group, 14, 188
Cost, 338
Coupled processes, 400
Coupling strength, 458
Critical mass, 509
Critical point, 346
Critical threshold, 394
Critical transmissibility, 125
Cross-immunity, 11, 111, 264, 323–328, 338, 344–347, 409, 510
group, 326
model, 324
- D**
- Damping, 72
Data, 362
Death, 364, 365, 372, 379
modulus, 437
rate, 100, 282, 404, 405
natural, 72
per capita, 403
- Decay rate, 553
Degree distribution, 118, 122, 124, 126
Delta function, 201, 202, 466
Demographic effects, 6, 27, 117, 312, 511
Demographic equation, 51
Demographic equilibrium, 51
Demographic model, 459
Demographic parameter, 400
Demographic process, 28
Demographics, 25, 63, 230
Demographic steady state, 277
Demographic time scale, 72
Dengue, 229, 230, 409, 414, 421, 491, 501
fever, 13
hemorrhagic fever, 409
pathogenicity, 409
shock syndrome, 409
spread, 422
- Density dependence, 436
Density-dependent birth rate, 29
Determinant, 32, 33, 67, 74, 88, 133, 188, 546, 585, 590, 593–595, 603
Deterministic model, 8, 23, 42, 122, 362
Diagnosis rate, 418
Diagonalization, 551
Diagonal matrix, 543, 551
Diarrhea, 3

- Difference equation, 121
 Differential-difference equation, 10, 24, 75
 Differential equation, 10, 24
 Differential infectivity, 134
 Diffusion, 467, 468, 471, 509
 coefficient, 465, 469
 instability, 471
 Diffusivity, 465
 Diphtheria, 71
 Direct contact rate, 419
 Direct host to host transmission, 414
 Direction field, 561, 562, 580
 Direct transmission, 414
 number, 415
 Discrete model, 217
 Discrete time system, 47
 Disease case, 321
 disease case, 331
 Disease compartment, 182–184
 Disease control, 337, 338
 Disease death, 31, 43, 315, 322
 Disease death rate, 44, 45, 100, 314, 469
 Disease debilitation, 21
 Disease dynamics, 407
 Disease-free equilibrium, 27
 Disease-induced death, 359
 Disease-induced mortality, 363
 Disease intervention, 96
 Disease mortality, 21, 100, 276, 380, 405
 Disease outbreak, 7, 23, 319, 334, 386
 Disease prevalence, 97, 252, 289, 406, 478, 511
 Disease progression, 182
 Disease progression rate, 448
 Disease stage, 218
 duration, 89
 Disease state, 187
 Disease strain, 510
 Disease survival rate, 314
 Disease symptoms, 315
 Disease transmission, 7, 14, 15, 166, 217, 273,
 275, 351, 394, 418, 477
 sexual, 418
 Dispersion coefficients, 52
 Distribution of stay, 117, 168, 209
 Distribution time of waiting time, 365
 Domain of attraction, 81, 88
 Drinking, 274
 Drug, 360
 injection, 286
 resistance, 334, 510, 511
 treatment, 252, 256
 compliance, 256
 use, 273
 Drug-resistant strain, 11, 251, 252
 Drug-sensitive strain, 251, 252, 266, 270
 Duration, 365, 381
 Dynamic behavior, 275, 365
 Dynamics of DENV-2, 422
- E**
 Early detection, 358
 Early diagnosis, 359
 Early outbreak data, 382
 Ebola, 6, 351, 352, 356, 359, 362, 380,
 382–384, 386, 387, 463, 480, 481,
 484, 501
 control, 358
 Ecology, 398
 Economic loss, 338
 Edge, 7, 8, 118, 459
 Effective contact, 131, 140, 142, 144
 Effective human population size, 242
 Effective population size, 485, 492
 Effective reproduction number, 204, 206, 224,
 351, 356, 359
 Effective sub-population, 204
 Efficacy, 337
 Eigenvalue, 32, 67, 71, 75, 84, 101, 130, 133,
 137, 184–186, 190, 191, 194, 205,
 215, 232, 233, 324, 400, 401, 406,
 416, 417, 463, 473, 548, 587, 592,
 602, 603
 dominant, 353
 double, 603
 Eigenvector, 184, 185, 215, 587, 590, 592
 Endemic, 21, 22, 230, 241, 249, 287, 332, 394
 disease, 11, 21, 26
 Environmental contamination, 107, 108
 Environmental risk, 477, 478
 Epidemic, 6, 14, 22, 23, 36, 44, 73, 117, 122,
 123, 133, 140, 159, 179, 211, 216,
 230, 312, 318, 331–333, 459, 471
 control, 359
 curve, 381
 death rate, 45
 duration, 173
 final size, 117, 159, 160, 164, 165, 173,
 174, 212, 214, 216, 352, 354, 381,
 419, 481–484, 494–498, 500, 511
 growth, 383
 outbreak, 351
 peak, 331
 size, 159, 160
 time, 159, 160
 recurrent, 102, 347
 recurring, 71
 of rumors, 56

- size, 419
spread, 457
threshold, 359
wave, 330, 332, 509
- Epidemic model, 539
discrete-time, 162
with general exposed and infectious periods, 146
with general quarantine/isolation, 146
with quarantine and isolation, 146, 162
with time-dependent infection rate, 159
- Epidemiological class, 368
- Epidemiological data, 102
- Epidemiological model, 539
- Epidemiological time scale, 72
- Equilibrium, 11, 26, 31–33, 36, 63, 72, 100, 109, 112, 121, 130, 133, 234, 253, 280, 283, 344, 384, 393, 458, 468, 571–575, 580–585, 591, 594–596
age distribution, 438
analysis, 26
asymptotically stable, 32–34, 64, 67, 68, 73, 77, 79, 82, 88, 101, 183, 185, 186, 269, 277, 278, 284, 402, 449, 459, 468, 472, 573, 583, 592
asymptotically stable boundary, 404
asymptotically stable disease-free, 234, 243, 265, 288, 304, 339, 400
asymptotically stable endemic, 12, 234, 243, 265, 281, 304, 340
asymptotically stable interior, 288, 404
boundary, 112, 345
coexistence, 290
disease-free, 7, 11, 32, 63, 64, 66, 68, 74–76, 79, 82, 84, 91, 130, 136, 182–186, 217, 223, 233, 264, 275, 277, 278, 284, 324, 352, 385, 392, 393, 417, 449, 450, 453, 458, 460
endemic, 7, 11, 28, 32–34, 63, 66, 67, 70, 71, 73–75, 77, 79, 84, 87, 91, 95, 101, 223, 252, 264, 265, 275, 278, 284, 339, 340, 392, 393, 401, 452, 458, 511
globally asymptotically stable, 478, 479
globally stable disease-free, 478
interior, 288–290
isolated, 581, 598
locally asymptotically stable, 595
mosquito population, 410
positive, 29, 265, 281
stable, 340, 345, 595
stable boundary endemic, 345
stable endemic, 263
stable non-trivial, 344
- stability theorem, 577, 594, 595
temporary, 356
unique asymptotically stable interior, 404
unique disease-free, 338
unstable, 77, 186, 278, 284, 288, 340, 573, 583, 595, 596
unstable endemic, 78, 81
unstable interior, 404
- Estimated rate of exponential rise, 358
- Estimated rate of rise of exponential fit, 358
- Estimation of initial growth, 351
- Eulerian perspective, 386, 461, 509
- EVD, 351, 354, 356–358, 360, 386
- Evolution, 398
- Excess degree, 119–121
distribution, 124
- Existence and uniqueness theorem, 560, 572
- Exogenous reinfection, 263, 264
- Experimental data, 334
- Explicit movement, 458
- Exponential distribution, 88, 90, 93, 99, 137, 150, 152, 158, 218, 367, 369–372, 376, 379
- Exponential growth, 30, 101, 102, 353, 356, 382, 385, 386, 555
- Exponentially distributed infective period, 146, 153
- Exponentially distributed period, 145, 156, 158
- Exponentially distributed stage, 167
- Exponentially distributed waiting time, 276
- Exponential rise, 357, 358
of epidemic curve, 356
- Exponential waiting time, 364
distribution, 278
- Exposed, 23, 37, 40, 64, 89–93, 95, 104, 129, 131, 135, 141, 142, 144, 146, 148, 154–156, 158, 159, 162–164, 170, 172–174, 182, 186, 189, 217, 230, 242, 275, 330, 337, 338, 347, 352, 361, 362, 395, 409, 420, 447, 448, 478, 481
period, 189
period distribution, 155
- Extinction, 252, 460
- F**
- Fall wave, 335, 336
peak, 336
- Family of functions, 558
- Family of solutions, 555
- Farr, William, 5
- Fast dynamics, 400
- Fast progression, 261, 268

- Fast system, 402
 Fast time scale, 400, 406
 Fast variable, 108
 Fatality rate, 330
 Fick's law, 465, 466
 Final compartment size, 326
 Final epidemic size, 174
 Final size, 241, 347, 348, 387
 of an heterogeneous mixing epidemic, 212
 equation, 318, 321
 inequality, 45
 relation, 10, 37, 39, 40, 132, 134, 137,
 139, 141, 180, 181, 186, 194, 195,
 212–214, 216, 313, 314, 318, 325,
 327, 328, 347, 387, 422, 509
 First wave, 330–332
 Fitness, 266, 403, 404
 Flavivirus, 409
 Flow diagram, 313, 317, 343
 Flux, 465
 density, 465
 Focus, 602
 Force of infection, 196, 202, 221, 252, 397, 458
 Fox population, 469
 Fox rabies, 464
 Functional equation, 10, 24, 437, 439
 Fundamental theorem of calculus, 566
- G**
 Galton–Watson, 7
 Gamma distribution, 88–90, 92, 93, 95, 98, 99,
 148–150, 152, 154, 156, 159, 167,
 218, 367, 369, 372, 374, 376, 377
 Garki project, 394
 Gauss–Jordan method, 544, 546
 Gaussian kernel, 202
 Gene, 402, 403
 General distribution, 374
 of stay, 141
 Generalized growth model, 382, 383
 General stage distribution, 166, 369
 Generating function, 119, 120, 122–124, 126
 moments, 119
 Generation time, 14, 201
 Genetic composition, 400
 Genetic variation, 323
 Genotype, 399, 400
 Geometric distribution, 165, 217, 218
 Gladwell, 509
 Global asymptotic stability, 26, 33, 53, 73, 277,
 278, 339, 404, 478, 479, 520
 Globally asymptotically stable, 277
 Global spread, 490
 Gonorrhea, 22, 23
 Graph, 7, 118, 119, 457, 459, 572
 Great Plague, 3, 40
 Growth rate, 553, 560
 Guillain–Barré syndrome, 414
- H**
 HAART, 273
 Half-life, 556
 Hamiltonian, 257
 Hazard of infection, 126
 Heat equation, 466
 Hepatitis, 67
 Herd immunity, 68, 346
 Heterogeneity, 14, 129, 162, 189, 204, 208,
 212, 213, 217, 219, 223, 224, 265,
 266, 275, 347, 477, 490
 of mixing, 161
 Heterogeneous contact rates, 477
 Heterogeneous mixing, 14, 65, 131, 188, 202,
 209, 212–214, 216, 219, 274, 347,
 429, 510
 Heterosexual transmission, 274
 Heterozygote, 398
 H5N1 influenza epidemic, 6
 High mobility, 484
 High risk, 221, 487, 490–492, 497
 area, 481, 482
 region, 488
 HIV/AIDS, 11, 14, 21, 67, 230, 249, 266, 267,
 273–276, 279, 284, 286, 291, 302,
 382
 dynamics, 274
 incidence, 286
 infections
 new, 286
 secondary, 287
 infective individuals, 286
 prevalence, 287
 sexually transmitted, 274
 transmission, 273, 286
 vertical transmission, 286
 Homoclinic bifurcation, 341, 342
 Homoclinic loop, 404
 Homoclinic orbit, 342
 Homogeneity, 265
 Homogeneous, 221
 mixing, 7, 14, 118, 161, 209, 210, 212–214,
 286
 Homosexual population, homogeneously
 mixing, 281
 Homosexual transmission, 273
 Homozygote, 398

- H1N1 influenza epidemic, 6
H1N1 strain, 311
Hopf bifurcation, 77, 78, 284, 339–341, 344
Hospitalization, 360, 363–365, 367, 372, 379–381
Host, 13, 229, 230, 346, 391
 asymptomatically infected, 405
 genetic structure, 398
 infective, 235
 population, 234, 393, 449, 494, 496
 population size, 394
Host–vector system, 500
Human host, 448
Hygiene, 161, 216
Hyperbola, 604
- I**
- IDEA model, 382, 384
Identity matrix, 544, 549
Immigration, 102, 183, 404
Immune, 447
Immune response, 39
Immunity, 4, 7, 23, 29, 44, 63, 73, 251, 284, 336
 against reinfection, 22, 394
 average period, 397
 boosting, 11
 loss, 397
 mean period, 397
 permanent, 405
 temporary, 23, 50, 73, 75, 396
Immunization, 23, 39, 68
 program, 208
Immunological change, 326
Immunological experience, 346
Immunological ideas, 10
Immunological model, 284
Immunology, 5
Incidence data, 361
 time series points, 358
Incidence function, 98
Incidence rate, 339
Incomplete treatment, 270
Incubation period, 274, 312, 352, 354, 357
 arbitrary, 304
Incubation period distribution, exponential, 275
Indefinite integral, 565
Index case, 184, 338
Infection, 319, 485
 age, 270, 281
 age density, 281
 age-dependent progression, 263
 asymptomatic, 405–407
 level, 291
 repeated, 405
 symptomatic, 406, 407
 transmission, 490
Infection period, fixed length, 396
Infection rate, 14, 242, 331, 448, 462, 485
 seasonally forced, 337
Infectious agent, 337
Infectiousness, 347
Infectious period distribution, 140
Infective, 5, 7, 8, 10–12, 14, 21–31, 33–35, 37, 39–41, 43–46, 48–50, 52, 55, 58, 63–69, 72, 73, 75, 77, 78, 81, 82, 89, 90, 93, 100–102, 104, 107, 108, 117, 118, 122, 124–127, 129–135, 137, 138, 141, 144–147, 149, 155–157, 159, 162, 164, 167–170, 173–175, 179–182, 184, 189, 190, 209, 215–217, 219, 230–232, 235, 242, 244, 251, 252, 264, 268, 269, 273, 275–280, 284, 287, 302–304, 312–315, 319, 321, 322, 332, 338, 352, 359, 360, 362–364, 367, 384, 386, 388, 392, 393, 410, 411, 413, 421, 439, 440, 447, 448, 458–462, 469, 472–474, 481, 510, 582
 compartment, 313
 stage, 189
Infective period, 25, 72, 137, 139, 146, 150, 156, 158, 183, 199, 219, 221, 274, 315, 319, 322, 335, 357, 458
 distribution, 144, 155, 158, 214, 216, 218
 fixed length, 146
Infectivity, 10, 11, 65, 131, 135, 139, 141, 144, 148, 168, 169, 190, 195, 209–211, 273, 282, 312, 313, 319, 323, 384
 distribution, 141
Influenza, 11, 22, 65, 131, 133, 154, 195, 313, 314, 346, 501, 511
 avian, 311
 cases, 53
 H1N1, 311, 331
 H1N1 strain, 330, 334
 H1N1 subtype, 323, 326
 H2N2 subtype, 326
 H3N2 subtype, 323, 326
 pandemic, 312, 327, 328, 330–332, 334, 461
 seasonal, 179, 328, 330, 332
 seasonal epidemic, 316, 323, 332
 seasonal outbreak, 326
 subtype, 323
 two strain, 338
 virus interaction, 347

- Infrectivity, 316
 In-hospital transmission, 162, 217
 Initial exponential growth, 352
 Initial boundary value problem, 466
 Initial condition, 190, 313, 317, 440, 448, 449, 466, 467, 554, 555, 559, 573
 Initial distribution, 440
 Initial exponential growth rate, 14, 39, 130, 131, 133, 137, 140, 141, 143, 148, 151, 153, 155, 156, 211, 212, 214, 233, 417, 418
 Initial growth, 354
 Initial growth rate, 353
 exponential, 10
 Initial infective distribution, 140
 Initial per capita growth, 402
 Initial population size, 404
 Initial secondary vertex, 121
 Initial size, 347, 348
 Initial stochastic phase, 382
 Initial time, 335, 354
 Initial value, 571
 problem, 554, 555, 559, 566
 Inoculation, 71
 Instability, 602
 Integral equation, 10, 24, 280, 367, 374, 452
 Integration by parts, 25
 Integro-differential equation, 263, 365, 367, 376
 Interaction, 196
 Intercohort mixing, 440
 Interepidemic period, 72
 Interpolation, 102
 Intervention, 346, 352, 354
 effectiveness, 338
 measures, 275
 program, 489
 strategy, 354
 Intracohort mixing, 440
 Invasion, 460
 criterion, 140, 211
 Inverse matrix, 545, 548
 Isolated, 359
 mixing, 206, 208
 Isolation, 23, 37, 89, 93, 99, 111, 134–136, 146–148, 162, 165–167, 173, 174, 216–218, 337–341, 343, 344, 346, 347, 359–361, 388, 510, 512
 Ivermectin, 241–243
- J**
 Jacobian, 217
 matrix, 32, 66, 352, 401, 582, 594, 595, 598
- Japanese encephalitis, 13, 229
 Joint dynamics, 284
- K**
 Kermack–McKendrick model, 6, 7, 10, 26, 35, 73, 100, 117, 118, 128, 129, 139, 151, 152, 582
 Kronecker delta, 196, 201
- L**
 Lagrangian approach, 477, 478, 481, 490
 Lagrangian framework, 477, 486
 Lagrangian model, 480, 501
 Lagrangian modeling perspective, 480
 Lagrangian perspective, 386, 421, 461, 477
 Laplace transform, 148
 Laplacian, 471
 Latency period, 249
 Latent, 249, 250
 class, 275
 Latent period, 195, 363
 distribution, 263, 268
 Latent stage, exponentially distributed, 269
 Legrand model, 361–367, 372, 378, 379
 Leishmaniasis, 13, 229
 Life expectancy, 4, 69, 70, 72
 Life span, 72
 Like-with-like mixing, 192
 Limit cycle, 596
 Limiting system, 281
 Linear approximation, 581
 Linear chain trick, 92, 218
 Linearization, 11, 31–33, 36, 49, 63, 66, 68, 71, 75, 76, 84, 87, 101, 130, 133, 136, 182, 183, 186, 233, 356, 385, 417, 581–585, 594, 595, 598, 602, 603
 theorem, 583, 593
 theory, 384
 Linear system, 539
 of equations, 539, 544
 Line of equilibria, 36
 Lipschitz structure, 258
 Logistic equation, 27, 560, 561, 568–571, 577
 Logistic model, 383
 Long range connection, 123
 Loss of immunity, 397
 Lotka characteristic equation, 434
 Lotka–Volterra model, 101
 Low mobility, 484
 Low risk, 221, 491, 492
 area, 481
 region, 483, 488

- Lyme disease, 13, 229
Lymphatic filariasis, 13, 229
- M**
- Major epidemic, 8, 118, 120, 122, 123, 125
Malaria, 3, 4, 6, 11, 13, 15, 21, 22, 229, 230, 391, 393, 397, 403, 407, 429, 447
asymptomatic, 405, 407
control, 6, 394
disease dynamics, 400
endemic, 394
falciparum, 398
incubation period, 394
outbreak, 394
prevalence, 407
recovery time, 399
transmission, 400
- Management model, 137
Management strategy, 311, 312, 315, 382
Maple, 562, 580
Mass action, 6, 122, 167, 234, 283, 404, 458
incidence, 24, 26, 43
Mathematica, 562, 580
Matlab, 562
Matrix, 133, 136, 137, 181, 183, 184, 187, 190, 194, 232, 233, 539, 581, 592, 595
algebra, 539
Jacobian, 186
non-negative, 185
operation, 542
product, 542
- Maximum, 40
epidemic size, 213, 214
number of infectives, 41
- McKendrick equation, 438
- Mean degree, 119, 126
- Mean duration, 366
- Mean excess degree, 120
- Mean exposed period, 158
- Mean generation time, 156–158
- Mean incubation period, 276
- Mean infective period, 25, 33, 44, 72, 104, 131, 138, 140, 146, 150, 153, 158, 166, 167, 172, 189, 275, 278, 280, 303, 361, 405
death-adjusted, 277, 280, 304
symptomatic, 405
- Mean infectivity, 139, 155, 156
- Mean number of sexual partners, 276–278, 304
- Mean sexual life expectancy, 276, 303
- Mean sojourn time, 91, 95
- Mean symptomatic attack rate, 315
- Mean transmissibility, 124
Mean transmission rate, 359
Measles, 3, 4, 15, 21, 68, 71, 72, 429, 461
vaccine, 4
- Mechanistic model, 382, 383
- Medicated class, 242
- Medicated host, 242
- Meningitis, 22
- Metapopulation, 197, 202, 204, 208, 223, 384, 457–461, 509
- Michaelis–Menten interaction, 43
- Microcephaly, 414, 490
- Microorganisms, 22
- Migration, 386, 461, 509
- Mild dengue fever, 409
- Minor outbreak, 8, 118, 122, 125
- Mixing, 191, 199, 208, 210, 223, 265, 266
function, 201, 202, 204
matrix, 204, 206
pattern, 14, 188, 199, 200
- M-matrix, 185
non-singular, 186
- Mobile, 497
- Mobility, 477, 480, 483–486, 488–500
behavior, 478
level, 484
matrix, 496
pattern, 500
regime, 500
restriction, 481
- Model
of dengue fever, 409
for Ebola, 351
endemic diseases, 63
linear age structured, 430
nonlinear age structured, 437
outcome, 365
prediction, 381
with quarantine, 174
Ross–MacDonald, 393
of Zika virus, 419
- Modeling assumption, 385
- Modeling framework, 338
- Morbidity, 311, 337
- Mortality, 311, 337
constant per capita, 398
fraction, 45
rate, 14, 278, 398
- Mosquito, 6, 391, 394
biting rate, 393, 397
control, 409
control program, 397
repellent, 414
- Movement restriction, 386

Multidrug-resistance, 250
 Multi-patch model, 480
 Multiple epidemic waves, 331
 Multiple pandemic waves, 334
 Multiple waves, 332
 Multivariable calculus, 579
 Mutation, 11, 73

N

Natural death, 459
 Natural death rate, 100, 251, 448, 479
 Natural mortality, 276
 Neighbor, 127
 Net growth rate, 465
 Network model, 8, 188, 189
 Network structure, 331
 Next generation approach, 463
 Next generation matrix, 182, 187, 188, 190, 193, 194, 197, 204, 210, 215, 217, 218, 232, 313, 317, 324, 387, 392, 395, 400, 406, 412, 413, 415–417, 458, 460
 with large domain, 184, 186–188, 193, 415
 1918 influenza epidemic, 3
 Nobel prize, 230
 Nobel Prize in Medicine, 394
 Node, 126, 128, 602–604
 Non-carrier, 406, 407
 Non-disease compartment, 182
 Normal form, 341
 Nosocomial transmission, 510
 Nullcline, 581

O

Occupied edge, 124
 Offspring, 434
 Onchocerciasis, 13, 229, 241
 Onset, 364, 365, 367
 Optimal allocation, 209
 Optimal control, 251, 257–260, 266
 Optimal strategy, 266
 Optimal treatment, 258
 Optimal treatment rate, 511
 Optimal vaccination strategy, 223, 266
 Optimal vaccine allocation, 205
 Optimal vaccine policy, 205
 Optimization, 205
 Orbit, 37, 102, 468, 580, 582, 595, 596
 Oscillation, 34, 71, 72, 77, 93, 332, 345
 damped, 71, 458
 Oscillatory solutions, 284

Outbreak, 72, 122, 140, 179, 356
 Overall sojourn, 379

P

Pair formation, 14, 43, 429, 445
 Pandemic, 214, 216, 311, 312, 315, 322, 323, 326–329, 331, 332, 334, 336, 511
 influenza, 326, 347
 1918 influenza, 326
 reproduction number, 327
 strain, 311, 326, 327, 329, 330, 348
 Parameter, 335
 Parasite, 107, 346
 Parasite load, 108
 Parasitemia, 399
 Partial derivative, 581
 Partial differential equation, 436, 464
 Partial fractions, 568
 Partial immunity, 323
 Particle, 465
 concentration, 465
 Partition, 436
 Partner interaction, 14
 Patch, 52, 386, 421, 457–461, 464, 471, 472, 477, 478, 480, 483, 485, 486, 489–492, 494–498, 500
 coupling, 458
 density, 500
 residence time, 477
 Patch-specific reproduction number, 479, 480
 Patch-specific risk, 491
 Pathogen, 11, 107, 108, 161, 167–171, 510
 Pathogenicity, 34
 Patient zero, 8, 119, 120
 Pattern in space, 467
 Peak size, 381, 382
 Peak time, 336
 Pearson chi-square statistics, 335
 Period, 71
 Periodic orbit, 77, 596
 stable, 243
 Periodic solution, 49, 339–342, 344, 345, 404, 580, 595
 Period of stay, 152, 158
 Perron–Frobenius theorem, 205
 Perturbation technique, 344
 Perturbed system, 341
 Phase line, 571, 574
 Phase locking, 458
 Phase plane, 580
 Phase portrait, 41, 468, 594, 599, 600, 602, 603

Phenomenological model, 382, 383
Piecewise exponential fit, 357, 358
Piecewise exponential rise, 357
Plague, 13, 229
Plasmodium, 399
Poincaré–Bendixson theorem, 596, 597
Poisson distribution, 50, 122, 123, 165
Policy decision, 322
Policy recommendation, 9
Polynomial equation, 33
Pontryagin maximum principle, 257
Population
 density, 33, 304, 399, 465, 497
 ecology, 466
 growth, 100, 102
 homosexually active, 275
 size, 189
 subgroup, 331
Population model, 577
 age structured, 429
Power series, 119
Pre-epidemic treatment, 317
Preferential mixing, 196, 200, 223
Preferred mixing, 191, 192, 204
Premedication class, 242
Pre-symptomatic case detection, 359
Pre-symptomatic infection, 359
Prevailing subtype, 326
Prevalence, 199, 406, 486, 488, 490
Prevalent, 291
Primary infected vertex, 122
Primary infection, 264
Primary resistance, 251
Prior immunity, 344
Probability, 8
 density, 142
 distribution, 122, 369
Progression, 485
 to active tuberculosis, 290
 to AIDS, 281, 290
 rate, 242, 252
 speed, 469
Properly posed problem, 31
Prophylaxis, 256, 261
Proportional death rate, 28, 29
Proportionate mixing, 191, 192, 195, 204, 206, 208, 211–213, 223, 445, 446
Public health infrastructure, 501
Public health measure, 161, 216, 334
Public health official, 386
Public health physicians, 6, 21
Public health policy, 312
Public transportation, 477

Q

Qualitative analysis, 31, 101, 570
Qualitative behavior, 29, 35, 44, 394, 568, 574, 591, 598
Qualitative property, 569, 570
Quantitative framework, 338
Quarantine, 21, 22, 37, 40, 42, 89, 93, 95, 98, 99, 134–136, 146, 148, 162, 164–167, 173, 174, 216, 337–341, 343, 354, 510, 512, 523
Quarantine–isolation model, 175
Quasi-steady-state, 234, 237, 239, 243

R

Rabies, 469, 471
Radioactive decay, 555
Random mixing, 446
Random partnership, 445
Rate of flow, 24, 465
Rate of transfer, 24
Reaction-diffusion, 464
Reaction-diffusion equation, 466–468
Reactivation, 488
Recovered, 447, 481
 compartment, 313
Recovery, 359, 364, 365, 367, 372, 388
Recovery rate, 10, 44, 45, 202, 324, 377, 388, 479, 480
 exponentially distributed, 35
Recruitment, 29, 281
Recruitment rate, 276, 278
Reduced system, 237–239
Reed–Frost model, 8
Region of asymptotic stability, 77
Region of attraction, 404
Reinfection, 22, 263, 266, 394, 488–490, 511
 potential, 486
 rate, 485
Relapse rate, 319
Relative infectivity, 324
Relative prevalence, 406, 407
Relative susceptibility, 407
Relative transmissibility, 359, 388
Removal rate, 278, 279, 305
 variable, 278
Removed, 10, 23, 31, 35, 39, 73, 90, 100, 189, 196, 230, 312, 313, 352, 359, 362, 386, 388, 396, 421, 439, 461
Renewal condition, 440
Renewal equation, 433
Replacement, 348

- Reproduction number, 96, 97, 125, 148, 154, 159, 166, 180–182, 185, 186, 190, 194, 197, 204, 205, 207, 208, 212, 213, 217–220, 222, 252–254, 262–264, 266, 269, 274, 275, 278, 287, 288, 291, 312, 318, 328, 329, 347, 354, 361, 374, 378, 387, 406, 417, 418, 420, 439, 452, 481, 509, 510
- HIV, 287
- with gamma distributed periods, 159
- with general infectious period distribution, 153
- stage-specific, 164
- Reproductive rate, 553
- Residence matrix, 478
- Residence time, 461, 478, 481, 484–486, 494, 496, 498, 500, 509
- Residence time matrix, 478
- Resident, 386, 459
- Resident population size, 461
- Resistance, 266
- Resistant strain, 252, 253, 256, 266, 269–271
- Respiratory infection, 3, 334
- Reversion from immune to susceptible, 396
- Richards model, 382, 383
- Rift Valley fever, 13, 229
- Rinderpest, 67, 68
- Risk, 276, 304, 337, 478, 485–487, 492, 494, 500, 501
- assessment, 500
 - factor, 287
 - heterogeneity, 496
 - of infection, 478
 - vector, 479
- Ross, Sir R.A., 6, 13, 229, 230, 391, 394, 509
- Ross–MacDonald model, 393, 398
- Routh–Hurwitz conditions, 33, 67
- Row operation, 544
- Rubella, 15, 22, 71, 429
- S**
- Saddle point, 88, 342, 468, 602–604
- Sanitation, 100
- SARS, 6, 14, 162, 216, 217, 337, 457, 461, 501, 509, 510
- Saturation, 43
- of contacts, 313
 - level, 277
- Scalar, 540, 548
- product, 540
- Scarlet fever, 340
- Schistomiasis, 21
- Schistosomiasis, 3, 13, 229
- Seasonal epidemic, 311, 315, 323, 326, 327, 331, 347
- Seasonal forcing, 160
- Seasonal influenza, 326, 347
- Seasonality, 347
- mosquito population, 394
- Seasonal outbreak, 11, 323, 328, 332
- Seasonal strain, 327, 329, 330, 348
- Seasonal variation, 72, 334
- Secondary cases, 120
- Secondary infected vertex, 120
- Secondary infection, 124, 164, 168, 182–185, 190, 220, 222, 277, 393, 463
- Second wave, 330–334
- Sensitive strain, 253, 269–271
- Sensitivity, 189
- analysis, 334
- Separable, 563, 565, 569, 570
- equation, 562
 - pair formation, 445, 446
- Separation of variables, 27, 101
- Separatrix, 88, 404, 604
- Serial interval, 157
- Seroconversion, 276, 304
- Serological study, 39
- Severity, 104, 312
- Sexual activity, 273, 274, 286, 418
- Sexual contact, 275, 303, 304, 418
- Sexually active, 281, 286, 302, 303
- population, 276, 304
 - population size, 282
- Sexually transmitted diseases, 14, 23, 188, 429, 445
- Sexual partner, 275–277, 304
- Sexual transmission, 286
- Sex workers, 274
- Sickle cell carrier, 407
- Sickle cell disease, 398, 406, 407
- Sickle cell genetics, 398
- Sickle cell trait, 405–407
- Sigmoid function, 105
- Sign structure, 74
- Singular perturbation, 236–239, 243, 400
- SIR model, 23, 53
- SIS model, 23, 49, 52
- Sleeping sickness, 3, 21
- Slow dynamics, 402
- Slower than exponential growth, 385
- Slow manifold, 402
- Slow progression, 261, 268
- Slow system, 108
- Slow variable, 108
- Smallpox, 4, 5, 69, 477

- Snow, John, 5
Social contact, 7, 118
Social interaction, 54
Sojourn distribution, 378, 379
Sojourn time, 163
Spatial distribution, 465
Spatial heterogeneity, 429
Spatial model, 458
Spatial pattern, 471
Spatial spread, 330, 464, 466, 509
Spatial structure, 384
Spectral bound, 185
Spectral radius, 184, 187, 205, 217, 324
Spiral, 596
 point, 603, 604
Stability, 11, 51, 53, 77, 263, 277, 283, 340,
 344, 345, 404, 584, 593, 595
 of equilibrium, 133
 index, 394
 region, 77, 345
 theorem, 593, 594
 transfer, 275
Stable age distribution, 434, 435, 449, 451
Stable manifold, 404
Stage distribution, 361, 367, 371
Staged progress, 170
Staged progression, 273
Stage duration, 379
 exponentially distributed, 268
Stages of infection, 316
Stage transition, 379
Standard incidence, 24, 43, 44, 283, 356, 404
Starting time, 333
Start time, 354
State at infection, 187
State equation, 258
Stay in compartment
 arbitrarily distributed, 7
 exponentially distributed, 7
STD, 274, 445
Steady state, 449
Stochastic branching process model, 125
Stochastic event, 7, 118
Stochastic model, 7, 8, 15, 42, 122, 189, 354,
 362
Stochastic phase, 10, 152, 154
Stochastic simulation, 312, 354
Strain, 251
Strategic model, 9, 330, 331, 333
Strategy, 321, 322
Strict barrier, 352
Subgroup, 275
Superinfection, 256, 347, 394
Superspread, 14, 123, 188
Superspreading, 122
Sub-population, 189, 196, 199, 202, 204, 208,
 210, 219, 221, 223, 266, 457
Surveillance, 354
Surveillance data, 331
Survival function, 365, 367, 369, 375, 377
Survival probability, 142, 372, 438
Survivorship function, 275, 279, 304
Susceptibility, 144, 179, 182, 190, 195, 312,
 316, 319, 323, 327, 334, 347, 384
Susceptible, 5–7, 10, 22–24, 26, 28–30, 33,
 35, 36, 39, 40, 44, 46, 48, 50, 52,
 55–59, 64, 65, 67, 70, 73, 75, 78,
 82, 83, 90, 100–102, 104, 107–109,
 111, 117, 118, 179–181, 185, 189,
 190, 195, 196, 202, 208, 210, 217,
 221, 222, 230, 232, 235, 251, 261,
 264, 266, 268, 270, 274–279, 281,
 284, 287, 293–295, 297–302, 304,
 312–314, 316–319, 323, 327, 328,
 330, 333, 336, 338, 342, 344, 346,
 352, 356, 358, 359, 362, 386, 393,
 396, 411, 421, 439–441, 443, 447,
 448, 458–462, 469, 473, 474, 477,
 479–481, 483, 491, 492, 494, 509,
 521–523, 582
Sustained oscillation, 34, 338, 341, 344
Symptom, 337
Symptomatic attack rate, 315
Symptomatic disease case, 318
System of differential equations, 539

T

- Tactical model, 9, 330, 331, 334
Taylor's theorem, 603
TB-HIV coinfection, 287
Technology, 501
Temporal behavior, 357
Temporal evolution, 356
Threshold, 6, 22, 28, 33, 63, 120, 230, 232,
 345, 418, 460, 485, 496
 conditions, 398
Time
 evolution, 356
 from onset to death, 375
 rescaled, 402
 scale, 243, 287, 400
 since onset, 367, 368
Tipping point, 28, 509
Total infectivity, 324, 328
Toxoplasmosis, 107
Trace, 32, 33, 67, 74, 88, 133, 187, 188,
 593–595

- Trait, 404
 Transcendental equation, 75
 Transcritical bifurcation, 275
 Transient subpopulations, 477
 Transition, 126, 326, 352, 365, 367, 368, 371, 372
 diagram, 162, 164, 221, 252, 361, 372, 376, 377
 probability, 222
 rate, 155, 219, 354, 371, 377
 time, 365
 Transmissibility, 123, 125, 273, 332
 Transmission, 6, 140, 159, 167, 334, 360, 387, 399, 480, 485, 489, 491
 coefficient, 324, 452, 460, 469, 471
 diagram, 253
 dynamics, 107, 263, 266, 274, 324, 359, 488, 496
 fitness, 510
 heterogeneity, 489
 intervention, 125
 mode, 273, 337
 network, 490
 probability, 190, 274, 278, 304, 420
 rate, 166, 270, 276, 281, 304, 331, 334, 352, 354, 356, 360, 363, 397, 398, 487, 492
 reduced, 352
 seasonality, 347
 seasonally forced, 159, 334
 Transpose, 547
 Transversality condition, 258
 Travel, 386, 387, 460, 461, 509
 rate, 459–461, 471
 restriction, 464
 Traveling wave, 464, 468–470
 Treatment, 65, 131, 159, 195, 252, 253, 261, 262, 264, 266, 312, 321, 322, 407
 class, 131, 132
 effort, 260, 452
 failure, 253, 254, 257, 262, 263
 fraction, 318
 model, 37, 144
 period distribution, 144, 145
 probability, 407
 rate, 252, 267, 319, 322, 407
 stage, 132
 Triangular matrix, 544
 Tuberculosis, 4, 13, 21, 22, 104, 230, 249–253, 256–258, 260, 261, 263, 265–269, 284, 337, 463, 480, 486, 501, 511
 active, 252, 286, 290
 antibiotic resistant, 252
 with distributed delay, 268
 drug-resistant, 262, 266, 270
 with drug-resistant strains, 252
 drug-sensitive, 270
 dynamics, 291
 with fast and slow progression, 260, 261
 HIV coexistence, 288
 HIV coinfection, 291
 infective, 286
 with optimal control, 256
 prevalence, 290, 291
 with reinfection, 263
 resistant, 251–254, 256, 257, 259, 260, 266
 resistant strain, 251
 sensitive, 252, 253, 270
 treatable, 284
 treatment, 287
 two-strain, 253, 266
 Tuberculosis cases, secondary, 287
 Tularemia, 13, 229
 Two-strain model, 256, 269, 342
 Typhoid, 5
 Typhus, 3, 21
- U**
 Unbounded solution, 573, 584
 Unstable, 11, 12, 26, 32–34, 63, 64, 77, 78, 80–82, 88, 91, 95, 100, 101, 109, 110, 130, 185, 186, 223, 254, 255, 265, 278, 280, 284, 288, 304, 339, 340, 342, 344, 345, 404, 406, 444, 445, 449, 450, 452, 460, 471, 583, 592, 594, 595, 597
 Untreated host, 242
- V**
 Vaccination, 9, 11, 15, 22, 48, 49, 56, 57, 60, 65, 68, 69, 82–84, 113, 114, 131, 159–162, 176, 179, 180, 182, 186, 188, 192, 194, 196, 197, 199, 204, 205, 208, 210, 216, 223, 224, 243, 266, 299, 300, 312, 316, 318, 319, 321, 330–332, 335, 336, 347, 440–443, 451, 454, 510
 pulse, 47, 48
 Vaccine, 4, 21, 69, 83, 114, 134, 159, 161, 179, 202, 205, 208, 216, 224, 299, 300, 311, 316, 319, 330–332, 336, 351, 360, 414
 Variable infectivity, 274
 Variolation, 4
 Vector, 12, 13, 229, 230, 241, 391, 421, 501, 539, 540, 592

- control, 490
population, 9, 234, 393, 447, 492, 494, 500
population size, 394
transmission, 230, 417, 420
 model, 391, 393, 414
 reproduction number, 415
Vector-borne disease, 13, 229, 501
Vector-matrix notation, 598
Vector-transmitted disease, 230, 232, 391
Vertex, 7, 118, 119, 122, 459
Vertical transmission, 12, 67, 273, 274, 410,
 414, 501
Viral agents, 22
Viral load, 359
Virus, 23, 332, 333
Vortex, 602
Vulnerability, 190, 210, 211
- W**
Waiting time, 364, 365, 367, 371, 375, 379
- exponential, 371
exponentially distributed, 366
Wave, 333, 471
West Nile fever, 13
West Nile virus, 13, 229, 230
Whooping cough, 71
Within-host system, 107, 108
- Y**
Yellow fever, 13, 229, 230
- Z**
Zika, 13, 414, 418, 419, 463, 480, 490, 501
 transmission through sexual contact, 414
virus, 229
ZIKV, 490, 491, 493, 494, 496–501
Z-sign pattern, 185