

**THE IMA VOLUMES IN MATHEMATICS  
AND ITS APPLICATIONS**

EDITORS

Carlos Castillo-Chavez  
with Sally Blower,  
Pauline van den Driessche,  
Denise Kirschner, and  
Abdul-Aziz Yakubu

# **Mathematical Approaches for Emerging and Reemerging Infectious Diseases**

Models, Methods, and  
Theory



Springer

# **The IMA Volumes in Mathematics and its Applications**

**Volume 126**

*Series Editor*  
Willard Miller, Jr.

# Institute for Mathematics and its Applications

## IMA

The Institute for Mathematics and its Applications was established by a grant from the National Science Foundation to the University of Minnesota in 1982. The IMA seeks to encourage the development and study of fresh mathematical concepts and questions of concern to the other sciences by bringing together mathematicians and scientists from diverse fields in an atmosphere that will stimulate discussion and collaboration.

The IMA Volumes are intended to involve the broader scientific community in this process.

Willard Miller, Jr., Professor and Director

\* \* \* \* \*

### IMA ANNUAL PROGRAMS

- |           |  |
|-----------|--|
| 1982–1983 | Statistical and Continuum Approaches to Phase Transition                                     |
| 1983–1984 | Mathematical Models for the Economics of Decentralized Resource Allocation                   |
| 1984–1985 | Continuum Physics and Partial Differential Equations   |
| 1985–1986 | Stochastic Differential Equations and Their Applications                                     |
| 1986–1987 | Scientific Computation   |
| 1987–1988 | Applied Combinatorics  |
| 1988–1989 | Nonlinear Waves  |
| 1989–1990 | Dynamical Systems and Their Applications   |
| 1990–1991 | Phase Transitions and Free Boundaries  |
| 1991–1992 | Applied Linear Algebra   |
| 1992–1993 | Control Theory and its Applications  |
| 1993–1994 | Emerging Applications of Probability   |
| 1994–1995 | Waves and Scattering   |
| 1995–1996 | Mathematical Methods in Material Science   |
| 1996–1997 | Mathematics of High Performance Computing  |
| 1997–1998 | Emerging Applications of Dynamical Systems   |
| 1998–1999 | Mathematics in Biology   |
| 1999–2000 | Reactive Flows and Transport Phenomena   |
| 2000–2001 | Mathematics in Multimedia  |
| 2001–2002 | Mathematics in the Geosciences   |
| 2002–2003 | Optimization   |
| 2003–2004 | Probability and Statistics in Complex Systems: Genomics, Networks, and Financial Engineering |

Continued at the back

Carlos Castillo-Chavez  
with  
Sally Blower, Pauline van den Driessche,  
Denise Kirschner, and Abdul-Aziz Yakubu  
Editors

Mathematical Approaches  
for Emerging and Reemerging  
Infectious Diseases:  
Models, Methods, and Theory

With 69 Illustrations



Springer

Carlos Castillo-Chavez, Director  
Mathematical and Theoretical Biology Institute  
and member, Departments of Biometrics, Statistics and Theoretical and Applied Mechanics  
Cornell University  
Ithaca, NY 14853-7801  
USA  
cc32@cornell.edu  
[http://www.biom.cornell.edu/Homepages/Carlos\\_Castillo-Chavez/](http://www.biom.cornell.edu/Homepages/Carlos_Castillo-Chavez/)

Sally Blower  
Department of Biomathematics  
UCLA School of Medicine  
Los Angeles, CA 90095-1766  
USA  
SBlower@biomath.medsch.  
ucla.edu

Pauline van den Driessche  
Department of Math and Stats  
University of Victoria  
Victoria, British Columbia  
V8W 3P4  
Canada  
pvdd@math.uvic.ca

Denise Kirschner  
Dept. of Microbiology and  
Immunology  
University of Michigan Medical  
School  
Ann Arbor, MI 48109-0620  
USA  
kirschne@umich.edu

Abdul-Aziz Yakubu  
Department of Mathematics  
Howard University  
Washington, DC 20059  
USA  
ayakubu@fac.howard.edu  
or Biometrics Department  
Cornell University  
Email: ay32@cornell.edu

*Series Editor:*  
Willard Miller, Jr.  
Institute for Mathematics and  
its Applications  
University of Minnesota  
Minneapolis, MN 55455  
USA

---

Mathematics Subject Classification (2000): 92-01, 92-02, 92-06, 92B05, 92D25, 92D30, 92D40, 37N25, 34C60,  
34D23, 37C75, 34K60, 39A11, 39-06, 45-06, 60J80

---

Library of Congress Cataloging-in-Publication Data  
Mathematical approaches for emerging and reemerging infectious diseases.  
Models, methods, and theory / editors, Carlos Castillo-Chavez . . . [et al.]

p. cm. — (The IMA volumes in mathematics and its applications ; 126)

Includes bibliographical references and index.

ISBN 978-1-4612-6550-4 ISBN 978-1-4613-0065-6 (eBook)

DOI 10.1007/978-1-4613-0065-6

1. Communicable diseases—Epidemiology—Mathematical models—Congresses. I. IMA  
volumes in mathematics and its applications ; v. 126.

RA643 .M352 2001

614.5'01'51—dc21

2001049541

Printed on acid-free paper.

© 2002 Springer Science+Business Media New York  
Originally published by Springer-Verlag New York, Inc. in 2002  
Softcover reprint of the hardcover 1st edition 2002

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher Springer Science+Business Media, LLC , except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by Springer Science+Business Media, LLC provided that the appropriate fee is paid directly to Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, USA (Telephone: (508) 750-8400), stating the ISBN number, the title of the book, and the first and last page numbers of each article copied. The copyright owner's consent does not include copying for general distribution, promotion, new works, or resale. In these cases, specific written permission must first be obtained from the publisher.

Production managed by A. Orrantia; manufacturing supervised by Jeffrey Taub.  
Camera-ready copy prepared by the IMA.

9 8 7 6 5 4 3 2 1

ISBN 978-1-4612-6550-4

SPIN 10850724

## **FOREWORD**

This IMA Volume in Mathematics and its Applications

### **MATHEMATICAL APPROACHES FOR EMERGING AND REEMERGING INFECTIOUS DISEASES: MODELS, METHODS AND THEORY**

is based on the proceedings of a successful one week workshop. The proceedings of the two-day tutorial which preceded the workshop “Introduction to Epidemiology and Immunology” appears as IMA Volume 125: Mathematical Approaches for Emerging and Reemerging Infectious Diseases: An Introduction. The tutorial and the workshop are integral parts of the September 1998 to June 1999 IMA program on “MATHEMATICS IN BIOLOGY.”

I would like to thank Carlos Castillo-Chavez (Director of the Mathematical and Theoretical Biology Institute and a member of the Departments of Biometrics, Statistics and Theoretical and Applied Mechanics, Cornell University), Sally M. Blower (Biomathematics, UCLA School of Medicine), Pauline van den Driessche (Mathematics and Statistics, University of Victoria), and Denise Kirschner (Microbiology and Immunology, University of Michigan Medical School) for their superb roles as organizers of the meetings and editors of the proceedings. Carlos Castillo-Chavez, especially, made a major contribution by spearheading the editing process. I am also grateful to Kenneth L. Cooke (Mathematics, Pomona College), for being one of the workshop organizers and to Abdul-Aziz Yakubu (Mathematics, Howard University) for serving as co-editor of the proceedings. I thank Simon A. Levin (Ecology and Evolutionary Biology, Princeton University) for providing an introduction.

Finally, I take this opportunity to thank the National Science Foundation (NSF), whose financial support of the IMA made the Mathematics in Biology program possible, and the National Institute of General Medical Sciences of the National Institutes of Health (NIH), for providing partial support for the workshop.

Willard Miller, Jr., Professor and Director  
Institute for Mathematics and its Applications  
University of Minnesota  
400 Lind Hall, 207 Church St. SE  
Minneapolis, MN 55455-0436  
612-624-6066, FAX 612-626-7370  
[miller@ima.umn.edu](mailto:miller@ima.umn.edu)  
World Wide Web: <http://www.ima.umn.edu>

## PREFACE

### MATHEMATICAL APPROACHES FOR EMERGING AND REEMERGING INFECTIOUS DISEASES: MODELS, METHODS AND THEORY

The research collected in the volume 125, *Mathematical Approaches for Emerging and Reemerging Infectious Diseases: An Introduction* and IMA Volume 126, *Mathematical Approaches for Emerging and Reemerging Infectious Diseases: Models, Methods and Theory*, grew out of the discussions and presentations instigated by the Workshop on Emerging and Reemerging Diseases sponsored by the Institute for Mathematics and its Applications (IMA) at the University of Minnesota in May 17–21, 1999. This workshop included a two day tutorial session directed to ecologists, epidemiologists, immunologists, mathematicians and scientists interested in being exposed to some of the modeling and mathematical approaches used in the study of disease dynamics. Tutorial papers presented at this workshop or requested, as a result of the nature of the contributions of workshop participants, have been included (mostly) in the introductory volume. On the other hand, this volume includes papers on applications that may require additional mathematical sophistication. The reader will find applications motivated by the study of diseases like Influenza, HIV, Measles and Tuberculosis. There are also two contributions to the study of macroparasitic diseases like schistosomiasis.

These volumes are dedicated to Fred Brauer and Kenneth Cooke on their retirement from their academic positions at the University of Wisconsin-Madison and Pomona College, respectively. The leading articles in each volume take a personal look at their contributions to mathematics, mathematical biology and epidemiology over the past five decades. It is but a small gesture of our appreciation as we continue to learn from our personal and professional association to these leading researchers and extraordinary teachers and mentors.

The last two contributions in IMA Volume 125 are co-authored by Angel Capurro who died suddenly in a car accident on December 10, 2000. Angel was a leading researcher at the Departamento de Investigaciones of the Universidad de Belgrano in Buenos Aires Argentina for the past six years. Angel, in a very short time, established one of the most active research groups on the study of tuberculosis. He was not only a leading researcher in the field of tuberculosis but also active in ecology with an appointment at the Universidad Nacional in Lujan, Argentina. Angel was a wonderful individual with friends all over the world. He was a close friend of many of us, and we will miss him.

We want to thank Simon A. Levin for writing the introduction and for encouraging and supporting the completion of these two volumes. We spent various amounts of time at the IMA in 1998 and 1999 and earlier during the preparation of many of the activities associated with the organization of the 1998-99 IMA Special Year in Mathematical Biology. We also want to express our gratitude to those individuals who were responsible not only for the support of the activities associated with the preparation of these volumes but also to those members of IMA's staff who made our visit to this institute a wonderfully productive experience. The leadership and support of Willard Miller and Fred Dulles are highly appreciated. We also want to acknowledge the help and friendliness of Kathy Boyer, Inés Foss, Michelle Glubke, Dzung Nguyen and all the IMA staff who are responsible for the IMA's first rate operation. Special thanks to Patricia V. Brick for her professionalism in the production of these two volumes.

The Workshop on Emerging and Reemerging Diseases and the preparation of these volumes were partially supported by NSF and NIH grants to IMA. We want to thank these agencies for their support although we must re-state that the views and research presented in these volumes are the direct responsibility of the authors and the editorial group, that is, they do not represent the views of the funding agencies or the IMA.

Carlos Castillo-Chavez (Cornell University)

Sally Blower (University of California at Los Angeles)

Pauline van den Driessche (University of Victoria)

Denise Kirschner (University of Michigan)

Abdul-Aziz Yakubu (Howard University)

## CONTENTS

Foreword .....	v
Preface.....	vii
New directions in the mathematics of infectious disease .....	1
<i>Simon A. Levin</i>	
Fred Brauer: The man and his mathematics .....	7
<i>Christopher M. Kribs-Zaleta</i>	
Kenneth L. Cooke: Researcher, educator par excellence .....	21
<i>P. van den Driessche</i>	
Maximal prevalence and the basic reproduction number in simple epidemics.....	31
<i>L. Esteva and K.P. Hadeler</i>	
The transition through stages with arbitrary length distributions, and applications in epidemics.....	45
<i>Horst R. Thieme</i>	
Measles outbreaks are not chaotic .....	85
<i>Ingemar Nåsell</i>	
Epidemics among a population of households .....	115
<i>Frank G. Ball and Owen D. Lyne</i>	
Infection transmission dynamics and vaccination program effectiveness as a function of vaccine effects in individuals.....	143
<i>Carl P. Simon and James S. Koopman</i>	
The influence of different forms of cross-protective immunity on the population dynamics of antigenically diverse pathogens .....	157
<i>Neil Ferguson and Viggo Andreasen</i>	
Dynamics of multiple strains of infectious agents coupled by cross-immunity: A comparison of models.....	171
<i>M. Gabriela M. Gomes and Graham F. Medley</i>	

Virulence evolution in macro-parasites.....	193
<i>Andrea Pugliese</i>	
Mathematical models for schistosomiasis with delays and multiple definitive hosts.....	215
<i>Jianhong Wu and Zhilan Feng</i>	
Infectious disease models with chronological age structure and epidemiological age structure.....	231
<i>Fred Brauer</i>	
Effects of genetic heterogeneity on HIV transmission in homosexual populations .....	245
<i>Shu-Fang Hsu Schmitz</i>	
Age-structured core group model and its impact on STD dynamics .....	261
<i>Carlos Castillo-Chavez and Wenzhang Huang</i>	
Global dynamics of tuberculosis models with density dependent demography .....	275
<i>Baojun Song, Carlos Castillo-Chavez, and Juan P. Aparicio</i>	
Global stability in some SEIR epidemic models .....	295
<i>Michael Y. Li and Liancheng Wang</i>	
The global stability analysis for an SIS model with age and infection age structures .....	313
<i>Yicang Zhou, Baojun Song, and Zhien Ma</i>	
Endemic threshold and stability in an evolutionary epidemic model .....	337
<i>Hisashi Inaba</i>	
Epilogue .....	361
List of tutorial/workshop participants .....	363
IMA volume 125 contents: Mathematical approaches for emerging and reemerging infectious diseases: an introduction.....	367

# NEW DIRECTIONS IN THE MATHEMATICS OF INFECTIOUS DISEASE

SIMON A. LEVIN\*

The study of infectious diseases represents one of the oldest and richest areas in mathematical biology. Infectious diseases have fascinated mathematicians for a century, and with good reason. Most seductive, of course, is the possibility of using mathematics to make a positive contribution to the world. In the study of infectious diseases, the essential elements are quickly grasped, and well-captured within mathematical representations. As new epidemics, from AIDS to bovine spongiform encephalopathy (mad-cow) to foot-and-mouth, make their appearances on the world stage, mathematical models are essential to inform decision-making. Governments and health agencies turn to the leading modelers for advice, and news services seek them out for the clarity they can bring. It is a rare opportunity for relevance for those who spend so much of their time in otherwise abstract and esoteric exercises.

Among the problems of biology, those involving the transmission dynamics of infectious diseases are among the most accessible to those not trained as biologists. There is a standard and well-established mathematical framework, expanded and given credibility by the work of many over the past century (see Anderson and May, 1991). For specific situations, this general framework always requires tailoring or more substantial modification, but that is part of the attraction for the researcher. The usual equations also have easily-identified control parameters, or intervention points, such as vaccination strategies or contact rates, which determine the dynamics and provide handles for decision-making. Furthermore, crisp, powerful and robust results, such as the existence of critical thresholds, emerge naturally from the analyses, and can be translated very directly into lessons for control.

Add to these attractive features the dynamical richness of epidemiological problems, and the appeal to mathematicians is irresistible. The basic epidemiological equations are sufficiently nonlinear to make things interesting, but still simple enough to provide sharp results. They provide wonderful vehicles for the techniques of dynamical systems theory, especially bifurcation analysis. Furthermore, straightforward elaborations of them, to account say for age structure or latent periods, take one immediately into the world of integral or partial differential equations, and into the study of delay equations. Management issues are naturals for control theory; and game theory is essential both for understanding the evolution

---

\*Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544-1003.

of infectious agents, as well as for a full treatment of management issues. Finally, disease transmission is a fundamentally stochastic process, meaning that the techniques of probability and statistics undergird the entire subject.

These two remarkable volumes explore the subject very broadly, from the classical roots to the frontiers of the subject. The impressive collection of contributors includes many who have been in the forefront of the development of the subject, and the warmth with which the volumes and individual articles have been dedicated to Fred Brauer and Ken Cooke is obvious. That is also not hard to understand. Not only have both of these outstanding mathematicians been seminal contributors to the literature, but both are wonderful human beings who evoke good feelings in all who interact with them. It is therefore my pleasure to join in the chorus, and to wish both Fred and Ken continued good health and productivity.

In the following sections, I will touch upon some of the problems at the frontier of the subject, as seen through my own idiosyncratic filter. Many of these are explored within the individual contributions in this volume; others undoubtedly will find their way into the sequels to these works.

**Frontiers in the study of infectious diseases.** Any selection of frontiers in any subject is necessarily somewhat speculative, a bit like trying to predict the next variant of a virus. Still, a number of classes of problems seem to me to be ripe for exploration, and fundamentally important in their implications. My own bias is often to look for problems at the ecotones between disciplines, because it is there that cross-fertilization has the potential to lead to rich harvests in short amounts of time. Many of these involve the integration of phenomena across scales, problems whose analysis has been facilitated by recent advances in methodology and technology (Levin et al., 1997).

One of the most important of the interfaces, of course, is that between epidemiology and immunology. This has received much attention recently by problems associated with the AIDS epidemic (Nowak et al., 1991; Perelson et al., 1996), and the interplay between viral evolution and the immune dynamics of hosts. It is, however, a fundamental aspect of understanding the dynamics of any disease. In the case of AIDS, the immune system must play a fundamental role in the progression of the disease, and mathematical models have elucidated the role of cytotoxic T cells (Nowak et al., 1995a; Nowak et al., 1995b). Evolutionary issues also intrude here, such as in the evolution of resistance in HIV to antiretroviral drugs, and the implications for treatment strategies (McLean and Nowak, 1992; Ho et al., 1995; Wei et al., 1995). Immunology has long been a rich but speculative area of theoretical biology, with milestones such as in the work of the late George Bell, a pioneer in scaling real and intellectual mountain peaks. In recent years, especially since the landmark contributions to AIDS research, the subject has achieved prominence, especially in terms of its interfaces with

epidemiology. Surely this activity will expand, and include the interplay of the immune system with hormonal systems, and how factors such as stress, nutrition and general health affect responses to infectious agents.

The evolutionary issues mentioned in the previous paragraph just scratch the surface. The importance of heterogeneous transmission in the dynamics of disease directs attention to genetic variation in infectivity and transmissibility, and ultimately to the evolution of these properties (Dushoff and Levin, 1995; Dushoff, 1999). Aspects of the evolution of infectious agents are central to understanding the dynamics of diseases such as malaria (Gupta et al., 1994) and influenza (Andreasen et al., 1997), with attention to the evolution of patterns of cross-reactivity and the self-organization of strain structure. Fitch et al. (1991; 1997) have done remarkable work on the evolution of influenza, with implications for the development of vaccines. The integration of all of these processes cross scales from the cellular to the ecological to the evolutionary, and require the efforts of a broad spectrum of scientists. Bacterial evolution poses particularly important problems: The evolution of antibiotic resistance is one of the greatest threats facing society today (Daily and Ehrlich, 1996). Finally, host-parasite systems provide ideal vehicles for the study of coevolution (Levin and Pimentel, 1981; Anderson and May, 1982; Levin, 1983a). The central issues certainly involve the interplay between host resistance and the pathogen infectivity (Anderson and May, 1982; Futuyma and Slatkin, 1983; Levin, 1983b; Levin et al., 1999), but also extend to issues such as plasmid evolution, the competitive dynamics of strains, and the interplay between interacting diseases.

Another interface likely to see more attention in the coming decade will be that between epidemiology and economics, with ethics also lurking not so far in the background. Like it or not, many of the management challenges have economic overtones, and furthermore introduce problems of the global commons. How should the cost of treatment be factored into decisions to vaccinate, or even to treat? What are the tradeoffs between using particular antibiotics to treat individuals, and the risks of the development of antibiotic resistance? What are the incentives for drug companies to develop genuinely novel antibiotics? Can we determine optimal global and regional management strategies for antibiotic use, accounting for the costs and benefits of multiple antibiotics, and balancing adequately the rights of individuals and the needs of societies? These non-traditional areas for mathematical investigation are likely to receive much more attention in the near future.

Climatology provides another venue for investigation of disease interactions. With global warming, there will undoubtedly be implications for the spread of disease. Changing climate will affect the diseases themselves, the health and habits of potential hosts, and most importantly the population biology of vectors. Malaria, dengue and cholera are all diseases

attracting considerable attention as regards the potential effects of global warming and habitat alteration.

The issue of habitat alteration is one aspect of a suite of spatial problems demanding attention. Understanding the spread of infectious diseases is a problem that long has been of interest to theoreticians, but more recently much has been learned about the coupling of the dynamics of disease among different cities (Bolker and Grenfell, 1993). The degree of synchronization among local oscillations can affect the persistence of diseases (Lloyd and May, 1996; Earn et al., 2000), and have profound implications for management. Furthermore, globalization and the changing connectivity of local populations can affect these dynamics, and also lead to reinterpretation of the critical thresholds for persistence. Indeed, many of the emergent and re-emergent diseases threatening our societies today were likely well-contained as endemics locally when communities were more isolated (see Levin, 1999).

All of these issues and more will receive increased attention in the coming years, and many have already been broached in these volumes. Castillo-Chavez and Yakubu examine the importance of patchy environments, and Allen and Ernest look at problems of spread. Ferguson and Andreasen look at links to the immune system, and patterns of cross-reactivity within evolving families of strains. Hsu Schmitz introduces genetic heterogeneity, and explores its consequences. It is clear that these are just a start, and that the study of infectious diseases will continue to be one of the most rewarding areas for mathematical investigation for years to come. These volumes provide an important step along that path.

## REFERENCES

- ANDERSON, R.M. AND MAY, R.M. 1982. Coevolution of hosts and parasites. *Parasitology*, **85**:411–426.
- ANDERSON, R.M. AND MAY, R.M. 1991. *Infectious Diseases of Humans: Dynamics and Control*. Oxford, UK: Oxford University Press.
- ANDREASEN, V., LIN, J., AND LEVIN, S.A. 1997. The dynamics of cocirculating influenza strains conferring partial cross-immunity. *Journal Mathematical Biology*, **35**:825–842.
- BOLKER, B.M. AND GRENFELL, B.T. 1993. Chaos and biological complexity in measles dynamics. *Proc. R. Soc. Lond. B*, **251**:75–81.
- DAILY, G.C. AND EHRLICH, P.R. 1996. Global change and human susceptibility to disease. *Ann. Rev. Energy Environ.*, **21**:125–144.
- DUSHOFF, J. 1999. Host heterogeneity and disease endemicity: A moment-based approach. *Theoretical Population Biology*, **56**.
- DUSHOFF, J. AND LEVIN, S.A. 1995. The effects of population heterogeneity on disease spread. *Mathematical Biosciences*, **128**:25–40.
- EARN, D.J.D., LEVIN, S.A., AND ROHANI, P. 2000. Coherence and conservation. *Science*, **290**:1360–64.
- FITCH, W.M., BUSH, R.M., BENDER, C.A., AND COX, N.J. 1997. Long term trends in the evolution of H(3) HA1 human influenza type A. *Proceedings of the National Academy of Sciences, USA*, **94**:7712–7718.

- FITCH, W.M., LEITER, J.M.E., LI, X., AND PALESE, P. 1991. Positive Darwinian evolution in human influenza A viruses. *Proceedings of the National Academy of Sciences, USA*, **88**:4270–4274.
- FUTUYAMA, D. AND SLATKIN, M., eds. 1983. *Coevolution*. Sunderland, MA: Sinauer.
- GUPTA, S., SWINTON, J., AND ANDERSON, R.M. 1994. Theoretical studies of the effects of heterogeneity in the parasite population on the transmission dynamics of malaria. *Proc. R. Soc. Lond., B*, **256**:231–238.
- HO, D.D., NEUMANN, A.U., PERELSON, A.S., CHEN, W., LEONARD, J.M., AND MARKOWITZ, M. 1995. Rapid turnover of plasma viroids and CD4 lymphocytes in HIV-1 infections. *Nature*, **373**.
- LEVIN, B.R., LIPSITCH, M., AND BONHOEFFER, S. 1999. Population biology, evolution, and infectious disease: Convergence and synthesis. *Science*, **283**:806–809.
- LEVIN, S.A. 1983a. Coevolution. In *Population Biology*, Lecture Notes in Biomathematics (H. Freedman and C. Strobeck, eds.), pp. 328–334, Berlin: Springer-Verlag.
- LEVIN, S.A. 1983b. Some approaches to the modelling of coevolutionary interactions. In *Coevolution* (M. Nitecki, ed.), pp. 21–65 Chicago, Illinois: University of Chicago Press.
- LEVIN, S.A. 1999. *Fragile Dominion: Complexity and the Commons*. Reading, MA: Perseus Books.
- LEVIN, S.A., GRENFELL, B., HASTINGS, A., AND PERELSON, A.S. 1997. Mathematical and computational challenges in population biology and ecosystem science. *Science*, **275**:334–343.
- LEVIN, S.A. AND PIMENTEL, D. 1981. Selection of intermediate rates of increase in parasite-host systems. *American Naturalist*, **117**:308–315.
- LLOYD AND MAY, R.M. 1996. Spatial heterogeneity in epidemic models. *J. Theoretical Biology*, **179**:1–11.
- MCLEAN, A.R. AND NOWAK, M.A. 1992. Competition between zidovudine sensitive and resistant strains of HIV. *AIDS*, **6**:71–79.
- NOWAK, M.A., ANDERSON, R.M., MCLEAN, A.R., WOLFS, T., GOUDSMIT, J., AND MAY, R.M. 1991. Antigenic diversity thresholds and the development of AIDS. *Science*, **254**:963–966.
- NOWAK, M.A., MAY, R.M., PHILLIPS, R.E., ROWLAND-JONES, S., LALOO, D.G., MCADAM, S., KLENERMAN, P., KÖPPE, B., SIGMUND, K., BANGHAM, C.R.M., AND McMICHAEL, A.J. 1995a. Antigenic oscillations and shifting immunodominance in HIV-1 infections. *Nature*, **375**:606–611.
- NOWAK, M.A., MAY, R.M., AND SIGMUND, K. 1995b. Immune responses against multiple epitopes. *J. Theoretical Biology*, **175**:325–353.
- PERELSON, A.S., NEUMANN, A.U., MARKOWITZ, M., LEONARD, J.M., AND HO, D.D. 1996. HIV-1 dynamics in vivo: Virion clearance rate, infected cell life-span, and viral generation time. *Science*, **271**:1582–1586.
- WEI, X., GHOSH, S.K., TAYLOR, M.E., JOHNSON, V.A., EMINI, E.A., DEUTSCH, P., LIFSON, J.D., BONHOEFFER, S., NOWAK, M.A., HAHN, B.H., SAAG, M.S., AND SHAW, G.M. 1995. Viral dynamics in HIV-1 infection. *Nature*, **373**:117–122.

## FRED BRAUER: THE MAN AND HIS MATHEMATICS

CHRISTOPHER M. KRIBS-ZALETA\*



**Fred Brauer<sup>†</sup>**

**Introduction.** When I first met Fred Brauer, he was warning a class of first-semester calculus students about the dangers of applying techniques without thinking. As an example, he wrote on the chalkboard the expression  $\frac{\sin x}{n}$ . He then proceeded to cancel the  $n$ 's from numerator and denominator:

$$\frac{\sin x}{\not{n}},$$

leaving *six*.

At the time, I, a first-semester graduate student, knew nothing of the vast body of work he had produced in rather more advanced mathematics,

---

\*Department of Mathematics, University of Texas at Arlington, Arlington, TX 76019-0408.

<sup>†</sup>Department of Mathematics University of British Columbia, Vancouver, B.C V6T 1Z2, Canada.

but his thoughtful approaches to solving problems, coupled with his sense of humor, appealed to me. He later encouraged those same calculus students, “Lions in the Serengeti Desert solve optimization problems — and they don’t even know calculus.”<sup>†</sup>

That semester we (I was one of the TAs assigned to his lecture) were using a new calculus text<sup>§</sup> which began somewhat unconventionally, using discrete-time population models. Perhaps that should have given me a clue as to what to expect a year later, when I went to him in search of a Ph.D. advisor. In the past fifty years, Fred’s research has spanned the gamut of population biology, from theoretical work in differential and integral equations to applied studies of the relationship between demographic processes and the spread of diseases. He has collaborated with dozens of mathematicians, written over 100 books and papers, performed research in such places as Brazil, Israel, and the Netherlands, and lectured in Italy, China, Mexico, and many places in between. In addition to his own doctoral students, he has fostered the work and careers of many young scientists. He is (to steal indiscriminately from Gilbert and Sullivan) the very model of a mathematics modeller.

Fred Brauer, who grew up with mathematics in his family, received his Ph.D. from MIT in 1956 under Norman Levinson. After teaching at the University of Chicago and the University of British Columbia, he came in 1960 to the University of Wisconsin. During his tenure in Madison he became a cornerstone of the department, serving as chair among other positions, and instituting the departmental seminar in “spherical trigonometry” to show that mathematicians can also play basketball. (He continues to do so.) In two terms as associate chair he made important developments to the undergraduate curriculum, including sequences of precalculus courses and mathematics courses for biologists as well as the infusion of differential equations into multivariable calculus, linear algebra and the upper-level curriculum. It was in Madison that he formed a long-term collaboration with John Nohel, also a student of Levinson; in the early 1960’s the two wrote what has become a classic textbook on qualitative theory of ODEs. A year or two later, Hans Schneider (also in the department) joined them to write a second text unifying linear algebra and linear systems of ODE’s. John and Fred wrote several more books together in the years that followed, and worked together on some of the aforementioned curriculum development.

Fred’s recent return to his native Canada with wife Esther brought them close not only to his old stomping grounds at UBC but to one of their three children — and, consequently, to grandchildren as well. It doesn’t hurt that Vancouver is a nice place to live — and to visit.

<sup>†</sup>Fred attributes this observation to Colin Clark.

<sup>§</sup>Frank Wattenberg’s *Calculus in a Real and Complex World*

**Differential equations.** With Norman Levinson as his advisor, it is not surprising that Fred's early work concentrated on properties of differential equations. From the completion of his dissertation in 1956 until the early 1970's, his papers addressed topics including spectral theory for linear systems, uniqueness and convergence of successive approximations, and asymptotic behavior of perturbed nonlinear systems. Three papers on the first topic in the late 1950's investigated the expansion in eigenfunctions of operators (with associated boundary conditions) for systems of linear differential equations. Three other papers in the same time period [1, 2, 3, 4] explored a duality between local uniqueness and global existence of solutions, by viewing global existence as a kind of uniqueness at infinity. These papers also compared local uniqueness with the convergence of successive approximations, of the type

$$x_j(t) = \int_0^t f[x_{j-1}(s), s] ds,$$

presenting various conditions for them and noting (via counterexamples) that although the two notions are independent — neither implies the other — most criteria which guarantee one also guarantee the other.

Four papers in the *Journal of Mathematical Analysis and Applications* from 1966 to 1972 [5, 6, 7, 8] described an alternative method for considering perturbations to nonlinear systems of DEs. Rather than using a Lyapunov function for the unperturbed equation, the articles adapted Alekseev's nonlinear variation of constants formula to determine the effects of perturbation on solutions, as a function of the stability properties of the solutions to the original system. The result was an extended hierarchical look at different kinds of stability. As Fred noted, some kind of stability information must be known about the solutions of the unperturbed system in order to know the effects of perturbations. A number of different types of (or criteria for) stability, several of which are identical for linear systems (and some of which were new), provided different results, and a ranking by implication:

$$\begin{aligned} \text{exponential stability} &\rightarrow \text{asymptotic stability in variation} \\ &\rightarrow \text{asymptotic stability} \\ &\rightarrow \text{uniform stability in variation} \\ &\rightarrow \text{integral stability} \\ &\rightarrow \text{uniform stability} \\ &\rightarrow \text{stability} \end{aligned}$$

In addition, uniform asymptotic stability also implies asymptotic stability, but is independent of asymptotic stability in variation (neither implies the other).

**Population biology.** It was in the early 1970's that Fred turned his eye to population biology, and for the next twenty years he studied problems related to predator-prey systems and harvesting. A series of papers, many co-authored by A. C. Soudack, considered managed nonlinear predator-prey systems of the form

$$\begin{aligned}x' &= x f(x, y) - F, \\y' &= y g(x, y) - G,\end{aligned}$$

where  $F$  and  $G$  represent constant harvesting (if positive) or stocking (if negative) terms,  $x$  the prey and  $y$  the predators, and with the biologically-motivated constraints  $f_x < 0$  (limited resources for prey),  $g_x > 0$  (predator reliance on prey),  $g_y \leq 0$  (limited resources for predators), and  $f_x g_y - f_y g_x > 0$  for  $x, y > 0$ . Results [9, 10] showed that excessive harvesting of predators would cause the predator population to crash (reach zero in finite time), and excessive harvesting of prey would cause both populations to crash. On the other hand, stocking had the possibility of destabilizing the system, and excessive “enrichment” of the prey environment (by raising the prey’s carrying capacity) could give rise to limit cycles. One case where the system was asymptotically stable only in a limited basin of attraction underscored the need for careful management of such populations.

Another research area involved the forms used to describe biological processes. Models in population biology have traditionally assumed exponential distributions in most of the processes involved — births, deaths, and even infections and recoveries in epidemic models — in order that the models be systems of differential equations. However, the quantitative biological justification for this assumption is weak, and Fred recognized that there were times when other forms would more accurately model the biological processes involved. In particular, processes with fixed duration can be represented by step functions, which lead to systems of difference equations, or delay differential equations. More generally, the models can be written as systems of Volterra integral equations, with the function describing the rate at which the process in question occurs serving as a kind of convolution kernel. Consequently, a number of Fred’s models considered the effects of delays and more general rate distributions. The analysis of such models is, of course, more complex, but the results are worth the fuss: If, on the one hand, results match those for the ODE model, a robust principle has been derived, which holds regardless of the form of the distribution. If, on the other hand, as conventional wisdom has it, delays are more conducive to instability of equilibria, then one must make some effort to determine how the process in question can best be described.

His first such papers extended earlier results, such as the application of Alekseev’s nonlinear variation of constants formula and the effects of constant-effort harvesting, to systems described by Volterra integral equations. In 1987 a paper in *Natural Resource Modeling* [11] described how delays and a generalized lifespan (mortality) distribution interact to affect the stability of an equilibrium: for the simple system

$$x(t) = p_0 \int_0^\infty g(x(t-s-\tau)) P(s) ds,$$

where  $g(x)$  is the birth rate as a function of population  $x$ ,  $\tau$  is age to adulthood ( $x(t)$  counts only adults), and  $p_0$  is the probability of survival to adulthood (age  $\tau$ ).  $P(s)$  is a nonnegative, nonincreasing, integrable function (with  $P(0) = 1$ ) representing the probability of surviving to age  $\tau + s$ , given survival to age  $\tau$ . The delay  $\tau$  and mortality distribution  $P$  interact in the following way:

- (i) If  $-1 < p_0 g'(x_\infty) \int_0^\infty P(s) ds < 1$ , then the equilibrium  $x_\infty$  is stable regardless of  $\tau$ ;
- (ii) if  $p_0 g'(x_\infty) \int_0^\infty P(s) ds > 1$ , then the equilibrium  $x_\infty$  is unstable regardless of  $\tau$ ;
- (iii) if  $p_0 g'(x_\infty) \int_0^\infty P(s) ds < -1$ , then there is a finite value of  $\tau$  past which the equilibrium  $x_\infty$  is unstable.

Here we see that the delay in recruitment and the expected lifetime do affect the stability of the population. Also in this paper an application to a previously studied epidemic model for an SIR (recovery with immunity) disease with fixed incubation and infective periods gave a hint of the direction in which Fred's research would next turn.

**Epidemic models.** Like the interaction between recruitment delay and mortality distribution in general population models, disease transmission and (other) demographics can affect each other in important ways. These effects have been one of the most prominent aspects of Fred's research in the realm of infectious diseases. In particular, models he studied with Carlos Castillo-Chávez, Jorge Velasco-Hernández, Steve Blythe and others focused on the influence of disease state on the size of the at-risk population, and the effects in turn of this recruitment on the disease itself. In the study of sexually transmitted diseases, the notion of a core group has been established [12] as central to understanding disease dynamics. It has also been observed that the perceived prevalence of the disease may affect individuals' decision whether or not to enter the at-risk population (by increasing their activity level). The aforementioned models considered both one-sex (homosexual) and two-sex (heterosexual) transmission of diseases — some universally fatal, such as HIV/AIDS, and others of more limited duration, such as gonorrhea — in which the rate at which new individuals entered the core was a (nonincreasing) function of the disease prevalence.

Initial results using exponentially distributed recovery times (as usual, leading to ODE models) and core recruitment influenced by the total number of infected individuals mirrored the prototypical  $R_0$  threshold behavior, but more detailed models [13, 14, 15], in which recruitment was a function of the *proportion* of either infected individuals, or treated infected individuals, or in which the noncore population was also active (but at a lower

rate than the core), or in which the recovery rate was not exponentially distributed, allowed for the possibility of Hopf bifurcations with  $R_0 > 1$ , which led to sustained oscillations. As these oscillations could have very long periods, their empirical observation might be difficult. In general, as the level of infection increased, the recruitment into the core would damp down, until the infection began to die out. Once the infection level was low, however, recruitment would return to a higher rate, with the infection level rising behind it. This sort of behavior and subsequent rise in infection prevalence has been observed in some cases. These results suggest that in order to develop robust principles regarding disease transmission and control, we must consider a variety of distributions for the biological processes we model (or, at least, consider which distributions are most appropriate, rather than most convenient), as well as different ways to model the influence of perceived disease prevalence.

**Conclusion.** In the past forty years Fred has been active not only in producing and fostering research but in developing college-level mathematics education, to a large extent on a local level, but also in developing courses and authoring textbooks which present relevant mathematics to biologists, such as probability and dynamical systems. Tailoring of mathematical content to biologists and those interested in investigating phenomena such as infectious diseases is a rising trend which helps to prepare — and perhaps attract — the mathematical biologists of tomorrow, and it is an important part of leadership for the mathematical biologists of today to take part in this development.

For all his contributions: research in systems of differential and integral equations, in population biology and in mathematical epidemiology; fostering in the large and on the individual level education and productivity in these areas; and for his keen humor and being an exemplary colleague, we make this dedication to Fred Brauer.

*Note: The references below are intended as a sample rather than an extensive list. They were chosen as representative of the results discussed in this article. Further references can be identified by search engines such as MathSciNet, or by checking citations in these articles.*

## REFERENCES

- [1] Fred Brauer and Shlomo Sternberg (1958). Local uniqueness, existence in the large, and the convergence of successive approximations, *Amer. J. Math.* 80: 421–430.
- [2] Fred Brauer and Shlomo Sternberg (1959). Errata to our paper “Local uniqueness, etc.”, *Amer. J. Math.* 81: 797.
- [3] Fred Brauer (1959). A note on uniqueness and convergence of successive approximations, *Canad. Math. Bull.* 2: 5–8.

- [4] Fred Brauer (1959). Some results on uniqueness and successive approximations, *Canad. J. Math.* 11: 527–533.
- [5] Fred Brauer (1966). Perturbations of nonlinear systems of differential equations, *J. Math. Anal. Appl.* 14: 198–206.
- [6] Fred Brauer (1967). Perturbations of nonlinear systems of differential equations. II, *J. Math. Anal. Appl.* 17: 418–434.
- [7] Fred Brauer and Aaron Strauss (1970). Perturbations of nonlinear systems of differential equations. III, *J. Math. Anal. Appl.* 31: 37–48.
- [8] Fred Brauer (1972). Perturbations of nonlinear systems of differential equations. IV, *J. Math. Anal. Appl.* 37: 214–222.
- [9] Fred Brauer and A. C. Soudack (1978). Response of predator-prey systems to nutrient enrichment and proportional harvesting, *Internat. J. Control* 27(1): 65–86.
- [10] Fred Brauer and A. C. Soudack (1982). On constant effort harvesting and stocking in a class of predator-prey systems, *J. Theor. Biol.* 95(2): 247–252.
- [11] Fred Brauer (1987). A class of Volterra integral equations arising in delayed-recruitment population models, *Natur. Resource Modeling* 2(2): 259–278.
- [12] Herbert W. Hethcote and James A. Yorke (1984). *Gonorrhea transmission dynamics and control*. Lecture Notes in Biomathematics 56. New York: Springer-Verlag.
- [13] S. P. Blythe, Fred Brauer, Carlos Castillo-Chávez and Jorge X. Velasco-Hernández (1995). Models for sexually transmitted diseases with recruitment, in *Mathematical Population Dynamics: Analyses of Heterogeneity* (O. Arino, D. Axelrod, M. Kimmel, M. Langlais, eds.). Winnipeg: Wuerz Publishing Co. Vol. 1, pp. 197–207.
- [14] Fred Brauer, Carlos Castillo-Chávez and Jorge X. Velasco-Hernández (1996). Recruitment effects in heterosexually transmitted disease models, *Int. J. App. Science & Computation* 3: 78–90.
- [15] Fred Brauer, Carlos Castillo-Chávez and Jorge X. Velasco-Hernández (1997). Recruitment with a core group and its effect on the spread of a sexually transmitted disease, in *Advances in Mathematical Population Dynamics — Molecules, Cells, and Man* (O. Arino, D. Axelrod, M. Kimmel, eds.). Singapore: World Scientific Press. pp. 477–486.

### Papers by Fred Brauer

1. Fred Brauer (1958). Singular self-adjoint boundary value problems for the differential equation  $Lx = \lambda Mx$ , *Trans. Amer. Math. Soc.* 88: 331–345.
2. Fred Brauer (1958). Spectral theory for the differential equation  $Lu = \lambda Mu$ , *Can. J. Math.* 10: 431–446.
3. Fred Brauer and Shlomo Sternberg (1958). Local uniqueness, existence in the large, and the convergence of successive approximations, *Amer. J. Math.* 80: 421–430; errata, same journal, 81: 797 (1959).
4. Fred Brauer (1959). A note on uniqueness and convergence of successive approximations, *Can. Math. Bulletin* 2: 5–8.
5. Fred Brauer (1959). Some results on uniqueness and successive approximations, *Can. J. Math.* 11: 527–533.
6. Fred Brauer (1960). Spectral theory for linear systems of differential equations, *Pacific J. Math.* 10: 17–34.
7. Fred Brauer (1961). Global behavior of solutions of ordinary differential equations, *J. Math. Anal. Appl.* 2: 145–158.

8. Fred Brauer (1962). Asymptotic equivalence and asymptotic behavior of linear systems, *Mich. Math. J.* 9: 33–43.
9. Fred Brauer (1963). Lyapunov functions and comparison theorems, in *International Symposium on Nonlinear Differential Equations and Nonlinear Mechanics* (J. P. LaSalle & S. Lefschetz, eds.). New York: Academic Press. pp. 435–441.
10. Fred Brauer (1963). Bounds for solutions of ordinary differential equations, *Proc. Amer. Math. Soc.* 14: 36–43.
11. Fred Brauer (1963). On the asymptotic behavior of Bessel functions, *Amer. Math. Monthly* 70: 954–957.
12. Fred Brauer (1964). Nonlinear differential equations with forcing terms, *Proc. Amer. Math. Soc.* 15: 758–765.
13. Fred Brauer (1964). On the completeness of biorthogonal systems, *Mich. Math. J.* 11: 379–383; errata, same journal, 12: 127–128 (1965).
14. Fred Brauer (1965). Some refinements of Lyapunov's second method, *Can. J. Math.* 17: 811–819.
15. Fred Brauer (1966). Perturbations of nonlinear systems of differential equations, *J. Math. Anal. Appl.* 14: 198–206.
16. Fred Brauer (1966). The use of comparison theorems for ordinary differential equations, in *Stability Problems of Solutions of Diff. Equations* (Proc. NATO Advanced Study Inst., Padua, 1965). Gubbio: Edizione “Oderisi”. pp. 29–50.
17. Fred Brauer (1966). The asymptotic behavior of perturbed nonlinear systems, in *Stability Problems of Solution of Diff. Equations* (Proc. NATO Advanced Study Inst., Padua, 1965). Gubbio: Edizione “Oderisi”. pp. 51–56.
18. Fred Brauer (1966). The solution of non-homogeneous systems of differential equations by undetermined coefficients, *Can. Math. Bull.* 9: 81–87.
19. Fred Brauer (1967). Green's functions for singular ordinary differential operators, *Can. J. Math.* 19: 571–582.
20. Fred Brauer (1967). Perturbations of nonlinear systems of differential equations, II., *J. Math. Anal. Appl.* 17: 418–434.
21. Fred Brauer (1968). A class of nonlinear eigenvalue problems, in *U.S.-Japan Seminar on Differential & Functional Equations*. New York: W. A. Benjamin, Inc. pp. 429–433.
22. Fred Brauer (1968). Nonlinear perturbations of Sturm-Liouville boundary-value problems, *J. Math. Anal. Appl.* 22: 591–598.
23. Fred Brauer and James S. W. Wong (1969). On asymptotic behavior of perturbed linear systems, *J. Diff. Equations* 6: 142–153.
24. Fred Brauer and James S. W. Wong (1969). On the asymptotic relationships between solutions of two systems of ordinary differential equations, *J. Diff. Equations* 6: 527–543.
25. Fred Brauer and Aaron Strauss (1970). Perturbations of nonlinear

- systems of differential equations, III, *J. Math. Anal. Appl.* 31: 37–48.
- 26. Fred Brauer (1972). Perturbations of nonlinear systems of differential equations, IV, *J. Math. Anal. Appl.* 37(1): 214–222.
  - 27. Fred Brauer (1972). A nonlinear variation of constants formula for Volterra equations, *Math. Systems Theory* 6: 226–235.
  - 28. Fred Brauer (1972). A nonlinear predator-prey problem, *Ordinary differential equations*, Proc. 1971 NRL-MRC Conference (L. Weiss, ed.). New York: Academic Press. pp. 371–377.
  - 29. Fred Brauer (1972). The nonlinear simple pendulum, *Ann. Math. Monthly* 79: 348–355.
  - 30. Fred Brauer (1974). On the populations of competing species, *Math. Biosci.* 19: 299–306.
  - 31. Fred Brauer (1975). On a nonlinear integral equation for population growth problems, *SIAM J. Math. Anal.* 6: 312–317.
  - 32. Fred Brauer and David A. Sánchez (1975). Constant rate population harvesting: equilibrium and stability, *Theor. Pop. Biol.* 8: 12–30.
  - 33. Fred Brauer and David A. Sánchez (1975). Some models for population growth with harvesting, in *Proc. International Conference on Differential Equations*, Univ. of Southern Calif., Los Angeles, 1974 (H. A. Antosiewicz, ed.). New York: Academic Press. pp. 53–64.
  - 34. Fred Brauer and David A. Sánchez (1976). Cosecha de poblaciones en competencia [Harvesting of competing populations], in *Mathematical Notes and Symposia, Vol. 2: Ecuaciones Diferenciales* (Proc. Third Mexico-U.S. Symposium) (C. Imaz, ed.). Mexico City: Fondo de Cultura Económica. pp. 171–176.
  - 35. Fred Brauer (1976). Constant rate harvesting of populations governed by Volterra integral equations, *J. Math. Anal. Appl.* 56: 18–27.
  - 36. Fred Brauer (1976). Some applications of the theory of ordinary differential equations to population growth problems, *Ann. Acad. Brasil de Ciencias* 48: 369–385.
  - 37. Fred Brauer (1976, October). Perturbations of the nonlinear renewal equation, *Advances in Math.* 22(1): 32–51.
  - 38. Fred Brauer (1976). Destabilization of predator-prey systems under enrichment, *Int. J. Control* 23: 541–552.
  - 39. Fred Brauer, A. C. Soudack and H. S. Jarosch (1976). Stabilization and destabilization of predator-prey systems under harvesting and nutrient enrichment, *Int. J. Control* 23: 553–573.
  - 40. Fred Brauer (1977). Stability of some population models with delay, *Math. Biosciences* 33: 345–358.
  - 41. Fred Brauer (1977). Periodic solutions of some ecological models, *J. Theor. Biol.* 69: 143–152.

42. Fred Brauer and A. C. Soudack (1978). Response of predator-prey systems to nutrient enrichment and proportional harvesting, *Int. J. Control* 27: 65–86.
43. Fred Brauer (1978). Asymptotic stability of a class of integro-differential equations, *J. Differential Equations* 28: 180–188.
44. Fred Brauer (1979). Harvesting strategies for population systems, *Rocky Mountain J. Math.* 19: 19–26.
45. Fred Brauer (1979). Decay rates for solutions of a class of differential-difference equations, *SIAM J. Math. Anal.* 10: 783–788.
46. Fred Brauer and A. C. Soudack (1979). Stability regions and transitions phenomena for harvested predator prey systems, *J. Math. Biol.* 7: 319–337.
47. Fred Brauer (1979). Boundedness of solutions of predator-prey systems, *Theor. Pop. Biol.* 15: 268–273.
48. Fred Brauer (1979). Characteristic return times for harvested population models with time-lag, *Math. Biosci.* 45: 295–311.
49. Fred Brauer and A. C. Soudack (1979). Stability regions in predator-prey systems with constant-rate prey harvesting, *J. Math. Biol.* 8: 55–71.
50. Fred Brauer and A. C. Soudack (1981). Constant-rate stocking of predator-prey systems, *J. Math. Biol.* 11(1): 1–14.
51. Fred Brauer and A. C. Soudack (1981). Coexistence properties of some predator-prey systems under constant rate harvesting and stocking, *J. Math. Biol.* 12: 101–114.
52. Fred Brauer and A. C. Soudack (1981). Constant-rate effort harvesting and stocking in a class of predator-prey systems, in *Differential Equations and Applications in Ecology, Epidemics and Population Problems* (Stavros N. Busenberg and Kenneth L. Cooke, eds.). New York: Academic Press. pp. 131–144.
53. Fred Brauer and A. C. Soudack (1982). On constant effort harvesting and stocking in a class of predator-prey systems, *J. Theor. Biol.* 95: 247–252.
54. Fred Brauer (1983). Constant-rate harvesting of age-structured populations, *SIAM J. Math Anal.* 14 (1983), 947–961.
55. Fred Brauer (1983). Nonlinear age-dependent population growth under harvesting, *Int. J. Computers and Math. with Appl.* 9: 345–352.
56. Fred Brauer (1984). The effect of harvesting on population systems, in *Trends in Theory & Practice of Nonlinear Differential Equations* (Proc., Arlington, TX, 1982, V. Lakshmikantham, ed.). New York: Marcel Dekker. pp. 81–89.
57. Fred Brauer (1984). Constant-yield harvesting of population systems, in *Mathematical Ecology*, Proc. Trieste 1982 (S. A. Levin and T. G. Hallam, eds.), Lec. Notes in Biomathematics 54. Berlin: Springer-Verlag. pp. 238–246.

58. Fred Brauer and A. C. Soudack (1985). Optimal harvesting in predator-prey systems, *Int. J. Control* 41(1): 111–128.
59. Fred Brauer and A. C. Soudack (1985). Mutualism models with nonlinear growth rates, *Int. J. Control* 41(6): 1601–1612.
60. Fred Brauer and Ma Zhien (1987). Stability of stage-structured population models, *J. Math. Anal. Appl.* 126: 301–315.
61. Fred Brauer (1987). A class of Volterra integral equations arising in delayed-recruitment population models, *Nat. Res. Modeling* 2(2): 259–278.
62. Fred Brauer (1987). Absolute stability in delay equations, *J. Differential Equations* 69: 185–191.
63. Fred Brauer (1987). A simplification of Taylor's theorem, *Amer. Math. Monthly* 94: 453–455.
64. Fred Brauer (1987). Harvesting in delayed-recruitment population models, in *Proc. Canadian Mathematical Society 1986 Seminar on Oscillation, Bifurcation and Chaos*. Providence, RI: Amer. Math. Soc. pp. 317–327.
65. Fred Brauer, Stephen Ellner and M. D. Krom (1988). Modeling and management for seawater fishponds, in *Proc. 1986 Trieste Research Conference on Mathematical Ecology* (T. G. Hallam, S. A. Levin, and L. J. Gross, eds.). Teaneck, NJ: World Scientific Press. pp. 215–235.
66. Fred Brauer (1988). Coexistence and survival of invading species, in *Proc. 1986 Trieste Research Conference on Mathematical Ecology* (T. G. Hallam, S. A. Levin, and L. J. Gross, eds.). Teaneck, NJ: World Scientific Press. pp. 599–610.
67. Fred Brauer, David Rollins and A. C. Soudack (1988). Harvesting in population models with delayed recruitment and age-dependent mortality, *Nat. Res. Modeling* 3(1): 45–62.
68. Fred Brauer (1988). Some topics in population biology involving delay equations, 38 pp.; translated into Chinese by Ma Zhien. Xian, China: Department of Mathematics, Xian Jiaotong University. 58 pp.
69. Fred Brauer (1989). Epidemic models in populations of varying size, in *Mathematical Approaches to Problems in Resource Management and Epidemiology* (C. Castillo-Chávez, S. Levin, and C. Shoemaker, eds.), Lecture Notes in Biomathematics 81. Berlin: Springer-Verlag. pp. 109–123.
70. Fred Brauer (1989). Multi-species interactions and coexistence, in *Proc. Int. Conf. on Theory and Applications of Differential Equations* (A.R. Aftabizadeh, ed.). Athens, OH: Ohio Univ. Press. pp. 91–96.
71. Fred Brauer (1990). Models for the spread of universally fatal diseases, *J. Math. Biology* 28: 451–462.
72. Fred Brauer (1990). Some infectious disease models with popula-

- tion dynamics and general contact rates, *J. Diff. Integral Equations* 5: 827–836.
- 73. Fred Brauer (1991). Stability of equilibria in some infectious disease models, in *Differential Equations: Stability and Control* (S. Elaydi, ed.). New York: Marcel Dekker. pp. 53–62.
  - 74. Fred Brauer and Hao Dun-Yuan (1991). Analysis of a characteristic equation, *J. Integral Equations and Appl.* 3: 239–254.
  - 75. Fred Brauer (1991). Models for the spread of universally fatal diseases II, in *Differential Equations Models in Biology, Epidemiology, and Ecology*, Proc. Claremont 1990 (Stavros Busenberg and Mario Martelli, eds.), Lec. Notes in Biomathematics 92. Berlin: Springer-Verlag. pp. 57–69.
  - 76. S. P. Blythe, Fred Brauer and Carlos Castillo-Chávez (1995). Demographic recruitment in sexually transmitted disease models, in *Computational Medicine, Public Health and Biotechnology: Building a Man in the Machine* (M. Witten, ed.). Singapore: World Scientific Press. pp. 1438–1457.
  - 77. Fred Brauer (1995). Models for diseases with exposed period, *Rocky Mountain J. Math.* 25: 57–66.
  - 78. S. P. Blythe, Fred Brauer, Carlos Castillo-Chávez, and Jorge X. Velasco-Hernández (1995). Models for sexually transmitted diseases with recruitment, in *Mathematical Population Dynamics: Analyses of Heterogeneity* (O. Arino, D. Axelrod, M. Kimmel, and M. Langlais, eds.). Vol 1. Winnipeg: Wuerz Publishing Co. pp. 197–207.
  - 79. Fred Brauer (1995). Models for diseases with vertical transmission and nonlinear population dynamics, *Math. Biosci.* 128: 13–24.
  - 80. Fred Brauer (1996). Variable infectivity in communicable disease models, in *Proc. First World Congress of Nonlinear Analysts*, Tampa, Florida, 1992 (V. Lakshmikantham, ed.). Berlin: de Gruyter. Vol 4, pp. 3201–3210.
  - 81. Fred Brauer, Carlos Castillo-Chávez and Jorge X. Velasco-Hernández (1996). Recruitment effects in heterosexually transmitted diseases, *Int. J. App. Science & Computation* 3: 78–90.
  - 82. Fred Brauer and Carlos Castillo-Chávez (1995). Basic models in epidemiology, in *Ecological Time Series* (T.M. Powell and J.H. Steele, eds.). Chapman & Hall. pp. 410–447.
  - 83. Fred Brauer, Jorge X. Velasco-Hernández and Carlos Castillo-Chávez (1996). Effects of treatment and prevalence-dependent recruitment on the dynamics of a fatal disease, *IMA J. Math. Applied to Medicine and Biology* 13: 175–192.
  - 84. Fred Brauer (1996). A characteristic equation arising in models for diseases with vertical transmission and without immunity, in *Differential Equations and Applications to Biology and to Industry* (M. Martelli, K. Cooke, E. Cumberbatch, B. Tang, and H. Thieme,

- eds.). Singapore: World Scientific Press. pp. 41–48.
85. Fred Brauer (1996). A model for populations with variable maturation period, *Dynamics of Continuous, Discrete, and Impulsive Systems* 2: 41–50.
  86. Fred Brauer, Carlos Castillo-Chávez and Jorge X. Velasco-Hernández (1997). Recruitment with a core group and its effect on the spread of a sexually transmitted disease, in *Advances in Mathematical Population Dynamics: Molecules, Cells, and Man* (O. Arino, D. Axelrod, and M. Kimmel, eds.). Singapore: World Scientific Press. pp. 477–486.
  87. Fred Brauer (1997). Continuous and discrete delayed, recruitment population models, *Dynamics of Continuous, Discrete and Impulsive Systems* 3: 245–252.
  88. Fred Brauer (1999). General recruitment models for sexually transmitted diseases, in *Differential Equations with Applications to Biology* (S. Ruan, G. S. K. Wolkowicz, and J. Wu, eds.), Fields Institute Communications No. 21. Providence, RI: Amer. Math. Soc. pp. 45–50.
  89. Fred Brauer (1999). Continuous and discrete population models with age-dependent mortality, *Dynamics of Continuous, Discrete, and Impulsive Systems* 5: 107–113.
  90. Fred Brauer (in press). Time lags in disease models with recruitment, *Mathematical and Computer Modelling*, to appear.
  91. Fred Brauer (in press). Infectious disease models with chronological age structure, this volume.
  92. Fred Brauer (in press). Basic ideas of mathematical epidemiology, this volume.
  93. Fred Brauer (in press). Extensions of the basic models, this volume.
  94. Fred Brauer and Pauline van den Driessche (in press). Models for the transmission of disease with immigration of infectives, to appear.
  95. Fred Brauer (in press). What goes up must come down, eventually, *American Mathematical Monthly*, to appear.
  96. Fred Brauer (in press). A model for an SI disease in an age-structured population, to appear.

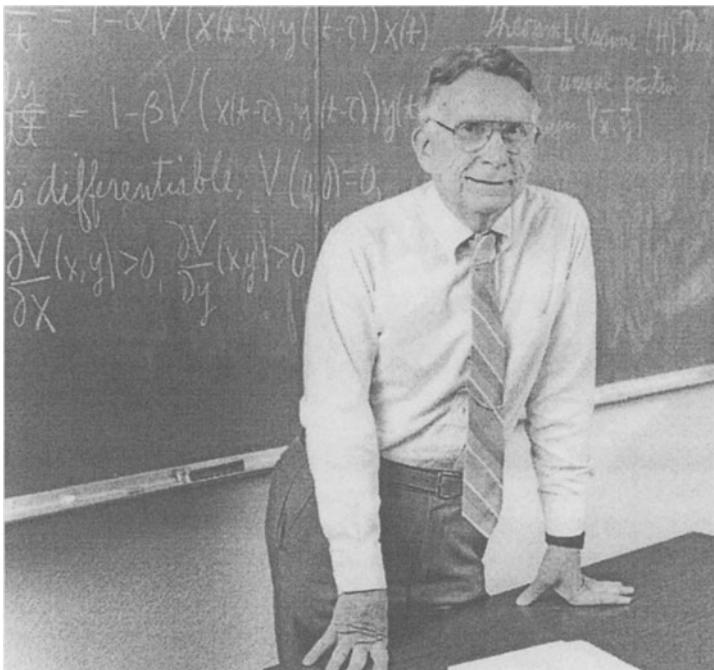
### Books by Fred Brauer

1. Fred Brauer and John A. Nohel (1967). *Ordinary Differential Equations: A first course*. New York: W. A. Benjamin, Inc. 1st ed. xvi + 457 pp. 2nd ed., 1973, ix + 470 pp.
2. Fred Brauer and John A. Nohel (1968). *Elementary Differential Equations: Principles, problems, solutions*. New York: W. A. Benjamin, Inc. xi + 222 pp.
3. Fred Brauer and John A. Nohel (1968). *Problems and Solutions in Ordinary Differential Equations*. New York: W. A. Benjamin,

- Inc. x + 267 pp.
- 4. Fred Brauer and John A. Nohel (1969). *Qualitative Theory of Ordinary Differential Equations*. New York: W. A. Benjamin, Inc. xi + 314 pp. Reprinted, Dover, 1989.
  - 5. Fred Brauer, John A. Nohel and Hans Schneider (1970). *Linear Mathematics*. New York: W. A. Benjamin, Inc. xii + 347 pp.
  - 6. Fred Brauer (1976). *Some Stability and Perturbation Problems for Differential and Integral Equations*, Monografias de Matemática no. 25. Rio de Janeiro: Instituto de Matemática Pura e Aplicada. iii + 163 pp.
  - 7. Fred Brauer and John A. Nohel (1985). *An Introduction to Differential Equations with Applications*. New York: Harper & Row. xii + 620 pages.
  - 8. Fred Brauer and Carlos Castillo-Chávez (2001). *Mathematical Models in Population Biology and Epidemiology*. (Texts in Applied Mathematics 40), Springer-Verlag, New York, (c) 2001, 416 pages, ISBN 0-387-98902-1.

KENNETH L. COOKE:  
RESEARCHER, EDUCATOR PAR EXCELLENCE

P. VAN DEN DRIESEN\*



**Kenneth L. Cooke<sup>†</sup>**

It was fifty years ago that Kenneth Cooke (Ken) started as an instructor at the State College of Washington while studying for his PhD degree at Stanford University. From there he went to Pomona College, Claremont, where he advanced through the ranks, and where he is now Professor Emeritus. During this time he has been Chairman of the Department, and he has held research positions and visited many institutions, including Brown University, IMA Minnesota, Cornell University, England, Italy, Brazil and Mexico.

Ken's professional work during these years has had a profound effect on the areas of functional differential equations (especially delay-differential

\*Department of Math and Stats, University of Victoria, Victoria, British Columbia, V8W 3P4 Canada.

<sup>†</sup>Mathematics Department, Pomona College, Claremont, CA 91711. Photograph is a courtesy of J.L. Harlan.

equations) and population biology (especially epidemic models). He has published five books and co-edited four volumes on these topics. His 1963 book *Differential-difference Equations*, co-authored with R. Bellman (his PhD advisor), has become a classic. His 1993 book *Vertically Transmitted Diseases*, co-authored with S. Busenberg, provides a marvellous single source for work in this recently developed area.

Sometime in the late 1960s, Ken's interests in differential equations expanded to include classes of equations that can be interpreted as models of biological processes. His 1967 paper *Functional differential equations: some models and perturbation problems* emphasized the "threshold" type equations and state dependent delays, with biological models as examples. His seminal 1973 paper with J. Yorke (which has been influential among modelers) studies the delay-differential equation

$$x' = g[x(t)] - g[x(t - L)]$$

and its generalization that allows for a distribution of life spans

$$x'(t) = g[x(t)] + \int_0^L g[x(t - s)]P'(s)ds.$$

Here  $x(t)$  is the number of individuals in a population at time  $t$ ,  $g[x(t)]$  is the number of births per unit time,  $P(a)$  is the proportion surviving to age  $a$ , and  $L$  is the maximum life span. Under realistic assumptions on  $g$  and  $P$ , they proved that any bounded solution approaches a limit as  $t \rightarrow \infty$ . More recently he has worked on various models of HIV/AIDS with several collaborators, including D.A. Allers, S. Busenberg, C. Castillo-Chavez, Y.-H. Hsieh, W. Huang (a former PhD student), S.A. Levin and H. Thieme.

Bellman and Cooke 1963 was one of my "bibles" when I met Ken at the University of Utah in 1978 at a summer school on Nonlinear Oscillations in Biology. One morning in the cafeteria, I plucked up courage to talk to this Professor whose book I so much admired, and was surprised to find him both approachable and enthusiastic about ideas. This conversation led to collaboration, and our first joint paper in 1986 on delay-differential equations. Lest it be misunderstood, it is not Ken who works slowly, but there were delays in our work! Since that time, I have come to value Ken as a collaborator and friend. He embodies superb qualities as a mathematician, which he applies to biological systems. On a more personal level, he shows patience, understanding and obviously enjoys many collaborations. He and his wife Margaret are very generous in hosting guests who visit Claremont to give talks or to attend meetings.

In January 1990, nearly two hundred research workers participated in an international meeting in Claremont for the celebration of Ken's 65<sup>th</sup> birthday. From this conference, two volumes of papers were published. One, devoted to models in mathematical biology, epidemiology and ecology, is published as volume 92 in the series Lecture Notes in Biomathematics.

The second, devoted to delay-differential equations, is published in the companion series Lecture Notes in Mathematics. The breadth, depth and interdisciplinary scope of Ken's work is well mirrored in these two volumes, and in the list of Ken's refereed articles given below.

In 1994, Ken was a member of the organizing committee of the Claremont International Conference "Differential Equations and Applications to Biology and to Industry". This was dedicated to the memory of his collaborator and great friend, Stavros Busenberg, who had tragically died the previous year. Ken gave generously in both time and effort to ensure the tremendous success of this conference and its proceedings.

This past year, Ken has been Chair of the Southern California Section of the Mathematical Association of America. This position fits his role as educator and leader. He has presented uncountably many excellent lectures, which are full of insight, enthusiasm and inspiration. He has advised six PhD students at Claremont Graduate University and been a mentor to countless other students and colleagues. Ken's humility and deep respect for all individuals and cultures is unparalleled. Hence, Ken's accomplishments cannot be completely understood without providing an explicit and typical example of the type of mentoring that he has carried out over the last fifty years. Carlos Castillo-Chavez vividly remembers his first interactions with Ken while both shared an office at Cornell in 1987. Carlos was a postdoc working with S.A. Levin when Ken came to spend his sabbatical leave. Ken almost immediately, actually on the second day, instigated a collaboration (that still continues) on the transmission dynamics of HIV. The long and variable period of incubation of HIV naturally brought the use of functional differential equations into play. *The expert* taught Carlos functional differential equations during their working sessions, and now, many years later, Carlos realizes that Ken taught him a lot more than functional differential equations.

For his role as researcher, educator, and for his generous, warm personality we make this dedication to Kenneth L. Cooke.

### **Refereed Articles of Kenneth L. Cooke**

- (1953): The asymptotic behavior of the solutions of linear and non-linear differential-difference equations. *Trans. Amer. Math. Soc.*, **75**, 80-105.
- (1954): The rate of increase of real continuous solutions of algebraic differential-difference equations of the first order. *Pacific J. Math.*, **4**, 483-501.
- (1955): A non-local existence theorem for systems of ordinary differential equations. *Rend. Circ. Mat. Palermo, Ser. II, Tomo 4*, 301-308.
- (1955): Forced periodic solutions of a stable non-linear differential-difference equation. *Annals of Math.* **61**, 387-391.
- (1957): Hadamard matrices. With J.L. Brenner, *Rev. Ci. Lima* **59**, 5-13.
- (1958): A symbolic method for finding integrals of linear difference and

- differential-difference equations. *Math. Magazine*, 121-126.
- (1959): Stability theory and adjoint operators for linear differential-difference equations. With R. Bellman, *Trans. Amer. Math. Soc.* **92**, 470-500.
- (1959): On the limit of solutions of differential-difference equations as the retardation approaches zero. With R. Bellman, *Proc. Nat. Acad. Sci.* **45**, 1026-1028.
- (1959): The rate of increase of real continuous solutions of certain algebraic functional equations. *Trans. Amer. Math. Soc.* **92**, 106-124.
- (1960): Stability and asymptotic theory for linear differential-difference equations. *Proc. Symposium on Ordinary Differential Equations at Mexico City, September 1959*, Boletin de la Sociedad Mat. Mexicana, 277-283.
- (1962): On transcendental equations related to differential-difference equations. *J. Math. Anal. Appl.* **4**, 65-71.
- (1963): Differential-difference equations, in *Nonlinear Differential Equations and Nonlinear Mechanics*, J.P. LaSalle and S. Lefschetz, eds., Academic Press, 155-171.
- (1965): Existence and uniqueness theorems in invariant imbedding - I: conservation principles. With R. Bellman, R. Kalaba, G.M. Wing, *J. Math. Anal. Appl.* **10**, 243-244.
- (1965): Existence and uniqueness theorems in invariant imbedding - II: convergence of a new difference algorithm. With R. Bellman, *J. Math. Anal. Appl.* **12**, 247-253.
- (1965): The condition of regular degeneration for singularly perturbed linear differential-difference equations. *J. Differential Equations* **1**, 39-94.
- (1965): Convergence of successive approximations in the shortest route problem. With D.L. Bentley, *J. Math. Anal. Appl.* **10**, 269-274.
- (1965): On the computational solution of a class of functional differential equations. With R. Bellman, *J. Math. Anal. Appl.* **12**, 495-500.
- (1966): The condition of regular degeneration for singularly perturbed systems of linear differential-difference equations. With K.R. Meyer, *J. Math. Anal. Appl.* **14**, 83-106.
- (1966): The shortest route through a network with time-dependent internodal transit times. With E. Halsey, *J. Math. Anal. Appl.* **14**, 493-498.
- (1966): Graphical solution of difficult crossing puzzles. With P. Detrick, R. Fraley, *Math. Mag.* **39**, 151-157.
- (1966): Functional differential equations close to differential equations. *Bull. Amer. Math. Soc.* **72**, 285-288.
- (1967): Functional differential equations with asymptotically vanishing lag. *Rend. Circ. Mat. di Palermo, series II*, XVI, 39-56.
- (1967): Functional differential equations: some models and perturbation problems, in *Differential Equations and Dynamical Systems*, J.K. Hale and J.P. LaSalle, eds., Academic Press, 167-183.
- (1967): Asymptotic theory for the delay-differential equation  $u'(t) = -au(t - r(u(t)))$ . *J. Math. Anal. Appl.* **19**, 160-173.
- (1967): Some recent work on functional-differential equations, in *Proc.*

- U.S. Japan Seminar on Differential and Functional Equations*, W.A. Harris, Jr., Y. Sibuya, eds., W.A. Benjamin, 27-47.
- (1968): Difference-differential equations and nonlinear initial-boundary value problems for linear hyperbolic partial differential equations. With D. Krumme, *J. Math. Anal. Appl.* **24**, 372-387.
- (1969): The Königsberg bridges problem generalized. With R. Bellman, *J. Math. Anal. Appl.* **25**, 1-7.
- (1970): Linear functional differential equations of asymptotically autonomous type. *J. Diff. Eq.* **7**, 154-174.
- (1971): A linear mixed problem with derivative boundary conditions. *Seminar on Differential Equations and Dynamical Systems, III*, D. Sweet and J.A. Yorke, eds., University of Maryland, College Park, 11-17.
- (1972): Equations modelling population growth, economic growth, and gonorrhea epidemiology. With J. Yorke, 1971 NRL-MRC Conference, *Ordinary Differential Equations*, L. Weiss, ed., Academic Press, New York.
- (1973): Some equations modelling growth processes and gonorrhea epidemics. With J. Yorke, *Math. Biosciences* **16**, 75-101.
- (1973): On a class of hereditary processes in biomechanics. With N. Distefano, B. Kashef, *Math. Biosciences* **16**, 359-373.
- (1974): A linear hyperbolic problem all of whose solutions are constant after finite time. *SIAM J. Math. Anal.* **5**, No. **3**, 482-488.
- (1975): Asymptotic equivalence of an ordinary and a functional differential equation. *J. Math. Anal. Appl.* **51**, 187-207, Abstract in *International Conference on Differential Equations*, H.A. Antonsiewicz, ed., Academic Press, New York, 808-809.
- (1975): A discrete-time epidemic model with classes of infectives and susceptibles. *Theor. Population Biol.* **7**, 175-196.
- (1975): A discrete-time epidemic model with classes of infectives and susceptibles (Summary). *Epidemiology*, Proceedings of SIMS Conference, D. Ludwig, K. Cooke, eds., SIAM, Philadelphia, 132-138.
- (1976): An epidemic equation with immigration. *Math. Biosciences* **29**, 135-158.
- (1976): A periodicity threshold theorem for epidemics and population growth. With J.L. Kaplan, *Math. Biosciences* **31**, 87-104.
- (1977): Stability or chaos in discrete epidemic models. With D.F. Calef, E.V. Level, *Nonlinear Systems and Applications*, V. Lakshmikantham, ed., Academic Press, New York, 73-93.
- (1978): Periodic solutions of a periodic nonlinear delay differential equation. With S. Busenberg, *SIAM J. Applied Math.* **35**, 704-721.
- (1979): Periodic solutions of delay differential equations arising in some models of epidemics. With S. Busenberg, *Applied Nonlinear Analysis*, V. Lakshmikantham, ed., Academic Press, New York, 67-78.
- (1979): Mathematical approaches to culture change, in *Transformations: Mathematical Approaches to Culture Change*, C. Renfrew, K. Cooke, eds., Academic Press, New York, 45-81.

- (1979): Stability analysis for a vector disease model. *Rocky Mountain J. of Math.* **9**, 31-42.
- (1979): An experiment on the simulation of culture changes. With C. Renfrew, in *Transformations: Mathematical Approaches to Culture Change*, C. Renfrew, K. Cooke, eds., Academic Press, New York, 327-348.
- (1980): The effect of integral conditions in certain equations modelling epidemics and population growth. With S. Busenberg, *J. Math. Biology* **10**, 13-32.
- (1981): On the construction and evaluation of mathematical models, in *Simulations in Archaeology*, J.A. Sabloff, ed., Univ. New Mexico Press, Albuquerque.
- (1981): Stability of a functional differential equation for the motion of a radiating charged particle. With J.L. Kaplan, M. Sorg, *Nonlinear Analysis: TMA* **5**, 1133-1139.
- (1981): Delay differential equations, in *Mathematics of Biology*, M. Iannelli, ed., Liguori editore, Napoli, 5-80.
- (1982): Vertically transmitted diseases. With S. Busenberg, in *Nonlinear Phenomena in the Mathematical Sciences*, V. Lakshmikantham, ed., Academic Press, New York, 189-197.
- (1982): Models of vertically transmitted diseases with sequential continuous dynamics. With S. Busenberg, in *Nonlinear Phenomena in the Mathematical Sciences*, V. Lakshmikantham, ed., Academic Press, New York, 179-187.
- (1982): Discrete delay, distributed delay and stability switches. With Z. Grossman, *J. Math. Anal. Appl.* **86**, 592-627.
- (1982): Models for endemic infections with asymptomatic cases I. One group. *Mathematical Modelling* **3**, 1-15.
- (1983): Analysis of a model of a vertically transmitted disease. With S. Busenberg, M.A. Pozio, *J. Math. Biol.* **17**, 305-329.
- (1983): Stability conditions for linear retarded functional differential equations. With J.M. Ferreira, *J. Math. Anal. Appl.* **96**, 480-504.
- (1984): Distributional and analytic solutions of functional differential equations. With J. Wiener, *J. Math. Anal. Appl.* **98**, 111-129.
- (1984): Retarded differential equations with piecewise constant delays. With J. Wiener, *J. Math. Anal. Appl.* **99**, 265-297.
- (1984): Stability conditions for linear non-autonomous delay differential equations. With S. Busenberg, *Quart. Appl. Math.*, **42**, 295-306.
- (1984): Mathematical models of vertical transmission of infection, in *Mathematical Ecology*, S.A. Levin, T.G. Hallam, eds., Lecture Notes in Biomathematics **54**, Springer-Verlag, 344-355.
- (1984): Stability of non-autonomous delay differential equations by Liapunov functionals, in *Infinite-Dimensional Systems*, F. Kappel, W. Schappacher, eds., Lecture Notes in Mathematics **1076**, Springer-Verlag.
- (1985): Infection models with asymptomatics, in *Mathematics and Computers in Biomedical Applications*, J. Eisenfeld, C. De Lisi, eds., Elsevier

Science Publ., 277-282.

- (1985): Stability of delay differential equations with applications in biology and medicine, in *Mathematics in Biology and Medicine*, V. Capasso, E. Grosso, S.L. Paveri-Fontana, eds., Lecture Notes in Biomathematics **57**, Springer-Verlag, 439-446.
- (1986): Stability regions for linear equations with piecewise continuous delay. With J. Wiener, Comp. & Math. with Appl. **12A**, 695-701.
- (1986): One-dimensional linear and logistic harvesting models. With M. Witten, Math. Modelling **7**, 301-340.
- (1986): On zeroes of some transcendental equations. With P. van den Driessche, Funkcialaj Ekvacioj **29**, 77-90.
- (1987): An equation alternately of retarded and advanced type. With J. Wiener, Proc. Amer. Math. Soc. **99**, 726-732.
- (1987): Neutral differential equations with piecewise constant arguments. With J. Wiener, Bollettino UMI **7**, 321-346.
- (1987): Analysis of the complicated dynamics of some harvesting models. With H.E. Nusse, J. Math Biol. **25**, 521-542.
- (1988): Harvesting procedures with management policy in iterative density-dependent population models. With R. Elderkin, M. Witten, Natural Resource Modeling **2**, No.3, 383-420.
- (1988): A nonlinear equation with piecewise continuous argument. With L.A.V. Carvalho, Diff. and Integral Eqns. **1**, 359-367.
- (1988): The population dynamics of two vertically transmitted infections. With S. Busenberg, Theoretical Population Biology **33**, 181-198.
- (1988): Endemic thresholds and stability in a class of age-structured populations. With S. Busenberg, M. Iannelli, SIAM J. Applied. Math. **48**, 1379-1395.
- (1988): Oscillation in systems of differential equations with piecewise constant argument. With J. Wiener, J. Math. Anal. Appl. **137**, 221-239.
- (1989): On the role of long incubation periods in the dynamics of acquired immunodeficiency syndrome (AIDS) Part 1: Single population models. With C. Castillo-Chavez, W. Huang, S.A. Levin, J. Math. Biol. **27**, 373-398.
- (1989): Stability and thresholds in some age-structured epidemics. With S. Busenberg and M. Iannelli, in *Mathematical Approaches to Problems in Resource Management and Epidemiology*, C. Castillo-Chavez, S.A. Levin, C.A. Shoemaker, eds., Lecture Notes in Biomathematics **81**, Spring-Verlag, 124-141.
- (1989): Results on the dynamics for models of the sexual transmission of the human immunodeficiency virus. With C. Castillo-Chavez, W. Huang, S.A. Levin, Appl. Math. Lett. **2**, No. 4, 327-331.
- (1989): On the modelling of epidemics. With C. Castillo-Chavez and S.A. Levin, in *High Performance Computing*, J.L. Delhaye, E. Glenebe, eds., North-Holland, Amsterdam-New York-Oxford-Tokyo, 389-402.
- (1989): The role of long periods of infectiousness in the dynamics of ac-

quired immunodeficiency syndrome (AIDS). With C. Castillo-Chavez, W. Huang, S.A. Levin, in *Mathematical Approaches to Problems in Resource Management and Epidemiology*, C. Castillo-Chavez, S.A. Levin, C.A. Shoemaker, eds., Lecture Notes in Biomathematics **81**, Springer-Verlag, 177-189.

(1989): On the role of long incubation periods in the dynamics of acquired immunodeficiency syndrome (AIDS) Part 2: Multiple group models. With C. Castillo-Chavez, W. Huang, S.A. Levin, in *Mathematical and Statistical Approaches to AIDS Epidemiology*, Lecture Notes in Biomathematics **83**, C. Castillo-Chavez, ed., 200-217.

(1990): Coexistence of analytic and distributional solutions for linear differential equations, I. With J. Wiener, *J. Math. Anal. Appl.* **148**, 390-421.

(1991): On dichotomic maps for a class of differential-difference equations. With L.A.V. Carvalho, *Proc. Roy. Soc. Edinburgh* **117A**, 317-328.

(1991): Mixing patterns in models of AIDS. With D.A. Allers and C. Castillo-Chavez, in *Mathematical Population Dynamics*, O. Arino, D.E. Axelrod, M. Kimmel, eds., Lecture Notes in Pure and Applied Mathematics **131**, Marcel Dekker, New York-Basel-Hong Kong, 297-309.

(1991): Demographic change and persistence of HIV/AIDS in a heterogeneous population. With S. Busenberg and H. Thieme, *SIAM J. Applied Math* **51**, 1030-1052.

(1991): Coexistence of analytic and distributional solutions for linear differential equations, II. With J. Wiener and S.M. Shah, *J. Math. Anal. Appl.* **159**, 271-289.

(1991): A survey of differential equations with piecewise continuous arguments. With J. Wiener, in *Delay Differential Equations and Dynamical Systems*, Lecture Notes in Mathematics **1475**, S. Busenberg and M. Martelli, eds., 1-15.

(1992): Stability and bifurcation for a multiple-group model for the dynamics of HIV/AIDS transmission. With W. Huang, C. Castillo-Chavez, *SIAM J. Appl. Math.* **52**, 835-854.

(1992): A theorem of George Seifert and an equation with state-dependent delay. With Wenzhang Huang, in *Delay and Differential Equations* (Proceedings in Honor of George Seifert on his Retirement), ed. A.M. Fink, Richard K. Miller, and Wolfgang Kliemann. World Scientific, Singapore-New Jersey-London-Hong Kong, 65-77.

(1992): Liapunov functionals and stability of certain differential-difference equations. With L.A.V. Carvalho, in *Functional Differential Equations*, T. Yoshizawa and J. Kato, eds., World Scientific, Singapore-New Jersey-London-Hong Kong, 32-42.

(1993): Spectral conditions and an explicit expression for the stabilization of hybrid systems in the presence of feedback delays. With J. Turi and G. Turner, *Quarterly Appl. Math.* **51**, 147-159.

(1993): Distributional and small solutions for linear time-dependent delay equations. With S.M. Verduyn Lunel, *Differential and Integral Equations*,

1101-1117.

- (1994): Numerical approximation of the solutions of delay differential equations on an infinite interval using piecewise constant arguments. With I. Györi, Computers Math. Applic. **28**, 81-92.
- (1994): Stability, instability in delay equations modeling human respiration. With J. Turi, J. Math. Biol. **32**, 535-543.
- (1995): A Model for HIV in Asia. With S. Busenberg and Y.-H. Hsieh, Math. Biosci. **128**, 185-210.
- (1995): Remembering Stavros Busenberg. With C. Castillo-Chavez and H. Thieme, Math. Biosci. **128**, 3-11.
- (1995): Two problems on differential-delay equations, in *Ordinary Differential Equations and their Applications*, Proceedings of International Meeting in Firenze, Italy, Sept. 20-24. 1993, published in Bologna.
- (1995): A boundary value problem for a functional-differential equation with a linearly transformed argument. With L.E. Rossovskii and A.L. Skubachevskii. Differential Equations **31**, 1294-1299. (Translated from Differentsial'nye Uravneniya **31** (August, 1995), 1348-1352.)
- (1996): On the sharpness of a theorem by Cooke and Verduyn Lunel. With G. Derfel, J. Math. Anal. Appl. **197**, 379-391.
- (1996): Applying Carvalho's method to find periodic solutions of difference equations. With L.A.C. Ladeira, J. Difference Equations & Applications, **2**, 105-115.
- (1996): Modeling antiviral immune response: a negative feedback control approach. With B. Tang and D. Babai, Int. J. of Appl. Sc. & Computations, **3** no. 1, 48-56.
- (1996): On the problem of linearization for state-dependent delay differential equations. With W. Huang. Proc. Amer. Math. Soc. **124**, 1417-1426.
- (1996): Analysis of an SEIRS epidemic model with two delays. With P. van den Driessche. J. Math. Biol. **35**, 240-260.
- (1997): Three lectures on mathematical modeling of diseases, in Three Lectures, Colóquio de Equações Diferenciais, L. A. V. Carvalho and L.A.V. Ladeira, eds. São Carlos, Brazil.
- (1998): Collapsible backward continuation and numerical approximations in a functional differential equation. With L. A.V. Carvalho. J. Differential Equations **143**, 96-109.
- (1998): Analyses of an antiviral immune response model with time delays. With Y. Kuang and B. Li. Canadian Appl. Math. Quarterly **6** no. 4, 321-354.
- (1999): Interaction of maturation delay and nonlinear birth in population and epidemic models. With P. van den Driessche and X. Zou. J. Math. Biology **39**, 332-352.
- (1999): On periodic solutions for a class of linear scaled differential equations I. With L.A.V. Carvalho and L.A.C. Ladeira. Communications in Applied Analysis **3**, no. 3, 399-413.
- (1999): On periodic solutions for a class of linear scaled differential equa-

- tions II. With L.A.V. Carvalho and L.A.C. Ladeira. Communications in Applied Analysis **3**, no. 3, 415-431.
- (2000): On the discretization of a delay differential equation. With A. F. Ivanov. Journal of Difference Equations and Applications, **6**, 105-119.
- (2000): Behaviour change and treatment of core groups: its effect on the spread of HIV/AIDS. With Ying-Hen Hsieh. IMA Journal Math. Appl. Med. Biol., Vol. 17: 213-241.

# MAXIMAL PREVALENCE AND THE BASIC REPRODUCTION NUMBER IN SIMPLE EPIDEMICS

L. ESTEVA\* AND K.P. HADELER†

**Abstract.** For the basic versions of the SIR and SIRS epidemic models estimates for the maximal prevalence are computed in terms of the basic reproduction number and other relevant quantities. Maximal prevalence is studied as a function of the rate of loss of immunity. Total size and total cost (total infected time) of the epidemic are estimated. For SIR models with demographic renewal it is investigated whether prevalence is a monotone function of the renewal rate.

**1. Maximal prevalence and basic reproduction number.** We consider simple epidemics in terms of the basic SIR or SIRS models for constant population size. There seems to be a general view that the degree of severity of an epidemic can be estimated either from the basic reproduction number  $R_0$  or from the prevalence and that these two quantities are closely related. Of course  $R_0$  has become ubiquitous in epidemic modelling after its introduction, Hethcote [18], Diekmann, Heesterbeek, Metz [12], see also [13], and Mollison [21] for stochastic models. Maximal prevalence is less prominent. It has been discussed in terms of stochastic models by Daniels [11]. Since  $R_0$  describes a stability property of the uninfected equilibrium and maximal or endemic prevalence is observed far from that equilibrium, there will be, generally speaking, no connection between these two quantities. In fact models with nonlinear transmission functions can produce almost any behavior, see, e.g. Liu et al. [20], Hethcote et al. [19]. The situation may be different within a fixed framework of basic models. The aim of the present note is to elucidate such connections. These may be useful since in practice the basic reproduction number as well as maximal prevalence are difficult to estimate due to delayed reporting etc.

The basic reproduction number  $R_0$  is the average number of new cases produced by one infected in a totally susceptible population. The prevalence measures the number of cases during the course of the epidemic or at an endemic equilibrium. In the following we shall clearly distinguish between maximal prevalence  $I^{max}$  and endemic prevalence  $\bar{I}$ . Maximal prevalence  $I^{max}$  is the maximal number of infecteds observed at any time during the time course of an outbreak starting from a totally susceptible population. If a (unique) endemic equilibrium exists then the level of infecteds at that equilibrium is the endemic prevalence. Of course maximal prevalence cannot be lower than endemic prevalence. But it is obvious, at least for cases where the prevalence oscillates near equilibrium, that

---

\*Departamento de Matemáticas, Facultad de Ciencias, UNAM, México, D.F. 04510;  
lesteva@servidor.unam.mx.

†Biomathematics, University of Tübingen, Auf der Morgenstelle 10, D-72076  
Tübingen; hadeler@uni-tuebingen.de.

maximal prevalence can be (much) higher than endemic prevalence. On biological grounds, maximal prevalence should be larger in the SIRS case than in the SIR case, because the additional supply of susceptibles from the  $R$  compartment leads to a higher number of infecteds. Later we shall even see that maximal prevalence depends on the reentrance rate in a monotone way.

The basic reproduction number  $R_0$  is a standardized bifurcation parameter describing the stability of the uninfected state and the possible bifurcation of an infected state from the uninfected state. The inequality  $R_0 < 1$  guarantees the local stability of the uninfected state, for  $R_0 > 1$  the uninfected state is unstable. The number  $R_0$  does not contain *any* information on the direction or slope of the bifurcation. Bifurcation can be forward (supercritical) or backward (subcritical). Of course, in standard models bifurcations are forward but in multigroup models backward bifurcation is a common phenomenon, see, e.g., Hadeler and Castillo-Chavez [16], Hadeler and van den Driessche [17]. In the epidemic case the basic reproduction number determines the possibility of an outbreak. If  $R_0 < 1$  then the number of infectious initially present will decrease, if  $R_0 > 1$  then it will increase.

We assume that the time scale of the epidemic is fast as compared to changes in population size, i.e., we assume constant population size normalized to 1. We consider the simplest version of the basic epidemic model

$$(1) \quad \begin{aligned} \dot{S} &= -\beta SI + \gamma R \\ \dot{I} &= \beta SI - \alpha I \\ \dot{R} &= \alpha I - \gamma R, \end{aligned}$$

which is, for a normalized total population  $S+I+R=1$ , equivalent with the system

$$(2) \quad \begin{aligned} \dot{S} &= -\beta SI + \gamma(1 - S - I) \\ \dot{I} &= \beta SI - \alpha I, \end{aligned}$$

to be studied in the triangle defined by the inequalities  $S \geq 0$ ,  $I \geq 0$ ,  $S + I \leq 1$ .

Here  $\beta > 0$  is the transmission rate,  $\alpha > 0$  is the recovery rate, and  $\gamma \geq 0$  is the reentrance rate (rate of loss of immunity). Thus, we have an SIRS model for  $\gamma > 0$ , and an SIR model in the limiting case  $\gamma = 0$ .

Usually the SIR and the SIRS cases are considered as describing different phenomena. The SIR model describes a simple passage of an epidemic through a population, leaving behind a number of susceptibles that is too small to sustain an epidemic. This is a typical situation for influenza. The SIRS model is connected to the idea of loss of immunity which in practice may result from the fact that there are several types of the infectious agent.

An example is common cold which is related to about a hundred different serotypes of virus, many of them types of rhinovirus. Bacterial infections like gonorrhoea also confer temporal immunity.

The SIR model is a structurally unstable system describing temporary outbreaks, the SIRS system is structurally stable (for  $R_0 \neq 1$ ) and it describes convergence to equilibrium. The present note shows that it makes sense to study the SIR and the SIRS system together: Although the phase portraits look very different, the initial part of the trajectory describing an outbreak depends monotonely on the rate  $\gamma$ .

The basic reproduction number for this model is

$$(3) \quad R_0 = \frac{\beta}{\alpha}.$$

The basic reproduction number does not depend on  $\gamma$ . This fact just indicates that at the onset of an epidemic the number of reentrance cases is negligibly small in comparison to the number of susceptibles.

In either case,  $\gamma > 0$  or  $\gamma = 0$ , if  $R_0 > 1$  then there is a distinguished trajectory  $M_\gamma$  ( $M_0$  in case  $\gamma = 0$ ) of the system (2) starting from the uninfected equilibrium  $(S, I) = (1, 0)$ , see Fig.1. The distinguished trajectory describes an epidemic that evolves from a very small group of infecteds ( $I = 0$ ) in a totally susceptible population ( $S = 1$ ). Since the trajectory starts at a stationary point, time must begin with  $t = -\infty$  rather than with  $t = 0$ . In the case  $\gamma > 0$  the point  $(1, 0)$  is an isolated stationary point. It is nondegenerate for  $R_0 \neq 1$ . For  $R_0 < 1$  it is a stable node and it attracts all trajectories. For  $R_0 > 1$  it is a saddle point. For  $\gamma = 0$  the point  $(1, 0)$  is one element in a continuum of stationary points of the form  $(S, 0)$ . If  $R_0 < 1$  then there is no unstable manifold. There is a single trajectory with  $I > 0$  approaching this point. If  $R_0 > 1$  then the point  $(1, 0)$  has an unstable manifold consisting of two trajectories, one with  $I > 0$ , which is  $M_0$ , and one with  $I < 0$ . The latter is of no interest in the present context.

Thus, in either case,  $\gamma > 0$  or  $\gamma = 0$ , for  $R_0 > 1$  there is a one-dimensional unstable manifold  $M_\gamma$  emanating from  $(S, I) = (1, 0)$  into  $I > 0$ . This trajectory describes the time course of the epidemic at the onset and in the sequel.

In the SIRS case the endemic equilibrium may be a node or a focus. From the general features of a transcritical bifurcation the following is obvious. If  $\alpha$  and  $\gamma$  are kept fixed then the stationary point is a node for  $R_0$  close to 1. Later we shall see that it becomes a focus when  $R_0$  is increased and that it changes back to a node for very large  $R_0$ . If the endemic state is a focus then  $M_\gamma$  winds around that state (counterclockwise) infinitely often. But even if the point is a node then  $M_\gamma$  may wind around the stationary point finitely many times. Whether, in the case of a node, this will happen, cannot be determined from local stability analysis.

In the SIR case, where there is no endemic equilibrium, for  $R_0 > 1$  the unstable manifold  $M_0$  eventually arrives at another stationary point  $(S_\infty, 0)$

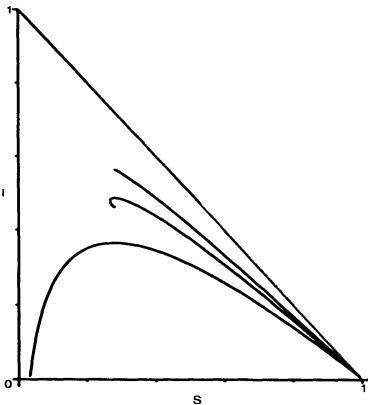


FIG. 1. *Unstable manifold  $M_\gamma$  of the saddle point  $(1,0)$  for  $\gamma = 0$ ,  $\gamma = 0.25$ ,  $\gamma = 0.5$ . The values of the other parameters are  $\beta = 0.5$ ,  $\alpha = 0.14$ . Then  $R_0 = 3.5$ ,  $S^* = 0.285$ .*

where the final size of the susceptible class  $S_\infty \in (0, 1)$  is the solution of the transcendental equation

$$(4) \quad S - \frac{1}{R_0} \log S = 1.$$

Along  $M_0$  the prevalence first increases from 0 to a maximum and then it decreases to 0. At maximal prevalence we have  $\dot{I} = 0$ . From the differential equation (2) we see  $\beta SI - \alpha I = 0$ , which, with  $I > 0$ , gives the explicit expression

$$(5) \quad S = \frac{\alpha}{\beta} = \frac{1}{R_0}.$$

Hence we have shown the following proposition.

**Proposition 1.** *In the SIR and in the SIRS model, maximal prevalence is the  $I$ -coordinate of the first intersection point of the unstable manifold  $M_\gamma$  with the line  $S = \alpha/\beta$ , or (this can occur only for SIRS) it is equal to the endemic prevalence in case  $M_\gamma$  does not reach the line  $S = \alpha/\beta$  in finite time.*

For  $\gamma > 0$  the infected stationary state  $(\bar{S}, \bar{I})$  is given by the formula

$$(6) \quad \bar{S} = \frac{\alpha}{\beta}, \quad \bar{I} = \frac{\gamma}{\alpha + \gamma} \left(1 - \frac{\alpha}{\beta}\right).$$

Thus, the infected state is feasible (i.e., positive) if and only if  $R_0 > 1$ .

From formula (6) we see that the endemic prevalence cannot be expressed as a function of  $R_0$  alone. If we fix the recovery rate  $\alpha$  and the rate

$\gamma$  for loss of immunity (or rather the quotient  $\gamma/\alpha$ ) then the equilibrium prevalence is a strictly increasing function of  $R_0$ ,

$$(7) \quad \bar{I} = \bar{I}(R_0) = \frac{\gamma}{\alpha + \gamma} \left(1 - \frac{1}{R_0}\right).$$

In a bifurcation diagram for  $\bar{I}$  versus  $R_0$  the bifurcation curve passes through the point  $R_0 = 1$ ,  $\bar{I} = 0$  with the slope

$$(8) \quad \frac{d\bar{I}}{dR_0} = \frac{\gamma}{\alpha + \gamma} > 0.$$

The same quantity  $\gamma/(\alpha + \gamma)$  is also the limit of  $\bar{I}$ , for large values of  $R_0$ . At any rate, the expression (7) is a lower bound for the maximal prevalence,  $I_{\gamma}^{max} \geq \bar{I}$  where the subscript  $\gamma$  indicates the dependence on  $\gamma$ .

Now consider the limiting case, the SIR model, i.e.,  $\gamma = 0$ . Then the system (2) has an invariant of motion

$$(9) \quad V(S, I) = S - \frac{\alpha}{\beta} \log S + I.$$

For the distinguished trajectory we have  $S(-\infty) = 1$ ,  $I(-\infty) = 0$ , and hence

$$(10) \quad I = 1 - S + \frac{\alpha}{\beta} \log S.$$

At maximal prevalence we have  $S = \alpha/\beta$  and hence

$$(11) \quad I_0^{max} = I_0^{max}(R_0) = 1 - \frac{1}{R_0} - \frac{1}{R_0} \log R_0$$

where the subscript 0 in  $I_0^{max}$  refers to  $\gamma = 0$ .

For the derivatives we find

$$(12) \quad \begin{aligned} \frac{dI_0^{max}}{dR_0} &= \frac{1}{R_0^2} \log R_0, & \frac{d^2I_0^{max}}{dR_0^2} &= \frac{1}{R_0^3} (1 - 2 \log R_0), \\ \frac{dI_0^{max}}{dR_0} \Big|_{R_0=1} &= 0, & \frac{d^2I_0^{max}}{dR_0^2} \Big|_{R_0=1} &= 1. \end{aligned}$$

Thus, for the epidemic or SIR case, the function  $I_0^{max}(R_0)$  looks very different from  $\bar{I}(R_0)$  as given by (7). The graph of  $\bar{I}(R_0)$  is a simple hyperbola. The graph of  $I_0^{max}(R_0)$  is shown in Fig.2 for two different scales. We state the following proposition.

**Proposition 2.** *For the SIR model the maximal prevalence  $I_0^{max}$  is zero for  $R_0 \leq 1$ , then it increases with zero slope, it continues in a sigmoid fashion with a single inflection point at  $R_0 = \sqrt{e}$ , at which point  $I_0^{max} = 1 - 3/(2\sqrt{e})$ , finally it approaches the value 1.*

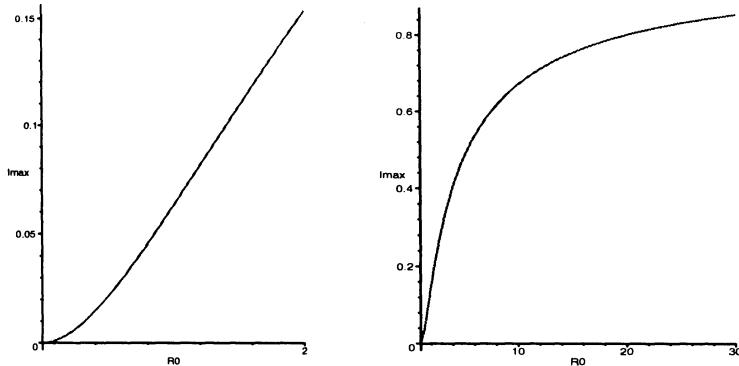


FIG. 2. The function  $I_0^{max}(R_0)$  for the ranges  $1 \leq R_0 \leq 2$  and  $1 \leq R_0 \leq 30$ . The function has zero slope at  $R_0 = 1$ , then increases about linearly and eventually approaches 1.

In view of the zero slope at  $R_0 = 1$ , the basic reproduction number  $R_0$  may clearly exceed 1 and still produce a rather low maximal prevalence  $I_0^{max}$ . For large  $R_0$  this effect does not play a role (the inflection point is at  $R_0 = 1.65$ ).

**2. Maximal prevalence and loss of immunity.** Now we fix  $R_0$  and  $\alpha$  and consider the maximal prevalence  $I^{max} = I_\gamma^{max}$  as a function of  $\gamma$ . We show the following monotonicity result which verifies the conjecture expressed before.

**Proposition 3.** *For the SIRS model the maximal prevalence  $I_\gamma^{max}$  is an increasing function of  $\gamma$  for  $0 \leq \gamma < \infty$ . In particular,  $I_0^{max}$ , as given by (11), is a lower bound for any  $I_\gamma^{max}$ .*

**Proof.** Choose some  $\bar{\gamma} > 0$ . Consider the unstable manifold  $M_{\bar{\gamma}}$  of the saddle point  $(1, 0)$  extended into the first orthant. The tangent vector along this manifold is

$$(13) \quad (-\beta SI + \bar{\gamma}(1 - S - I), \beta SI - \alpha I)^T.$$

Hence a normal vector along  $M_{\bar{\gamma}}$  is

$$(14) \quad (\beta SI - \alpha I, \beta SI - \bar{\gamma}(1 - S - I))^T.$$

Now, we restrict to the domain  $S > \alpha/\beta$ ,  $I > 0$ ,  $I + S < 1$ . In this domain the first component of the normal vector is positive. Now choose  $\gamma > \bar{\gamma}$  and consider the vector field (the right hand side of equation (2))

$$(15) \quad (-\beta SI + \gamma(1 - S - I), \beta SI - \alpha I)^T$$

along  $M_{\tilde{\gamma}}$ . The inner product of (14) and (15) is

$$(16) \quad (\gamma - \tilde{\gamma})(\beta S - \alpha)I(1 - S - I) > 0.$$

Now consider the wedge domain between  $M_{\tilde{\gamma}}$  and the line  $S + I = 1$ . Trajectories of the system (2), as long as they stay in  $S > \alpha/\beta$ , cannot leave this wedge. Therefore, the unstable manifold of  $(1, 0)$  with respect to (2) stays in this wedge until it arrives at  $S = \alpha/\beta$ . Consequently we have  $I_{\gamma}^{max} \geq I_{\tilde{\gamma}}^{max}$ .  $\square$

Thus, for  $\gamma > 0$  we have two lower bounds for the first maximum  $I_{\gamma}^{max}$ , namely

$$I_0^{max} = 1 - \frac{1}{R_0} - \frac{1}{R_0} \log R_0 \leq I_{\gamma}^{max}$$

and

$$\bar{I}_{\gamma} = \frac{\gamma}{\alpha + \gamma} \left(1 - \frac{1}{R_0}\right) \leq I_{\gamma}^{max}.$$

For  $\gamma$  small, the first bound is better, the second bound is better for large  $\gamma$ . The two bounds agree for

$$(17) \quad \gamma^* = \left(\frac{R_0 - 1}{\log R_0} - 1\right)\alpha.$$

The right hand side in (17) is indeed positive for  $R_0 > 1$ .

**3. Node versus focus.** Here we address the problem of whether the endemic equilibrium in the SIRS model is a node or a focus. It is rather easy to discuss the behavior in terms of  $R_0$ . The Jacobian of the system (2) at the endemic state is

$$(18) \quad \begin{pmatrix} -\frac{\alpha\gamma}{\alpha + \gamma}(R_0 - 1) - \gamma & -(\alpha + \gamma) \\ \frac{\alpha\gamma}{\alpha + \gamma}(R_0 - 1) & 0 \end{pmatrix}.$$

The discriminant is

$$(19) \quad D = x^2 - 2(2 + \tilde{\gamma})x + \tilde{\gamma}^2$$

with

$$(20) \quad x = \frac{\gamma}{\alpha + \gamma}(R_0 - 1), \quad \tilde{\gamma} = \frac{\gamma}{\alpha}.$$

The polynomial (19) has always two positive zeros  $x_1 < x_2$ ,

$$(21) \quad x_{1,2} = 2 + \tilde{\gamma} \pm 2\sqrt{1 + \tilde{\gamma}}.$$

If  $R_0$  runs from 1 to  $\infty$  then the endemic state is a node up to  $x_1$ , then turns into a focus at  $x_2$ , and turns again to a node at  $x_2$ .

Actually we would be more interested in the behavior as a function of  $\gamma$ , for  $R_0$  fixed. Unfortunately this behavior is rather complicated. Let  $R_0 > 1$  and  $\alpha$  be fixed. For small  $\gamma$  the endemic equilibrium is always a focus. This observation agrees with the fact that for  $\gamma = 0$  the unstable manifold of  $(1, 0)$  returns to the axis  $I = 0$ . For very large  $\gamma$  the endemic equilibrium is a node. We cannot exclude situations where there are three (but not more) changes between focus and node. It is not evident that these details of qualitative behavior have any biological significance.

**4. Total size and total cost.** Maximal prevalence is one characteristic feature of an epidemic, others are total size and total infected time of an epidemic. The two notions make sense only in the SIR case. For the stochastic SIR model the problem of determining the distribution of the total size of the epidemic has attracted much attention, see Daniels [9] [10] [11], Barbour [7], Ball [2] [3] [4], Ball and Clancy [5], Picard and Lefèvre [22], Scalia-Tomba [23], Ball and Nåsell [6], Diekmann et al. [14], Svensson [25], Startsev [24], Andersson and Djehiche [1]. Of course, for a finite population, the total size is the number of all individuals ever infected. The stochastic problem is complicated because of the threshold phenomenon. For  $R_0 > 1$  there is a positive probability of “major” outbreaks, but for  $R_0 < 1$  there may be “minor” outbreaks. The mathematical difficulties seem to arise because the desired distribution has to account for these very distinct features. Indeed, in spite of much effort, the produced explicit expressions are still very complicated, see Ball [3]. In the deterministic case, i.e., for the system (2) with  $\gamma = 0$ , the total size  $T$  of the epidemic can be obtained by integrating over the incidence, i.e., by integrating over all individuals entering the infected state. Hence the total size is

$$(22) \quad T = \int_{-\infty}^{\infty} \beta SI dt.$$

The total infected time of the epidemic (which has also been called the total cost of the epidemic, see Gani and Jerwood [15]) is the quantity

$$(23) \quad C = \int_{-\infty}^{\infty} Idt.$$

Using the differential equations (2), one checks directly the formulae

$$(24) \quad T = 1 - S_{\infty}$$

and

$$(25) \quad C = \frac{1}{\alpha}(1 - S_{\infty}).$$

These formulae can be interpreted as follows: The total size is the complement of those individuals which have always remained uninfected. In view of the recovery law,  $1/\alpha$  is the average time spent in the infected state, and hence  $C = T/\alpha$ .

We can get more detailed information by integrating the second equation of (2) from  $-\infty$  to some finite time  $\tau$ . Since time runs from  $-\infty$ , any  $\tau$  would do, e.g.  $\tau = 0$ . However, here it makes sense to think of a fixed parametrization by the time variable in order that different time points can be compared and that  $S(t)$  and  $I(t)$  are coupled via the time variable. Integration yields

$$\begin{aligned} \int_{-\infty}^{\tau} \beta S I dt &= \int_{-\infty}^{\tau} I dt + \alpha \int_{-\infty}^{\tau} I dt \\ &= I(\tau) - \frac{\alpha}{\beta} \int_{-\infty}^{\tau} \frac{\dot{S}}{S} dt \\ &= I(\tau) - \frac{\alpha}{\beta} \log S(\tau). \end{aligned}$$

For  $\tau = \infty$ , using (4), we get again (24). On the other hand, assume that  $\tau$  is the time when prevalence is maximal. Then  $I(\tau)$  is given by (10) and  $S(\tau) = \alpha/\beta$ , hence

$$(26) \quad \int_{-\infty}^{\tau} \beta S I dt = 1 - \frac{1}{R_0}.$$

**Proposition 4.** *Consider the SIR model ( $\gamma = 0$ ). At the time when prevalence is maximal the accumulated fraction (measuring the accumulated number of cases) is given by equation (26).*

Taking the argument in reverse we get the following.

**Proposition 5.** *Within the framework of the SIR model ( $\gamma = 0$ ) the basic reproduction number  $R_0$  can be estimated from the peak prevalence via equation (11) or from the number of accumulated cases at the time of peak prevalence via equation (26).*

**5. Demographic renewal.** Loss of immunity, as in the SIRS system (1), is one mechanism to raise an epidemic to an endemic state. If the basic reproduction number is greater than one and if there is some however small loss of immunity then the disease becomes endemic. Demographic turnover, even in an SIR situation, shows exactly the same features, and it is practically more important, in particular with respect to typical childhood diseases. Again we consider the simplest possible example, an SIR model for a population with constant renewal rate  $\mu > 0$ . The discussion

could be carried through in terms of a model for variable population size as in Busenberg and Hadeler [8] but here we refrain from the additional technical effort. Such treatment would be necessary in case disease-related differential mortality is taken into account. We assume constant total population size 1. The system has the form

$$(27) \quad \begin{aligned} \dot{S} &= \mu - \beta SI - \mu S \\ \dot{I} &= \beta SI - \alpha I - \mu I \\ \dot{R} &= \alpha I - \mu R. \end{aligned}$$

Again it suffices to consider the first two equations in the triangle  $S \geq 0$ ,  $I \geq 0$ ,  $S + I \leq 1$ , hence the system

$$(28) \quad \begin{aligned} \dot{S} &= \mu - \beta SI - \mu S \\ \dot{I} &= \beta SI - \alpha I - \mu I. \end{aligned}$$

The basic reproduction number for (28) is

$$(29) \quad R_0 = \frac{\beta}{\alpha + \mu}.$$

In comparison to (3) also  $\mu$  appears in the denominator because recovery and death remove infecteds from the population. If  $R_0 \leq 1$  then  $(S, I) = (1, 0)$  is the only stationary point. If  $R_0 > 1$  then there is an endemic equilibrium  $(\bar{S}, \bar{I})$ ,

$$(30) \quad \bar{S} = \frac{1}{R_0}, \quad \bar{I} = \frac{\mu}{\alpha + \mu} \left(1 - \frac{1}{R_0}\right) = \frac{\mu}{\alpha + \mu} \left(1 - \frac{\alpha + \mu}{\beta}\right).$$

Now we analyze this system in terms of the demographic parameter  $\mu$ . We let  $\mu$  run from 0 to  $\infty$ . Assume

$$(31) \quad \frac{\beta}{\alpha} > 1$$

saying that  $R_0 > 1$  at least for small  $\mu$ . If  $\mu = 0$  then there is no infected stationary state (the expression for  $\bar{I}$  in (30) is zero). The unstable manifold of  $(1, 0)$  returns to  $I = 0$  at  $S = S_\infty$ . If  $\mu$  is very small then  $\mu$  can be neglected in comparison to  $\alpha$ , the basic reproduction number (29) is still larger than one. There is an endemic equilibrium (30) with  $\bar{I}$  positive, but  $\bar{I}$  is very small. Thus, an arbitrarily small turnover rate  $\mu$  is sufficient to create an endemic state.

Now we keep  $\beta$  and  $\alpha$  fixed, always assuming (31), and consider the endemic prevalence  $\bar{I}$  as a function of  $\mu \in (0, \infty)$ . First  $\bar{I}$  increases approximately linearly. But for large  $\mu$ , in the expression (30) for  $\bar{I}$ , the decrease of  $R_0$  dominates, eventually  $R_0$  becomes smaller than one, and the endemic

state disappears. Hence  $\bar{I}$ , as a function of  $\mu$ , first increases, then decreases. It attains its maximum at

$$(32) \quad \mu^* = \alpha \left( \sqrt{\frac{\beta}{\alpha}} - 1 \right),$$

the maximum being

$$(33) \quad \bar{I}^{max} = \left( 1 - \sqrt{\frac{\alpha}{\beta}} \right)^2.$$

Thus, it is evident that maximal prevalence does not depend on  $\mu$  in a monotone way.

Nevertheless we can show *some* monotonicity properties. Again we fix  $\beta, \alpha$  and we assume (31). In the following it is useful to imagine the two equilibria (uninfected and infected) and the (distinguished) trajectory for different values of  $\mu$  as being represented in the same phase plane or rather phase triangle. Consider, for any  $\mu$ , the endemic equilibrium (30). If  $\mu$  runs from 0 to  $\infty$ , the infected equilibria form a curve starting at the point  $(\alpha/\beta, 0)$  and ending at  $(1, 0)$ , see Fig.3. This curve can also be parametrized by the variable  $S$  and then it assumes the form

$$(34) \quad I \equiv \Psi(S) = \frac{(\beta S - \alpha)(1 - S)}{\beta S}.$$

Of course the curve itself is independent of  $\mu$ . We find

$$(35) \quad \frac{d\Psi}{dS} = \frac{\alpha - \beta S^2}{\beta S^2}, \quad \frac{d\Psi}{dS}|_{S=1} = \frac{\alpha}{\beta} - 1.$$

Now consider the system (28) for a fixed value of  $\mu > 0$  such that  $R_0 > 1$  ( $R_0$  given by (29)). Along the unstable manifold of the point  $(1, 0)$  (i.e., the distinguished trajectory) there is a point of maximal prevalence. If we follow the distinguished trajectory from the uninfected equilibrium to the point of maximal prevalence, then  $S$  is decreasing and  $I$  is increasing, see Fig.3. After the point of maximal prevalence, the distinguished trajectory spirals down to the infected equilibrium, see Fig.3 (or maximal prevalence is assumed at equilibrium). The next proposition shows the following. For any given  $\mu$ , as long as the unstable manifold of  $(1, 0)$  has not reached the point with maximal prevalence, it is located above the curve  $I = \Psi(S)$ , see Fig.3.

**Proposition 6.** *Consider the model with demographic renewal (27). As long as the epidemic does not reach maximal prevalence, the inequality*

$$(36) \quad I > (S - \frac{\alpha}{\beta}) \frac{1 - S}{S}$$

*holds.*

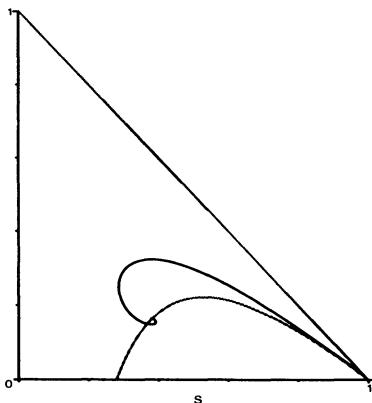


FIG. 3. The curve of equilibria  $\Psi(S)$  and the distinguished trajectory for  $\mu = 0.05$ . This trajectory approaches the endemic equilibrium corresponding to  $\mu = 0.05$ . The other parameters are  $\beta = 0.5$ ,  $\alpha = 0.14$ . Then  $R_0 = 2.63$ .

We rephrase the result in biological terms. Suppose we observe the time course of an epidemic starting from very few infecteds in a totally susceptible population. At a given moment  $t$ , before maximal prevalence is reached, we observe the *actual prevalence*  $I(t)$  and the *actual number of susceptibles*  $S(t)$  (some individuals are in the recovered or immune class). With  $\alpha$  and  $\beta$  given, the number  $S(t)$  can be interpreted as the number of susceptibles at the endemic equilibrium for some  $\mu > 0$ , i.e.,  $S(t) = \bar{S}(\mu)$ . At that equilibrium we would have a number of infecteds  $\bar{I}(\mu) = \Psi(S)$  as given by (34). Proposition 6 says  $I(t) > \bar{I}(\mu)$ .

It is somewhat astonishing that the curve constructed as the set of equilibria for variable parameters gives information on the dynamical behavior for fixed parameters.

**Proof of Proposition 6.** We use similar ideas as in the proof of proposition 1. For  $S$  between  $\alpha/\beta$  and 1 the curve  $S \rightarrow (S, \Psi(S))$  forms a concave arc. At the point  $(S, \Psi(S))$  the tangent vector is  $(\beta S^2, \alpha - \beta S^2)$  and hence the normal vector is  $(\beta S^2 - \alpha, \beta S^2)$ . Near  $(1, 0)$  this normal vector has both components positive. We form the inner product with the vector field of (28) which gives

$$(37) \quad \alpha\beta(S - \frac{\alpha + \mu}{\beta})(1 - S)^2.$$

As long as  $S > 1/R_0$ , this expression is positive. An easy calculation shows that at  $(1, 0)$  the curve  $I = \Psi(S)$  and the unstable manifold have the same tangent vector  $(1, (\alpha + \mu)/\beta - 1)$ . Hence the unstable manifold of  $(1, 0)$  stays above this curve.  $\square$

## REFERENCES

- [1] Andersson, H. and Djehiche, B., Limit theorem for the total size of a spatial epidemic. *J. Appl. Prob.* **34**, 689–710 (1997).
- [2] Ball, F., A unified approach to the distribution of the trajectories of infectives in epidemic models. *Adv. Appl. Prob.* **18**, 289–310 (1986).
- [3] Ball, F., A note on the total size distribution of epidemic models. *J. Appl. Prob.* **23**, 832–836 (1986).
- [4] Ball, F., A note on the total size distribution of carrier-borne epidemic models. *J. Appl. Prob.* **27**, 908–912 (1990).
- [5] Ball, F. and Clancy, D., The final size and severity of a generalized stochastic multitype epidemic model. *Adv. Appl. Prob.* **25**, 721–736 (1993).
- [6] Ball, F. and Nåsell, I., The shape of the size distribution of an epidemic in a finite population. *Math. Biosc.* **123**, 167–181 (1994).
- [7] Barbour, A.D., A note on the maximum size of a closed epidemic. *J. Roy. Statist. Soc. B* **37**, 459–460 (1975).
- [8] Busenberg, S.N. and Hadeler, K.P., Demography and epidemics. *Math. Biosc.* **101**, 63–74 (1990).
- [9] Daniels, H.E., The deterministic spread of a simple epidemic. *Perspectives Prob. Stat. Papers Honour M.S. Bartlett Occ. 65th birthday*, 373–386 (1975).
- [10] Daniels, H.E., The maximum size of a closed epidemic. *J. Adv. Appl. Prob.* **6**, 607–621 (1979).
- [11] Daniels, H.E., The time of occurrence of the maximum of a closed epidemic. In: J.-P. Gabriel, C. Lefèvre, P. Picard, *Stochastic Processes in Epidemic Theory*, pp. 129–136. *Lect. Notes in Biomathematics* **86**, Springer Verlag 1990.
- [12] Diekmann, O., Heesterbeek, J.A.P., and Metz, J.A.J., On the definition and the computation of the basic reproduction ratio  $R_0$  in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**, 365–382 (1990).
- [13] Diekmann, O. and Heesterbeek, Hans, *Mathematical epidemiology of infectious diseases. Model building, analysis and interpretation*. Wiley, Chichester 1999.
- [14] Diekmann, O., de Koeijer, A.A., and Metz, J.A.J., On the final size of epidemics within herds. *Canad. Appl. Math. Quart.* **4**, 21–30 (1996).
- [15] Gani, J. and Jerwood, D., The cost of a general stochastic epidemic. *J. Appl. Probab.* **9**, 257–269 (1972).
- [16] Hadeler, K.P. and Castillo-Chavez, C., A core group model for disease transmission. *Math. Biosc.* **128**, 41–55 (1995).
- [17] Hadeler, K.P. and van den Driessche, P., Backward bifurcation in epidemic control. *Math. Biosc.* **146**, 15–35 (1997).
- [18] Hethcote, H.W., Qualitative analysis for communicable disease models. *Math. Biosc.* **28**, 335–356 (1976).
- [19] Hethcote, H.W., Stech, H.W., and van den Driessche, P., Periodicity and stability in epidemic models: a survey. In: S.N. Busenberg, K. Cooke (eds) *Differential equations and applications in ecology, epidemics, and population problems*, Proc. Conf. Claremont/Calif. 1981, Academic Press, 65–82 (1981).
- [20] Liu, Wei-min, Hethcote, H.W., and Levin, S.A., Dynamical behavior of epidemiological models with nonlinear incidence rates. *J. Math. Biol.* **25**, 359–380 (1987).
- [21] Mollison, D., Dependence of epidemic and population velocities on basic parameters. *Math. Biosc.* **107**, 255–287 (1991).
- [22] Picard, P. and Lefèvre, C., A unified analysis of the final size and severity distribution in collective Reed-Frost epidemic processes. *Adv. Appl. Prob.* **22**, 269–294 (1990).
- [23] Scalia-Tomba, G., On the asymptotic final size distribution of epidemics in heterogeneous populations. *Lect. Notes in Biomathematics* **86**, 189–196, Springer-Verlag, Heidelberg, 1990.

- [24] Startsev, A.N., On the distribution of the size of an epidemic in a non-Markovian model. *J. Theor. Prob. Appl.* **41**, 730–740 (1996).
- [25] Svensson, A., On the asymptotic size and duration of a class of epidemic models. *J. Appl. Prob.* **32**, 11–24 (1995)

# THE TRANSITION THROUGH STAGES WITH ARBITRARY LENGTH DISTRIBUTIONS, AND APPLICATIONS IN EPIDEMICS\*

HORST R. THIEME†

## Contents

- Introduction
- 1. The duration of a stage (life, e.g.)
  - 1.1. The sojourn function
  - 1.2. Mean sojourn time, expectation of life. The variance of sojourn time
  - 1.3. Remaining sojourn time and its expectation
  - 1.4. Fixed stage durations
  - 1.5. Per capita exit rates (mortality rates)
  - 1.6. Exponentially distributed stage durations
  - 1.7. Log-normally distributed stage durations
- 2. Stage dynamics with given inputs
  - 2.1. Stage input and stage-age density
  - 2.2. The PDE formulation
  - 2.3. Stage content and average stage duration
  - 2.4. Average stage age and average expectation of remaining sojourn
  - 2.5. Stage exit rate and fundamental law of stage dynamics
  - 2.6. Stage outputs
- 3. Application to epidemic modeling
  - 3.1. An epidemic model with chronological age structure
  - 3.2. Epidemic models with arbitrarily distributed infection stages
- Bibliographic remarks
- Acknowledgement
- A. Appendix: Stieltjes integration
- References

---

\*Research partially supported by NSF grants DMS-9403884 and DMS-9706787.

†Department of Mathematics, Arizona State University, Tempe, AZ 85287-1804;  
Email: h.thieme@asu.edu.

**Introduction.** Life occurs in stages. Human (earthly) life is the stage between conception and death, clearly divided into two separate stages by birth. There are more stages that are less clearly separated: childhood, youth, adulthood, senescence. Invertebrate animals may have more distinctly separated stages, like egg, larva (or instar), pupa, imago. There may be even several larval stages. Infectious diseases take their course through various stages: latent period, infectious period without symptoms (the two together usually form the incubation period), infectious period with symptoms, and often an immunity period. With some diseases, HIV/AIDS e.g., the infectious period can be further subdivided according to the progress of the disease.

For demographic and epidemic modeling to take advantage of these stages the following issues need to be addressed:

How long does it take to go through a certain stage? What is the average duration (mean sojourn time) and how large is the variation? Given that one already is in the stage for a certain time  $t$ , how long can one expect to remain in the stage? Related to this question is the average expectation of remaining sojourn.

How is the number of individuals in a stage (stage content) and the rate of people leaving the stage (stage output) related to the number of individuals entering the stage (stage input)?

Often stage transition is addressed in the spirit of compartmental modeling (*Jacquez, 1978*), i.e., individuals are assumed to leave the stage at a fixed per capita rate. This leads to very special length distributions of the stage, exponential distributions, with the curious feature that the average stage duration, the average expectation of remaining sojourn, and the standard deviation of the stage duration are all identical, namely the reciprocal of the per capita exit rate. The per capita exit rate enters crucial parameters that decide about the qualitative behavior of a system and therefore should be interpreted in non-mathematical and general terms. This raises the dilemma whether it should be interpreted as reciprocal of the mean duration or of the average expectation of remaining sojourn or of the standard variation.

A competing, but also over-simplifying assumption is fixed length of the stage, then the standard deviation is zero and the expectation of remaining sojourn is half the mean sojourn time. Both assumptions are extreme: It is a common feature that even under identical conditions there is variation in individual stage sojourns, but that the standard deviation of sojourn rarely is of same magnitude as the mean sojourn time. The latent, incubation, and infectious periods of most infectious diseases, e.g., have coefficients of variation that are considerably closer to 0 than to one (Section 1.7).

In its first part, this article develops the notions of mean sojourn time, expected remaining sojourn, and average expectation of remaining sojourn for general sojourn functions. In spite of some stochastic language, it is

written from a deterministic point of view ignoring the underlying random variables. For the stochastic connection, see the note of *Kim and Aron* (1989) from which much inspiration has been drawn. In the second part we derive the integral equations that relate stage contents, stage exit rates and stage outputs to stage inputs. The first two parts of this article are condensations of notes from several courses the author taught at Arizona State University; some theorems are stated without proof, full proofs of the theorems can be found in the original course notes.

The third part deals with applications to epidemic modeling. We first show how expected remaining life naturally appears in the basic reproduction number of a disease that spreads in a population with chronological age structure. Then we illustrate how epidemic models with arbitrarily many and arbitrarily distributed periods of infection can be formulated. An analysis of these models can be found in *Feng, Thieme* (2000ab). Recently *Keeling and Grenfell* (1997, 1998) have demonstrated that infectious periods with small variance give better agreement with persistence data for measles than exponential distributions. We will explain how the length of the inter-epidemic period and the stability threshold for the endemic equilibrium change, when the length distributions of the infection stages are changed, but the mean lengths are kept the same. In particular we will make the point that constant per capita stage exit rates (i.e., exponentially distributed stage durations) can suggest misleading interpretations.

Arbitrarily distributed periods of infection lead to integral equations rather than differential equations. Integral equations have been my first mathematical love to which I always seem to return and to which I have been drawn by a paper by Ken Cooke (1967) which I first got to know through two other papers which it triggered and which were co-authored by Fred Brauer's first Ph.D student (*Hoppensteadt, Waltman*, 1970, 1971). As testified by the references (*Brauer*, 1990, 1992, 1996; *Castillo-Chavez, Cooke et al.*, 1989, *Busenberg, Cooke, Iannelli*, 1988), Fred Brauer's and Ken Cooke's work has been an inspiration for this paper, and I am pleased that it is part of a volume dedicated to these two pioneers in integral equation epidemic modeling.

**1. The duration of a stage (life, e.g.)** Assume that you have entered a certain stage. You would like to know how long you are going to remain in this stage. In other words, you would like to know the sojourn time in or the duration (length) of the stage. If the stage under consideration is life, you would like to know for how long you are going to live.

**1.1. The sojourn function.** All relevant information about the duration of the stage is contained in a function

$$\mathcal{F} : [0, \infty) \rightarrow [0, \infty),$$

where  $\mathcal{F}(a)$  gives the probability to be still in the stage,  $a$  time units after you have entered it. In case life is the stage under consideration,  $\mathcal{F}(a)$  gives

the probability of being still alive at age  $a$ . In this case,  $\mathcal{F}$  is sometimes called the *survival function* or the *survivorship*. If we consider an arbitrary stage, we call  $\mathcal{F}$  the *sojourn function* of the stage. If the stage is not life, but a certain part of life, e.g., then we call the variable a *stage age* or *class age* which is the time that has passed since entering the stage. With this understanding, we can call the sojourn time (or the stage duration) *stage age at exit*.

$\mathcal{F}$  has two obvious properties:

- $\mathcal{F}(0) = 1$ .
- $\mathcal{F}$  is a monotone non-increasing function.

We can introduce the function  $\mathcal{G}(a) = 1 - \mathcal{F}(a)$ ,  $a \geq 0$ .  $\mathcal{G}(a)$  is the probability of already having left the stage (e.g., be dead) at stage age  $a$ .  $\mathcal{G}$  has the following properties:

$$\mathcal{G}(0) = 0.$$

$\mathcal{G}$  is a monotone non-decreasing function.

If  $0 \leq a < b$ , then  $\mathcal{F}(a) - \mathcal{F}(b) = \mathcal{G}(b) - \mathcal{G}(a)$  is the probability of leaving the stage in the stage age interval from  $a$  to  $b$ .

$\mathcal{F}$  and  $\mathcal{G}$  are not necessarily continuous functions. Since they are monotone, their limits from the right and left exist, however:

$$\mathcal{F}(a+) = \lim_{a < b \rightarrow a} \mathcal{F}(b), \quad \mathcal{F}(a-) = \lim_{a > r \rightarrow a} \mathcal{F}(r).$$

For instance, if there is a non-zero chance of dying immediately after birth, one has  $1 = \mathcal{F}(0) > \mathcal{F}(0+)$  and  $\mathcal{F}$  is not continuous at 0. Similarly, if a stage is harvested at a particular stage age  $a_0 > 0$ , this leads to a discontinuity at  $a_0$ .  $\mathcal{G}(a+)$  is often called the *probability distribution function of the stage sojourn time*.

The monotonicity of  $\mathcal{F}$  also implies that the limit  $\mathcal{F}(\infty) = \lim_{a \rightarrow \infty} \mathcal{F}(a)$  exists. If  $\mathcal{F}(\infty) = 0$ , everybody finally has to leave the stage; if  $\mathcal{F}(\infty) > 0$ , it is possible to stay in the stage forever.

Let us introduce the *maximum stage age*  $c$  (or *maximum stage duration*, *maximum sojourn time*) as follows: If  $\mathcal{F}(a) > 0$  for all  $a \geq 0$ , then  $c = \infty$ ; otherwise  $c$  is the uniquely determined number such that  $\mathcal{F}(a) > 0$  for all  $a \in [0, c)$  and  $\mathcal{F}(a) = 0$  for all  $a > c$ .

**1.2. Mean sojourn time, expectation of life. The variance of sojourn time.** In order to determine the mean sojourn time in the stage, let us first assume that we have a finite maximum sojourn time  $c$ . Choose  $b > c$  and a partition  $0 = t_0 < \dots < t_{n+1} = b$  such that  $t_{j+1} - t_j$  is very small for all  $j$ . As we have seen before,

$$p_j = \mathcal{F}(t_j) - \mathcal{F}(t_{j+1})$$

gives the probability of leaving the stage in the time interval from  $t_j$  to  $t_{j+1}$ . In this case, the sojourn time is some number  $s_j$  between  $t_j$  and  $t_{j+1}$ .

Notice that

$$\sum_{j=0}^n p_j = \sum_{j=0}^n \mathcal{F}(t_j) - \sum_{j=0}^n \mathcal{F}(t_{j+1}) = \mathcal{F}(0) - \mathcal{F}(b) = 1.$$

Since the differences  $t_{j+1} - t_j$  are very small, the mean (or average) sojourn time (expectation of life), which is denoted by  $D$ , is close to

$$(1.1) \quad \sum_{j=0}^n s_j p_j = \sum_{j=0}^n s_j (\mathcal{F}(t_j) - \mathcal{F}(t_{j+1})).$$

Taking the sums apart and changing the index of summation in the first sum and recombining them (recall  $\mathcal{F}(t_{n+1}) = 0$ ), we obtain

$$\sum_{j=0}^n s_j p_j = s_0 \mathcal{F}(0) + \sum_{j=0}^{n-1} \mathcal{F}(t_{j+1})(s_{j+1} - s_j).$$

Notice that  $t_{j+1} \in [s_j, s_{j+1}]$ . If  $s_n = b$  and  $s_0 = 0$ , the  $s_j$  form a partition of the interval from 0 and  $b$ ; otherwise we introduce  $s_{-1} = 0$  and/or  $s_{n+1} = b$ ,  $t_{n+1} = b$  and can write

$$\sum_{j=0}^n s_j p_j = \sum_{j=-1}^n \mathcal{F}(t_{j+1})(s_{j+1} - s_j).$$

If the differences  $t_{j+1} - t_j$  tend to 0 for every  $j$ , so do the differences  $s_{j+1} - s_j$  and the sums converges towards the Riemann integral

$$(1.2) \quad \int_0^b \mathcal{F}(t) dt = \int_0^\infty \mathcal{F}(a) da =: D$$

and  $D$  is the *mean sojourn time* in the stage or the *mean duration* of the stage or the *mean exit (stage) age*. This integral exists as an improper Riemann integral because  $\mathcal{F}$  is monotone and  $\mathcal{F}(a) = 0$  for all  $a \geq b > c$ . Alternatively, as the partitions gets finer, the sums (1.1) converge to the Stieltjes integral

$$-\int_0^b t d\mathcal{F}(t) = D,$$

and we obtain formula (1.2) by integrating by parts. Formula (1.2) can be extended to the case  $c = \infty$  by taking the improper Riemann integral or the Lebesgue integral on  $[0, \infty)$  which both exist if  $\mathcal{F}(a)$  converges to 0 fast enough as  $a \rightarrow \infty$ .

More generally, if  $f$  is a continuous function on  $[0, \infty)$  and  $f$  or  $\mathcal{F}$  have compact support,

$$\sum_{j=0}^n f(s_j) p_j \longrightarrow -\int_0^\infty f(t) d\mathcal{F}(t) = -\int_0^\infty f(t) \mathcal{F}(dt).$$

The first integral is a Riemann Stieltjes integral and the second a Lebesgue Stieltjes integral with  $\mathcal{F}(dt)$  denoting integration with the Lebesgue Stieltjes measure associated with  $-\mathcal{F}$ . See the Appendix and Lemma A.1. The Lebesgue Stieltjes integral is also defined, if  $f$  is Borel measurable and bounded. If  $\mathcal{F}$  is continuously differentiable,

$$-\int_0^\infty f(t)d\mathcal{F}(t) = -\int_0^\infty f(t)\mathcal{F}'(a)da.$$

If  $f$  is continuously differentiable, we have the integration by parts formula

$$-\int_0^\infty f(t)d\mathcal{F}(t) = f(0) + \int_0^\infty f'(t)\mathcal{F}(t)dt.$$

**The variance of the sojourn time, moments and central moments.** The variance of the sojourn time,  $V$ , is approximated by sums

$$\sum_{j=0}^n (s_j - D)^2 p_j = \sum_{j=0}^n (s_j - D)^2 (\mathcal{F}(t_j) - \mathcal{F}(t_{j+1}))$$

with partitions  $0 = t_0 < \dots < t_{n+1} < \infty$  where, as before,  $p_j$  is the probability of leaving the stage between  $t_j$  and  $t_{j+1}$  and  $s_j$  is some number in  $[t_j, t_{j+1}]$ . This means that the variance of the sojourn time is given by the Stieltjes integral

$$V = -\int_0^\infty (a - D)^2 d\mathcal{F}(a).$$

The approximation above also tells us that  $V \geq 0$ . Integrating by parts,

$$(1.3) \quad 0 \leq V = 2 \int_0^\infty a \mathcal{F}(a) da - D^2.$$

The  $n^{th}$  moment of the sojourn time is defined by

$$-\int_0^\infty t^n d\mathcal{F}(t) = n \int_0^\infty t^{n-1} \mathcal{F}(t) dt,$$

and the  $n^{th}$  central moment by

$$C_n = -\int_0^\infty (t - D)^n d\mathcal{F}(t).$$

Notice  $V = C_2$ . The *coefficient of variation*,  $\gamma_0$ , the *coefficient of skewness*,  $\gamma_1$ , and the *coefficient of excess (kurtosis)*,  $\gamma_2$ ,

$$\gamma_0 = \frac{\sqrt{V}}{D}, \quad \gamma_1 = \frac{C_3}{C_2^{3/2}}, \quad \gamma_2 = \frac{C_4}{C_2^2},$$

are scale invariant measures of the dispersion, asymmetry and flatness of the distribution of sojourn times, respectively (Patel *et al.*, 1976).

**1.3. Remaining sojourn time and its expectation.** Similarly to the mean duration, one can calculate the time one can still expect to remain in the stage under the condition one has already stayed there till stage age  $a$ . We call this time the *expected remaining sojourn* at stage age  $a$  and denote it by  $D(a)$ . If we consider life, we can call this the *expected remaining life* at age  $a$ .

The probability to be still in the stage at time  $a + t$ , provided one is still in the stage at time  $a$ , is given by

$$(1.4) \quad \mathcal{F}(t|a) = \frac{\mathcal{F}(t+a)}{\mathcal{F}(a)}, \quad 0 \leq a < c, t \geq 0.$$

We set

$$\begin{aligned} \mathcal{F}(t|a) &= 0, \quad \text{whenever } a \geq c, t > 0, \\ \mathcal{F}(0|a) &= 1, \quad a \geq c. \end{aligned}$$

Notice that, for fixed but arbitrary  $a \geq 0$ ,  $\mathcal{F}(\cdot|a)$  has the properties of a sojourn function:  $\mathcal{F}(0|a) = 1$  and  $\mathcal{F}(\cdot|a)$  is a non-increasing function. Indeed,  $\mathcal{F}(\cdot|a)$  is the sojourn function of the substage of the original state that starts at age  $a$  and we call it the *conditional sojourn function* at stage age  $a$ .

Analogously to the above consideration the expected remaining sojourn at stage age  $a$  can be determined by

$$D(a) = \int_0^\infty \mathcal{F}(t|a) dt.$$

Using the definition of  $\mathcal{F}(t|a)$  in (1.4),

$$(1.5) \quad D(a) = \int_0^\infty \frac{\mathcal{F}(t+a)}{\mathcal{F}(a)} dt = \frac{1}{\mathcal{F}(a)} \int_a^\infty \mathcal{F}(t) dt, \quad 0 \leq a < c.$$

Obviously

$$D(a) = 0, \quad a \geq c, \quad D(0) = D.$$

At first glance, one might think that  $D(a)$  should be a monotone non-increasing function on  $a$ . But a moment of reflection will tell us that, in a country where infant mortality is considerably higher than the mortality in later age classes, the expected remaining life at age  $a$ ,  $D(a)$ , may actually increase for small  $a$ . See Section 1.4 and the bibliographic remarks at the end of this section for references.

We use the letter  $D$  both for the average sojourn time and for the function whose values provide the expected remaining sojourn. To avoid confusion we make the following convention: An isolated  $D$  usually denotes the average sojourn time and sometimes we rather write  $D(0)$ . If we refer to the function  $D$ , we usually write  $D(\cdot)$ .

Since  $\mathcal{F}$  is non-increasing, from (1.5),

$$(1.6a) \quad \mathcal{F}(a)(a-s) \leq \mathcal{F}(s)D(s) - \mathcal{F}(a)D(a) \leq \mathcal{F}(s)(a-s) \quad a \geq s \geq 0.$$

This implies that  $D(a) + a$  is a non-decreasing function of  $a \geq 0$  and, evaluating (1.6a) for  $s = 0$ , the estimate

$$(1.6b) \quad \frac{D-a}{\mathcal{F}(a)} \leq D(a) \leq \frac{D}{\mathcal{F}(a)} - a.$$

Without some restrictions on  $\mathcal{F}$ ,  $D(a)$  may be an unbounded function of  $a$ . An example is given by sojourn functions of Pareto type,  $\mathcal{F}_\kappa(a) = [1+a]^{-\kappa}$ ,  $a \geq 0$ , with parameter  $\kappa > 2$ .

**Proposition 1.1.** *Let  $c = \infty$ . Then the following three statements are equivalent:*

- (a)  *$D(a)$  is a bounded function of  $a \in [0, \infty)$ .*
- (b)  *$\frac{\mathcal{F}(a+t)}{\mathcal{F}(a)} \rightarrow 0$  for  $t \rightarrow \infty$ , uniformly in  $a \in [0, \infty)$ .*
- (c) *There exist  $\epsilon > 0, M \geq 1$  such that  $\mathcal{F}(t|a) \leq M e^{-\epsilon t}$  for all  $t, a \geq 0$ .*

If life is the stage under consideration, obviously the function  $D(\cdot)$  is of interest to retirement and pension plans and life insurance. Data for  $D(\cdot)$  seem to go farther back than for the survival function  $\mathcal{F}$  itself, actually even to the Roman empire (*Smith, Keyfitz, 1977*, p.5). So it is of some interest to recover  $\mathcal{F}$  from  $D(\cdot)$ . Set  $u(a) = D(a)\mathcal{F}(a)$ . From (1.6), by the fundamental theorem of calculus,  $u'(a) = -\mathcal{F}(a) = -\frac{u(a)}{D(a)}$ . Integrating and using the definition of  $u$  we have the following result:

**Proposition 1.2.** *The following relation holds between the sojourn function and the expected remaining sojourn:*

$$\mathcal{F}(a) = \frac{D(0)}{D(a)} \exp \left( - \int_0^a \frac{ds}{D(s)} \right), \quad 0 \leq a < c.$$

But not every function  $D : [0, \infty) \rightarrow [0, \infty)$  qualifies. We have already seen from (1.6a) that  $D(a) + a$  necessarily is a non-decreasing function of  $a$ . The following converse holds.

**Remark.** Let  $D : [0, c) \rightarrow (0, \infty)$  and  $\mathcal{F}$  be related by the formula in Proposition 1.2 and  $D$  be continuous on  $[0, c)$ . Then  $\mathcal{F}$  is monotone non-increasing if and only if  $D(a) + a$  is a monotone non-decreasing function of  $a \in [0, c)$ .

*Proof.* If  $D$  is differentiable, this follows from differentiating  $\ln \mathcal{F}(a)$  in Proposition 1.2. Otherwise define

$$D_k(a) = k \int_a^{a+(1/k)} D(s) ds = \int_0^1 D(a + (s/k)) ds$$

and notice  $D_k(a) \rightarrow D(a)$  as  $k \rightarrow \infty$  uniformly for  $a$  in every closed interval in  $[0, c)$ . If  $D(a) + a$  is non-decreasing in  $a \geq 0$ ,  $D'_k(a) + 1 \geq 0$ . Then the

function  $\mathcal{F}_k$  related to  $D_k$  via the formula in Proposition 1.2 is monotone non-increasing. Since  $\mathcal{F}_k$  converges towards  $\mathcal{F}$ , this property is inherited by  $\mathcal{F}$ .  $\square$

**Average expectation of remaining sojourn.** In order to find the average expectation of remaining sojourn,  $E$ , we average the expected remaining sojourns  $D(a)$  over all stage ages  $a$ . Since the probability to be still in the stage at stage age  $a$  is  $\mathcal{F}(a)$ , we need to form the average weighed by  $\mathcal{F}(a)$ ,

$$(1.7) \quad E = \frac{\int_0^\infty D(a)\mathcal{F}(a)da}{\int_0^\infty \mathcal{F}(a)da} = \int_0^\infty \frac{D(a)}{D(0)} \mathcal{F}(a)da.$$

Substituting  $D(a)$  from (1.5) and interchanging the order of integration we obtain the following formula for the *average expectation of remaining sojourn*,

$$(1.8) \quad E = \frac{\int_0^\infty a\mathcal{F}(a)da}{D(0)} = \frac{\int_0^\infty a\mathcal{F}(a)da}{\int_0^\infty \mathcal{F}(a)da}.$$

Formula (1.7) and Proposition 1.2 reveal another way in which the average expectation of remaining sojourn averages the expected remaining sojourn:

$$E = \int_0^c \exp\left(-\int_0^a \frac{1}{D(s)} ds\right) da.$$

Formula (1.3) implies the following relation between the mean and the variance of the stage duration and the expected remaining sojourn:

$$(1.9) \quad V = D(2E - D).$$

This can be rewritten in various useful ways,

$$2\frac{E}{D} = 1 + \frac{V}{D^2}, \quad V - D^2 = 2D(E - D).$$

From the second equation we learn that  $V - D^2$  has the same sign as  $E - D$ .

**Proposition 1.3.**  $D/2 \leq E \leq c/2$ .

*Proof.* Let  $0 < a < c < \infty$ . By (1.6b) and (1.7),  $E \leq c - E$ .  $E \leq c/2$  is trivially true if  $c = \infty$ . The other inequality follows from (1.9) and  $V \geq 0$ .  $\square$

We will see in the next section that  $E = D/2$  if and only if the variance of the stage duration is 0, i.e., if the sojourn function is the step function equaling 1 for  $a < c$  and 0 for  $a > c$ . In this case the maximum stage duration  $c$  is finite and coincides with the mean stage duration. In Sections 1.4 and 1.7 we will see examples in which  $E/D$  is arbitrarily large, actually it can be any number greater than 1/2.

**1.4. Fixed stage durations.** There are two extreme but widely used assumptions about the duration of a stage, the first being that it is fixed and the second that it is exponentially distributed. Here we consider a slight generalization of the first assumption, namely that

$$\mathcal{F}(a) = \begin{cases} 1 & a = 0 \\ p & 0 < a < c \\ 0 & a > c \end{cases}$$

where  $p \in [0, 1]$ . At  $a = c$ ,  $\mathcal{F}(a)$  can be any number between 0 and  $p$ . If  $p = 1$ , everybody leaves the stage at stage age  $c$ . If  $p = 0$ , everybody leaves the stage immediate after having entered it. If  $p \in (0, 1)$ , one either leaves the stage immediately after having entered it (this happens with probability  $1 - p$ ) or one leaves the stage exactly at the maximum possible stage age  $c$  (with probability  $p$ ). If the stage is life, the jump at age  $a = 0$  idealizes a very short period after birth with high infant mortality. Though  $\mathcal{F}$  is an extreme idealization of a realistic sojourn function, it is widely used in mathematical models which would become too complicated otherwise. Here we want to reflect upon what such an idealization means in terms of average sojourn time, expected remaining sojourn, and average expectation of remaining sojourn. From (1.5) one readily sees that

$$D(a) = \begin{cases} pc & a = 0 \\ c - a & 0 < a < c \\ 0 & a \geq c. \end{cases}$$

We notice that the expected remaining sojourn jumps upward immediately after birth, but then declines steadily till it becomes 0 when reaching the maximum sojourn time. From formula (1.8) we obtain that  $E = \frac{c}{2}$ . So we have the relations

$$D = pc, \quad E = \frac{c}{2} = \frac{1}{2p}D.$$

We see that, by adjusting  $p \in (0, 1]$ ,  $E/D = \frac{1}{2p}$  can take any value between  $1/2$  and infinity. Interestingly enough, these relations characterize fixed stage durations.

**Remark.** (a) If the maximum stage duration,  $c$ , is finite and the average expectation of remaining sojourn satisfies  $E = c/2$ , then  $\mathcal{F}$  is constant on  $(0, c)$ .

(b) If  $E = D/2$ , then the maximum stage duration is finite and equals  $D$ , and  $\mathcal{F}(a) = 1$  for all  $a \in [0, D)$  and  $\mathcal{F}(a) = 0$  for all  $a > D$ .

**1.5. Per capita exit rates (mortality rates).** The conditional probability to exit the stage in the interval between stage age  $a$  and stage age  $a + h$ , under the condition to be still in the stage at stage age  $a$ , is

$$\frac{\mathcal{F}(a) - \mathcal{F}(a + h)}{\mathcal{F}(a)} = 1 - \mathcal{F}(h|a).$$

If  $\mathcal{F}$  is differentiable at  $a$ , then

$$(1.10) \quad \mu(a) = \lim_{h \searrow 0} \frac{1}{h} \frac{\mathcal{F}(a) - \mathcal{F}(a+h)}{\mathcal{F}(a)} = -\frac{\mathcal{F}'(a)}{\mathcal{F}(a)} = -\frac{d}{da} \ln \mathcal{F}(a)$$

can be interpreted as the (per capita) *exit rate* at stage age  $a$ . We can rewrite (1.10) as

$$\mathcal{F}'(a) = -\mu(a)\mathcal{F}(a).$$

If  $\mathcal{F}$  is differentiable on  $(0, c)$ , we can recover  $\mathcal{F}$  from  $\mu$  as

$$(1.11) \quad \mathcal{F}(a) = \exp \left( - \int_0^a \mu(s) ds \right), \quad 0 < a < c.$$

If  $0 \leq b < c$  and  $\mathcal{F}$  is differentiable on  $(b, c)$ , we obtain from (1.4) that

$$(1.12) \quad \mathcal{F}(t|a) = \exp \left( - \int_a^{a+t} \mu(s) ds \right) = \exp \left( - \int_0^t \mu(a+s) ds \right), \\ b < a < c, \quad t \geq 0, \quad a + t < c.$$

$\mathcal{F}$  is differentiable on  $(0, c)$  if and only if the expected remaining sojourn  $D(\cdot)$  is differentiable, as it can be seen from (1.5) and Proposition 1.2. From the subsequent Remark and (1.5) we discover the relations

$$(1.13) \quad D(a) = \int_0^{c-a} \exp \left( - \int_0^t \mu(a+s) ds \right), \quad \mu(a) = -\frac{D'(a) + 1}{D(a)}, \\ 0 \leq a \leq c.$$

The figures in the literature suggest that there is some opposite relationship between the graphs of  $\mu(\cdot)$  and  $D(\cdot)$ . See Hutchinson (1978), Fig. 27, p.45, and Fig. 32, p. 52, e.g. This is not accidental, because (1.13) has the following consequences.

**Proposition 1.4.** *Let  $b \in [0, c]$ .*

- a) *Assume that  $\mathcal{F}$  is differentiable on  $[0, b)$  and  $D(\cdot)$  is non-decreasing and concave on  $[0, b)$ . Then  $\mu$  is non-increasing on  $[0, b)$ .*
- b) *Assume that  $\mathcal{F}$  is differentiable on  $(b, c)$  and that  $\mu$  is non-decreasing on  $(b, c)$ . Then  $D(a)$  is non-increasing in  $a \in (b, \infty)$ .*
- c) *Assume that  $c = \infty$  and  $\mu$  is non-increasing on  $(b, \infty)$ . Then  $D$  is non-decreasing on  $(b, \infty)$ .*

**Proposition 1.5.** a) *Assume that  $\mathcal{F}$  is differentiable on  $(0, c)$  and that  $\mu$  is non-decreasing on the interval  $(0, c)$ . Then  $E \leq D$ .*

b) *Assume that  $c = \infty$  and  $\mu$  is non-increasing on  $(0, \infty)$ . Then  $E \geq D$ .*

*Proof.* a) By Proposition 1.4 (a),  $D(a) \leq D(0) = D$ . The statement now follows from (1.7). The proof of part (b) is similar.  $\square$

If  $\mu(a)$  is the age-dependent per capita mortality rate, it typically decreases first and then increases as a function of age. See Hutchinson (1978), Fig. 32, p. 52.

**1.6. Exponentially distributed stage durations.** An important special case is the (per capita) exit rate  $\mu$  being a constant. Then

$$\mathcal{F}(a) = e^{-\mu a}$$

and we say that the duration of the stage is *exponentially distributed*. Obviously  $\mathcal{F}(a+s) = \mathcal{F}(a)\mathcal{F}(s)$ , which can be written more suggestively as in terms of the conditional sojourn function,

$$\mathcal{F}(s|a) = \mathcal{F}(s) \quad \forall a, s \geq 0,$$

i.e., the probability of staying longer in the stage is independent of the present stage age. Actually the converse is also true: If  $\mathcal{F}(a+s) = \mathcal{F}(a)\mathcal{F}(s)$  for all  $a, s \geq 0$ , then either  $\mathcal{F}(a) = 0$  for all  $a > 0$  or there exists some  $\mu \in \mathbf{R}$  such that  $\mathcal{F}(a) = e^{-\mu a}$ .

If the exit rate is a constant  $\mu$ , then the mean duration of the stage (mean sojourn time in the stage) is given by  $D = \frac{1}{\mu} = D(a)$  for all  $a \geq 0$ . The last equality is a special case of Proposition 1.4, but also follows easily from formula (1.5). Actually the following characterization follows from relation (1.13).

**Proposition 1.6.** *The sojourn time is exponentially distributed if and only if the expected remaining sojourn  $D(a)$  at age  $a$  does not depend on  $a$ .*

By (1.8),  $E = \frac{1}{\mu} = D$  and, from (1.9),  $V = D(2E - D) = (1/\mu)^2$ .

**Proposition 1.7.** *If the sojourn time in a stage is exponentially distributed, the mean sojourn time, the expected remaining sojourn at certain age, the average expectation of remaining sojourn, and the standard deviation of the sojourn time are all identical and equal the reciprocal of the per capita exit rate.*

By definition, the standard deviation is the square root of the variance. In particular, if life duration is exponentially distributed, life expectation and average expectation of remaining life coincide and are the reciprocal of the constant mortality rate.

In later models we will make the idealizing assumption that the life duration is exponentially distributed though it this is not the case in reality. This section teaches us that this assumption makes the life expectation, the average expectation of remaining life, and the standard deviation of life length all the same, though they typically do not coincide in reality. It also puts us into a dilemma. Shall we choose the constant per capita mortality rate as the reciprocal of the life expectation or of the average expectation of remaining life. This will depend on what features we would like to be preserved. See Section 2.

**1.7. Log-normally distributed stage durations.** Presumably the most commonly used probability distribution is the normal distribution. From a theoretical point of view, it is not very satisfying to use the normal distribution for stage durations (sojourn times), because its domain is

the whole real line, while stage durations are non-negative by their very nature. Further it is symmetric and so does not fit asymmetric distributions very well. Nevertheless it has been used, e.g., to determine the length and variance of the latency in measles (*Bailey*, 1975, Chapter 15), mainly because many techniques are available to estimate its mean and standard deviation. Theoretically it is more satisfying to assume that not the sojourn time itself, but its natural logarithm is normally distributed. *Sartwell* (1950, 1966) finds that the incubation periods of various infectious diseases often are reasonably well fit by log-normal distributions (*Sartwell*, 1950, Fig.2) and at any rate better than by normal distributions. The duration of a stage is log-normally distributed, if its probability distribution function satisfies

$$\mathcal{G}(a) = \mathcal{N}(\ln a)$$

where  $\mathcal{N}$  is the probability distribution function of a normal distribution, in other words,

$$\mathcal{F}(a) = \mathcal{F}_{m\sigma}(a) = \int_{\ln a}^{\infty} \exp\left(-\frac{1}{2}\left[\frac{x - \ln m}{\sigma}\right]^2\right) \frac{1}{\sigma\sqrt{2\pi}} dx$$

with  $\ln m$  being the mean and  $\sigma$  the standard deviation of the normal distribution. Following Sartwell we call  $m$  the *median* and  $e^\sigma$  the *dispersion factor* of the distribution. By a couple of substitutions we obtain

$$(1.14) \quad \mathcal{F}(a) = \int_{\frac{\ln(a/m)}{\sigma}}^{\infty} \exp\left(-\frac{1}{2}y^2\right) \frac{1}{\sqrt{2\pi}} dy = 1 - \mathcal{N}_0\left(\frac{\ln(a/m)}{\sigma}\right)$$

where  $\mathcal{N}_0$  is the probability distribution function of a normal distribution with mean 0 and standard deviation 1. Notice that  $\mathcal{F}_{m\sigma}(m) = \frac{1}{2}$  and that  $m$  is the only argument for which this value is taken. Further  $\mathcal{F}_{m\sigma}$  pointwise approaches the sojourn function for stages with fixed duration  $m$  as  $\sigma \rightarrow 0$ .  $\mathcal{F}$  is differentiable and the probability density of stage duration is given by

$$-\mathcal{F}'(a) = \exp\left(-\frac{1}{2}\left[\frac{\ln(a/m)}{\sigma}\right]^2\right) \frac{1}{a\sigma\sqrt{2\pi}}.$$

One of the nice features of the log-normal distribution is the easy calculation of moments:

$$(1.15) \quad \begin{aligned} n \int_0^{\infty} a^{n-1} \mathcal{F}(a) da &= - \int_0^{\infty} a^n \mathcal{F}'(a) da \\ &= m^n \int_{-\infty}^{\infty} e^{nt\sigma} \exp\left(-\frac{1}{2}t^2\right) \frac{1}{\sqrt{2\pi}} dt = m^n e^{(1/2)(n\sigma)^2}. \end{aligned}$$

The average duration of the stage,  $D$ , and the average expectation of remaining sojourn,  $E$ , are given and related by

$$(1.16) \quad D = m e^{(1/2)\sigma^2}, \quad E = \frac{m}{2} e^{(3/2)\sigma^2}, \quad \frac{E}{D} = \frac{1}{2} e^{\sigma^2},$$

so log-normal distributions are flexible enough to fit all  $D$  and  $E$  values in the feasible range  $E/D \geq 1/2$  except fixed stage durations which are the limiting case for  $\sigma \rightarrow 0$ . The variance and coefficient of variation are

$$V = m^2(e^{2\sigma^2} - e^{\sigma^2}) = D^2(e^{\sigma^2} - 1), \quad \gamma_0 = \sqrt{e^{\sigma^2} - 1}.$$

Log-normally distributed stage durations are always positively skewed, the skewness  $\gamma_1 = (e^{\sigma^2} + 2)(e^{\sigma^2} - 1)^{1/2}$  increases with  $\sigma$  as does the kurtosis,  $\gamma_2 = e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{\sigma^2} - 6$ .

Since the normal distribution is symmetric, its mean and its median coincide which in our case is  $\ln m$ . Since medians are preserved under transformation,  $m$  inherits being the median of stage sojourn, but not being the mean, which is  $D$ . The fact that  $D > m$  is not surprising, because the log-normal distribution is positively skewed. It is known from the properties of the normal distribution that about 16% of the logarithmic sojourn times lie above  $\ln m + \sigma$  and again about 16% below  $\ln m - \sigma$ . This means that about 16% of the sojourn times will exceed  $me^\sigma$  and about 16% will fall below  $me^{-\sigma}$ . This nice relation has motivated Sartwell (1950, 1966) to list the dispersion factor  $e^\sigma$  in addition to the median for the incubation times of many infectious diseases. Unfortunately the dispersion factor  $e^\sigma$  itself is not so useful for our purposes, and we have to recover  $\sigma$  from Sartwell's tables.

From the data of Stillerman and Thalhimer (1944) on 199 cases of measles, Sartwell (1950) estimates a median of 12.2 days and a dispersion factor of 1.18 for the incubation period (actually the time interval between the onsets of rash in the first and second cases in households). This gives  $e^{\sigma^2} \approx 1.028$  and  $e^{(1/2)\sigma^2} \approx 1.014$ . So the mean duration of the incubation period,  $D$ , would be about 12.4 days, while the average expectation of remaining sojourn,  $E$ , would be about 6.4 days. Further  $\sqrt{V}/D \approx 0.17$  for the coefficient of variation.

For Sartwell's estimate of another set of measles data (Goodall, 1931), one obtains  $\sqrt{V}/D \approx 0.28$ . Remember that for an exponential distribution this number would be one. Very few of the many diseases reported by Sartwell have dispersion factors above 1.6. If the dispersion factor is below 1.6, then  $\sqrt{V}/D \leq 0.5$  and  $2E/D = e^{\sigma^2} \leq 1.25$ . Recall that always  $2E/D \geq 1$ , with equality for fixed durations, and  $E = D$  for exponential distributions. For comparison, Bailey (1975, Chapter 15), estimates the latency and infectious period for three sets of measles data, assuming that the latency period is normally distributed while the infectious period has fixed length. The values for the mean length and the variance are different from set to set, but for all three the coefficient of variation satisfies

$\sqrt{V}/D \approx 0.21$  for the latency period. By formula (1.9),  $2E/D \approx 1.044$ . *Gough* (1977) and *Becker* (1989) rework one of these sets of measles data (collected by *Hope Simpson*; see *Bailey*, 1975, Table 15.1; *Becker*, 1989, Table 4.1). *Becker* makes the same assumptions as *Bailey*, but his technical procedure and his results are slightly different; still he basically finds the same coefficient of variation, 0.200 compared to 0.206. *Gough* assumes that both the latent and infectious period are  $\chi^2$  distributed, mentioning that there is no empirical or theoretical evidence for this assumption, and finds the coefficient of variation to be approximately 0.17 for the latent period and 0.1 for the infectious period. *Gough's* estimates provide longer average periods of latency and infectiousness than *Bailey's*. At the end of Section 3.2 we will realize the relevance of these relations in determining a threshold parameter for stability loss of the endemic equilibrium.

**2. Stage dynamics with given input.** In order to describe the dynamics in a particular stage, we stratify the number  $N(t)$  of individuals that are in the stage at time  $t$  along stage age,

$$(2.1) \quad N(t) = \int_0^\infty u(t, a) da.$$

Here  $u(t, a)$  is the density of individuals at time  $t$  that have stage age  $a$ , i.e., the integral  $\int_a^b u(t, r) dr$  gives the number of individuals that are in the stage at time  $t$  and have a stage age between  $a$  and  $b$ . We recall that *stage age* of an individual in a certain stage is the time the individual has already spent in the stage. We call  $u(t, \cdot)$  the *stage-age density* and  $N(t)$  the *stage content* at time  $t$ .

**2.1. Stage input and stage-age density.** Assume that  $B(t)$  is the input into the stage at time  $t$ , i.e., the rate of individuals entering the stage at time  $t$ , i.e., the number of individuals entering the stage at time  $t$  per unit of time. If the stage is life,  $B$  is the birth rate of the population under consideration.

Let 0 be the time at which we start to consider the development of the stage, and  $u_0(a)$  the age density of individuals that are in the stage at time 0. As before, let  $\mathcal{F}(a)$  be the probability to be still in the stage at stage age  $a$ .

Recall the definition of the maximum stage age  $c$ . If  $\mathcal{F}$  is strictly positive on  $[0, \infty)$ , then  $c = \infty$ . Otherwise  $0 < c < \infty$  is determined by the property that  $\mathcal{F}(a) > 0$  for  $a < c$  and  $\mathcal{F}(a) = 0$  for  $a > c$ . Consistently with the interpretation of maximum stage age we assume that  $u_0(a) = 0$  for  $a > c$ .

The stage age density  $u$  can now be expressed as follows:

$$(2.2) \quad \begin{aligned} u(t, a) &= B(t - a)\mathcal{F}(a), & t > a \geq 0, \\ u(t, a) &= u_0(a - t) \frac{\mathcal{F}(a)}{\mathcal{F}(a - t)}, & 0 \leq t < a < c, \\ u(t, a) &= 0, & t \geq 0, a > c. \end{aligned}$$

The first equation comes from the following consideration: An individual that has stage age  $a$  at time  $t > a$ , has entered the stage at time  $t - a$ . The probability that it is still in the stage is  $\mathcal{F}(a)$ .

As for the second equation the reasoning goes as follows: An individual that has stage age  $a$  at time  $t < a$ , already was in the stage at time 0 and, at time 0, had the stage age  $a - t$ . The probability to be still in the stage at time  $t$ , with stage age  $a$ , is then the conditional probability  $\frac{\mathcal{F}(a)}{\mathcal{F}(a-t)}$ , the probability to still be in the stage at stage  $a$  under the condition one was in the stage at stage age  $a - t$ .

One typically leaves  $u(t, a)$  undefined for  $t = a$ , because it is not clear whether one should choose  $B(0)\mathcal{F}(a)$  or  $u_0(0)\mathcal{F}(a)$  unless  $B(0) = u_0(0)$ . But there is no biological reason why such an equality should hold. The following convention will make our formalism much easier:

$$\frac{u_0(a)}{\mathcal{F}(a)} := 0, \quad a > c.$$

We have from (2.1) that

$$N(t) = \int_0^t B(t-a)\mathcal{F}(a)da + \int_t^\infty \mathcal{F}(a) \frac{u_0(a-t)}{\mathcal{F}(a-t)} da.$$

This means that we can split up the stage content into two parts, one consisting of individuals that were in the state already at time  $t = 0$  and another consisting of individuals that have entered later. The size of the first part is denoted by  $N_0(t)$  and the size of the second by  $N_1(t)$ ; so after substitution

$$(2.3) \quad \begin{aligned} N(t) &= N_0(t) + N_1(t), \\ N_0(t) &= \int_0^c u_0(a) \frac{\mathcal{F}(a+t)}{\mathcal{F}(a)} da, \\ N_1(t) &= \int_0^t B(t-a)\mathcal{F}(a)da. \end{aligned}$$

**2.2. The PDE formulation.** Assume that  $B$  and  $u_0/\mathcal{F}$  are differentiable. Then, by (2.2),

$$\frac{u(t, a)}{\mathcal{F}(a)} = \begin{cases} B(t-a); & t > a \geq 0, \\ \frac{u_0(a-t)}{\mathcal{F}(a-t)}; & 0 \leq t < a < c, \end{cases}$$

is differentiable for  $t \neq a$  and satisfies the PDE

$$\frac{\partial}{\partial t} \frac{u(t, a)}{\mathcal{F}(a)} + \frac{\partial}{\partial a} \frac{u(t, a)}{\mathcal{F}(a)} = 0, \quad t \neq a.$$

From (2.2) and  $\mathcal{F}(0) = 1$ , we also have the boundary condition

$$u(t, 0) = B(t),$$

and the initial condition

$$u(0, a) = u_0(a).$$

Now assume that  $B, u_0$  and  $\mathcal{F}$  are differentiable. Then  $u$  is differentiable for  $t \neq a$  and

$$0 = \frac{\partial}{\partial t} u(t, a) + \frac{\partial}{\partial a} u(t, a) - u(t, a) \frac{\mathcal{F}'(a)}{(\mathcal{F}(a))^2}.$$

Multiplying this equation by  $\mathcal{F}$  and using the per capita exit rate

$$\mu(a) = -\frac{\mathcal{F}'(a)}{\mathcal{F}(a)},$$

we obtain the following PDE for  $u$  with boundary and initial conditions:

$$(2.4) \quad \begin{aligned} \frac{\partial}{\partial t} u(t, a) + \frac{\partial}{\partial a} u(t, a) + \mu(a)u(t, a) &= 0, & t \neq a, \\ u(t, 0) &= B(t), \\ u(0, a) &= u_0(a). \end{aligned}$$

This equation is nowadays called *McKendrick's equation* (*McKendrick*, 1926), after it was associated with *Von Foerster* (1959) for a while. Equation (2.2) can be recovered from it integrating along characteristic curves.

**2.3. Stage content and average stage duration.** In this section we want to relate stage content, stage input, and average stage duration under the assumption that the stage input stabilizes. We start with the observation that the individuals that were in the state already at time 0 vanish as time tends to infinity. Recall formula (2.3).

We assume that the function  $u_0$  on  $[0, c)$  is non-negative and that

$$N_0(0) = \int_0^c u_0(a)da$$

exists and is finite. This means that the initial stage content has finite size.

**Proposition 2.1.** *Let  $N_0(0) < \infty$  and  $\mathcal{F}(\infty) = 0$ . Then  $N_0(t) \leq N_0(0) < \infty$  for  $t > 0$  and*

$$N_0(t) \rightarrow 0, \quad t \rightarrow \infty.$$

*Proof.* The first statement follows from  $\mathcal{F}$  being non-increasing, the second statement from Lebesgue's theorem of dominated convergence.  $\square$

If the stage input approaches a finite limit, so does the stage content and the two limits are related to each other by the average stage duration (or mean sojourn time),  $D$ .

**Theorem 2.2.** Assume  $N_0(0) < \infty$  and  $D < \infty$  and that  $B(t)$  tends to a finite strictly positive limit  $B(\infty)$  as  $t \rightarrow \infty$ . Then

$$N(t) \rightarrow DB(\infty)$$

where  $D$  is the mean sojourn time in the stage.

*Proof.* By assumption,

$$\infty > D = \int_0^\infty \mathcal{F}(a)da \geq \int_0^t \mathcal{F}(a)da \geq t\mathcal{F}(t).$$

Hence

$$\mathcal{F}(t) \leq D/t \rightarrow 0, \quad t \rightarrow \infty.$$

By Proposition 2.1,  $N_0(t) \rightarrow 0$  as  $t \rightarrow \infty$ . So it is sufficient to prove that  $N_1(t) \rightarrow DB(\infty)$  as  $t \rightarrow \infty$  which follows from Lebesgue's theorem of dominated convergence.  $\square$

If we consider the dynamics of a population, this theorem means the following: If the population birth rate stabilizes to some limit  $B(\infty)$ , the population size stabilizes to this limit times the life expectation.

As for the stage age density, if the input converges, we have from (2.2) that  $n(t, a) \rightarrow B(\infty)\mathcal{F}(a)$  for  $t \rightarrow \infty$ , pointwise in  $a \geq 0$ .

**2.4. Average stage age and average expectation of remaining sojourn.** The mean duration of a stage (expected sojourn time in a stage) has to be discriminated from the mean time one already spent in a stage, the average stage (or class) age. If the stage is life, the average stage age coincides with the average age of a population. Unfortunately, the average stage age depends very much on the input into the stage, i.e., the way in which individuals enter the stage.

The *average stage age* is the average of the individual stage ages taken over all individuals in the stage. At time  $t$ , it is given by

$$A(t) = \frac{1}{N(t)} \int_0^\infty au(t, a)da.$$

By (2.2),

$$\begin{aligned} A(t) &= A_0(t) + A_1(t), \\ A_0(t) &= \frac{1}{N(t)} \int_0^\infty (a+t)u_0(a) \frac{\mathcal{F}(a+t)}{\mathcal{F}(a)} da, \\ A_1(t) &= \frac{1}{N(t)} \int_0^t aB(t-a)\mathcal{F}(a)da. \end{aligned}$$

If the input stabilizes, the average stage age approaches the average expectation of remaining sojourn under reasonable assumptions (cf. Kim, Aron, 1989).

**Theorem 2.3.** Assume that  $B(t)$  tends to a finite strictly positive limit  $B(\infty)$  as  $t \rightarrow \infty$  and  $D, E$ , and  $N_0(0)$  are finite.

- (a) Then  $A_1(t) \rightarrow E$  as  $t \rightarrow \infty$ .
- (b) If  $A_0(0)$  is finite as well and there exists a constant  $k > 0$  such that

$$\mathcal{F}(a+t) \leq k \frac{1+a}{1+t} \mathcal{F}(a) \quad \forall t, a \geq 0,$$

then  $A(t) \rightarrow E$  as  $t \rightarrow \infty$ .

*Proof.* The proof of (a) is identical to the proof of Theorem 2.2 mutandis mutatis.

(b) By (a) it sufficient to show that  $A_0(t) \rightarrow 0$  as  $t \rightarrow \infty$ . The result obviously holds if  $c < \infty$ , so we assume  $c = \infty$ . The proof that

$$\int_0^\infty a u_0(a) \frac{\mathcal{F}(a+t)}{\mathcal{F}(a)} da \rightarrow 0 \quad t \rightarrow \infty,$$

follows from  $A_0(0) < \infty$  and Lebesgue's Theorem of dominated convergence and  $\mathcal{F}(t) \rightarrow 0$  as  $t \rightarrow \infty$ . Further

$$t \frac{\mathcal{F}(a+t)}{\mathcal{F}(a)} \leq k(1+a) \frac{t}{1+t}$$

and

$$\int_0^\infty u_0(a)(1+a) \frac{t}{1+t} da \leq N_0(0) + A_0(0).$$

Finally, for  $t \geq 1$ ,

$$t\mathcal{F}(a+t) \leq t\mathcal{F}(t) \leq \frac{2}{t} \int_0^t s\mathcal{F}(s)ds \leq \frac{2}{t}ED \rightarrow 0 \quad t \rightarrow \infty.$$

So Lebesgue's theorem of dominated convergence implies

$$\int_0^\infty t u_0(a) \frac{\mathcal{F}(a+t)}{\mathcal{F}(a)} da \rightarrow 0 \quad t \rightarrow \infty.$$

□

Let us return to the dilemma we face when we want to replace a realistic survival distribution by an exponentially distributed one (Proposition 1.5). This dilemma originates from  $D = E = 1/\mu$  for an exponentially distributed survivorship, while the equality  $D = E$  does not hold in reality, creating the question whether one should choose  $\mu = 1/D$  or  $\mu = 1/E$  as the per capita mortality rate. From Theorem 2.3 and Theorem 2.5 we have the relations

$$N(t) \rightarrow DB(\infty), \quad A(t) \rightarrow E, \quad t \rightarrow \infty.$$

Most of the time, it is the first relation one wants to be preserved and so one typically chooses  $\mu$  as the reciprocal of the life expectation, and not as the reciprocal of the expectation of remaining life (or of the standard variation of life length).

**2.5. Stage exit rate and fundamental law of stage dynamics.** The exit rate of the stage,  $C$ , is the rate at which individuals leave the stage, i.e., the number of individuals leaving the stage per unit of time. Let us assume that  $\mathcal{F}$  is differentiable on  $[0, \infty)$  (or, more generally, absolutely continuous), i.e., the per capita exit rate  $\mu(\cdot)$  is defined for (almost) all ages,

$$\mu(a) = -\frac{\mathcal{F}'(a)}{\mathcal{F}(a)}, \quad a \in [0, \infty).$$

Then the stage exit rate is given by

$$(2.5) \quad \begin{aligned} C(t) &= \int_0^\infty \mu(a)u(t, a)da = C_1(t) + C_0(t), \\ C_1(t) &= \int_0^t \mu(a)B(t-a)\mathcal{F}(a)da, \\ C_0(t) &= \int_t^\infty \mu(a)\mathcal{F}(a)\frac{u_0(a-t)}{\mathcal{F}(a-t)}da. \end{aligned}$$

Here we have used formula (2.2) for the stage age density  $u$ .  $C_0$  is the part of the exit rate associated with the individuals that were in the state initially, while  $C_1$  is the part associated with those that have entered later.

We show that the state content  $N(t)$  is absolutely continuous in  $t$  and

$$(2.6) \quad N'(t) = B(t) - C(t), \text{ for a.a. } t > 0,$$

i.e., the rate of change of the population size is the difference between the stage input rate and the stage exit rate.

If the stage is life,  $N$  can be interpreted as the population size,  $B$  as the population birth rate and  $C$  as the population mortality rate.

With this interpretation (2.6) is called the *fundamental law of population dynamics* for closed population, i.e., without immigration and emigration. For a general stage, we call it the *fundamental law of stage dynamics*.

**Theorem 2.4.**

$$N(t) = N(0) + \int_0^t B(s)ds - \int_0^t C(s)ds, \quad t \geq 0.$$

*Proof.* Substituting, changing the order of integration and using that  $\mu(a)\mathcal{F}(a) = -\mathcal{F}'(a)$ ,

$$\begin{aligned} \int_0^t C_1(s)ds &= \int_0^t \left( \int_0^s \mu(s-a)\mathcal{F}(s-a)B(a)ds \right) dt \\ &= \int_0^t \left( \int_a^t \mu(s-a)\mathcal{F}(s-a)ds \right) B(a)dt \\ &= \int_0^t (1 - \mathcal{F}(t-a))B(a)da = \int_0^t B(a)da - N_1(t) \end{aligned}$$

and

$$\begin{aligned}
 \int_0^t C_0(s)ds &= \int_0^t \left( \int_0^\infty \mu(a+s)\mathcal{F}(a+s) \frac{u_0(a)}{\mathcal{F}(a)} da \right) ds \\
 &= \int_0^\infty \left( \int_0^t \mu(a+s)\mathcal{F}(a+s)ds \right) \frac{u_0(a)}{\mathcal{F}(a)} da \\
 &= \int_0^\infty (\mathcal{F}(a) - \mathcal{F}(a+t)) \frac{u_0(a)}{\mathcal{F}(a)} da \\
 &= N(0) - N_0(t).
 \end{aligned}$$

Adding the equations yields the result. See (2.5).  $\square$

As we can see from the proof of the last proposition or directly from (2.5), we can write the exit rate as

$$\begin{aligned}
 C(t) &= C_0(t) + C_1(t), \\
 C_1(t) &= - \int_0^t B(t-a)\mathcal{F}'(a)da, \\
 C_0(t) &= - \int_t^\infty \mathcal{F}'(a) \frac{u_0(a-t)}{\mathcal{F}(a-t)} da.
 \end{aligned}$$

In case that  $\mathcal{F}$  is not differentiable, we can formally express the stage exit rate by Stieltjes integrals

$$C(t) = - \int_0^t B(t-a)\mathcal{F}(da) - \int_t^\infty \frac{u_0(a-t)}{\mathcal{F}(a-t)} \mathcal{F}(da)$$

with the usual understanding that

$$\frac{u_0(a)}{\mathcal{F}(a)} = 0, \quad a \geq c,$$

if  $c < \infty$ . This formula can be given a rigorous meaning, with the fundamental law of stage dynamics remaining valid, by interpreting  $\mathcal{F}(da)$  as integration with the Lebesgue Stieltjes measure associated with  $\mathcal{F}$ . See the Appendix, Lemma A.2 and A.3.

**2.6. Stage outputs.** If a stage does not comprehend the whole life but only a part, the stage exit rate is made of individuals leaving by death and of individuals leaving alive. We consider the second part the *output* of the stage, and we may be particularly interested in this one. If one looks at various stages of an insect species, the input of the pupal stage, e.g., is the living part of the exit rate of the larval stage, i.e., the output of the larval stage.

In order to address the output, we factor the sojourn function  $\mathcal{F}$  as

$$\mathcal{F}(a) = \mathcal{F}_s(a)\mathcal{F}_d(a),$$

where  $\mathcal{F}_s(a)$  is the probability to be still alive at stage age  $a$  and  $\mathcal{F}_d(a)$  is the probability of not having left the stage before stage age  $a$  in other ways than dying. Both  $\mathcal{F}_s$  and  $\mathcal{F}_d$  are non-negative, non-increasing functions on  $[0, \infty)$ ,  $\mathcal{F}_s(0) = \mathcal{F}_d(0) = 1$ .

While we have not done so before, we now distinguish between stage duration and stage sojourn, if a stage is only a part of life. *Stage duration* is the length of the stage discounting death, *stage sojourn* may also be terminated by death. We call  $\mathcal{F}_s$  the (stage) *survival function*,  $\mathcal{F}_d$  the *stage duration function*, and the product  $\mathcal{F}$  the *stage sojourn function*.

Assume that  $\mathcal{F}_s$  and  $\mathcal{F}_d$  are differentiable. Then

$$\mu_s(a) = -\frac{\mathcal{F}'_s(a)}{\mathcal{F}_s(a)}, \quad \mu_d(a) = -\frac{\mathcal{F}'_d(a)}{\mathcal{F}_d(a)}$$

are the per capita mortality rate,  $\mu_s$ , and the per capita rate of leaving the stage alive,  $\mu_d$ . The (living) output of the stage is given by

$$\begin{aligned} C^d(t) &= \int_0^\infty \mu_d(a) u(t, a) da \\ &= \int_0^t B(t-a) \mu_d(a) \mathcal{F}(a) da + \int_t^\infty \frac{u_0(a-t)}{\mathcal{F}(a-t)} \mu_d(a) \mathcal{F}(a) da. \end{aligned}$$

Using the formulas for  $\mu_s$  and  $\mu_d$ ,

$$C^d(t) = - \int_0^t B(t-a) \mathcal{F}_s(a) \mathcal{F}'_d(a) da - \int_t^\infty \frac{u_0(a-t)}{\mathcal{F}(a-t)} \mathcal{F}_s(a) \mathcal{F}'_d(a) da.$$

If  $\mathcal{F}_s$  or  $\mathcal{F}_d$  are not differentiable, then we still have the Stieltjes formulation

$$\begin{aligned} (2.7) \quad C^d(t) &= C_1^d(t) + C_0^d(t), \\ C_1^d(t) &= - \int_0^t B(t-a) \mathcal{F}_s(a) \mathcal{F}_d(da), \\ C_0^d(t) &= - \int_t^\infty \frac{u_0(a-t)}{\mathcal{F}(a-t)} \mathcal{F}_s(a) \mathcal{F}_d(da), \\ C^d(t) &= \int_0^\infty \frac{u(t, a)}{\mathcal{F}_d(a)} \mathcal{F}_d(da). \end{aligned}$$

with  $C_1^d$  being the part originating from the stage input and  $C_0^d$  being the part originating from the initial stage content.  $\mathcal{F}_d(da)$  is to be interpreted as integration with the Lebesgue Stieltjes measure associated with  $\mathcal{F}_d$ . Recall that

$$\frac{u_0(a)}{\mathcal{F}(a)} = 0, \quad a \geq c.$$

A similar formula holds for the stage mortality rate,  $C^s(t)$ . You just need to interchange the indices  $s$  and  $d$ . The following refined version of the

fundamental law of stage dynamics holds (see the Appendix, Corollary A.5).

**Theorem 2.5.** *Assume that  $\mathcal{F}_s$  and  $\mathcal{F}_d$  have no discontinuities in common. Then  $N$  is absolutely continuous and*

$$N' = B(t) - C^s - C^d.$$

Lebesgue's theorem of dominated converges provides a relation between input and output of a stage.

**Theorem 2.6.** *Let*

$$p = - \int_0^\infty \mathcal{F}_s(a) \mathcal{F}_d(da) < \infty.$$

- a) If  $B(t) \rightarrow B(\infty)$  as  $t \rightarrow \infty$ , then  $C_1^d(t) \rightarrow pB(\infty)$ .
- b) If  $u_0/\mathcal{F}$  is bounded on  $[0, c)$ , then  $C_0^d(t) \rightarrow 0$  as  $t \rightarrow \infty$ .

**Remark.** Notice that

$$(2.8) \quad p = - \int_0^\infty \mathcal{F}_s(a) \mathcal{F}_d(da).$$

is a number between 0 and 1. It can be interpreted as the probability to get through the stage alive as can be seen by approximating the right hand side of (2.8) by Stieltjes sums. Integration by parts shows that

$$(2.9) \quad 1 - p = - \int_0^\infty \mathcal{F}_d(a) \mathcal{F}_s(da)$$

is the probability of dying while in the stage.

**Example 2.7.** Let the stage have a fixed duration  $c \in (0, \infty)$ , i.e.,  $\mathcal{F}_d(a) = 1$  for  $a \in [0, c]$  and  $\mathcal{F}_d(a) = 0$  for  $a > c$ , and let the stage survival function  $\mathcal{F}_s$  be defined on some interval  $[0, b)$  with  $b > c$  and be continuous at  $c$ . Then  $p = \mathcal{F}_s(c)$ .

*Proof.* The Lebesgue Stieltjes measure associated with  $-\mathcal{F}_d$  is the point (or Dirac) measure concentrated at  $c$ .  $\square$

**3. Application to epidemic modeling.** Stage transition has two aspects in epidemic modeling at least. First, while contracting and transmitting the disease, the individuals of the population are also passing through life, and one may like to model this passage as realistically as possible, in particular if one wants to figure out optimal ages of vaccination. We present a very simple model in Section 3.1 in order to illustrate how expected remaining life naturally appears in the basic reproduction number of the disease. Secondly, one would like to take account of the various stages of the disease (Section 3.2), with the (living) output of each stage being the input into the next one.

**3.1. An epidemic model with chronological age structure.** Following Busenberg *et al.* (1988) in a special case, we consider a non-fatal infectious disease of  $S \rightarrow I$  type with no recovery in a population with constant size  $N$  and stationary age density  $n(\cdot)$ ,

$$(3.1) \quad N = \int_0^\infty n(a)da, \quad n(a) = B\mathcal{F}(a).$$

$B$  is the constant birth rate and  $\mathcal{F}$  the survival function. Cf. equation (2.2). We restrict ourselves to the so-called inter-cohort case, where a susceptible individual can be infected by infectives of any age with an equal per capita rate,  $\kappa(a)$ , that only depends on the age of the susceptible,  $a$ . Here we only want to determine the basic reproductive number,  $\mathcal{R}_0$ ; the most direct way consists in the next generation approach of Diekmann *et al.* (1990). We consider one average infectious individual that is introduced into the population which is completely susceptible and calculate the number of infectives it produces. The age-density of infectives,  $i(a)$ , produced by the one initial infective, satisfies the differential equation

$$\frac{\partial}{\partial a} \frac{i(a)}{\mathcal{F}(a)} = \kappa(a) \frac{n(a)}{\mathcal{F}(a)} = \kappa(a)B, \quad i(0) = 0,$$

and the basic reproductive number is

$$\mathcal{R}_0 = \int_0^\infty i(a)da.$$

Integrating twice and changing the order of integration,

$$(3.2) \quad \begin{aligned} \mathcal{R}_0 &= \int_0^\infty \left( \int_0^a B\kappa(s)ds \right) \mathcal{F}(a)da \\ &= \int_0^\infty B\kappa(s) \left( \int_s^\infty \mathcal{F}(a)da \right) ds \\ &= \int_0^\infty \kappa(s)n(s) \frac{\int_s^\infty \mathcal{F}(a)da}{\mathcal{F}(s)} ds. \end{aligned}$$

Recalling the definition of expected remaining life at age  $a$ ,  $D(a)$ , in (1.5),

$$\mathcal{R}_0 = \int_0^\infty \kappa(s)n(s)D(s)ds.$$

This formula makes sense:  $\kappa(s)n(s)$  is the rate at which individuals become infective at age  $s$  and  $D(s)$  is the expected time they have left to live as infectives.

If the susceptibility does not depend on age, this formula reduces to

$$\mathcal{R}_0 = \kappa \int_0^\infty B\mathcal{F}(s)D(s)ds = \kappa NE,$$

with  $E$  being the average expectation of remaining life introduced in (1.7) (cf. *Kim, Aron*, 1989). Here we have used that  $N = BD$  for a population with constant input. A model with age-independent per capita mortality rate  $\mu$  would have given  $\mathcal{R}_0 = \kappa N(1/\mu)$  and may have led us to misinterpret  $1/\mu$  as life expectation rather than the average expectation of remaining life. If  $\kappa$  is constant, we see from (3.2) and (3.1) that

$$\mathcal{R}_0 = \kappa \int_0^\infty an(a)da = \kappa NA,$$

with  $A$  being the mean age of the population. As we have seen in Section 2.4, the mean age of a population of constant size and stationary age-distribution equals the average expectation of remaining life.

**3.2. Epidemic models with arbitrarily distributed infection stages.** We consider the spread of an infectious disease in a population the epidemiologically relevant part of which has size  $N(t)$  at time  $t$ . The epidemiologically relevant part is the whole population for diseases like influenza or rubella, while it is the sexually active part of the population for sexually transmitted diseases. We divide the population into susceptible and infected individuals, the number of susceptible individuals at time  $t$  is denoted by  $S(t)$ . The infected individuals are further divided into  $n$  stages of infection,

$$N(t) = S(t) + \sum_{j=1}^n I_j(t)$$

with  $I_j$  denoting the number of individuals in the  $j^{th}$  stage. A possible division is  $n = 4$  with  $I_1(t) = E(t)$  denoting the exposed individuals (those in the latency period who are infected, but not yet infectious),  $I_2(t)$  denoting the infectious individuals,  $I_3(t) = Q(t)$  the individuals in quarantine who are potentially infectious, but have been isolated, and  $I_4(t) = R(t)$  denoting the recovered individuals who are no longer infectious. In modeling a disease like HIV/AIDS one may like to further divide the infectious stage according to disease progression (*Hethcote, Van Ark*, 1992; *Simon, Jacquez*, 1992).

The change of the susceptible population obeys the following law:

$$\begin{aligned}\dot{S}(t) &= \Lambda - \mu S(t) - B_0(t) + B_n(t), & S(0) &= \check{S}, \\ B_0(t) &= f(S(t), I(t)), & I(t) &= (I_1(t), \dots, I_n(t)).\end{aligned}$$

Here  $\Lambda$  is the (constant) influx (or recruitment) rate of new individuals into the epidemiologically relevant part of the population, all freshly entering individuals are assumed to be susceptible. If the whole population is epidemiologically relevant,  $\Lambda$  is the population birth rate.  $\mu$  is the per

capita mortality rate due to not disease-related causes.  $B_0(t)$  is the incidence, i.e., the infection rate at time  $t$ , which is a function of the number of susceptibles and the number of individuals in the various infected classes.  $B_n(t)$  is the rate of individuals who have recovered from the disease, but lose their immunity and return into the susceptible class. The function  $f$  describes how the incidence depends on the number of susceptibles and the number of infected individuals in the various stages.

Following our approach in Section 2.1, we stratify the individuals in the  $j^{th}$  stage of infection as

$$I_j(t) = \int_0^{a_j} u_j(t, a) da$$

where  $u_j(t, \cdot)$  denotes the stage age density at time  $t$  and  $a_j > 0$  is the maximum sojourn time in stage  $j$ .  $a_j$  may be finite or infinite.

We introduce functions  $P_j, \mathcal{F}_j : [0, \infty) \rightarrow [0, 1]$  which describe the duration of the  $j^{th}$  stage and the disease-related mortality in the  $j^{th}$  stage. More precisely  $P_j(a)$  is the probability that the  $j^{th}$  stage lasts longer than  $a$  time units. Further  $1 - \mathcal{F}_j(a)$  gives the probability to die from disease related causes during the  $j^{th}$  stage before reaching stage age  $a$ . Using the language of Section 2.6,  $P_j$  is the duration function of stage  $j$ , while  $\mathcal{F}_j$  is the disease survival function. The overall survival function of stage  $j$  is given by  $e^{-\mu a} \mathcal{F}_j(a)$ .  $\mathcal{F}_j$  and  $P_j$  are non-negative, non-increasing functions on  $[0, \infty)$ ,

$$\mathcal{F}_j(0) = 1 = P_j(0).$$

The average duration of the  $j^{th}$  stage,  $D_j$ , and the average sojourn time,  $T_j$ , are given by

$$D_j = \int_0^\infty P_j(a) da, \quad T_j = \int_0^\infty e^{-\mu a} \mathcal{F}_j(a) P_j(a) da.$$

We assume that  $D_j < \infty$  for all stages  $j$  except possibly for the last stage,  $j = n$ .

Notice that  $P_j$  and  $\mathcal{F}_j$  are not necessarily continuous or absolutely continuous. This allows us to include the case that a stage has a fixed duration. We assume, however, that  $P_j$  and  $\mathcal{F}_j$  have no joint discontinuities.

In order to describe the stage dynamics, let  $B_{j-1}(t)$  be the rate of individuals entering the  $j^{th}$  stage. For  $j = 2, \dots, n$ , this is also the rate of individuals leaving the  $(j-1)^{st}$  stage alive. In other words, the output of one stage is the input of the next. Then, following section 2.1,

$$u_j(t, a) = \begin{cases} B_{j-1}(t-a) \mathcal{F}_j(a) P_j(a) e^{-\mu a}; & 0 \leq a < t, \\ \check{u}_j(a-t) \frac{\mathcal{F}_j(a) P_j(a)}{\mathcal{F}_j(a-t) P_j(a-t)} e^{-\mu t}; & 0 < t \leq a < a_j, \\ u_j(t, a) = 0; & t \geq 0, a > a_j. \end{cases}$$

To close our model we must still describe  $B_j(t)$ , the rate at which individuals leave stage  $j$  and enter stage  $j + 1$  if  $j < n$  or return to the susceptible class if  $j = n$ . Using the language of Section 2.6,  $B_j$  is the output of stage  $j$ ,

$$B_j(t) = - \int_0^{a_j} \frac{u_j(t, a)}{P_j(a)} P_j(da),$$

where  $P_j(da)$  denotes integration with the Lebesgue Stieltjes measure associated with  $-P_j(a)$ . Cf. (2.7) and Appendix A. In Appendix A (Corollary A.6), we will also show that  $I_j$  is absolutely continuous and

$$I'_j(t) = B_{j-1}(t) - \mu I_j(t) + \int_0^{a_j} \frac{u_j(t, a)}{\mathcal{F}_j(a)} \mathcal{F}_j(da) - B_j(t) \quad \text{a.e.}$$

with  $B_{j-1}$  and  $B_j$  given as above and the following interpretation: The rate of change of the number of individuals in the  $j^{\text{th}}$  stage is the input into the stage minus the rate of deaths due to disease-unrelated causes, the rate of deaths from the disease, and the output of the stage.

On the one hand one can reduce this system of equations to a system of Volterra integral equations for  $S, I_1, \dots, I_n$ , and on the other hand one can show that

$$\Theta(t, \check{S}, \check{u}_1, \dots, \check{u}_n) = (S(t), u_1(t, \cdot), \dots, u_n(t, \cdot))$$

defines a continuous dynamical system (semiflow)  $\Theta$  on  $[0, \infty) \times L_+^1[0, a_1] \times \dots \times L_+^1[0, a_n]$ . The first representation is useful to show existence and uniqueness of solutions, the second to study qualitative behavior like stability of equilibria and disease extinction or persistence. It is possible to identify a basic reproduction number  $\mathcal{R}_0$  for the disease such that an endemic equilibrium exists if and only if  $\mathcal{R}_0 > 1$ . Further the disease goes extinct if  $\mathcal{R}_0 < 1$  and persists if  $\mathcal{R}_0 > 1$  (Feng, Thieme, 2000a).

The endemic equilibrium is locally asymptotically stable if there is no return from the last infection stage to the susceptible class and the function  $f$  has the property that, at the equilibrium value  $S^*$ , the partial derivatives  $\frac{\partial f}{\partial I_j}(S^*, I)$ ,  $j = 1, \dots, n$ , are non-negative and monotone non-increasing in  $I \in (0, \infty)^n$  with the canonical coordinate-wise order. (Details can be found in Feng, Thieme, 2000a).

**Diseases without fatalities and fast disease dynamics.** In addition to no return of infected individuals to the susceptible class, let us assume that there are no disease fatalities and that the life expectation is very long compared to the length of the various periods of infection. Then the condition for the stability of the endemic equilibrium can be written in explicit though complex terms (Feng, Thieme, 2000b). These formulas become much simpler if there is only one stage, let us say stage  $m$ , during which the infected individuals are actually infective. The stages before are exposed stages and the stages after quarantine or immune stages.

**Standard or mass action incidence.** Let  $f$  describe mass action or standard incidence,

$$f(S, I) = \kappa S I_m, \quad \text{or} \quad f(S, I) = \kappa \frac{S I_m}{N}.$$

Since under our demographics,  $N(t) \rightarrow \frac{\Lambda}{\mu}$  as  $t \rightarrow \infty$  because there are no disease fatalities, there is little difference between the two as far as the stability conditions are concerned. In both cases  $f$  satisfies the assumptions mentioned earlier and the endemic equilibrium is locally asymptotically stable. Under the assumption that the life expectation is much longer than the length of the disease periods, one obtains the additional information that perturbations away from the endemic equilibrium die out in damped oscillations the quasi-periods (in this connection called *inter-epidemic periods*) of which are approximately given by

$$\frac{2\pi}{\sqrt{A\bar{D}}} \quad \text{with} \quad \bar{D} = E_m + \sum_{j=1}^{m-1} D_j,$$

where  $D_j$  is the average length of the  $j^{\text{th}}$  period (which is an exposed period) and  $E_m$  the average expectation of remaining duration of the infectious period. In the past, this analysis was only done for infection periods with exponential distribution which, since  $E_m = D_m$  in this case, had led to the suggestion that  $\bar{D}$  is sum of the average lengths of all stages before and including the infectious stage (Dietz, 1976, Anderson, May, 1982), though Dietz cautioned that this formula may have to be modified if other distribution are considered. Indeed, the use of general distributions shows that for the infectious period the average expectation of remaining sojourn is the relevant parameter. Bailey (1976) found for measles that assuming a fixed length of the infectious period leads to good fits to data, while the studies of Sartwell (1950, 1966) have shown that log-normal distributions with quite small dispersions approximate the length distributions of the incubation periods of many diseases well and one may assume that the same holds for the infectious period. In both scenarios,  $E_m$  should be much closer to  $D_m/2$  than  $D_m$ .

**A model with an isolated class.** Let us consider a model of  $S \rightarrow E \rightarrow I \rightarrow Q \rightarrow R$  type, i.e., we have a latent period with exposed individuals ( $E$ ), an infectious period, a quarantine (isolation) period, and a recovery period with permanent immunity. A certain kind of isolation occurs automatically, when children with symptoms stay at home and do not frequent the most important epidemic scene for childhood diseases, school. Again we assume that nobody dies from the disease and that the life expectation is much larger than the mean durations of the latent, infectious and quarantine period. Since the quarantined individuals do not take part in the mixing of the population, the standard incidence is modified as

$$f(S, E, I, Q, R) = \kappa \frac{SI}{N - Q} = \kappa \frac{SI}{S + E + I + R}.$$

In this model the endemic equilibrium becomes unstable, if the average length of the quarantine period,  $D_Q$ , exceeds a threshold length  $D_Q^*$ . It may regain stability for very long quarantine periods (*Feng, Thieme, 1995*). The approximate value for the threshold length is (*Feng, Thieme, 2000b*)

$$\begin{aligned} D_Q^* &\approx \frac{\mathcal{R}_0(\mathcal{R}_0 + 1)}{2(\mathcal{R}_0 - 1)}(E_I + D_L) + \frac{\mathcal{R}_0}{2} \frac{W_I + V_L}{E_I + D_L} \\ &= \frac{\mathcal{R}_0}{2}(D_L + E_I) \left[ \frac{\mathcal{R}_0 + 1}{\mathcal{R}_0 - 1} + \frac{V_L + W_L}{(D_L + E_I)^2} \right]. \end{aligned}$$

Here  $\mathcal{R}_0$  is the basic reproduction number of the disease, and  $D_L$  and  $D_I$  are the average lengths of the latent and infectious periods respectively and  $E_I$  the average expectation of remaining duration of the infectious period.  $V_L$  is the variance of the length of the latent period, and  $W_I$  is some kind of second central moment about  $E_I$ ,

$$W_I = \frac{1}{D_I} \int_0^\infty (a - E_I)^2 P_I(a) da,$$

with  $P_I$  being the duration function of the infectious period. The estimate

$$D_Q^* \geq \frac{\mathcal{R}_0(\mathcal{R}_0 + 1)}{2(\mathcal{R}_0 - 1)} \left( D_L + \frac{D_I}{2} \right)$$

shows for most childhood diseases (with the possible exemption of scarlet fever) that the endemic equilibrium in our model is locally asymptotically stable in realistic parameter ranges (see *Anderson, May, 1991*, Table 3.1 and Table 4.1) such that isolation alone presumably is not the cause of undamped oscillations. Here we would like to make the point that, while the fixed average lengths  $D_L$  and  $D_I$ , are kept fixed, different length distributions give different values for  $D_Q^*$ . This is not of practical consequence for the application of our model to childhood diseases, but could easily be for other models.

One calculates that

$$W_I = \frac{1}{3} D_I^2 - (D_I - E_I)^2 + \frac{C_I^3}{D_I} = D_I^2 \left( \frac{1}{12} + \gamma_0^2 \left( \frac{1}{2} - \gamma_0^2 \right) + \gamma_1 \gamma_0^3 \right),$$

where  $C_I^3$  is the third central moment of the length distribution of the infectious period and  $\gamma_0$  and  $\gamma_1$  are the coefficients of variation and skewness for the infectious period. For the second equality we have used (1.9) or rather one of the subsequent formulas. This shows that, for infectious periods whose lengths are positively skewed and which satisfy  $E_I \leq \frac{3}{2}D_I$  (remember that  $E_I \geq \frac{1}{2}D_I$  is always satisfied), or  $\gamma_0 \leq \sqrt{1/2}$ , we have

$$W_I \geq \frac{1}{12} D_I^2,$$

which is the value for the special case of fixed length, as we will see below. We want to compare the formula for  $D_Q^\sharp$  for the cases that the latent and infectious periods are both exponentially distributed, both have fixed durations, and both are log-normally distributed.

*Exponential distributions.* Most deterministic epidemic models assume that the infection stages are exponentially distributed. Then  $E_I = D_I$ ,  $V_L = D_L^2$  and  $W_I = D_I^2$ , and

$$D_Q^\sharp \approx \frac{\mathcal{R}_0}{2}(D_L + D_I) \left[ \frac{\mathcal{R}_0 + 1}{\mathcal{R}_0 - 1} + \frac{D_L^2 + D_I^2}{(D_L + D_I)^2} \right].$$

*Fixed lengths.* Now  $E_I = D_I/2$ ,  $V_L = 0$ , and

$$W_I = \frac{1}{D_I} \int_0^{D_I} \left( a - \frac{D_I}{2} \right)^2 da = \frac{2}{D_I} \int_0^{D_I/2} a^2 da = \frac{D_I^2}{12}.$$

So

$$D_Q^\sharp \approx \frac{\mathcal{R}_0}{2} \left( D_L + \frac{D_I}{2} \right) \left[ \frac{\mathcal{R}_0 + 1}{\mathcal{R}_0 - 1} + \frac{1}{12} \frac{D_I^2}{(D_L + \frac{D_I}{2})^2} \right].$$

It follows from our consideration for  $W_I$  above, that this is the smallest value for  $D_Q^\sharp$  under the condition that the length distribution of the infectious period is positively skewed and  $E_I \leq \frac{3}{2}D_I$  which we conjecture is true for any infectious disease fitting our assumptions.

*Log-normal distributions.* From (1.15) we derive

$$\begin{aligned} W_I &= \frac{1}{D_I} \int_0^\infty a^2 P_I da - E_I^2 = \frac{1}{D_I} \frac{1}{3} m^3 e^{(9/2)\sigma_I^2} - E_I^2 \\ &= \frac{1}{3} m^2 e^{(8/2)\sigma_I^2} - E_I^2 = E_I^2 \left( \frac{4}{3} e^{\sigma_I^2} - 1 \right) \end{aligned}$$

and we recall

$$V_L = D_L^2 (e^{\sigma_L^2} - 1).$$

Recall that  $e^{\sigma_L}$  and  $e^{\sigma_I}$  are the dispersion factors of the log-normal length distributions for the latent and the infectious period (Section 1.7). Substituting this into the formula for  $D_Q^\sharp$ ,

$$D_Q^\sharp \approx \frac{\mathcal{R}_0}{2} (D_L + E_I) \left[ \frac{\mathcal{R}_0 + 1}{\mathcal{R}_0 - 1} + \frac{D_L^2 (e^{\sigma_I^2} - 1) + E_I^2 \left( \frac{4}{3} e^{\sigma_I^2} - 1 \right)}{(D_L + E_I)^2} \right].$$

Also recall  $2E = De^{\sigma^2}$ . For  $e^{\sigma^2} = 2$ , the log-normal distribution shares the property of exponential distributions that  $E = D = V^{1/2}$  and we obtain

an almost identical formula for  $D_Q^{\frac{1}{2}}$ , though the values are slightly larger which seems to be related to the fact that the log-normal distribution has a larger positive skewness in this case. The findings of Sartwell (1950, 1966) which we discussed at the end of Section 1.7 suggest that for most childhood diseases  $e^{\sigma^2}$  is considerably closer to 1 than to 2 which means that we obtain similar values for  $D_Q^{\frac{1}{2}}$  as for fixed lengths.

### Bibliographic remarks.

**Chapter 1.** *Hutchinson* (1978), Chapter 2, displays many figures of survivorships. Most of them are on semilogarithmic paper such that it is difficult to realize the concavity and convexity properties of  $\mathcal{F}$ . An exception are survivorships for the waterflea *Daphnia pulex* at various population densities (Fig. 36, p. 57). There  $\mathcal{F}(a)$  switches back and forth between concavity and convexity several times. *Impagliazzo* (1985) shows the survivorship of Danish females, 1967, (Fig. 1.5, p. 23) and the associated per capita mortality rates (Fig. 1.12, p. 32). *Keyfitz, Flieger* (1971; Fig. 10, p. 43) compare male survivorships and the associated per capita mortality rates from the USA and from England and Wales, 1966-1968. These survival functions are strictly concave except for very small and large ages.

*Sartwell* (1950, Fig. 1) shows discrete versions of  $-\mathcal{F}'$  for the sojourn functions of the incubation periods of food-borne streptococcal sore throat and serum hepatitis.

Examples for  $D$ , in the meaning of life expectation, can be found in *Keyfitz, Flieger* (1971), Table D, for the USA and Europe, and Table E, for five European countries and their principal cities, separately for men and women. Figures displaying  $D(a)$ , in the meaning of expected remaining life at age  $a$ , for various countries at various times, can be found in *Hutchinson* (1978), Fig. 27, p. 45, *Impagliazzo* (1985), Fig. 1.3, p. 11, and Fig. 1.8, p. 24, and *Smith, Keyfitz* (1977), p. 5. *Hutchinson* (1987), Fig. 32, p. 52, also shows the associated mortality rates. In some cases,  $D(a)$  is a decreasing function of  $a$ , but in other cases  $D(a)$  first increases and then decreases. Figure 9 in *Keyfitz, Flieger* (1971) shows how  $D(a)$  has changed in the USA from 1920 to 1967 for  $a = 0, 50, 70$ , separately for men and women.

The log-normal distribution is also used in describing how populations partition resources (*Roughgarden*, 1979, Chap. 24) and to model insurance claim severity or investment returns (*Hassett, Stewart*, 1999, Chap. 4). *Hassett, Stewart* (1999) presents several probability distribution functions that can be adapted to serve as sojourn functions.

In this paper, we do not model stages the lengths of which do not only vary from individual to individual, but depend on environmental factors, like temperature for ectotherms, or on nutrients for individuals which start reproducing when reaching a certain minimum body size (*Daphnia magna*). See *Gurney et al.* (1986) and *Schumacher, Thieme* (1988) and the references therein.

**Chapter 2.** In describing the stage dynamics we have preferred the older *Lotka* (1907) approach (formula (2.2)) over the somewhat more recent *McKendrick* (1926) approach (the partial differential equation (2.4)) because it appears to be more elementary and is more general, at least for the time-autonomous case. While the partial differential equation is readily extended to the non-autonomous case, this extension is somewhat more clumsy for the Lotka approach. Equation (2.4) may in particular be more communicable when one wants to incorporate nonlinear effects. Still the Lotka approach has definite advantages even here, in particular if populations are not structured with respect to age, but to body size or other characteristics. See *Diekmann, Gyllenberg, Metz, Thieme* (1998). If life is the stage under consideration, model (2.2) can be extended to a demographic model by coupling the stage input, which is now the birth rate, to the stage density via per capita birth rates. See *Frauenthal* (1986), *Hoppensteadt* (1974, 1975), *Iannelli* (1995), *Impagliazzo* (1985), and *Webb* (1985).

**Chapter 3.** There is vast literature on epidemic models for populations with chronological age-structure (see *Anderson, May*, 1991); *Andreasen*, 1995; *Castillo-Chavez et al.*, 1994; *Diekmann et al.*, 1995; *Iannelli*, 1995; and *Thieme*, 1991; for references), with a natural interest in optimal vaccination strategies (e.g., *Greenhalgh, Dietz*, 1994; *Müller*, 1998).

*Hoppensteadt* (1974, 1975) has introduced a model framework that incorporates almost arbitrary length distributions of various disease stages (the distributions must have a density) and seems to have coined the notion of class age (or stage age, as we say).

Quite a few models have been considered where one infection period is arbitrarily distributed, typically the infectious or the immune period:

*Stech and Williams* (1981) show a remarkable global stability result for the endemic equilibrium in a model with an arbitrarily distributed immunity period, their result was recently extended to diseases that cause fatalities (*Thieme, van den Driessche*, 1999). *Lin and van den Driessche* (1992) prove threshold results for models with an arbitrarily distributed immunity period and a nonlinear incidence. *Castillo-Chavez et al.* (1989) study the global stability of the disease-free and the local stability of the endemic equilibrium for an AIDS model with arbitrarily distributed infectivity period and a general contact function. *Brauer* (1990, 1991, 1996) analyzes the local stability of the endemic equilibrium in models with arbitrarily distributed infectious period incorporating varying population size, disease fatalities, and vertical transmission.

There seem to be very few papers that analyze models where at least two infection stages are arbitrarily distributed. *Hethcote and Tudor* (1980) first consider an SIR model with an arbitrarily distributed infectious period for which they establish stability results for the disease-free and the endemic equilibrium. Then they present an SEIR model with arbitrarily distributed latent and infectious periods for which they prove the global

stability of the disease-free equilibrium in the subthreshold case and show that the characteristic equation associated with the endemic equilibrium has roots with strictly negative real part only (which implies local asymptotic stability if one adds some dynamical system theory which was not available at that time). Both models include vaccination. Hethcote *et al.* (1981) formulate an SEIS model with arbitrarily distributed latent and infectious periods and show global stability of the disease-free equilibrium in the subthreshold case and local or global asymptotic stability of the endemic equilibrium for fixed stage durations or for mixtures of one arbitrary and one exponential stage distribution.

If there is not a constant influx of susceptibles or if there are disease fatalities, mass action and standard incidence may lead to different dynamics. They can also make a difference in the effect of vaccination (*Gao et al.*, 1995, 1996, and the references therein). These two laws are interpolated by saturating contact functions (see, e.g., *Castillo-Chavez et al.*, 1989; *Greenhalgh, Das*, 1995; *Thieme*, 1992).

*Kermack and McKendrick* (1927, 1932, 1933) have provided an alternative, in a sense more general, mathematical framework for the analysis of infectious diseases than many and arbitrarily distributed infection stages; they consider just one infection stage, but allow the infectivity of an infected individual to depend on its age of infection (i.e. the time that has elapsed since the moment of infection). While the Kermack-McKendrick model allows for powerful mathematics leading to important theoretical insight (see *Diekmann et al.*, 1995, for a survey; see also *Hethcote, Thieme*, 1985; *Thieme, Castillo-Chavez*, 1993), it connects to data that are more difficult to collect than data concerning the lengths of the various periods mentioned above (though this is already cumbersome enough).

Besides quarantine, there are many mechanisms that can destabilize the endemic equilibrium, at least in theory, while it is often difficult to check whether instability occurs in a realistic parameter range. For reviews and references see *Dietz, Schenzle* (1985), *Hethcote, Levin* (1989), *Feng, Thieme* (1995), *Gao et al.* (1995, 1996).

**Acknowledgement.** I thank Herb Hethcote for reading the course notes that have been the basis for the first two chapters of this article and for his many helpful suggestions and for additional references. I thank Maia Martcheva for useful remarks.

## APPENDIX

**A. Stieltjes integration.** Let  $g : [0, \infty) \rightarrow \mathbf{R}$  be a non-decreasing function. Then we denote by  $m_g$  the Borel measure on  $[0, \infty)$  which is uniquely determined by

$$m_g([0, a]) = m_g([0, a]) = g(a) - g(0)$$

whenever  $a \geq 0$  and  $g$  is continuous at  $a$ .

If  $g$  is non-increasing, we consider  $m_g = m_{-g}$ .

See, e.g., McDonald, Weiss (1999), Sec. 4.7. Notice that our construction is different from Gripenberg et al. (1990), Section 3.7. We deviate in order to have the following result.

**Lemma A.1.** *Let  $u : [0, a] \rightarrow \mathbf{R}$  be continuous and  $g$  as above and continuous at  $a$ . Then*

$$\int_{[0, a)} u(s)m_g(ds) = \int_0^a u(s)dg(s)$$

where the second integral is the Riemann-Stieltjes integral.

*Proof.* It is sufficient to prove the assertion for  $u$  that is continuously differentiable on  $[0, a]$ , because continuous  $u$  can be uniformly approximated by such functions.

$$\begin{aligned} \int_{[0, a)} u(s)m_g(ds) &= \int_{[0, a)} \left( u(0) + \int_0^s u'(r)dr \right) m_g(ds) \\ &= u(0)m_g([0, a)) + \int_{[0, a)} u'(r) \left( \int_{[r, a)} m_g(ds) \right) dr \\ &= u(0)(g(a) - g(0)) + \int_0^a u'(r)(g(a) - g(r))dr \\ &= \int_0^a u(r)dg(r). \end{aligned}$$

The last equality follows from the integration by parts formula for Riemann-Stieltjes integrals.

Lemma A.1 gives conditions under which the Riemann-Stieltjes integral equals the corresponding Lebesgue Stieltjes integral (the integral on the left hand side of the equation). In the following, the Stieltjes integral is always to be interpreted as Lebesgue Stieltjes integral and we write

$$\int_{[a, b)} u(s)m_g(ds) = \int_a^b u(s)g(ds).$$

□

**Lemma A.2.** *Let  $u$  be integrable on  $[0, \tau]$  and  $P$  non-increasing. Then the convolution*

$$(u * P)(r) = \int_0^r u(r-s)P(s)ds$$

is absolutely continuous and, outside a set of Lebesgue measure 0,

$$(u * P)'(r) = u(r)P(0) + \int_0^r u(r-s)P(ds).$$

Recall that

$$-\int_0^r u(r-s)P(ds) = \int_{[0,r)} u(r-s)m(ds)$$

where  $m([0,s)) = P(0) - P(s)$  at all points  $s$  where  $P$  is continuous.

*Proof.* Changing the order of integration (*McDonald, Weiss, 1999, Sec. 4.9*),

$$\begin{aligned} \int_0^t \left( \int_{[0,r)} u(r-s)m(ds) \right) dr &= \int_{[0,t)} \left( \int_s^t u(r-s)dr \right) m(ds) \\ &= \int_{[0,t)} \left( \int_0^{t-s} u(r)dr \right) m(ds) = \int_0^t u(r) \left( \int_{[0,t-r)} m(ds) \right) dt \\ &= \int_0^t u(r)(P(0) - P(t-r))dt = P(0) \int_0^t u(r)dr - (u * P)(t). \end{aligned}$$

□

**Lemma A.3.** Let  $u \in L_+^1[0, \infty)$  and  $P : [0, \infty) \rightarrow [0, \infty)$  non-increasing,  $P(a) \rightarrow 0$  as  $a \rightarrow \infty$ , then the function  $w$ ,

$$w(t) = \int_t^\infty u(s-t)P(s)ds,$$

is absolutely continuous and, outside a set of Lebesgue measure 0,

$$w'(t) = \int_{[t,\infty)} u(s-t)P(ds).$$

In particular  $w$  is non-increasing and

$$w(t) \leq w(0) = \int_0^\infty u(s)P(s)ds, \quad \int_0^\infty |w'(t)|dt \leq w(0).$$

*Proof.* Changing the order of integration,

$$\begin{aligned} \int_r^\infty \left( \int_{[t,\infty)} u(s-t)m(ds) \right) dt &= \int_{[r,\infty)} \left( \int_r^s u(s-t)dt \right) m(ds) \\ &= \int_{[r,\infty)} \left( \int_0^{s-r} u(t)dt \right) m(ds) = \int_0^\infty u(t) \left( \int_{[t+r,\infty)} m(ds) \right) dt \\ &= \int_0^\infty u(t)P(r+t)dt = \int_r^\infty u(t-r)P(t)dt. \end{aligned}$$

□

The next results follow from Lemma A.2 and A.3 and Theorem (21.67) in *Hewitt, Stromberg (1969)*.

**Proposition A.4.** Let  $u \in L_+^1[0, \infty)$  and  $P, Q : [0, \infty) \rightarrow [0, \infty)$  non-increasing,  $P(a)Q(a) \rightarrow 0$  as  $a \rightarrow \infty$ .

a) Then  $v$ , defined by

$$v(t) = \int_0^t u(t-s)Q(s)P(s)ds,$$

is absolutely continuous and, outside a set of Lebesgue measure 0,

$$v'(t) = u(t)P(0)Q(0) + \int_0^t u(t-s)Q(s-)P(ds) + \int_0^t u(t-s)P(s+)Q(ds)$$

with  $P(0-) = P(0)$ .

b) Further  $w$ , defined by

$$w(t) = \int_t^\infty u(s-t)P(s)Q(s)ds$$

is absolutely continuous and, outside a set of Lebesgue measure 0,

$$w'(t) = \int_t^\infty u(s-t)[P(s+)Q(ds) + Q(s-)P(ds)].$$

*Proof.* Define

$$P_0(a) = P(a)Q(a), \quad P_1(a) = \int_{[0,a]} P(s+)Q(ds), \quad P_2(a) = \int_{[0,a]} Q(s-)P(ds).$$

By Theorem (21.67) in *Hewitt, Stromberg* (1969),

$$P_0(a+) - P_0(0) = P_1(a) + P_2(a).$$

This implies that  $m_0 = m_1 + m_2$  where  $m_j$  are the Lebesgue Stieltjes measures associated with  $P_j$ . The assertion now follows from Lemma A.2 and A.3.  $\square$

**Corollary A.5.** Let  $u \in L_+^1[0, \infty)$  and  $P, Q : [0, \infty) \rightarrow [0, \infty)$  non-increasing,  $P(a)Q(a) \rightarrow 0$  as  $a \rightarrow \infty$ . Assume that  $P$  and  $Q$  have not points of discontinuity in common.

a) Then  $v$ , defined by

$$v(t) = \int_0^t u(t-s)Q(s)P(s)ds,$$

is absolutely continuous and, outside a set of Lebesgue measure 0,

$$v'(t) = u(t)P(0)Q(0) + \int_0^t u(t-s)Q(s)P(ds) + \int_0^t u(t-s)P(s)Q(ds).$$

b) Further  $w$ , defined by

$$w(t) = \int_t^\infty u(s-t)P(s)Q(s)ds$$

is absolutely continuous and, outside a set of Lebesgue measure 0,

$$w'(t) = \int_t^\infty u(s-t)[P(s)Q(ds) + Q(s)P(ds)].$$

*Proof.*  $P$  and  $Q$  are discontinuous at countably many points only. Since  $P$  and  $Q$  have no points of discontinuity in common, the points of discontinuity of  $P$  form a set of measure 0 for the Lebesgue Stieltjes measure associated with  $Q$  and vice versa. The assertion now follows from Proposition A.4.  $\square$

**Corollary A.6.** Let  $u \in L_+^1[0, \infty)$  and  $P, \mathcal{F} : [0, \infty) \rightarrow [0, \infty)$  non-increasing,  $P(a)Q(a) \rightarrow 0$  as  $a \rightarrow \infty$ ,  $\mu > 0$ . Assume that  $P$  and  $Q$  have not points of discontinuity in common.

Then  $v$ , defined by

$$v(t) = \int_0^t u(t-s)e^{-\mu s}\mathcal{F}(s)P(s)ds$$

is absolutely continuous and, outside a set of Lebesgue measure 0,

$$\begin{aligned} v'(t) &= u(t)P(0)\mathcal{F}(0) - \mu v(t) + \int_0^t u(t-s)e^{-\mu s}[\mathcal{F}(s)P(ds) + P(s)\mathcal{F}(ds)] \\ &\leq u(t)P(0)\mathcal{F}(0) - \mu v(t) + \int_0^t u(t-s)e^{-\mu s}\mathcal{F}(s)P(ds). \end{aligned}$$

b) Further  $w$ , defined by

$$w(t) = \int_t^\infty u(s-t)e^{-\mu t}P(s)\mathcal{F}(s)ds$$

is absolutely continuous and, outside a set of Lebesgue measure 0,

$$\begin{aligned} w'(t) &= -\mu w(t) + e^{-\mu t} \int_t^\infty u(s-t)[\mathcal{F}(s)P(ds) + P(s)\mathcal{F}(ds)] \\ &\leq -\mu w(t) + e^{-\mu t} \int_t^\infty u(s-t)\mathcal{F}(s)P(ds). \end{aligned}$$

## REFERENCES

- [1] ANDERSON, R.M. AND MAY, R.M. (1982): Directly transmitted infectious diseases: Control by vaccination. *Science* **215**, 1053–1060.
- [2] ANDERSON, R.M. AND MAY, R.M. (1991): *Infectious Diseases of Humans*. Oxford University Press.
- [3] ANDREASEN, V. (1995): Instability in an SIR-model with age-dependent susceptibility. *Mathematical Population Dynamics: Analysis of Heterogeneity*. Volume One: Theory of Epidemics (O. Arino, D. Axelrod, M. Kimmel, M. Langlais; eds.), 3–14. Wuerz.

- [4] BAILEY, N.T.J. (1975): *The Mathematical Theory of Infectious Diseases and its Applications*. Griffin.
- [5] BECKER, N.G. (1989): *Analysis of Infectious Disease Data*. Chapman and Hall.
- [6] BRAUER, F. (1990): Models for the spread of universally fatal diseases. *J. Math. Biol.* **28**, 451–462.
- [7] BRAUER, F. (1992): Models for the spread of universally fatal diseases, II. *Differential Equation Models in Biology, Epidemiology, and Ecology* (S. Busenberg and M. Martelli, eds.), 57–69. Lecture Notes in Biomathematics **92**, Springer.
- [8] BRAUER, F. (1996): A characteristic equation arising in models for diseases with vertical transmission and without immunity. *Differential Equations and Applications to Biology and Industry* (M. Martelli, K. Cooke, E. Cumberbatch, H. Thieme; eds.), 41–48. World Scientific.
- [9] BUSENBERG, S.; COOKE, K.L.; AND IANNELLI, M. (1988): Endemic thresholds and stability in a class of age-structured epidemics. *SIAM J. Appl. Math.* **48**, 1379–1395.
- [10] CASTILLO-CHAVEZ, C.; COOKE, K.L.; HUANG, W.; AND LEVIN, S.A. (1989): On the role of long incubation periods in the dynamics of acquired immunodeficiency syndrome (AIDS). Part 1: Single population models. *J. Math. Biol.* **27**, 373–398.
- [11] CASTILLO-CHAVEZ, C.; VELASCO-HERNANDEZ, J.X.; AND FRIDMAN, S. (1994): Modeling contact structures in biology. *Frontiers in Mathematical Biology* (S.A. Levin, ed.), 454–491. Lecture Notes in Biomathematics **100**. Springer.
- [12] COOKE, K.L. (1967): Functional differential equations: some models and perturbation problems. *Differential equations and Dynamical Systems* (J.K. Hale, J.P. LaSalle; eds.), 167–183. Academic Press.
- [13] DIEKMANN, O.; GYLLENBERG, M.; METZ, J.A.J.; AND THIEME, H.R. (1998): On the formulations and analysis of general deterministic structured population models. I. Linear Theory. *J. Math. Biol.* **36**, 349–388.
- [14] DIEKMANN, O.; HEESTERBEEK, J.A.P.; AND METZ, J.A.J. (1990): On the definition and the computation of the basic reproduction rate  $R_0$  in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**, 365–382.
- [15] DIEKMANN, O.; HEESTERBEEK, H.; AND METZ, J.A.J. (1995): The legacy of Kermack and McKendrick. *Epidemic Models: Their Structure and Relation to Data* (D. Mollison, ed.), 95–115. Cambridge Univ. Press.
- [16] DIETZ, K. (1976): The incidence of infectious diseases under the influence of seasonal fluctuations. *Mathematical Models in Medicine* (Berger, J.; Bühler, W.; Repges, R.; Tautu, P; eds.), 1–15. Lecture Notes in Biomathematics **11**. Springer.
- [17] DIETZ, K. AND SCHENZLE, D. (1985): Mathematical models for infectious disease statistics. *A Celebration of Statistics* (A.C. Atkinson, S.E. Fienberg; eds.), 167–204. Springer.
- [18] FENG, Z. (dissertation): *A mathematical model for the dynamics of childhood diseases under the impact of isolation*. Ph.D Thesis, Arizona State University, 1994.
- [19] FENG, Z. AND THIEME, H.R. (1995): Recurrent outbreaks of childhood diseases revisited: the impact of isolation. *Math. Biosci.* **128**, 93–130.
- [20] FENG, Z. AND THIEME, H.R. (2000a): Endemic models with arbitrarily distributed periods of infection I. General theory. *SIAM J. Appl. Math.* **61**, 803–833.
- [21] FENG, Z. AND THIEME, H.R. (2000b): Endemic models with arbitrarily distributed periods of infection II. Fast disease dynamics and permanent recovery. *SIAM J. Appl. Math.* **61**, 903–1012.
- [22] FRAVENTHAL, J.C. (1986): Analysis of age-structure models. *Mathematical Ecology* (T.G. Hallam, S.A. Levin; eds.), 117–147.
- [23] GAO, L.Q.; MENA-LORCA, J.; AND HETHCOTE, H.W. (1995): Four SEI endemic models with periodicity and separatrices. *Math. Biosci.* **128**, 157–184.

- [24] GAO, L.Q.; MENA-LORCA, J.; AND HETHCOTE, H.W. (1996): Variations on a theme of SEI endemic models. *Differential Equations and Applications to Biology and Industry* (M. Martelli, K.L. Cooke, E. Cumberbatch, B. Tang, H.R. Thieme, eds.), 191–207. World Scientific.
- [25] GOODALL, E.W. (1931): Incubation period of measles. (Letter to editor.) *Brit. Med. J.* **1**, 73–74.
- [26] GOUGH, K.J. (1977): The estimation of latent and infectious periods. *Biometrika* **64**, 559–565.
- [27] GREENHALGH, D. AND DAS, R. (1995): Modelling epidemics with variable contact rates. *Theor. Pop. Biol.* **47**, 129–179.
- [28] GREENHALGH, D. AND DIETZ, K. (1994): Some bounds on estimates for reproductive ratios from the age-specific force of infection. *Math. Biosci.* **124**, 9–57.
- [29] GRIPENBERG, G.; LONDON, S.-O.; AND STAFFANS, O. (1990): *Volterra Integral and Functional Equations*. Cambridge Univ. Press.
- [30] GURNEY, W.S.C.; NISBET, R.M.; AND BLYTHE, S.P. (1986): The systematic formulation of models of stage-structured populations. *The Dynamics of Physiologically Structured Populations* (J.A.J. Metz, O. Diekmann; eds.), 474–494. Lecture Notes in Biomathematics **68**, Springer.
- [31] HASSETT, M.J. AND STEWART, D.G. (1999): *Probability for Risk Management*. Actex Publications.
- [32] HETHCOTE, H.W. AND VAN ARK, J.W. (1992): *Modeling HIV Transmission and AIDS in the United States*. Lecture Notes in Biomathematics **95**. Springer.
- [33] HETHCOTE, H.W. AND LEVIN, S.A. (1989): Periodicity in epidemiological models. *Applied Mathematical Ecology* (S.A. Levin, T.G. Hallam, L.J. Gross; eds.), 193–211. Springer.
- [34] HETHCOTE, H.W.; STECH, H.W.; AND VAN DEN DRIESSCHE, P. (1981): Stability analysis for models of diseases without immunity. *J. Math. Biol.* **13**, 185–198.
- [35] HETHCOTE, H.W. AND THIEME, H.R. (1985): Stability of the endemic equilibrium in epidemic models with subpopulations. *Math. Biosci.* **75**, 205–227.
- [36] HETHCOTE, H.W. AND TUDOR, D.W. (1980): Integral equation models for endemic infectious diseases. *J. Math. Biol.* **9**, 37–47.
- [37] HEWITT, E. AND STROMBERG, K. (1969): *Real and Abstract Analysis*. Springer.
- [38] HOPPENSTEADT, F. (1974): An age dependent epidemic problem. *J. Franklin Inst.* **297**, 325–333.
- [39] HOPPENSTEADT, F. (1975): *Mathematical Theories of Populations: Demographics, Genetics, and Epidemics*. Regional Conference Series in Applied Mathematics **15**. SIAM.
- [40] HOPPENSTEADT, F. AND WALTMAN, P. (1970): A problem in the theory of epidemics. *Math. Biosci.* **9**, 71–91.
- [41] HOPPENSTEADT, F. AND WALTMAN, P. (1971): A problem in the theory of epidemics, II. *Math. Biosci.* **12**, 133–145.
- [42] HUTCHINSON, G.E. (1978): *An Introduction to Population Ecology*. Yale University Press.
- [43] IANNELLI, M. (1995): *Mathematical Theory of Age-Structured Population Dynamics*. Giardini Editori e Stampatori.
- [44] IMPAGLIAZZO, J. (1985): *Deterministic Aspects of Mathematical Demography*. Springer.
- [45] JACQUEZ, J.A. (1978): *Compartmental Analysis in Biology and Medicine*, 2<sup>nd</sup> ed. University of Michigan Press.
- [46] KEELING, M.J. AND GRENfell, B.T. (1997): Disease extinction and community size: Modeling the persistence of measles. *Science* **275**, 65–67.
- [47] KEELING, M.J. AND GRENfell, B.T. (1998): Effect of variability in infection period on the persistence and spatial spread of infectious diseases. *Math. Biosci.* **147**, 207–226.
- [48] KERMACK, W.O. AND MCKENDRICK, A.G. (1927): A contribution to the mathematical theory of epidemics. *Proc. Roy. Soc. A* **115**, 700–721.

- [49] KERMACK, W.O. AND MCKENDRICK, A.G. (1932): Contributions to the mathematical theory of epidemics. II. The problem of endemicity. *Proc. Roy. Soc. A* **138**, 55–83.
- [50] KERMACK, W.O. AND MCKENDRICK, A.G. (1933): Contributions to the mathematical theory of epidemics. III. Further studies of the problem of endemicity. *Proc. Roy. Soc. A* **141**, 94–122.
- [51] KEYFITZ, N. AND FLIEGER, W. (1971): *Population. Facts and Methods of Demography*. Freeman and Co.
- [52] KIM, Y.J. AND ARON, J.L. (1989): On the equality of average age and average expectation of remaining life in a stationary population. *Siam Review* **31**, 110–113.
- [53] LIN, X. AND VAN DEN DRIESSCHE, P. (1992): A threshold result for an epidemic model. *J. Math. Biol.* **30**, 647–654.
- [54] LOTKA, A.L. (1907): Relation between birth rates and death rates. *Science, N.S.* **26**, 21–22. Reprinted in *Mathematical Demography* (D. Smith, N. Keyfitz; eds.), 93–95. Springer 1977.
- [55] McDONALD, J.N. AND WEISS, N.A. (1999): *A Course in Real Analysis*. Academic Press.
- [56] MCKENDRICK, A.G. (1926): Applications of mathematics to medical problems. *Proc. Edinb. Math. Soc.* **44**, 98–130.
- [57] MÜLLER, J. (1998): Optimal vaccination patterns in age-structured populations. *SIAM J. Appl. Math.* **59**, 222–241.
- [58] Patel, J.K.; Kapadia, C.H.; Owen, D.B. (1976): *Handbook of Statistical Distributions*. Dekker.
- [59] ROUGHGARDEN, J. (1979): *Theory of Population Genetics and Evolutionary Ecology: An Introduction*. Macmillan.
- [60] SARTWELL, P.E. (1950): The distribution of incubation periods of infectious diseases. *Am. J. Hyg.* **51**, 310–318.
- [61] SARTWELL, P.E. (1966): The incubation period and the dynamics of infectious disease. *Am. J. Epid.* **83**, 204–318.
- [62] SCHUMACHER, K. AND THIEME, H.R. (1988): Some theoretical and numerical aspects of modelling dispersion in the development of ectotherms. *Comput. Math. Appl.* **15**, 565–594.
- [63] SIMON, C.P. AND JACQUEZ, J.A. (1992): Reproduction numbers and the stability of equilibria of SI models for heterogeneous populations. *SIAM J. Appl. Math.* **52**, 541–576.
- [64] SMITH, D. AND KEYFITZ, N. (1977): *Mathematical Demography*. Springer.
- [65] STECH, H.S. AND WILLIAMS, M. (1981): Stability in a class of cyclic epidemic models with delay. *J. Math. Biol.* **11**, 95–103.
- [66] STILLERMAN, M. AND THALHIMER, W. (1944): Attack rate and incubation period of measles. *Am. J. Dis. Child.* **67**, 15.
- [67] THIEME, H.R. (1991): Stability change of the endemic equilibrium in age-structured models for the spread of S-I-R type infectious diseases. *Differential Equations. Models in Biology, Epidemiology and Ecology* (S. Busenberg, M. Martelli, eds.), 139–158. Lecture Notes in Biomathematics **92**, Springer.
- [68] THIEME, H.R. (1992): Epidemic and demographic interaction in the spread of potentially fatal diseases in growing populations. *Math. Biosci.* **111**, 99–130.
- [69] THIEME, H.R. AND CASTILLO-CHAVEZ, C. (1993): How may infection-age dependent infectivity affect the dynamics of HIV/AIDS? *SIAM J. Appl. Math.* **53**, 1447–1479.
- [70] THIEME, H.R. AND VAN DEN DRIESSCHE, P. (1999): Global stability in cyclic epidemic models with disease fatalities. *Fields Institute Communications* **21**, 459–472.
- [71] VONFOERSTER, H. (1959): Some remarks on changing populations. *The Kinetics of Cellular Proliferation* (F. Stohlman, ed.), Grune and Stratton.
- [72] WEBB, G.F. (1985): *Theory of Nonlinear Age-Dependent Population Dynamics*. Dekker.

# MEASLES OUTBREAKS ARE NOT CHAOTIC

INGEMAR NÅSELL\*

**Abstract.** A fully stochastic model is used to study the phenomena of recurrence and extinction associated with childhood infections. This model is also used to investigate the properties of the corresponding deterministic model. We conclude that recurrent epidemics have such intricate behaviour that they cannot be studied with deterministic models; recognition of demographic stochasticity is necessary to understand them. The damping associated with the oscillations appearing in the deterministic model is shown to be a measure of the stochastic variability of the time between successive outbreaks. Furthermore, chaotic deterministic models for recurrent epidemics are shown to be unrealistic. The mechanism that drives chaos in these models is an unjustified mathematical approximation introduced in going from the stochastic to the deterministic formulation.

**Key words.** Chaos, stochastic model, recurrent epidemics, population biology model, demographic stochasticity.

**AMS(MOS) subject classifications.** 60J27, 92D40.

**1. Introduction.** Deterministic models are used on a broad scale for building models and shaping theories in population biology. This work has been highly successful. Early contributions by Lotka and Volterra are now classical. They exemplify the ability of the theory to deal with nonlinearities and bifurcation phenomena and to draw qualitative conclusions. The range of phenomena that can be analyzed in a deterministic setting was further widened after May (1976) announced his discovery of chaos in ecological models. Chaos is a fascinating mathematical phenomenon, where deterministic behaviour has a random character. It may appear that stochastic models have no role to play in this setting. However, the aim of the present paper is to show the opposite. We claim that an insight into the phenomena associated with childhood epidemics like measles can only be gained through formulation and analysis of fully stochastic models.

Stochastic ingredients in population biology models can take at least two different forms, demographic and environmental stochasticity, as discussed by May (1973). Populations come in discrete units, and are treated accordingly in a model that accounts for demographic stochasticity. In contrast, any deterministic model has the serious weakness that it deals with population sizes on a continuous scale. This is an approximation that may be acceptable for large population sizes, but that can lead to serious erroneous conclusions if the population size ever gets close to zero. Environmental stochasticity has a different character. It is based on the interpretation that parameters that appear in a deterministic model not are constants, but random variables. Our analysis in this paper of stochastic models is exclusively devoted to demographic stochasticity.

---

\*Department of Mathematics, The Royal Institute of Technology, S-100 44 Stockholm, Sweden. Email: [ingemar@math.kth.se](mailto:ingemar@math.kth.se).

The models that we are concerned with have a simple structure, with the deterministic models leading to a set of differential equations, while the corresponding stochastic models take the form of Markov chains with continuous time and discrete state space. There is an important conceptual relation between any such deterministic model and its stochastic counterpart, namely that the deterministic model can be derived as an approximation of the stochastic model under the assumption that the population size approaches infinity. For any finite population size  $N$  this leads us to consider the mathematical problem of determining if  $N$  is sufficiently large for the deterministic approximation to be acceptable. This may appear as a fine point if we are dealing with a really large population. We shall however see for the chaotic model of recurrent epidemics that even a population size of 1 million is much too small for the deterministic approximation of the stochastic one to be valid. Note that analysis of the stochastic model is required in order to draw this kind of conclusion concerning the deterministic model.

Stochastic models are mathematically more difficult to analyze than their deterministic counterparts. It is therefore common in the analysis of a stochastic model to use the properties of the corresponding deterministic model to acquire some insight into the behaviour of the stochastic model. In this paper we use the opposite approach. We use the stochastic model formulation to analyze the corresponding deterministic model.

Our main conclusion is that deterministic models of recurrent epidemics are unacceptable approximations of the corresponding stochastic models. This holds both for the phenomena of recurrence and extinction, and for the important threshold conditions.

**2. Recurrent epidemics.** Observations of the number of cases of childhood infections, such as measles, are available for a large number of cities from the time before vaccination was introduced. These data show two interesting properties. One is that there is a tendency for recurring epidemic outbreaks, and the other one is that the infection tends to disappear spontaneously in small populations. It is a challenge in mathematical epidemiology to understand the mechanisms that cause both the epidemic outbreaks and the spontaneous disappearance. The establishment of a model that allows one to study these two phenomena is of importance for theoretical epidemiology, as discussed by Dietz (1995).

The number of susceptible individuals falls after an epidemic outbreak to such a low level that a new outbreak cannot occur. It was claimed already by Hamer (1906) that an inflow of new susceptible individuals is required before an additional outbreak can take place. This means that the outbreaks are a result of the combined influence of epidemic and demographic forces. This basic idea has been used in all succeeding models.

The aim of this paper is to study basic qualitative properties of deterministic models (both chaotic and nonchaotic) for recurrent epidemics. We

consider therefore only the simplest model formulations, and avoid a discussion of more realistic models where additional assumptions concerning spatial distribution, heterogeneity, age dependence, isolation, and duration of infection are introduced. Such more complex models are discussed e.g. by Schenzle (1984), Grenfell, Bolker, and Kleczkowski (1995), Feng and Thieme (1995), and Keeling and Grenfell (1997).

The first deterministic model, based on Hamer's ideas, was introduced by Soper (1929). He used a so-called SIR model, with the population partitioned into the three categories Susceptible, Infected, and Recovered. Infected individuals are assumed infective, and recovered individuals are assumed permanently immune. Soper's model leads to a pair of differential equations for the number of susceptible individuals  $S$  and the number of infected individuals  $I$ . The solution of these equations shows damped oscillations for both these state variables. The fact that this tendency toward periodicity occurred without any assumed periodicity in the parameters was rightly regarded as a step forward. On the other hand, the damping did not agree with the behaviour observed in the data. We show in Subsection 4.2 below that this damping can be understood by viewing the deterministic model as an approximation of the stochastic one. A further weakness with the Hamer-Soper model is that it cannot explain the spontaneous disappearance of the infection. Indeed, this phenomenon is purely stochastic and cannot be studied with any deterministic model. Any phenomenon that depends on the population size requires stochastic modelling, as further emphasized in the discussion of the parameters of the model of Section 3.

A real breakthrough in this situation occurred when Bartlett (1956, 1957, 1960a-c) claimed that the stochastic version of Soper's model could account for both quasi-periodic outbreaks without damping and for spontaneous disappearance of the infection. The Bartlett model takes the form of a bivariate Markov chain with discrete state space and continuous time. It can be referred to as a bivariate birth-death process since the only allowed transitions in the Markov chain are to neighboring states. Bartlett used Monte Carlo simulation to show that his model allowed recurring outbreaks without damping. Furthermore, he derived an analytic approximation of the time to extinction. His results showed that the time to extinction was an increasing function of the population size. He used this to introduce the important concept of "critical community size", which is such that the infection will go extinct rather quickly if the population size is below the critical community size, while it will last for a long time if the population size is larger. The critical community size expresses a threshold behaviour for the stochastic model. It differs from the thresholds for deterministic models by involving the population size. A further discussion of thresholds for stochastic models is given by Nåsell (1995). A good description of the models by Soper and Bartlett is given by Bailey (1975). A detailed analysis of Bartlett's model is given by Nåsell (1999). This analysis shows that

Bartlett's approximation of the time to extinction fails in an important part of the parameter space, and it gives an insight into the reason for this failure. However, Bartlett's claim that stochastic modelling can account for both recurrence and extinction is strongly supported.

It is surprising that Bartlett's important papers remained the only ones devoted to stochastic analysis of recurrent epidemics for decades. The search for realistic models that could explain periodic or quasi-periodic behaviour continued, but it was limited to a search in the deterministic domain. Examples are given by Hethcote and Levin (1989). Chaos was discovered by Schaffer (1985) and Schaffer and Kot (1985) in a deterministic model quite similar to the model analyzed by Bartlett. This discovery attracted a lot of interest. One purpose of the present paper is to redirect the attention to Bartlett's claim that the phenomena associated with recurrent epidemics can be understood by stochastic modelling. A second purpose is, as already mentioned, to show that the chaotic models are unrealistic.

The model in which chaos appears differs from the deterministic version of Bartlett's model (i.e. from Soper's model) in three ways. The first difference concerns the treatment of the demographic mechanism. Soper's model is unrealistic in the sense that it allows an inflow of new susceptibles, but it does not allow for death of susceptible or infective individuals. One undesirable consequence of this is that the number of susceptible individuals grows exponentially with time if no infection is present. This unrealistic feature is readily remedied by assuming a constant death rate per individual. The deterministic model where this assumption is made has been studied by Martini (1921), Lotka (1923, 1956), Hethcote (1974, 1976), and Anderson and May (1991). The stochastic version of this model has been used to study time to extinction in recurrent epidemics by Nåsell (1999). The corresponding model for a host-vector situation has been studied by Dietz (1975).

The second difference is that the chaos model is an SEIR model, while the Soper model and the Martini model are SIR models. This means simply that a state E (for Exposed) is inserted between the susceptible and the infective states. An individual who is infected but not infective is called exposed.

The third difference is that the contact rate is assumed to be periodic with a period of one year. This is a simple way to reflect the important annual pattern of aggregation of children in schools. Models with this feature have been analyzed by London and Yorke (1973), Dietz (1976), Aron and Schwartz (1984), and Schenzle (1984).

Useful reviews of measles models are given by Hastings *et al.* (1993) and Isham (1993). Hastings *et al.* emphasizes chaotic models, but includes also a brief discussion of stochastic models. Isham is concerned with the statistical aspects of chaos.

**3. Model formulation.** We study both deterministic and stochastic versions of the SEIR model briefly described above in three cases. The first case deals with a constant contact rate in an isolated population. The second case is concerned with an isolated population with periodic contact rate. In this case, chaos appears in the deterministic version of the model if the amplitude of the contact rate is sufficiently large. The third case, finally, deals with a population that is exposed to a small amount of external infection, and where the contact rate is periodic. We proceed to describe the model used in the third case, which is the most general of the three.

We use three state variables, denoted  $S(t)$ ,  $E(t)$ , and  $I(t)$ , to denote the number of susceptible, exposed, and infective individuals, respectively, at time  $t$ . The deterministic version of the model leads to the following system of differential equations for the state variables:

$$(3.1) \quad S' = (\mu - \epsilon_1 - \epsilon_2)N - \frac{\beta}{N}SI - \mu S,$$

$$(3.2) \quad E' = \epsilon_1 N + \frac{\beta}{N}SI - (\gamma_1 + \mu)E,$$

$$(3.3) \quad I' = \epsilon_2 N + \gamma_1 E - (\gamma_2 + \mu)I,$$

where we allow the contact rate  $\beta$  to be periodic with a period of one year. By using a time unit of one year we find that the contact rate can be written as follows:

$$(3.4) \quad \beta = \beta_0(1 + \beta_1 \cos 2\pi t).$$

If the population is isolated, i.e. if  $\epsilon_1 = \epsilon_2 = 0$ , and if also the contact rate is constant, i.e. if  $\beta_1 = 0$ , then this system of differential equations has two critical points. One of them,  $(N, 0, 0)$ , corresponds to absence of infection, while the other one corresponds to an endemic infection level. The components of this second critical point can be written as follows:

$$(3.5) \quad \bar{S} = \frac{N}{R_0},$$

$$(3.6) \quad \bar{E} = \frac{1}{\alpha_1} \frac{R_0 - 1}{R_0} N,$$

$$(3.7) \quad \bar{I} = \frac{\alpha_1 - 1}{\alpha_1 \alpha_2} \frac{R_0 - 1}{R_0} N,$$

where we are using the following reparametrization:

$$(3.8) \quad R_0 = \frac{\gamma_1 \beta_0}{(\gamma_1 + \mu)(\gamma_2 + \mu)},$$

$$(3.9) \quad \alpha_1 = \frac{\gamma_1 + \mu}{\mu},$$

$$(3.10) \quad \alpha_2 = \frac{\gamma_2 + \mu}{\mu}.$$

We refer to  $R_0$  as the basic reproduction ratio. It is straightforward to show that the deterministic model with  $\beta_1 = \epsilon_1 = \epsilon_2 = 0$  has its threshold at  $R_0 = 1$ . The deterministic model predicts that if any amount of infection is introduced into a community, then the infection will establish itself at the endemic infection level given by  $(\bar{S}, \bar{E}, \bar{I})$  if  $R_0 > 1$ , while it will ultimately disappear if  $R_0 \leq 1$ .

The parameters  $\epsilon_1$  and  $\epsilon_2$  that determine the immigration rates of exposed and infective individuals are not independent. We assume that immigrating infected individuals arrive from one or several endemic areas. Hence it is natural to assume that the ratio  $\epsilon_1/\epsilon_2$  equals the ratio  $\bar{E}/\bar{I}$  of the endemic levels of exposed to infective individuals in an isolated population with constant contact rate. By putting  $\epsilon_1/\epsilon_2 = \alpha_2/(\alpha_1 - 1)$  we find that we can write

$$(3.11) \quad \epsilon_1 = \frac{\alpha_2}{\alpha_1 + \alpha_2 - 1} \delta \mu,$$

$$(3.12) \quad \epsilon_2 = \frac{\alpha_1 - 1}{\alpha_1 + \alpha_2 - 1} \delta \mu,$$

where the parameter  $\delta$  is introduced as a measure of the intensity of immigration of infected individuals.

The stochastic version of the model is a three-dimensional birth-death process with discrete state space and continuous time. The transition rates that determine the stochastic model can be read off from the various terms in the right-hand sides of the differential equations in (3.1)-(3.3). Thus, the rate of growth due to immigration of susceptible individuals in the deterministic model,  $(\mu - \epsilon_1 - \epsilon_2)N$ , equals the rate of transition in the stochastic model from any state  $(S, E, I)$  to  $(S + 1, E, I)$ . This means that the probability for such a transition in the time interval  $(t, t + \Delta t)$  is equal to the transition rate multiplied by the length of the time interval,  $\Delta t$ , plus a term, which, when divided by  $\Delta t$ , approaches zero as  $\Delta t$  goes toward zero. Similarly, the rate at which the number of susceptible individuals in the deterministic model decreases due to infection,  $\beta S I / N$ , equals the rate of transition in the stochastic model from any state  $(S, E, I)$  to  $(S - 1, E + 1, I)$ . By identification of the nine transition rates one can by standard techniques derive a system of differential equations, the Kolmogorov forward equations, for the state probabilities

$$(3.13) \quad p_{sei}(t) = P\{S(t) = s, E(t) = e, I(t) = i\}.$$

This is an infinite system of linear differential equations. The mathematical problems associated with the stochastic model are appreciably more difficult than those that go with the deterministic model.

The original model formulation introduced eight parameters, namely the total population size  $N$ , the death rate  $\mu$ , the mean contact rate  $\beta_0$ , the

amplitude of the time variation of the contact rate  $\beta_1$ , the rate of transition from exposed to infective  $\gamma_1$ , the rate of loss of infection  $\gamma_2$ , and the rates of immigration of exposed and infective individuals per individual  $\epsilon_1$  and  $\epsilon_2$ .

The number of parameters after the above assumption concerning immigration of infected individuals and reparametrization is equal to seven. Five of these, namely  $R_0$ ,  $\alpha_1$ ,  $\alpha_2$ ,  $\beta_1$ , and  $\delta$ , are dimensionless. They are all essential for the study of the model. The remaining two parameters are the death rate per individual  $\mu$  and the population size  $N$ . The death rate  $\mu$  has the dimension inverse time. Its inverse is the expected life length for an individual. It identifies a natural time constant for the model. If one wishes, one can dispense with  $\mu$  completely by introducing a scaled and dimensionless time as the product  $\mu t$ . Because of this, we refer to  $\mu$  as an innocent parameter. The population size  $N$  counts the number of individuals in the population. It has no physical dimension, but it has character reminiscent of parameters with such a dimension. One can refer to it as having a quasi-dimension, as in Nåsell (1985, Section 3.1.2). This parameter is quite special, since it is essential for the stochastic model and innocent for the deterministic model. A heuristic explanation for this is that the deterministic model can be derived from the stochastic one by scaling the state variables and taking the limit as  $N \rightarrow \infty$ . In this way,  $N$  disappears from the scene when we go from the stochastic to the deterministic setting.

The parameters  $\alpha_1$  and  $\alpha_2$  give the ratios of average life length to average length of time in the exposed and infective states, respectively. They are therefore large for the childhood infections that we are concerned with. The parameter  $\delta$  equals the ratio of the immigration rates of infected and susceptible individuals, and is small. The reparametrization is a necessary preparation for the asymptotic approximations that are carried out in the study of the quasi-stationary distribution in Subsection 4.4. It allows us to use the facts that  $\alpha_1$  and  $\alpha_2$  are large and that  $\delta$  is small in the analysis.

The process is ergodic if the population is exposed to external infection, i.e. if the parameter  $\delta$  is positive. On the other hand, if the population is isolated, i.e. if  $\delta = 0$ , then the states  $(s, 0, 0)$  form an absorbing class, while the states  $(s, e, i)$  with  $\max(e, i) > 0$  are transient. Eventual absorption is then certain and the time to absorption is an important random variable. This means that if the infection ever disappears from the population, i.e. if  $e = i = 0$ , then the number of infected individuals will remain equal to zero for all time thereafter. The quasi-stationary distribution is supported on the set of transient states.

Mathematical analysis of the model formulated is difficult in all three cases that we consider, and most of our results are based on numerical evaluations. We use Monte Carlo simulations for the stochastic model, and numerical solution of the system of differential equations for the deterministic model. In all three cases we shall use the standard measles parameters

given by Grenfell *et al.* (1995). They report  $\mu = 0.02$ ,  $\gamma_1 = 45.6$ ,  $\gamma_2 = 73.0$ , and  $\beta_0 = 1010.7$ . Hence  $\alpha_1 = 2281$ ,  $\alpha_2 = 3651$ , and  $R_0 = 13.84$ . Most of the numerical work reported in the following three sections has been done using Matlab. Exceptions are given by Figures 5–7 and 12, where a large number of Monte Carlo simulations have been done to evaluate the quasi-stationary distribution and the time to extinction. In these cases we have used Fortran instead of Matlab to speed up the computations. All computations have been run on a PC.

The SIR model used by Nåsell (1999) to study recurrent epidemics can be derived as a special case of the SEIR model that we are concerned with here by letting the parameter  $\alpha_1$  approach infinity. The stochastic versions of these two models have qualitatively similar behaviours.

**4. The isolated population with constant contact rate.** We study both the stochastic and the deterministic versions of the SEIR model formulated above with constant contact rate  $\beta$  and without external infection. In the first subsection we show that an approximation of the quasi-period of both the stochastic and the deterministic models can be determined from the deterministic formulation. In the second subsection we indicate how the damping associated with the deterministic model can be understood as a measure of the variability in the time between successive outbreaks in the stochastic model. The third subsection is devoted to the important quasi-stationary distribution and its relation with the time to extinction in the stochastic model. The fourth subsection, finally, gives a diffusion approximation of the quasi-stationary distribution.

**4.1. The quasi-period.** The number of infective individuals  $I(t)$  is shown as a function of  $t$  in Figures 1 and 2 for two different values of the population size  $N$ , with  $N$  equal to 1 000 000 in Figure 1, and  $N$  equal to 400 000 in Figure 2. The remaining parameters are the standard measles parameters given above. Each figure contains two plots of  $I(t)$ , one for the deterministic model and the other one for the stochastic model. The smooth curve is associated with the deterministic model, and the ragged one with the stochastic model.

The classical result that the solution of the deterministic model shows damped oscillations is illustrated in both figures. As time increases, the value of  $I(t)$  in the deterministic model approaches the endemic level  $\bar{I}$  given by (3.7). With the standard measles parameters we find that the endemic level  $\bar{I}$  equals 254 in Figure 1 and 102 in Figure 2.

The angular frequency  $\omega_0$  of the damped oscillations can be approximated by the standard method of linearization of the right-hand side of the system of differential equations (3.1)–(3.3) about the critical point  $(\bar{S}, \bar{E}, \bar{I})$  corresponding to the endemic infection level. By using the reparametrization in (3.8)–(3.10) and putting  $\beta_1 = \delta = 0$  we find that the Jacobian matrix  $J$ , evaluated at the critical point  $(\bar{S}, \bar{E}, \bar{I})$ , equals

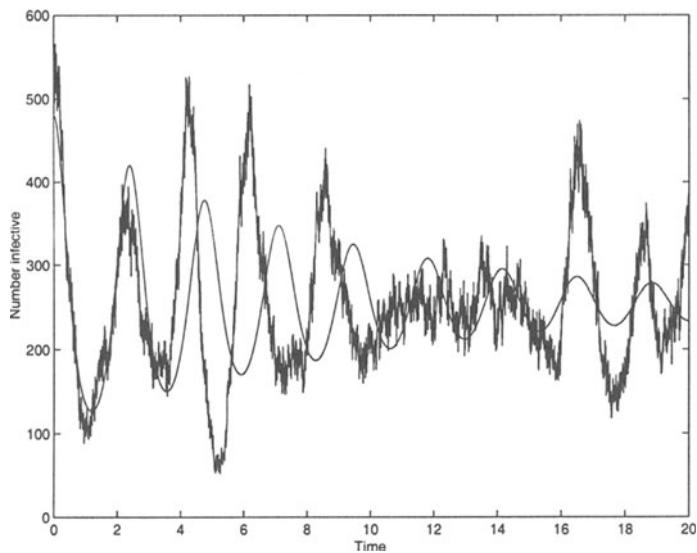


FIG. 1. *The number of infective individuals as functions of time for both the stochastic and deterministic versions of the SEIR model, with constant contact rate in an isolated population, and with  $N = 1\,000\,000$ . Standard measles parameters are used, as in all figures, with  $R_0 = 13.84$ ,  $\alpha_1 = 2281$ ,  $\alpha_2 = 3651$ , and  $\mu = 0.02$ .*

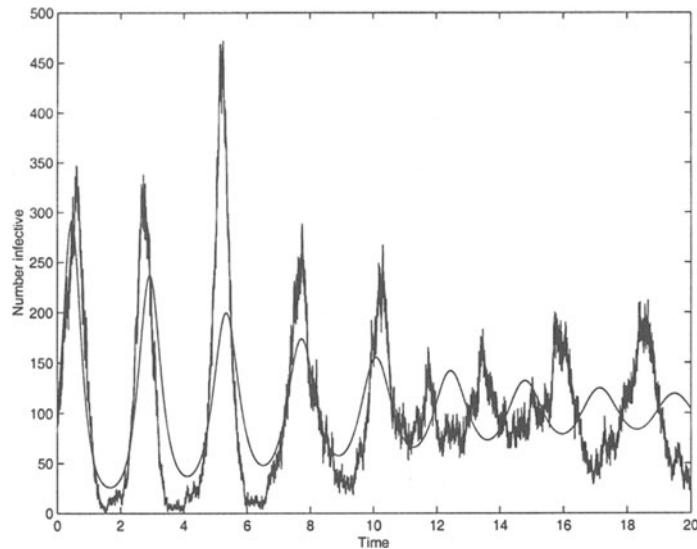


FIG. 2. *The number of infected individuals as functions of time for both the stochastic and deterministic versions of the SEIR model, with constant contact rate in an isolated population, and with  $N = 400\,000$ .*

$$(4.1) \quad J = \mu \begin{pmatrix} -R_0 & 0 & -\frac{\alpha_1 \alpha_2}{\alpha_1 - 1} \\ R_0 - 1 & -\alpha_1 & \frac{\alpha_1 \alpha_2}{\alpha_1 - 1} \\ 0 & \alpha_1 - 1 & -\alpha_2 \end{pmatrix}.$$

The characteristic equation for the eigenvalues  $\lambda$  of this matrix can be written

$$(4.2) \quad \lambda^3 + \mu(\alpha_1 + \alpha_2 + R_0)\lambda^2 + \mu^2(\alpha_1 + \alpha_2)R_0\lambda + \mu^3\alpha_1\alpha_2(R_0 - 1) = 0.$$

This equation can be solved explicitly, but the resulting expressions for the eigenvalues are too complicated to be useful. We derive instead an asymptotic approximation for the complex eigenvalue  $a_0 + i\omega_0$  under the assumption that  $\alpha_1 \rightarrow \infty$  and that  $\alpha_2$  is of the same order as  $\alpha_1$ . It is straightforward to derive the following result:

$$(4.3) \quad a_0 \sim -\mu \left( \frac{R_0}{2} - \frac{R_0 - 1}{2} \frac{\alpha_1 \alpha_2}{(\alpha_1 + \alpha_2)^2} \right),$$

$$(4.4) \quad \omega_0 \sim \mu \sqrt{\frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}} (R_0 - 1).$$

With the standard measles parameters we find from the latter of these two expressions that the quasi-frequency is approximately equal to  $\omega_0 \approx 2.685$ . (A numerical solution of the characteristic equation (4.2) gives  $\omega_0 = 2.682$ .) The corresponding quasi-period is  $T_0 = 2\pi/\omega_0 \approx 2.34$ . Figures 1 and 2 indicate that this value of  $T_0$  is close to the quasi-period for the deterministic model.

The simulations of the stochastic model in Figures 1 and 2 show recurrent outbreaks without any sign of damping. This confirms the claim by Bartlett (1956) that recurrence can be understood as a stochastic phenomenon. This holds even though our model differs somewhat from the one that Bartlett used. The time intervals between the outbreaks vary stochastically. We note that  $T_0$ , which was derived from the deterministic model, is also a good measure of the time between successive peaks in the stochastic model. One exception to this occurs in Figure 1 in the time interval from 10 to 16 years, where the number of infective individuals hovers around the deterministic endemic level  $\bar{I}$ , and no epidemic outbreak takes place. This increases the waiting time until the next outbreak, and hence adds to the variability in the length of the time intervals between outbreaks. This behaviour can be understood by reference to the realizations of the stochastic model.

Each such realization can be interpreted as a random movement in the state space of nonnegative integer values of  $S$ ,  $E$ , and  $I$ . This movement continues until extinction of the infection occurs at  $E = I = 0$ . The random movement is subjected to a drift in the direction of the trajectory

of the deterministic model. The drift is strong far away from the point  $(\bar{S}, \bar{E}, \bar{I})$  that corresponds to the endemic infection level in the deterministic model, and weak close to  $(\bar{S}, \bar{E}, \bar{I})$ . These two quite different behaviours are illustrated in Figures 3 and 4. Figure 3 shows the projection onto the S-I-plane of part of one realization of the stochastic model starting far away from  $(\bar{S}, \bar{E}, \bar{I})$ , and part of the corresponding deterministic path with the same initial point. The strong drift is clearly illustrated. As a contrast, Figure 4 indicates that the direction of the random movement appears largely unaffected by the direction of the deterministic path when the initial point is close to  $(\bar{S}, \bar{E}, \bar{I})$ .

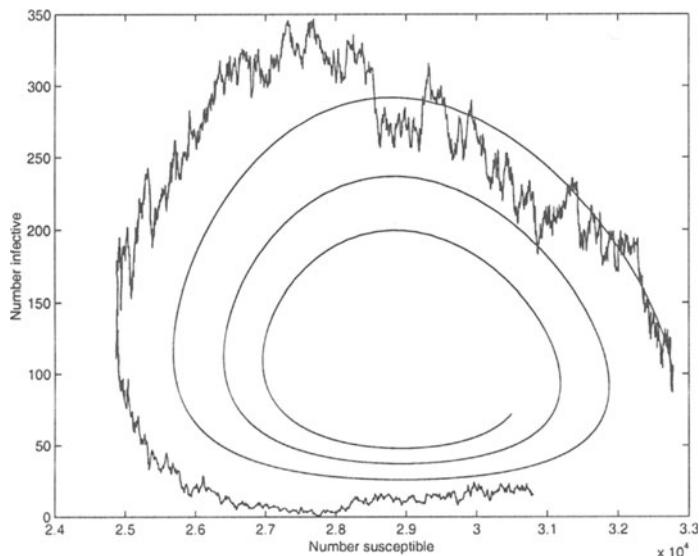


FIG. 3. An orbit of the deterministic model and a realization of the stochastic model projected onto the S-I-plane, with  $\beta_1 = \delta = 0$  and  $N = 400\,000$ . The initial point is far away from the deterministic endemic equilibrium, and the stochastic model realization is interpreted as a random movement with strong drift.

Chance will cause the stochastic realization to occasionally move into the region close to  $(\bar{S}, \bar{E}, \bar{I})$ , where the drift is weak. A stochastically determined delay will then occur before the realization leaves this inner region and the drift again grows strong and a new outbreak becomes possible. This explains the behaviour of the stochastic model realization in the time interval between 10 and 16 years in Figure 1.

**4.2. The damping.** It is a well-known fact that the solution of the deterministic model approaches the endemic infection level through damped oscillations. The damping in the model is not realistic. It has caused

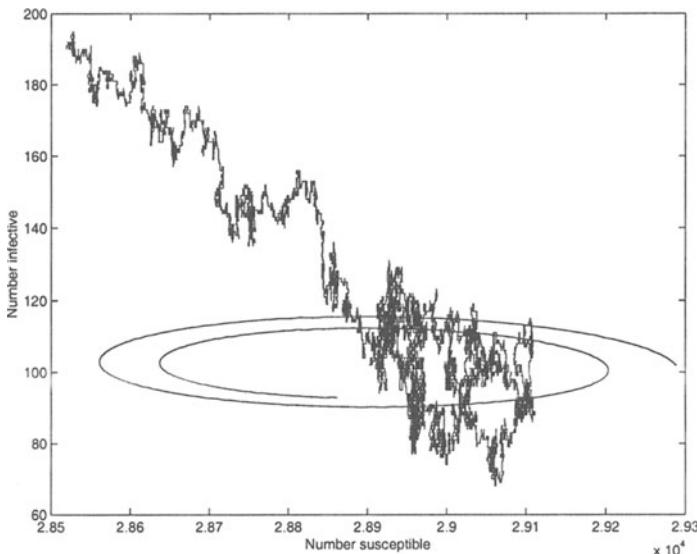


FIG. 4. An orbit of the deterministic model and a realization of the stochastic model projected onto the S-I-plane, with  $\beta_1 = \delta = 0$  and  $N = 400\,000$ . The initial point is close to the deterministic endemic equilibrium, and the stochastic model realization is interpreted as a random movement with weak drift.

a search for alternative deterministic model formulations with undamped oscillations. It has been shown that deterministic models with periodic contact rate, with age structure, and with isolation of infective individuals all can account for undamped periodic solutions. We claim that no changes in hypotheses are necessary; the damping can be understood from the stochastic version of the model that we study here.

Let us consider a number of realizations of the stochastic model, all with the same initial values (see Figures 1 and 2). The first outbreak will occur at nearly the same time in most of the realizations, but this synchronization of outbreaks will decrease as time increases, exactly because the lengths of the time intervals between the outbreaks vary stochastically. The expectation over these realizations will therefore show decreasing maxima and increasing minima as time increases, just as in the deterministic curve. But this is exactly what would be expected on theoretical grounds: The solution of the deterministic model is an approximation of the expectation over all realizations of the stochastic model. We arrive at the conclusion that the damping in the deterministic model is caused by the fact that the expectation over all realizations in the stochastic model does not mimic the individual realizations. This observation has previously been made by Alun Lloyd (pers. comm., 1997). This in turn means that the amount of

damping in the deterministic model is a measure of the stochastic variability in the lengths of the time intervals between successive outbreaks in the stochastic model.

It is useful to note that the stochastic model realizations also show a tendency toward damping in the case where the drift in the random movement is strong, as in Figure 3. However, this tendency toward damping is overshadowed by stochastic variability when the drift in the random movement is weak, as in Figure 4.

**4.3. Quasi-stationarity and the time to extinction.** As noted in the previous section, the states  $(s, 0, 0)$  form an absorbing class for the stochastic model. If the process ever reaches a state where both the number of exposed individuals  $E(t)$  and the number of infective individuals  $I(t)$  are equal to zero, then both of these state variables will remain equal to zero thereafter, and the infection has gone extinct. The term stochastic fade-out is also used to refer to the extinction phenomenon. The time to extinction is an important random variable.

Another concept of great importance in this situation is that of quasi-stationarity. The so-called quasi-stationary distribution is a conditional stationary distribution. The condition entering into its definition is that absorption has not yet occurred. If the process has been going on for a long time, and if absorption has not yet occurred, then the state of the process is well approximated by the quasi-stationary distribution. It turns out that the quasi-stationary distribution contains information about the time to extinction in a special but important case, namely the one where the initial distribution is equal to the quasi-stationary distribution.

The quasi-stationary distribution and the time to extinction are studied in some detail for the related SIR model by Nåsell (1999). By using the approach used in this reference, we can readily derive the following results for the SEIR model.

The quasi-stationary distribution is derived by a consideration of the conditional probabilities

$$(4.5) \quad \begin{aligned} q_{sei}(t) &= P\{S(t) = s, E(t) = e, I(t) \\ &= i | \max(E(t), I(t)) > 0\} = \frac{p_{sei}(t)}{1 - p_{00}(t)}, \end{aligned}$$

where

$$(4.6) \quad p_{ei}(t) = \sum_{s=0}^{\infty} p_{sei}(t) = P\{E(t) = e, I(t) = i\}$$

denotes the unconditional marginal distribution of the number of exposed and infective individuals at time  $t$ . The conditional probabilities  $q_{sei}(t)$  are defined on the set of transient states, i.e. for all nonnegative values of  $s$ ,  $e$ , and  $i$  except  $e = i = 0$ .

By using the Kolmogorov equations for the state probabilities  $p_{sei}(t)$  we can derive a differential equation for  $p_{00}(t)$ . By applying (4.5) and using this equation for  $p_{00}(t)$  and the Kolmogorov equations for  $p_{sei}(t)$  we can derive a system of differential equations for the conditional state probabilities  $q_{sei}(t)$ . This system of equations contains infinitely many nonlinear equations. Its stationary solution  $\{q_{sei}\}$  is the quasi-stationary distribution of the process.

It is not possible to solve for the quasi-stationary distribution in explicit form. We use two methods for getting information about it. One is via Monte Carlo simulation, and the other is through the diffusion approximation discussed in the next subsection.

The time to extinction is a random variable with important epidemiological interpretations. It is used to quantify the three related concepts of persistence of the infection, the critical community size, and the persistence threshold; see Nåsell (1995). An important property of the time to extinction is that it has an exponential distribution if the initial distribution of the process is equal to the quasi-stationary distribution. Furthermore, the expectation of the time to extinction can in this case be determined from the quasi-stationary distribution through the formula

$$(4.7) \quad E\tau_Q = \frac{1}{\mu q_{10} + \alpha_2 q_{01}}.$$

Here,  $\tau_Q$  is used to denote the time to extinction from the quasi-stationary distribution. Note that transition of the process from the transient to the absorbing class can only take place through one-step transitions from one of the boundary states  $(s, 1, 0)$  or  $(s, 0, 1)$  to  $(s, 0, 0)$  for some  $s \geq 0$ . In the first case, absorption results by death of the one exposed individual, while in the second case it follows from death or recovery of the one infective individual. The expression for  $E\tau_Q$  above contains the corresponding transition rates weighted by the corresponding probabilities in the quasi-stationary distribution.

**4.4. Diffusion approximation of the quasi-stationary distribution.** The main result in this subsection is that the quasi-stationary distribution is approximated by a three-dimensional normal distribution if  $N$  is sufficiently large. The mean values that describe this normal distribution are equal to the components of the endemic critical point for the deterministic model given by (3.5)–(3.7), while we indicate how to derive the covariance matrix that is needed to specify this normal distribution.

We consider the process  $x(t) = (x_1(t), x_2(t), x_3(t))$ , whose components are the scaled state variables of the original process:  $x_1(t) = S(t)/N$ ,  $x_2(t) = E(t)/N$ , and  $x_3(t) = I(t)/N$ . The critical point of the correspondingly scaled deterministic model that corresponds to an endemic infection level for  $R_0 > 1$  is denoted  $\hat{x} = (\hat{x}_1, \hat{x}_2, \hat{x}_3) = (1/R_0, (R_0 - 1)/(\alpha_1 R_0), (\alpha_1 - 1)(R_0 - 1)/(\alpha_1 \alpha_2 R_0))$ .

The changes in the scaled state variables  $x_i$  during the time interval from  $t$  to  $t + \Delta t$  are denoted  $\Delta x_i$ . By using the hypotheses for the original process we determine the mean and the covariance of the vector with components  $\Delta x_i$ . For the mean we get

$$(4.8) \quad E \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \end{pmatrix} = \mu \begin{pmatrix} 1 - \frac{\alpha_1 \alpha_2 R_0 x_1 x_3}{\alpha_1 - 1} - x_1 \\ \frac{\alpha_1 \alpha_2 R_0 x_1 x_3}{\alpha_1 - 1} - \alpha_1 x_2 \\ (\alpha_1 - 1)x_2 - \alpha_2 x_3 \end{pmatrix} \Delta t + o(\Delta t)$$

$$= b(x)\Delta t + o(\Delta t).$$

The vector  $b(x)$  gives the rate of growth of this expectation. The Jacobian matrix of this vector with respect to  $x$  is denoted  $B(x)$ :

$$(4.9) \quad B(x) = \frac{\delta b(x)}{\delta x} = \mu \begin{pmatrix} -\frac{\alpha_1 \alpha_2 R_0 x_3}{\alpha_1 - 1} - 1 & 0 & -\frac{\alpha_1 \alpha_2 R_0 x_1}{\alpha_1 - 1} \\ \frac{\alpha_1 \alpha_2 R_0 x_3}{\alpha_1 - 1} & -\alpha_1 & \frac{\alpha_1 \alpha_2 R_0 x_1}{\alpha_1 - 1} \\ 0 & \alpha_1 - 1 & -\alpha_2 \end{pmatrix}.$$

We approximate the matrix  $B(x)$  by evaluating at the deterministic critical point  $\hat{x}$ :

$$(4.10) \quad B(\hat{x}) = \mu \begin{pmatrix} -R_0 & 0 & -\frac{\alpha_1 \alpha_2}{\alpha_1 - 1} \\ R_0 - 1 & -\alpha_1 & \frac{\alpha_1 \alpha_2}{\alpha_1 - 1} \\ 0 & \alpha_1 - 1 & -\alpha_2 \end{pmatrix}.$$

The covariance matrix of the vector of changes  $\Delta x_i$  is found to be as follows:

$$(4.11) \quad \text{Cov} \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \end{pmatrix}$$

$$= \frac{\mu}{N} \begin{pmatrix} 1 + A + x_1 & -A & 0 \\ -A & A + \alpha_1 x_2 & -(\alpha_1 - 1)x_2 \\ 0 & -(\alpha_1 - 1)x_2 & (\alpha_1 - 1)x_2 + \alpha_2 x_3 \end{pmatrix} \Delta t + o(\Delta t)$$

$$= \frac{S(x)}{N} \Delta t + o(\Delta t),$$

where

$$(4.12) \quad A = \frac{\alpha_1 \alpha_2 R_0 x_1 x_3}{\alpha_1 - 1}.$$

The matrix  $S(x)$  is approximated by evaluating it at the critical point  $\hat{x}$ :

$$(4.13) \quad S(\hat{x}) = \frac{\mu(R_0 - 1)}{\alpha_1 R_0} \begin{pmatrix} 2\alpha_1 R_0 / (R_0 - 1) & -\alpha_1 & 0 \\ -\alpha_1 & 2\alpha_1 & -(\alpha_1 - 1) \\ 0 & -(\alpha_1 - 1) & 2(\alpha_1 - 1) \end{pmatrix}.$$

The process  $N^{1/2}\{x(t) - \hat{x}\}$  is approximated for large  $N$  by a three-dimensional Ornstein-Uhlenbeck process with local drift matrix  $B(\hat{x})$  and local covariance matrix  $S(\hat{x})$ . Its stationary distribution approximates the quasi-stationary distribution. It is approximately normal with mean 0 and covariance matrix  $\Sigma$ , where the matrix  $\Sigma$  is determined from the matrices  $B(\hat{x})$  and  $S(\hat{x})$  by solving the matrix equation

$$(4.14) \quad B(\hat{x})\Sigma + \Sigma B^T(\hat{x}) = -S(\hat{x}),$$

where the superscript  $T$  denotes transpose.

We have used the symbolic computation package Maple to solve this matrix equation; to solve the equation by hand would for most people be both very tedious and highly error prone. The Maple results show that each entry in the covariance matrix  $\Sigma$  can be written as the ratio of two polynomials in  $\alpha_1$  of degree at most 4, and whose coefficients are polynomials in  $\alpha_2$  and  $R_0$ . These expressions are too complicated to be useful. We have therefore derived asymptotic approximations for each term in  $\Sigma$ , again with the aid of Maple, under the assumption that  $\alpha_1 \rightarrow \infty$ , also assuming that  $\alpha_1$  and  $\alpha_2$  are of the same order. By writing  $\theta = \alpha_2/\alpha_1$  we find that the results of these calculations can be expressed as follows:

$$(4.15) \quad \Sigma \sim C \begin{pmatrix} \theta(\theta + 1)\alpha_1 & -\theta(\theta R_0 + 1) & -(\theta + 1)R_0 \\ -\theta(\theta R_0 + 1) & \theta^2(R_0 - 1) & \theta(R_0 - 1) \\ -(\theta + 1)R_0 & \theta(R_0 - 1) & R_0 - 1 \end{pmatrix},$$

where

$$(4.16) \quad C = \frac{1}{R_0[R_0(\theta^2 + \theta + 1) + \theta]}.$$

The correlation coefficient  $\rho_{23}$  is of importance for determining the joint marginal distribution of  $E$  and  $I$ , which in turn appears in the expression (4.7) for the expected time to extinction from quasi-stationarity. It is defined by  $\rho_{23} = s_{23}/\sqrt{s_{22}s_{33}}$ , where  $s_{ij}$  denotes the entry in row  $i$  and column  $j$  of the matrix  $\Sigma$ . If we approximate the covariances that go into the definition of  $\rho_{23}$  by their asymptotic approximations given in (4.15), we readily find that  $\rho_{23} \sim 1$ . One additional term in the asymptotic approximation of  $\rho_{23}$  is therefore needed. By again using Maple, we find that

$$(4.17) \quad \rho_{23} \sim 1 - \frac{\theta R_0 + (1 + 1/\theta)^2 R_0 + 1}{2\alpha_1}.$$

By using the results in (4.15) we find that the marginal distributions of  $S$ ,  $E$ , and  $I$  in quasi-stationarity are approximately normal. The expectations of the marginal distributions are given in (3.5) - (3.7). The last of them is simplified by its asymptotic approximation as  $\alpha_1 \rightarrow \infty$ . Thus we get

$$(4.18) \quad \mu_S = \frac{N}{R_0},$$

$$(4.19) \quad \mu_E = \frac{(R_0 - 1)N}{\alpha_1 R_0},$$

$$(4.20) \quad \mu_I \sim \frac{(R_0 - 1)N}{\alpha_2 R_0}.$$

The standard deviations of the three marginal distributions have the following asymptotic approximations:

$$(4.21) \quad \sigma_S \sim \sqrt{\frac{(\alpha_1 + \alpha_2)\alpha_1\alpha_2 N}{R_0(R_0(\alpha_1^2 + \alpha_1\alpha_2 + \alpha_2^2) + \alpha_1\alpha_2)}},$$

$$(4.22) \quad \sigma_E \sim \alpha_2 \sqrt{\frac{(R_0 - 1)N}{R_0(R_0(\alpha_1^2 + \alpha_1\alpha_2 + \alpha_2^2) + \alpha_1\alpha_2)}},$$

$$(4.23) \quad \sigma_I \sim \alpha_1 \sqrt{\frac{(R_0 - 1)N}{R_0(R_0(\alpha_1^2 + \alpha_1\alpha_2 + \alpha_2^2) + \alpha_1\alpha_2)}}.$$

A criterion for the validity of these approximations can be read off from the results. The derivation of this criterion will lead to a reparametrization that can be used to identify two parameter regions with qualitatively different behaviours of the quasi-stationary distribution. The two regions are described as “ $R_0$  is distinctly large” and “ $R_0$  is in the transition region”, respectively. The quasi-stationary distribution is approximately normal in the first region, but deviates appreciably from normality in the second one.

Since we are dealing with nonnegative random variables, it is clear that a normal distribution cannot be a good approximation of any of the marginal distributions unless the ratio of mean to standard deviation of the marginal distribution is sufficiently large. Provisionally we introduce therefore a new parameter  $\rho$  by putting  $\rho = \mu_I/\sigma_I$ , and note that the same result would be achieved if instead we put  $\rho = \mu_E/\sigma_E$ . By using the asymptotic approximations of  $\mu_I$  and  $\sigma_I$  above we get the following relation between  $R_0$  and  $\rho$ :

$$(4.24) \quad R_0 \sim 1 + \frac{\alpha_1^2 \alpha_2^2 \rho^2}{\alpha_1^2 + \alpha_1\alpha_2 + \alpha_2^2 + \frac{\alpha_1\alpha_2}{R_0}} \frac{1}{N}.$$

Note that  $R_0 \rightarrow 1$  as  $N \rightarrow \infty$  if  $\rho$ ,  $\alpha_1$ , and  $\alpha_2$  are kept fixed. By replacing  $R_0$  in the right-hand side by its one-term asymptotic approximation we are led to the following relation between  $\rho$  and  $R_0$ :

$$(4.25) \quad R_0 = 1 + \left( \frac{\alpha_1 \alpha_2 \rho}{\alpha_1 + \alpha_2} \right)^2 \frac{1}{N}, \quad N \rightarrow \infty.$$

The boundary between the two parameter regions is for practical purposes determined by setting  $\rho = 3$ . This choice of  $\rho$  is related to the fact that a normally distributed random variable takes values larger than its mean minus three times its standard deviation with high probability. The new parameter  $\rho$  is used instead of  $R_0$  in the transition region, but not in the region where  $R_0$  is distinctly large.

An alternate interpretation of these results is that we will wind up in the region where  $R_0$  is distinctly large if the parameters  $R_0 > 1$ ,  $\alpha_1$ , and  $\alpha_2$  are fixed as the population size  $N$  is increased to sufficiently large values. For the standard measles parameters we find that it is necessary for  $N$  to exceed 1.4 million individuals in order to reach the parameter region where the marginal distributions of  $E$  and  $I$  in quasi-stationarity are reasonably close to normal.

Figures 5 and 6 show the marginal distribution of the number of infective individuals in quasi-stationarity. The cases treated here are those with standard measles parameters and population sizes equal to 400 000 and 1 000 000, respectively. By using (4.25) we find that the corresponding  $\rho$ -values are 1.6 and 2.6, respectively. This means that for both cases we are in the transition region, where the distributions are expected to be non-normal. Two curves are shown in each figure. The smooth one gives the normal approximation whose moments are given above, modified by truncation at -0.5 to account for continuity correction. The rugged curve is the result of Monte Carlo simulations. The agreement between the two curves improves as we move toward the parameter region where  $R_0$  is distinctly large.

A comparison between the stochastic realizations in Figures 1 and 2 shows that the minima between the outbreaks are appreciably lower in Figure 2 than in Figure 1. This in turn means that the possibility of extinction is higher in Figure 2, where the population size is lower, and that therefore the time to extinction is shorter. This expectation is borne out by the results of Monte Carlo simulations that have been used to measure the time to extinction. The results are shown in Figure 7. The initial values for all of the simulations have been chosen at integer values close to the deterministic model critical point  $(\bar{S}, \bar{E}, \bar{I})$  that corresponds to the endemic levels. The estimates of the coefficient of variation of the time to extinction all lie in the interval from 0.68 to 0.81, and the number of simulations was 500 in each case. This means that the 90% confidence intervals for all of the results lie between  $\pm 5\%$  and  $\pm 6\%$ .

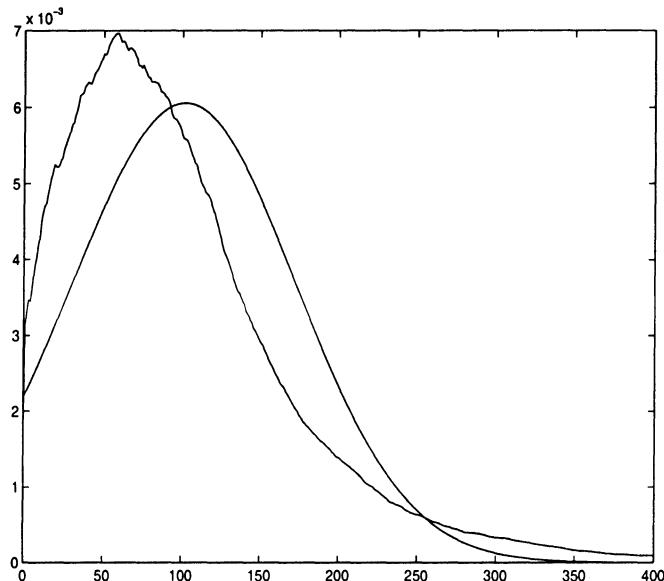


FIG. 5. *The marginal distribution of the number of infective individuals in quasi-stationarity: Simulation results compared with the truncated normal approximation. The population size is 400 000.*

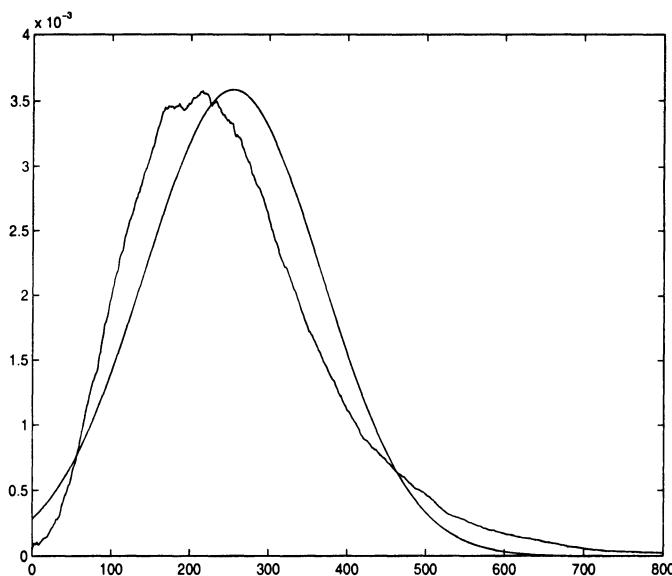


FIG. 6. *The marginal distribution of the number of infective individuals in quasi-stationarity: Simulation results compared with the truncated normal approximation. The population size is 1 000 000.*

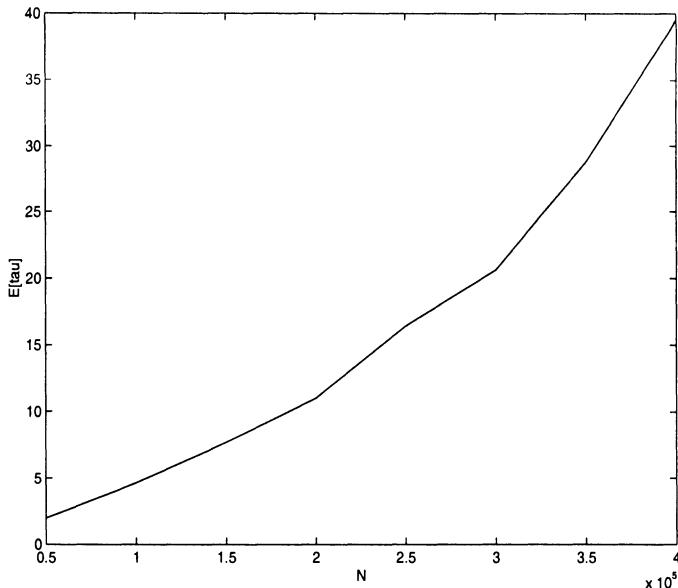


FIG. 7. *The expected time to extinction as a function of the population size  $N$ .*

The theoretical result in (4.7) gives the expected time to extinction from quasi-stationarity in terms of the two probabilities  $q_{10}$  and  $q_{01}$  from the joint marginal distribution of exposed and infective individuals in quasi-stationarity. This result can be combined with our normal approximation of the quasi-stationary distribution to give an explicit approximation of the expected time to extinction from quasi-stationarity. Unfortunately, the resulting approximation gives a serious overestimate of the time to extinction. The reason for this is that even though the marginal distributions of the number of exposed and infective individuals are reasonably well approximated by our normal result, the joint distribution is not. The conditional distribution of  $I$  given  $E$  has a constant variance in our normal approximation, while simulation results show that this variance is much smaller than the value given by the normal approximation when  $E$  is small. This behaviour may be related to the fact that the joint marginal distribution of  $E$  and  $I$  is close to singular since the correlation coefficient is asymptotically equal to 1. Further investigation of this matter is needed in order to improve the approximation of the joint marginal distribution of  $E$  and  $I$  in quasi-stationarity.

It is well-known that the extinction phenomenon cannot be captured by the deterministic model. One reason for this is that the time to extinction depends on the population size  $N$ , and this parameter is not essential for the deterministic model. Another, related, reason is that extinction is associated with individual realizations of the stochastic model; different

realizations will go extinct at different times. Now the expected value of the extinction time over the different realizations grows without bound as  $N \rightarrow \infty$ , as can be seen from Figure 7, and the condition that  $N \rightarrow \infty$  is essential in deriving the deterministic model from the stochastic one.

**5. The chaotic case: An isolated population with periodic contact rate.** The case with periodic contact rate and  $\beta_1 = 0.28$  leads to chaos in the deterministic model. Illustrations giving the number of infective individuals  $I(t)$  as function of  $t$  are given in Figures 8 and 9. The two figures show the same result of numerically solving the deterministic system of differential equations, with a linear vertical scale in Figure 8 and a logarithmic vertical scale in Figure 9. The initial values are chosen so that the transients that appear before the chaotic behaviour is apparent are avoided.

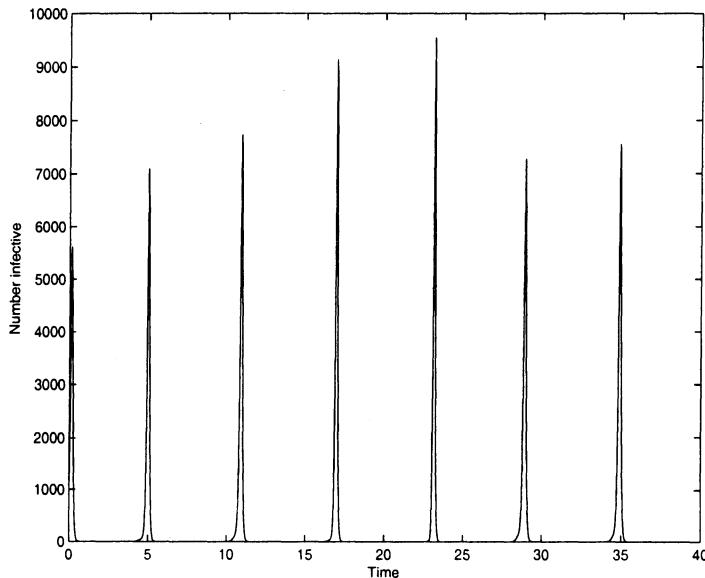


FIG. 8. A chaotic solution of the deterministic version of the SEIR model with periodic contact rate ( $\beta_1 = 0.28$ ) for an isolated population with  $N = 1\,000\,000$ .

The behaviour of  $I(t)$  can be interpreted as a succession of epidemic outbreaks. The time intervals between successive outbreaks are not constant, and the peak number of infective individuals is variable, and in these senses the chaotic model gives an improvement over the case with the deterministic model with constant contact rate.

A serious weakness with the chaotic model is, however, that the number of infective individuals dips down to unrealistically low values between the outbreaks. This observation is not new; remarks to this effect have been made by a number of authors, including Grenfell (1992), Bolker and

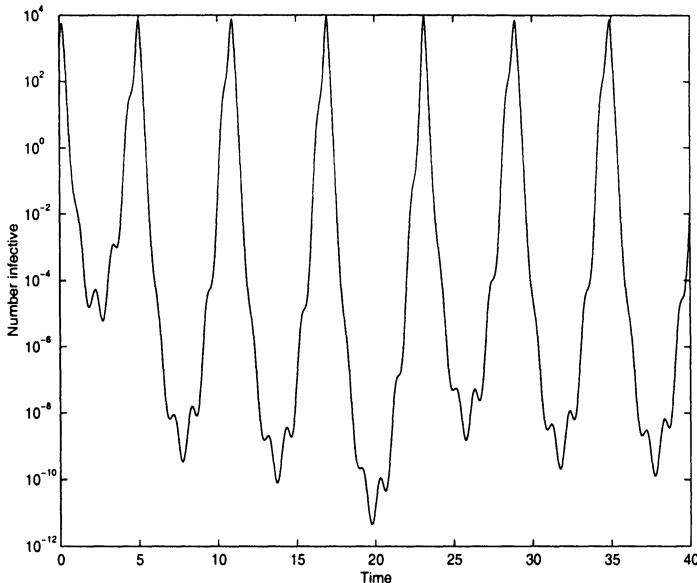


FIG. 9. The same chaotic results as in Figure 8, but with a logarithmic vertical scale.

Grenfell (1993), Grenfell *et al.* (1994). Figure 9 shows that the minimum number of infective individuals after about 20 years goes down below  $10^{-11}$  individuals in a population of 1 million individuals! (This low value resembles the “atto-fox” discovered by Mollison (1991) in a deterministic model for rabies. Mollison defines one atto-fox to be equal to  $10^{-18}$  foxes.) In order for this model to be reasonably realistic we would require the population size to be at least of the order of  $10^{12}$  million individuals (sic!). It is a well-known fact that the deterministic model is an acceptable approximation of the stochastic one if the population size  $N$  is sufficiently large. The difficulty with applying this requirement in any particular situation is that for practical purposes one needs to find out what is meant by “sufficiently large”. Our finding here is a rather surprising quantification of this requirement.

As in the previous subsection we study the fully stochastic counterpart of this deterministic model to get some understanding for the behaviour of the chaotic model. Comparisons of deterministic orbits with stochastic realizations are given in Figures 10 and 11. Figure 10 shows the projection onto the S-I-plane of part of one orbit of the deterministic model, with the same initial point as in Figure 8, and the projection of one realization of the stochastic model with the same initial point. This figure indicates that the drift for the stochastic model is strong. The figure also shows that the infection in the stochastic model goes extinct at a time when the number of infective individuals in the deterministic model reaches the vicinity of 1.

Indeed, a large majority of all stochastic realizations with the same initial value will go extinct at about the same time. One can use the fact that the minimum "number" of infective individuals in the deterministic model is about  $10^{-5}$  to conjecture that only one out of  $10^5$  stochastic realizations will survive to give a second epidemic outbreak. Figure 11 compares a deterministic orbit with a stochastic realization when the components of the initial point are integer values close to the critical point that corresponds to an endemic infection when the contact rate is constant ( $\beta_1 = 0$ ). This figure shows that the variation with time of the contact rate causes the deterministic orbit to quickly move away from the vicinity of the initial point, quite opposite to the case illustrated in Figure 4. One consequence of this is that the drift in the stochastic model is stronger than when the contact rate is constant as in Figure 4.

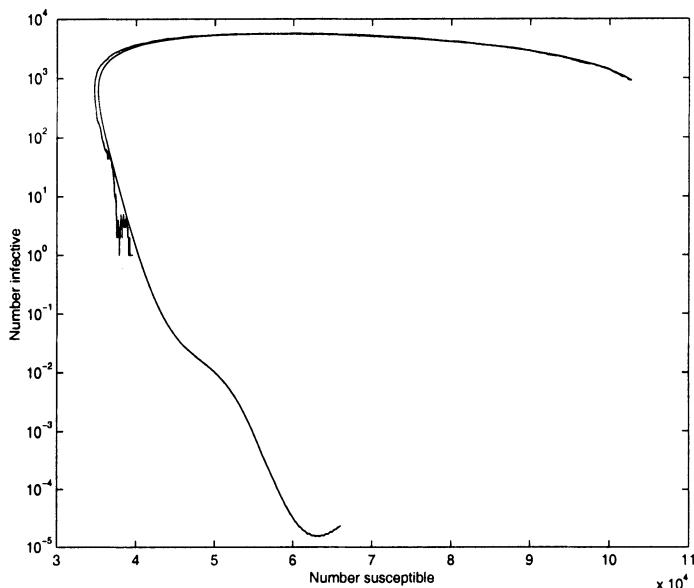


FIG. 10. An orbit of the deterministic model and a realization of the stochastic model projected onto the S-I-plane, with  $\beta_1 = 0.28$ ,  $\delta = 0$ , and  $N = 10^6$ .

The result in Figure 10 indicates that the time to extinction for the stochastic model with  $\beta_1 = 0.28$  is much shorter than in the case where the contact rate is constant. To further illustrate this fact we have run Monte Carlo simulations of the stochastic model to determine the dependence on  $\beta_1$  of the expected time to extinction. The results for a population size of 300 000 individuals are shown in Figure 12. The initial values used for these runs are integer values close to  $(\bar{S}, \bar{E}, \bar{I})$ . The figure shows that the expected time to extinction decreases sharply from over 20 years to less than 1 year as the parameter  $\beta_1$  is increased from zero to 0.5.

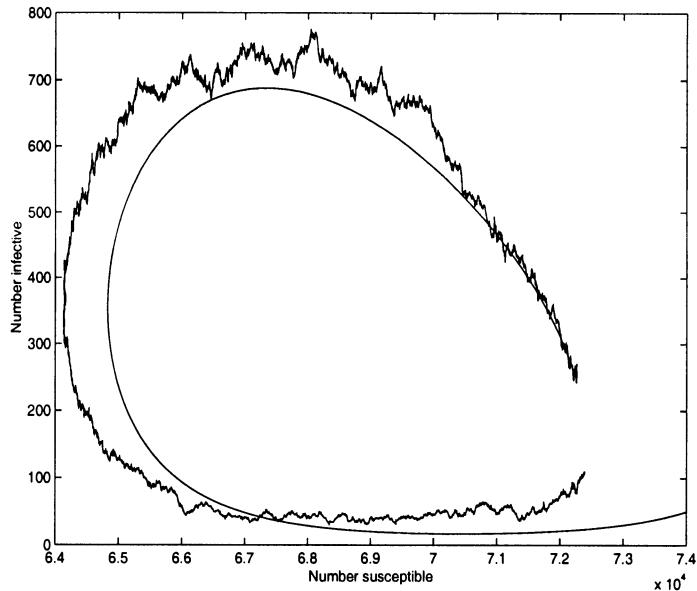


FIG. 11. An orbit of the deterministic model and a realization of the stochastic model projected onto the S-I-plane, with  $\beta_1 = 0.28$ ,  $\delta = 0$ , and  $N = 10^6$ .

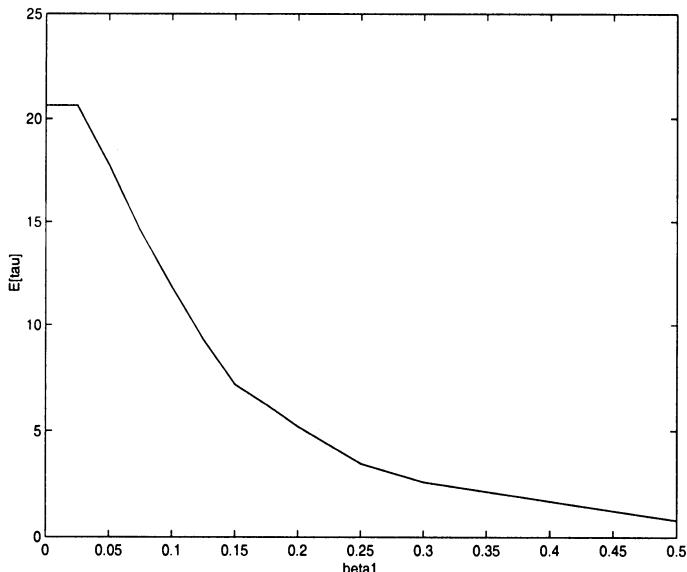


FIG. 12. The expected time to extinction as a function of  $\beta_1$  for an isolated population of 300 000 individuals.

The results that we have shown allow us to claim that the deterministic model in this case is an unacceptable approximation of the stochastic model, and therefore also an unacceptable model of the real world phenomenon that we are concerned with. The chaos in the deterministic model is associated with an unrealistic behaviour of the deterministic model in the region where  $I(t)$  is below the value 1. This in turn is an artifact with no correspondence in the fully stochastic model, nor in the real world.

A second conclusion that can be drawn is that the stochastic model with  $\beta_1 = 0.28$  is unrealistic; it predicts much shorter times to extinction than what has been observed. One way to possibly remedy this situation is to allow for external infection in the model. Some comments on deterministic and stochastic models with periodic contact rate and external infection are given in the next subsection.

#### **6. The case with periodic contact rate and external infection.**

The model that we have studied so far deals with a community that is completely isolated. This is of course unrealistic. We turn therefore to a brief study of the full model established in Section 3, allowing for a small inflow of infection from one or several other infected communities. Figure 13 shows the number of infective individuals as a function of time for the deterministic model. The parameters in this case are  $\beta_1 = 0.28$ ,  $\delta = 5 \cdot 10^{-6}$ , and  $N = 10^6$ . This corresponds to the minute number of  $\mu\delta N = 0.1$  infected individuals immigrating per year into a population of 1 million individuals. The figure shows that the chaos that appeared when  $\delta = 0$  has disappeared and been replaced by a periodic solution with period 11. It may at first sight be surprising to find that this minute amount of external infection is sufficient to give a qualitative change in the solution of the deterministic model. However, it is easy to understand heuristically that even this small amount of external infection may have a substantial influence on the solution of the chaotic model, since the external infection rate is large compared to the extremely small values for the “number” of infective individuals achieved by the chaotic model in the troughs between outbreaks. Figure 13 shows that the external infection has caused the minimum values of  $I(t)$  to increase by a factor of the order of  $10^9$  from the minimum values observed in the chaotic case. The deterministic model is still unrealistic for a population of 1 million, since the minimum values are far below 1.

Some numerical experiments show that the deterministic model with this minute amount of external infection will again show chaotic behaviour if  $\beta_1$  is increased to the value 0.30. If the number of immigrating infected individuals is increased to 1 per year, then chaos appears when  $\beta_1$  is increased to the value 0.39. A further increase of the external infection to 4 infected individuals per year requires one to increase  $\beta_1$  to 0.58 to find chaos. Finally, if as many as 10 infected individuals immigrate per year, then the deterministic model appears to be non-chaotic even for the maximum possible value of  $\beta_1$ , namely 1.0.

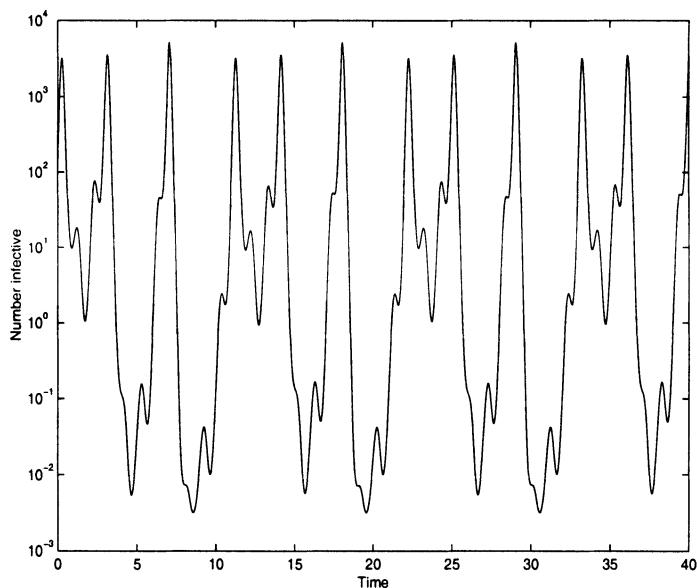


FIG. 13. The number of infective individuals as function of time for the deterministic SEIR model with periodic contact rate ( $\beta_1 = 0.28$ ) and a minute external infection rate. One infected individual immigrates every ten years into a population of 1 000 000. The solution is periodic with period 11.

One conclusion of this brief study is that the deterministic model with periodic contact rate is very sensitive to external infection. It is obvious that this sensitivity also holds for the stochastic version of the model.

**7. Concluding Comments.** One of the key papers claiming epidemic chaos is Olsen, Truty and Schaffer (1988). They develop a Monte Carlo procedure that they refer to as a “probabilistic analog” of the chaotic differential equations describing the SEIR model. At the end of their paper, the authors claim that “the Monte Carlo simulation method has several advantages over the traditional approach which entails solving differential equations. First of all, only integers are allowed. Thus, whereas in the differential equations,  $I(t)$  can drop to numbers as small as  $10^{-12}$ , in the Monte Carlo simulations, the disease can actually become extinct. To prevent this in our calculations, we allowed for a small, constant influx of infectives.” These arguments in favor of Monte Carlo simulations coincide with our arguments for a stochastic model. The main difference between our standpoint and that of Olsen, Truty and Schaffer is of a basic, conceptual nature. They appear to regard the deterministic model (with chaos) as the fundamental one, while we claim that the stochastic model is basic, and that the deterministic model is an approximation of the stochastic one, which is acceptable only when certain conditions are met.

Similar comments can be made with regard to Engbert and Drepper (1994). They are also concerned with the distinction between chaos and stochasticity, using the SEIR model as a concrete example. Their awareness of the shortcomings of the deterministic model is clearly indicated by the following quote: "Another problem of the deterministic description is the existence of attractors, where the fraction of infectives  $I$  drops down to below  $10^{-7}$ . This is a problematic property of the dynamics of the SEIR model, because even in the biggest population centres the deterministic description breaks down. The integer structure of the host population has to be taken into account for such low incidence levels. The problem will be addressed using Monte-Carlo simulations based on a stochastic formulation of the model." In spite of this, the authors claim that their "analysis lends support to the plausibility of chaos". It appears again that an acceptance of our stand with regard to the relative importance of deterministic and stochastic models is all that is needed to claim that the Engbert-Drepper quote coincides with our arguments for a stochastic model.

The question of distinguishing chaos from stochastically generated noise is addressed by Sugihara and May (1990), Hastings *et al.* (1993), and May (1995). The stochastic ingredient considered by these authors is however very different from the demographic stochasticity that we are concerned with. They refer to it as representing measurement error. The distinction between deterministic chaos and measurement error is that the predictability deteriorates with increasing prediction-time interval in the chaotic case, while it is roughly constant for the stochastic case. This contrasts sharply with our case of demographic stochasticity, where the stochastic model realizations can be interpreted as a random movement with drift. It is clear the predictability of such realizations will decrease as the length of the prediction-time interval is increased in that part of state space where the drift is strong.

A heuristic insight into the behaviour of a stochastic model whose deterministic approximation has periodic (or rather damped periodic) solutions is given by the following quote from Diekmann *et al.* (1995): "A general feeling is that the possibility of extinction due to demographic stochasticity is strongly enhanced if the deterministic dynamics is characterized by oscillations, rather than by a stable endemic steady state". One interpretation of this statement is that the time to extinction is much shorter for an endemic SIR model or endemic SEIR model (where the endemic critical point in the deterministic model is a stable spiral) than for an endemic SIS model (where the endemic critical point in the deterministic model is a stable node). A comparison of the results that we have discussed here with the results for the SIS model (Nåsell (1996,1999)) give strong support to this statement.

The discussion in the present paper identifies some challenging open problems in the stochastic realm. For the model of an isolated popula-

tion with constant infection rate we note that the recurrence properties of the endemic SIR or SEIR model are described by short-term predictions, while the quasi-stationary distribution is concerned with the long-term predictions. The time intervals between successive outbreaks are of obvious short-term interest, while the time to extinction from the quasi-stationary distribution is related to the quasi-stationary distribution, and therefore dependent on the long-term predictions.

The originator of a stochastic formulation for recurrent epidemics, Bartlett, discussed stochastic versus chaotic models in Bartlett (1990), where he gave arguments for the stochastic formulation. His arguments were supported by Dietz and Schenzle (1990). Our discussion in the preceding three sections can be seen as a further support and extension of these arguments.

**Acknowledgements.** I thank Klaus Dietz for discussions that stimulated me to undertake this study. I express my gratitude to IMA and to Carlos Castillo-Chavez for inviting me to present this paper at the IMA Workshop on Mathematical Approaches for Emerging and Reemerging Infectious Diseases at the University of Minnesota in Minneapolis on May 17, 1999. I am grateful to Frank Ball, Maurice Bartlett, Ben Bolker, Jonathan Dushoff, Bryan Grenfell, Johannes Müller, Rolf Sundberg, and Hans Thunberg for useful comments.

## REFERENCES

- [1] R.M. Anderson and R.M. May, *Infectious Diseases of Humans: Dynamics and Control* (1991), Oxford: Oxford University Press.
- [2] J.L. Aron and I.B. Schwartz, *Seasonality and period-doubling bifurcations in an epidemic model*, J. Theor. Biol. **110** (1984), 665–679.
- [3] N.T.J. Bailey, *The Mathematical Theory of Infectious Diseases and its Applications* (1975), London: Griffin.
- [4] M.S. Bartlett, *Deterministic and stochastic models for recurrent epidemics*, In Proc. 3rd Berkeley Symp. Math. Stat. and Prob., **4** (1956), 81–109, Berkeley: University of California Press.
- [5] M.S. Bartlett, *Measles periodicity and community size (with discussion)*, J. Roy. Stat. Soc. A, **120** (1957), 48–70.
- [6] M.S. Bartlett, *Stochastic Population Models in Ecology and Epidemiology* (1960a), London: Methuen.
- [7] M.S. Bartlett, *Some stochastic models in ecology and epidemiology*, In Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling (eds I. Olkin, S.G. Ghurye, W. Hoeffding, W.G. Madow and H.B. Mann) (1960b), 89–96, Stanford: Stanford University Press.
- [8] M.S. Bartlett, *The critical community size for measles in the United States*, J. Roy. Stat. Soc. A, **123** (1960c), 37–44.
- [9] M.S. Bartlett, *Chance or chaos? (with discussion)*, J. Roy. Stat. Soc. A, **153** (1990), 321–347.
- [10] B.M. Bolker and B.T. Grenfell, *Chaos and biological complexity in measles dynamics*, Proc. Roy. Soc. Lond. B, **251** (1993), 75–81.

- [11] O. Diekmann, J.A.P. Heesterbeek, and J.A.J. Metz, *The legacy of Kermack and McKendrick*, In Epidemic Models: Their Structure and Relation to Data, edited by D. Mollison, Publications of the Newton Institute, Cambridge University Press, Cambridge, 1995.
- [12] K. Dietz, *Transmission and control of arbovirus diseases*, In Epidemiology (eds. D. Ludwig and K.L. Cooke) (1975), 104–121. Philadelphia: Society for Industrial and Applied Mathematics.
- [13] K. Dietz, *The incidence of infectious diseases under the influence of seasonal fluctuations*, Lecture Notes in Biomathematics **11** (1976), 1–15. Berlin: Springer Verlag.
- [14] K. Dietz, *Some problems in the theory of infectious disease transmission and control*, In Epidemic Models: Their Structure and Relation to Data, edited by D. Mollison, Publications of the Newton Institute, Cambridge University Press, Cambridge, 1995.
- [15] K. Dietz and D. Schenzle, *Discussion on Chance or chaos? (by M.S. Bartlett)*, J. Roy. Stat. Soc. A, **153** (1990), p. 338.
- [16] S. Ellner, A.R. Gallant, and J. Theiler, *Detecting nonlinearity and chaos in epidemic data*, in Epidemic Models: Their Structure and Relation to Data (ed. D. Mollison) (1995), 229–247. Cambridge: Cambridge University Press.
- [17] R. Engbert and F.R. Drepper, *Chance and chaos in population biology — Models of recurrent epidemics and food chain dynamics*, In Chaos, Solitons & Fractals, **4** (1994), 1147–1169.
- [18] Z. Feng and H.R. Thieme, *Recurrent outbreaks of childhood diseases revisited: The impact of isolation*, Math. Biosci., **128** (1995), 93–130.
- [19] B.T. Grenfell, *Chance and chaos in measles dynamics*, J. Roy. Stat. Soc. B, **54** (1992), 383–398.
- [20] B.T. Grenfell, A. Kleczkowski, S.P. Ellner, and B.M. Bolker, *Measles as a case study in nonlinear forecasting and chaos*, Phil. Trans. Roy. Soc. Lond. A, **348** (1994), 515–530.
- [21] B.T. Grenfell, B. Bolker, and A. Kleczkowski, *Seasonality, demography and the dynamics of measles in developed countries*, In Epidemic Models: Their Structure and Relation to Data, edited by D. Mollison, Publications of the Newton Institute, Cambridge University Press, Cambridge, 1995.
- [22] W.H. Hamer, *Epidemic disease in England — The evidence of variability and of persistence of type*, Lancet (1906), 733–739.
- [23] A. Hastings, C.L. Hom, S. Ellner, P. Turchin, and H.C.J. Godfray, *CHAOS IN ECOLOGY: Is Mother Nature a Strange Attractor?*, Annu. Rev. Ecol. Syst. **24** (1993), 1–33.
- [24] H.W. Hethcote, *Asymptotic behavior and stability in epidemic models*, Lecture Notes Biomathematics **2**, (1974). Berlin: Springer Verlag.
- [25] H.W. Hethcote, *Qualitative analysis of communicable disease models*, Math. Biosci., **28** (1976), 335–356.
- [26] H.W. Hethcote and S.A. Levin, *Periodicity in epidemiological models*, In Applied Mathematical Ecology (ed S.A. Levin, T.G. Hallam, and L.J. Gross), Biomathematics **18**, Berlin: Springer Verlag.
- [27] V. Isham, *Statistical aspects of chaos: A review*, In Chaos and Networks: Statistical and Probabilistic Aspects (eds. O.E. Barndorff-Nielsen, J.L. Jensen, and W.S. Kendall), London: Chapman and Hall (1993), 124–200.
- [28] M.J. Keeling and B.T. Grenfell, *Disease extinction and community size: Modelling of the persistence of measles*, Science **275** (1997), 65–67.
- [29] W.P. London and J.A. Yorke, *Recurrent outbreaks of measles, chickenpox and mumps. I. Seasonal variations in contact rates*, Am. J. Epidemiol. **98** (1973), 453–468.
- [30] A.J. Lotka, *Martini's equations for the epidemiology of immunising diseases*, Nature **111** (1923), 633–634.

- [31] A.J. Lotka, *Elements of Mathematical Biology* (1956), New York: Dover Publications.
- [32] E. Martini, *Berechnungen und Beobachtungen zur Epidemiologie und Bekämpfung der Malaria* (1921), Hamburg: Gente.
- [33] R.M. May, *Stability and Complexity in Model Ecosystems* (1973), Princeton: Princeton University Press.
- [34] R.M. May, *Simple mathematical models with very complicated dynamics*, Nature **261** (1976), 459–476.
- [35] R.M. May, *Necessity and chance: Deterministic chaos in ecology and evolution*, Bull. Amer. Math. Soc. **32** (1995), 291–308.
- [36] D. Mollison, *The dependence of epidemic and population velocities on basic parameters*, Math. Biosci. **107** (1991), 255–287.
- [37] I. Nåsell, *Hybrid Models of Tropical Infections*, Lecture Notes in Biomathematics **59** (1985). Berlin: Springer Verlag.
- [38] I. Nåsell, *The threshold concept in stochastic epidemic and endemic models*, In *Epidemic Models: Their Structure and Relation to Data* (ed. D. Mollison), Publications of the Newton Institute, Cambridge University Press, Cambridge, 1995.
- [39] I. Nåsell, *The quasi-stationary distribution of the closed endemic SIS model*, Adv. Appl. Prob., **28** (1996), 895–932.
- [40] I. Nåsell, *On the quasi-stationary distribution of the stochastic logistic epidemic*, Math. Biosci., **156** (1999a), 21–40.
- [41] I. Nåsell, *On the time to extinction in recurrent epidemics*, J. R. Stat. Soc. B **61** (1999b), 309–330.
- [42] L.F. Olsen and W.M. Schaffer, *Chaos versus noisy periodicity: hypotheses for childhood epidemics*, Science, **249** (1990), 499–504.
- [43] L.F. Olsen, G.L. Truty, and W.M. Schaffer, *Oscillations and chaos in epidemics: A nonlinear study of six childhood diseases in Copenhagen, Denmark*, Theor. Pop. Biol. **33** (1988), 344–370.
- [44] D.A. Rand and H.B. Wilson, *Chaotic stochasticity: a ubiquitous source of unpredictability in epidemics*, Proc. Roy. Soc. Lond. B **246** (1991), 179–184.
- [45] W.M. Schaffer, *Can nonlinear dynamics elucidate mechanisms in ecology and epidemiology?* IMA J. Math. Appl. Med. Biol. **2** (1985), 221–252.
- [46] W.M. Schaffer and M. Kot, *Nearly one dimensional dynamics in an epidemic*, J. Theor. Biol. **112** (1985), 403–427.
- [47] D. Schenzle, *An age-structured model of pre- and post-vaccination measles transmission*, IMA J. Math. Appl. Biol. Med. **1** (1984), 169–191.
- [48] H.E. Soper, *The interpretation of periodicity in disease prevalence (with discussion)*, J. Roy. Stat. Soc. A **92** (1929), 34–73.
- [49] G. Sugihara and R.M. May, *Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series*, Nature **344** (1990), 734–741.

# EPIDEMICS AMONG A POPULATION OF HOUSEHOLDS\*

FRANK G. BALL<sup>†</sup> AND OWEN D. LYNE<sup>†</sup>

**Abstract.** This paper considers SIR and SIS epidemics among a population partitioned into households. This heterogeneity has important implications for the threshold behaviour of epidemics and optimal vaccination strategies. It is shown that taking into account household structures when modelling public health problems is valuable. An overview of households models is given, including a determination of threshold parameters, the probability of a global epidemic and some new results on vaccination strategies for SIS households epidemics. Simulation and numerical studies are presented which exemplify the results discussed.

**Key words.** SIR and SIS epidemics; household structure; threshold behaviour; optimal vaccination strategy; metapopulation models.

**AMS(MOS) subject classifications.** Primary 92D30.

**1. Introduction.** By far the most important result to come out of mathematical epidemic theory is the celebrated threshold theorem, which broadly states that a major epidemic can occur only if the initial susceptible population (numbers for stochastic model; density or proportion for deterministic model) is sufficiently large (see, for example, Heesterbeek and Dietz (1996)). The result is important because it implies what proportion of susceptibles need to be vaccinated in order to prevent an epidemic occurring (see Section 5). However, for it to be practically relevant, it is necessary that the assumptions underlying the model on which it is based adequately reflect what happens in real-life epidemics.

Most epidemic models assume that the population among which a disease is spreading is locally, as well as globally, large. By locally large, it is meant that if the population is partitioned into a number of groups, for example by age, sex, and/or geographical location, then each of these groups (and not just the total population) are also large. This assumption is implicit in deterministic models, which require locally large populations to provide a good approximation to their more realistic stochastic counterparts (c.f. Kurtz (1970), (1971)). It is also made when determining the threshold behaviour of most stochastic models from a suitable branching process approximation (see, for example, Whittle (1955), Williams (1971) and Ball (1983)). However, locally large populations are clearly inappropriate for many human and animal epidemics, since such populations are usually partitioned into small groups or households. For some animal diseases, and also many plant diseases, it is natural to explicitly model the spatial structure of the population and spatial epidemic models have a long

---

\*This research was supported in part by the UK Engineering and Physical Sciences Research Council, under research grant number GR/L56282.

<sup>†</sup>School of Mathematical Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, UK; fgb@maths.nott.ac.uk, odl@maths.nott.ac.uk.

history, see for example the review papers of Mollison (1977) and Mollison and Levin (1995). However, the mixing behaviour of human populations is typically far more complicated than that provided by a lattice structure and there has been considerable interest recently in models for the spread of an epidemic among a population comprising a large number of small households, with different contact rates for within- and between-household infection.

The earliest paper that we are aware of which considers an epidemic model with two levels of mixing is Rushton and Mautner (1955), which analysed a deterministic model for a so-called simple epidemic, i.e. one without removal of infectives. Although a number of papers on both stochastic and deterministic epidemics among stratified populations appeared in the intervening period, the next papers directly relevant to the present households setting are Watson (1972) and Bartoszyński (1972). Watson considered both deterministic and stochastic epidemics for a fixed number of large groups; see Daley and Gani (1994) for a more detailed analysis of the deterministic model. Bartoszyński's paper is the first one to consider epidemics among a population of *small* households. Bartoszyński considered a discrete-time epidemic model, in which the spread within different households followed independent arbitrary but specified processes, and at each time point infectious individuals independently infected a number of new households, all of which were assumed to have previously contained only susceptibles. The process of infected households in this model follows a branching process, which enabled Bartoszyński to derive an associated threshold theorem. Note that although Bartoszyński's model is defined in terms of individual behaviour, the key epidemiological unit for its analysis is a household. This is the case for all the households models described in this paper.

Models for the spread of epidemics within small households have a long history, including the classical Greenwood and Reed-Frost chain-binomial models (see, Bailey (1975), chapter 8). These models have been extended to incorporate the possibility of community infection, but without explicitly modelling this phenomenon and assuming that the spread of infection within distinct households is independent, see, for example, Becker (1989) and Addy et al. (1991). During the programme on Epidemic Models, held at the Isaac Newton Institute for Mathematical Sciences (University of Cambridge) in 1993, Klaus Dietz posed the question of how the reproductive ratio  $R_0$  (see, for example, Heesterbeek and Dietz (1996)) should be defined for such models when community infection is explicitly incorporated, prompted by which two works developed in parallel. One, originating in Becker and Dietz (1995), has been primarily concerned with evaluating vaccination strategies. The other, originating in Ball et al. (1997) but first reported in the discussion to Mollison et al. (1994), has been primarily concerned with providing a general modelling framework for such epidemics, together with associated methods of analysis. A related modelling ap-

proach, which lies outside the framework considered in this paper, is that of Aparicio et al. (2000) for social clustering in tuberculosis transmission.

The aim of this paper is to provide a brief overview of models for epidemics among a population of households. The SIR (susceptible → infective → removed) model of Ball et al. (1997), which includes the model of Becker and Dietz (1995) for highly infectious diseases as a special case, is introduced in Section 2. The threshold behaviour of this SIR model is determined by considering an approximating branching process, which in fact is the limit of the epidemic process as the number of households tends to infinity and is related to the model of Bartoszyński (1972). The final outcome in the event of the epidemic taking off is outlined briefly. Simulations that illustrate the theoretical results are also described. The simplest class of epidemic models that can display endemic behaviour, namely that of SIS (susceptible → infected → susceptible) epidemics, is considered in a households setting in Section 3. The threshold behaviour of such households SIS epidemics is determined and compared with that of corresponding SIR epidemics. In contrast to standard epidemic models, the SIR and corresponding SIS households epidemic models can have differing threshold behaviours. Standard deterministic models are inappropriate for a population of households unless all of the households are large in size. However, as outlined in Section 4, it is still possible to develop appropriate deterministic (meta-population) models. Vaccination strategies for diseases spreading among a community of households are considered in Section 5, primarily along the lines developed by Becker and his co-workers. Finally, a few concluding comments are given in Section 6, which also contains a brief discussion on possible extensions.

## 2. SIR epidemics among a community of households.

**2.1. Model.** Consider the following model, analysed in detail in Ball et al. (1997), for the spread of an SIR epidemic among a closed population, partitioned into small groups or households. Suppose that, for  $n = 1, 2, \dots$ , the population contains  $m_n$  households of size  $n$ . Let  $m = \sum_{n=1}^{\infty} m_n$  and  $N = \sum_{n=1}^{\infty} nm_n$  denote respectively the total number of households and individuals, both of which are assumed to be finite. The infectious periods of different infectives are independent and identically distributed according to a random variable  $T_I$ , having an arbitrary but specified distribution. Throughout its infectious period, a given infective makes (global) contacts with a given susceptible at the points of a homogeneous Poisson process having rate  $\lambda_G/N$  (thus  $\lambda_G$  is the total rate that a given infective makes global contacts) and it makes (local) contacts with a given susceptible in its own household at the points of a Poisson process having rate  $\lambda_L$ . All the Poisson processes describing contacts (whether or not either or both of the individuals involved are the same), as well as the random variables describing infectious periods, are assumed to be mutually independent. A contacted susceptible becomes infected and is immediately able to infect

other individuals, so there is no latent period. At the end of its infectious period an infected individual is immune to further infection. The epidemic is initiated by a number of individuals becoming infected at time  $t = 0$  and it terminates as soon as there are no infectives present in the population.

For ease of exposition, it has been assumed that there is no latent period and that an infectious individual can make both local and global contacts with susceptibles in its own household. These are no real restrictions given the purpose of this paper. The threshold behaviour of an SIR epidemic model is a function of its final outcome, the distribution of which is invariant to very general assumptions concerning a latent period (see Ludwig (1975), Ball (1986) and, in the present households setting, Ball et al. (1997)). It may seem more natural to formulate the model so that global contacts can only occur between individuals from different households. Consider such a formulation, with global and local contacts between pairs of given individuals occurring at rate  $\lambda'_G/N$  and  $\lambda'_L$ . Then, provided  $\lambda'_L \geq \lambda'_G/N$  (i.e. the individual to individual contact rate is greater for within-household contacts than for between-household contacts — a very plausible assumption in practice) such a formulation is equivalent to our model with  $\lambda_G = \lambda'_G$  and  $\lambda_L = \lambda'_L - \lambda'_G/N$ .

Note that letting  $\lambda_L \rightarrow \infty$  in our model yields the model of Becker and Dietz (1995) for highly infectious diseases.

**2.2. Threshold behaviour.** Suppose that the number of households  $m$  is large. Then, during the early stages of an epidemic with a small number of initial infectives, the probability that a global contact is with an individual residing in a previously infected household is small. Thus the initial growth of the epidemic can be approximated by a process in which each global contact is with an individual in an otherwise completely susceptible household. The process of infected households in this approximate process follows a branching process, whose offspring distribution can be defined as follows.

For  $n = 1, 2, \dots$ , let  $\alpha_n = m_n/m$  be the proportion of households of size  $n$  and let  $\tilde{\alpha}_n = nm_n/N$  ( $= n\alpha_n / \sum_{k=1}^{\infty} k\alpha_k$ ) be the probability that a global contact is with an individual residing in a household of size  $n$ . Consider a completely susceptible household of size  $n$  that is contacted globally. The initially infected individual in that household will start a realisation of a single household epidemic (without outside infection as we are assuming that all global contacts are with completely susceptible households). Let  $T_n$  denote the total number of individuals (including the initial case) infected by that single household epidemic and  $A_n$  be the sum of the infectious periods of those  $T_n$  infectives, i.e. the so-called severity of the epidemic. Each infectious individual makes global contacts at total rate  $\lambda_G$ , so the total number of global contacts emanating from the above single household epidemic,  $R_n$  say, follows a Poisson distribution with random mean  $\lambda_G A_n$ . Moreover, each such global contact is with a completely

susceptible household. Hence the number of infected households emanating from a typical infected household,  $R$  say, is a mixture of  $R_1, R_2, \dots$  with respective mixing probabilities  $\tilde{\alpha}_1, \tilde{\alpha}_2, \dots$ . Thus  $R$  is the offspring random variable for the approximating branching process.

The above approximation of the epidemic process by a branching process can be made mathematically fully rigorous by considering a sequence of epidemics in which  $m \rightarrow \infty$  and using a coupling argument, see Ball (1996) and Ball (1999) for epidemics with equal and unequal household sizes, respectively. A threshold theorem for the epidemic process can then be obtained by saying that a *global epidemic* occurs if in the limit as  $m \rightarrow \infty$  the epidemic infects infinitely many households, i.e. if the branching process does not go extinct. Let  $R_* = \mathbb{E}(R)$  and  $f(s) = \mathbb{E}(s^R)$  be the mean and probability generating function of  $R$ , respectively. Then by standard branching process theory, a global epidemic can occur only if  $R_* > 1$ . Moreover, if  $R_* > 1$  and the epidemic is started by  $a$  initial cases chosen independently and at random from the population then the probability of global epidemic is  $1 - p^a$ , where  $p$  is the root of  $f(s) = s$  in  $(0, 1)$ . For other configurations of initial infectives, the calculation of the probability of a global epidemic is more difficult since the initial individuals in the branching process have different offspring laws to subsequent individuals. In such circumstances the probability of a global epidemic is most easily calculated by conditioning on the size of the first generation in the branching process; see Ball et al. (1997) for details.

The above discussion indicates that  $R_*$  is a threshold parameter for the SIR households epidemic model, corresponding to the proliferation of infected households. It is also possible to define other threshold parameters (Becker and Dietz (1995), (1996)), such as for the proliferation of infected individuals. Of course, all appropriate threshold parameters pass through the threshold value of 1 simultaneously.

In order to compute the threshold parameter  $R_*$ , note that

$$(2.1) \quad R_* = \mathbb{E}(R) = \sum_{n=1}^{\infty} \tilde{\alpha}_n \mathbb{E}(R_n) = \sum_{n=1}^{\infty} \tilde{\alpha}_n \lambda_G \mathbb{E}(A_n),$$

since  $R_n$  follows a Poisson distribution with random mean  $A_n$ . Further,  $\mathbb{E}(A_n) = \mathbb{E}(T_n)\mathbb{E}(T_I)$ , by a Wald's identity for epidemics (Ball (1986)). Let  $\mu_n = \mathbb{E}(T_n)$  ( $n = 1, 2, \dots$ ), then

$$R_* = \lambda_G \mathbb{E}(T_I) \sum_{n=1}^{\infty} \tilde{\alpha}_n \mu_n,$$

which takes the form  $R_* = R_0 \mu$ , where  $R_0 = \lambda_G \mathbb{E}(T_I)$  is the usual reproductive ratio (see, for example, Heesterbeek and Dietz (1996)) for the model in which all the households are of size 1 (i.e. the standard homogeneously mixing epidemic model) and  $\mu = \sum_{n=1}^{\infty} \tilde{\alpha}_n \mu_n$  is a mean amplification factor owing to internal spread within a household. Finally,  $\mu_n$  ( $n = 1, 2, \dots$ )

can be computed as follows. For  $n, a = 1, 2, \dots$ , let  $\mu_{n,a}$  be the mean final size (including the initial infectives) of a single household epidemic with initially  $a$  infectives and  $n$  susceptibles, so  $\mu_n = \mu_{n-1,1}$ . (This extra notation will be useful in the next subsection.) Also let  $\phi(\theta) = \mathbb{E}(e^{-\theta T_I})$  ( $\theta \geq 0$ ) denote the moment generating function of  $T_I$ . Then it follows from (2.25) and (2.26) of Ball (1986) that

$$\mu_{n,a} = n + a - \sum_{k=0}^n \binom{n}{k} \beta_k \phi(\lambda_L k)^{n+a-k},$$

where  $\beta_0, \beta_1, \dots$  are determined recursively by

$$\sum_{i=0}^k \binom{k}{i} \beta_i \phi(\lambda_L i)^{k-i} = k \quad (k = 0, 1, \dots).$$

To determine the probability of a global epidemic, a formula for  $f(s)$  is required. Now

$$f(s) = \mathbb{E}(s^R) = \sum_{n=1}^{\infty} \tilde{\alpha}_n \mathbb{E}(s^{R_n}).$$

Further,  $R_n$  is Poisson with random mean  $\lambda_G A_n$ , so

$$(2.2) \quad \mathbb{E}(s^{R_n}) = \mathbb{E}\left[\mathbb{E}(s^{R_n} | A_n)\right] = \mathbb{E}\left[e^{-\lambda_G A_n(1-s)}\right] = \phi_{n-1,1}(\lambda_G(1-s)),$$

where  $\phi_{n,a} = \mathbb{E}(e^{-\theta A_{n,a}})$  and  $A_{n,a}$  is the severity of a single household epidemic with initially  $n$  susceptibles and  $a$  infectives. Finally, it follows from Theorem 2.5 of Ball (1986) that

$$\phi_{n,a}(\theta) = \sum_{k=0}^n \binom{n}{k} \gamma_k(\theta) \phi(\theta + \lambda_L k)^{n+a-k},$$

where  $\gamma_0(\theta), \gamma_1(\theta), \dots$  are determined recursively by

$$\sum_{i=0}^k \binom{k}{i} \gamma_i(\theta) \phi(\theta + \lambda_L i)^{k-i} = 1 \quad (k = 0, 1, \dots).$$

**2.3. Final outcome in the event of a global epidemic.** Still assuming that the number of households  $m$  is large and the initial number of infectives is small, suppose that a global epidemic occurs and let  $z$  be the proportion of individuals that are ultimately infected by the epidemic. This  $z$  is also the probability that a randomly chosen individual is infected by the epidemic. Such an individual resides in a household of size  $n$  with probability  $\tilde{\alpha}_n$  and, with probability tending to 1 as  $m \rightarrow \infty$ , it also resides in a household that did not contain any initial infectives. Let  $\bar{T}_n$  denote the

total number of individuals that are ultimately infected in that household. Thus, given that it resides in a household of size  $n$ , the probability that the randomly chosen individual is ultimately infected is also given by  $n^{-1}\mathbb{E}(\tilde{T}_n)$ .

To determine  $\mathbb{E}(\tilde{T}_n)$  note that the total person units of infection present throughout the epidemic is approximately  $Nz\mathbb{E}(T_I)$  and hence the probability that a given individual avoids global infection is approximately given by

$$(2.3) \quad \pi = \exp\left(-\frac{\lambda_G}{N}Nz\mathbb{E}(T_I)\right) = \exp(-\lambda_G z\mathbb{E}(T_I)).$$

(This is because a given infective globally contacts a given susceptible at the points of a Poisson process with rate  $\lambda_G/N$  so if the given individual is to avoid global infection throughout the epidemic there must be no points of a Poisson process with rate  $\lambda_G/N$  in an interval of length  $Nz\mathbb{E}(T_I)$ , which happens with the above probability.) Further, distinct individuals avoid global infection approximately independently of each other, so  $\tilde{T}_n$  is approximately distributed as the final size of a single household epidemic with a binomial number of initial infectives. Thus

$$\mathbb{E}(\tilde{T}_n) = \sum_{k=1}^n \binom{n}{k} (1-\pi)^k \pi^{n-k} \mu_{n-k,k}$$

and hence  $z$  satisfies

$$(2.4) \quad z = \sum_{n=1}^{\infty} \tilde{\alpha}_n n^{-1} \sum_{k=1}^n \binom{n}{k} (1-\pi)^k \pi^{n-k} \mu_{n-k,k},$$

which together with (2.3) gives an implicit equation for  $z$ .

Clearly  $z = 0$  is a root of (2.4). If  $R_* \leq 1$  then  $z = 0$  is the only root in  $[0, 1]$ , whilst if  $R_* > 1$  there is a unique second root in  $[0, 1]$  which gives the proportion of individuals ultimately infected by a global epidemic (see Ball et al. (1997) for details). Note that the probability that a household of size  $n$  avoids global infection is  $\pi^n = \exp(-n\lambda_G z\mathbb{E}(T_I))$  so the proportion of households that ultimately remain uninfected is therefore  $\sum_{n=1}^{\infty} \alpha_n \exp(-n\lambda_G z\mathbb{E}(T_I))$ . The above approximation becomes exact in the limit as  $m \rightarrow \infty$  and, as shown in Ball et al. (1997), the above heuristic argument can be made fully rigorous.

Suppose that all the households are of size 1. Then (2.3) and (2.4) reduce to

$$(2.5) \quad z = 1 - \exp(-\lambda_G \mathbb{E}(T_I) z),$$

the equation governing the final size of single population SIR deterministic epidemic.

**2.4. Simulations.** To illustrate the results of Section 2 a simulation study was carried out using a population of 300 households of size 5. The simulations were done using an infectious period distribution that was used in Addy et al. (1991). They considered a gamma infectious period distribution, written  $T_I \sim \text{gamma}(2, 2.05)$ , with probability density function  $f_{T_I}(t) = c^2 \exp(-ct)$ , ( $t > 0$ ), where  $c = 2/4.1$  and thus  $\mathbb{E}(T_I) = 4.1$ . For each of 9 pairs of parameter values  $(\lambda_L, \lambda_G)$ , 1000 simulations were conducted. Each epidemic was started with 1 infective and 1499 susceptibles.

Figure 1 shows the total number of people infected at the end of the epidemic. The two columns of figures are approximately matched for  $R_*$ , the left-hand column shows increasing  $\lambda_L$  with  $\lambda_G$  fixed, the right-hand column shows increasing  $\lambda_G$  with  $\lambda_L$  fixed. Both columns start with the same set of parameter values. The importance of the threshold parameter  $R_*$  is clearly shown. Large epidemics never occur unless  $R_* > 1$ , and, given  $R_*$ , the precise values of  $\lambda_L$  and  $\lambda_G$  have little effect on either the number of people infected or the probability of a large epidemic occurring.

Figure 2 is arranged in the same fashion as Figure 1 and shows the numbers of households that were affected by the epidemics, that is, the number of households which contained an infective at some point during the epidemic. Whereas the two columns on Figure 1 look very similar, the two columns of Figure 2 show the differing character of epidemics driven by local and global infection, respectively. An epidemic driven by local infection totally infects many households and leaves many others untouched, whereas an epidemic driven by global infection affects many more households though infecting those households less heavily.

Using equations (2.3) and (2.4) it is possible to calculate the expected proportion  $z$  of individuals ultimately infected by the above simulations and, using (2.2), to calculate the probability  $p$  of a global epidemic. It is also possible to convert the expected proportion  $z$  of individuals ultimately infected into expected numbers of people infected and houses affected, using the results of Section 2.3. Arranging the results in the same way as the figures gives Table 1 which compares calculated predictions and simulated realisations.

The values in the Table 1 define a global epidemic to be one affecting at least one quarter of the households, that is 75 households or more. The expected proportion infected, people infected and houses affected were calculated from only those simulations which did affect 75 households or more. For the  $R_* = 1$  row this means only 56 and 43 simulations have been used for the locally driven and globally driven epidemics respectively.

The predictions agree well with the simulations for  $R_* = 0.7, 1.3$  or  $1.6$ . However there are problems with  $R_* = 1$  and  $1.15$ . This is because there can be a difficulty defining a global epidemic for a finite population. For large values of  $R_*$  the distribution of final size is bimodal with a large gap between the two modes, but for smaller  $R_* > 1$ , there is an overlap between the two parts of the distribution. Thus an arbitrary value had

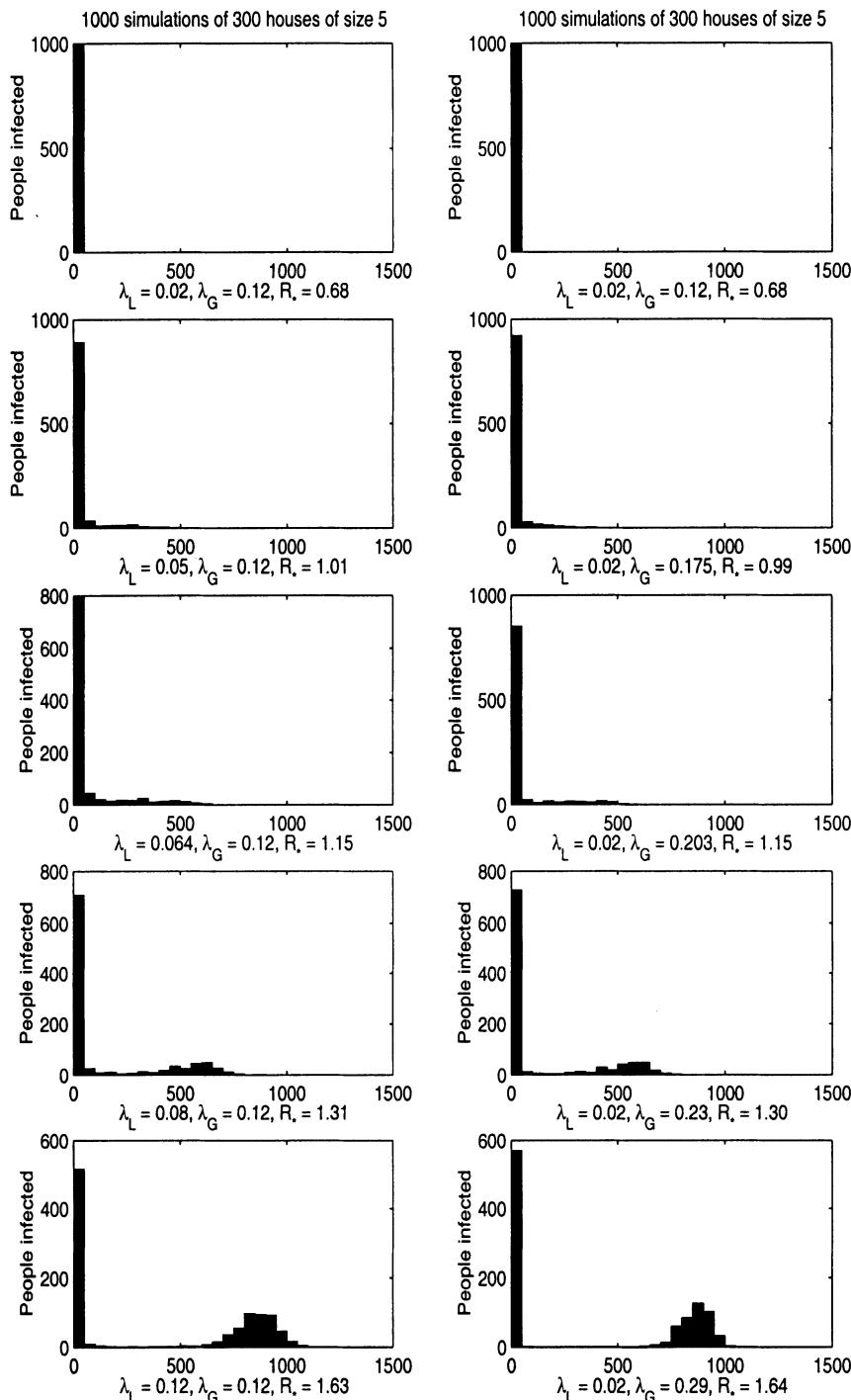


FIG. 1. Number of people infected in each set of 1000 simulations,  $T_I \sim \text{gamma}(2, 2.05)$ , population consisting of 300 households of size 5.

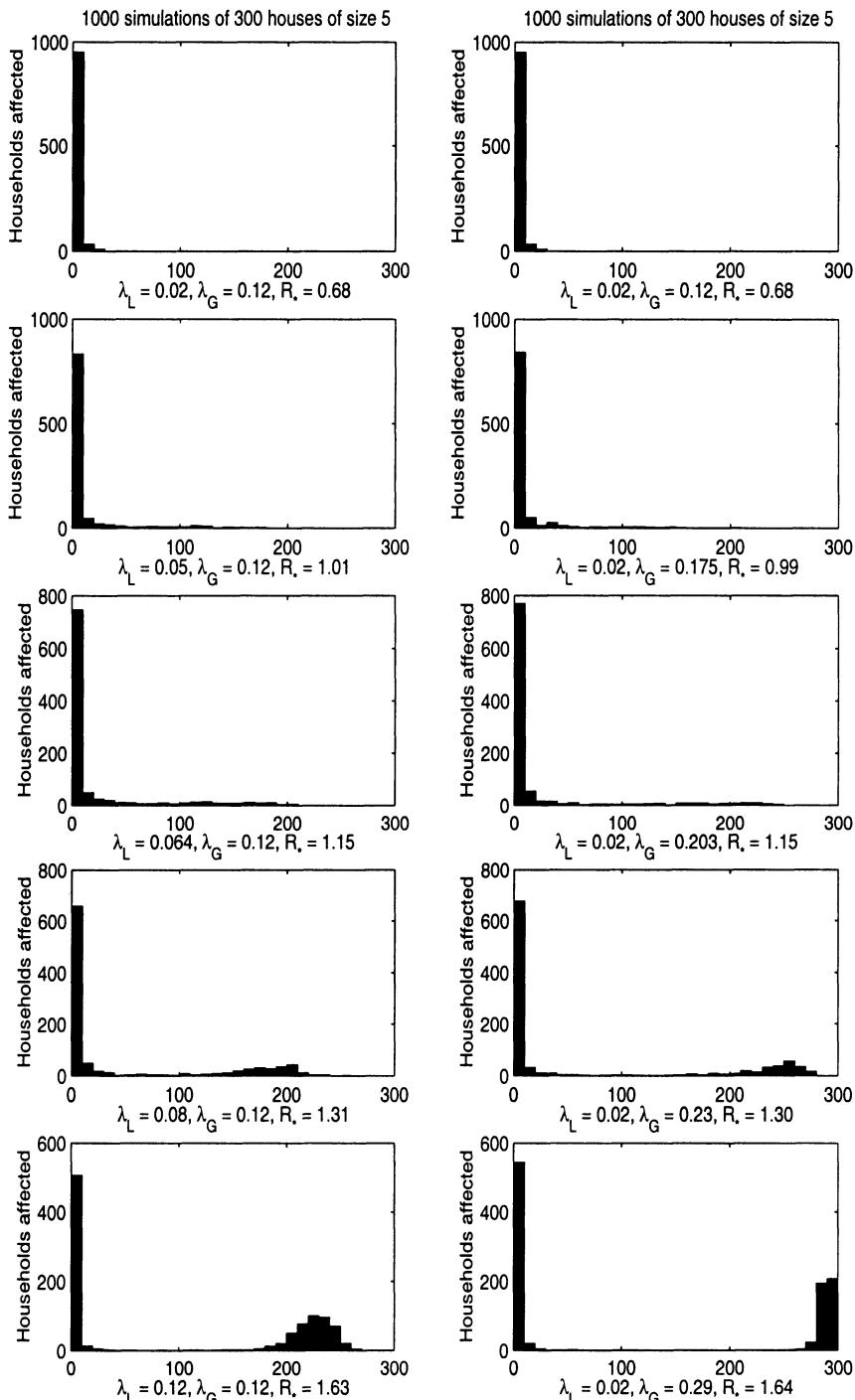


FIG. 2. Number of households affected in each set of 1000 simulations,  $T_1 \sim \text{gamma}(2, 2.05)$ , population consisting of 300 households of size 5.

TABLE 1

*Table of calculated predictions and simulated realisations of households SIR epidemics.*

Calculated	Locally driven				Globally driven				
	$R_*$	$p$	$z$	People	Houses	$p$	$z$	People	Houses
0.70	1	0	0	0	0	1	0	0	0
1.00	0.99	0.01	12	6	6	1	0	0	0
1.15	0.86	0.20	294	115	115	0.86	0.20	302	170
1.30	0.74	0.35	523	173	173	0.75	0.35	531	243
1.60	0.57	0.58	860	227	227	0.58	0.58	865	290
Simulated	Locally driven				Globally driven				
$R_*$	$p$	$z$	People	Houses	$p$	$z$	People	Houses	
0.70	1	0	0	0	0	1	0	0	0
1.00	0.94	0.19	282	119	119	0.96	0.13	200	120
1.15	0.87	0.25	374	138	138	0.88	0.22	327	174
1.30	0.75	0.36	539	175	175	0.74	0.34	513	232
1.60	0.53	0.57	851	224	224	0.57	0.57	862	290

to be chosen to define a global epidemic to produce the simulation results given in Table 1. Simulations on large populations agree very well with the calculated values, but it is clear that the asymptotics cannot exactly hold for smaller populations. There is also a problem for the asymptotics of the number of households affected, which can be seen from the bottom right graph of Figure 2. Here the globally driven epidemic affects most of the households, so the distribution is truncated from above by 300, the number of households.

It is possible to prove asymptotic normality for all these calculated quantities and to calculate the associated variances (c.f Ball et al. (1997)). These agree well with simulations except for the above-mentioned truncation from above or below.

### 3. SIS epidemics among a community of households.

**3.1. Model and threshold behaviour.** The SIR model of Section 2.1 can be turned into an SIS model by assuming that at the end of its infectious period an infective individual becomes susceptible again. The assumption that there is no latent period is again made; this is necessary for some of the calculations carried out in this section. Suppose also that successive infectious periods of the same individual are independent, each distributed according to  $T_I$ . This gives a special case of a model studied in Ball (1999), which allowed for successive infectious periods of the same individual to be both dependent and non-identically distributed and also incorporated a more general infecting mechanism.

The threshold behaviour of the households SIS epidemic can be determined using similar arguments to those used for the SIR epidemic in

Section 2.2, the only difference being that  $T_n$  and  $A_n$  are now the total size (counting successive infections of the same individual as separate cases) and severity of a single household SIS epidemic with initially 1 infective and  $n - 1$  susceptibles.

In order to give an explicit expression for the threshold parameter  $R_*$  and to determine the probability of a global epidemic, it is assumed further that the infectious period  $T_I$  follows a negative exponential with mean  $\gamma^{-1}$ , written  $T_I \sim \text{NE}(\gamma)$ , and hence has probability density function  $f_{T_I}(t) = \gamma e^{-\gamma t}$  ( $t > 0$ ). Thus, during the early stages of an epidemic, when global contacts are ignored, the spread of infection within a typical household of size  $n$  follows the standard Markov single population stochastic SIS epidemic (see, for example, Kryscio and Lefèvre (1989)) with infection rate  $\lambda_L$ , removal rate  $\gamma$ , 1 initial infective and  $n - 1$  initial susceptibles. Thus to determine  $R_*$  (using (2.1)), it is necessary to compute  $\mathbb{E}(A_n)$  (i.e. the mean severity) for such an epidemic. This can be done in several ways. The following method uses a direct probabilistic argument which gives an interpretation for the different terms appearing in the resulting formula.

Consider the above Markov SIS single household epidemic and for  $t \geq 0$ , let  $Y_n(t)$  denote the number of infectives present at time  $t$ , so  $Y_n(0) = 1$ . For  $k = 1, 2, \dots, n$ , let  $N_k$  and  $N^{(k)}$  denote the total number of sojourns of  $\{Y_n(t); t \geq 0\}$  in state  $k$  and the set of states  $\{k, k+1, \dots, n\}$ , respectively. Thus  $N_k$  and  $N^{(k)}$  are respectively the total number of distinct periods that there are  $k$  infectives and at least  $k$  infectives in the SIS epidemic.

Consider the initial sojourn of the epidemic in which there are 1 infective and  $n - 1$  susceptibles. During that period infections occur at rate  $(n - 1)\lambda_L$  and the infective is being removed at rate  $\gamma$ . Thus the probability that the initial sojourn ends with the removal of the infective is given by  $p_1 = \gamma/(\gamma + (n - 1)\lambda_L)$ . If the initial sojourn in state 1 ends with an infection then  $Y_n(t)$  will eventually return to 1 again and, by the Markov property, the second sojourn in state 1 independently has probability  $p_1$  of ending with a removal. Continuing like this one obtains that  $\mathbb{P}(N_1 = i) = (1 - p_1)^{i-1} p_1$  ( $i = 1, 2, \dots$ ), so  $N_1$  has a geometric distribution with parameter  $p_1$ , written  $N_1 \sim \text{geom}(p_1)$ . Also, since two successive sojourns of  $\{Y_n(t); t \geq 0\}$  at level 1 are separated by a single sojourn at level  $\geq 2$ ,  $N^{(2)} = N_1 - 1$ .

Now, for fixed  $k \in \{2, 3, \dots, n - 1\}$ ,  $\{Y_n(t); t \geq 0\}$  has  $N^{(k)}$  distinct sojourns at level  $\geq k$ . Consider the  $i$ th such sojourn. It will start with a period in which there are  $k$  infectives and  $n - k$  susceptibles, during which infections and removals occur at rates  $k(n - k)\lambda_L$  and  $k\gamma$ , respectively. Thus the probability that that sojourn at level  $k$  ends with a removal (and thus also ends the sojourn at level  $\geq k$ ) is given by  $p_k = \gamma/(\gamma + (n - k)\lambda_L)$ . Hence, arguing as before, if  $N_{ki}$  denotes the number of sojourns

at level  $k$  within the  $i$ th sojourn at level  $\geq k$ , then  $N_{ki} \sim \text{geom}(p_k)$  ( $i = 1, 2, \dots, N^{(k)}$ ) independently. Further,

$$(3.1) \quad N_k = \sum_{i=1}^{N^{(k)}} N_{ki} \quad (k = 1, 2, \dots, n)$$

and

$$(3.2) \quad N^{(k+1)} = \sum_{i=1}^{N^{(k)}} (N_{ki} - 1) \quad (k = 1, 2, \dots, n-1).$$

(Note that it has previously been shown that (3.1) and (3.2) hold for  $k = 1$ , since  $N^{(1)} \equiv 1$ . Also,  $N_n = N^{(n)}$ , so (3.1) holds for  $k = n$  since  $N_{ni} \equiv 1$  as  $p_n = 1$ .)

Now,  $\mathbb{E}(N_{ki}) = p_k^{-1}$  ( $k = 1, 2, \dots, n; i = 1, 2, \dots, N^{(k)}$ ), so taking expectations of (3.1) and (3.2) yields  $\mathbb{E}(N_k) = \mathbb{E}(N^{(k)})/p_k$  ( $k = 1, 2, \dots, n$ ) and  $\mathbb{E}(N^{(k+1)}) = \frac{1-p_k}{p_k} \mathbb{E}(N^{(k)})$ , from which it follows that

$$(3.3) \quad \mathbb{E}(N^{(k)}) = \prod_{i=1}^{k-1} (1 - p_i)/p_i \quad (k = 1, 2, \dots, n)$$

and

$$\mathbb{E}(N_k) = \frac{1}{p_k} \prod_{i=1}^{k-1} (1 - p_i)/p_i \quad (k = 1, 2, \dots, n),$$

where the products are 1 if vacuous.

For  $k = 1, 2, \dots, n$ , let  $A_n^{(k)} = \int_0^\infty Y_n(t) 1_{\{Y_n(t)=k\}} dt$  (where  $1_{\{Y_n(t)=k\}}$  is the indicator function of the event  $\{Y_n(t) = k\}$ ) be the contribution to the severity of  $\{Y_n(t); t \geq 0\}$  coming from periods when there are  $k$  infectives present. Let  $S_k$  denote the length of a typical sojourn of  $\{Y_n(t); t \geq 0\}$  at level  $k$ . Then  $S_k \sim \text{NE}(k\gamma + k(n-k)\lambda_L)$ . This sojourn contributes  $kS_k \sim \text{NE}(\gamma + (n-k)\lambda_L)$  to  $A_n^{(k)}$ , so the mean contribution of such a sojourn to  $A_n^{(k)}$  is  $(\gamma + (n-k)\lambda_L)^{-1} = \gamma^{-1}p_k$ . Hence  $\mathbb{E}(A_n^{(k)}) = \gamma^{-1}p_k \mathbb{E}(N_k) = \gamma^{-1}\mathbb{E}(N^{(k)})$  ( $k = 1, 2, \dots, n$ ). Now  $A_n = \sum_{k=1}^n A_n^{(k)}$  and, since each removal in  $\{Y_n(t); t \geq 0\}$  corresponds to the end of a sojourn at level  $\geq k$ , for some  $k$ , the total number of removals in  $\{Y_n(t); t \geq 0\}$  is given by  $T_n = \sum_{k=1}^n N^{(k)}$ . Thus  $\mathbb{E}(A_n) = \gamma^{-1}\mathbb{E}(T_n)$  and the Wald's identity also holds for the single household SIS epidemic. Further, using (3.3) and the definition of  $p_k$  ( $k = 1, 2, \dots, n$ ),

$$(3.4) \quad \mu_n = \mathbb{E}(T_n) = \frac{(n-1)!}{\rho^{n-1}} \sum_{k=0}^{n-1} \frac{\rho^k}{k!}$$

where  $\rho = \gamma/\lambda_L$  denotes the relative removal-rate. The threshold parameter  $R_*$  follows using (2.1) with  $\mathbb{E}(T_I) = \gamma^{-1}$  and  $\mu_n$  as above. Note that the  $k$ th term in (3.4) is the mean number of removals occurring when there are  $n - k$  infectives present in the SIS epidemic.

In order to calculate the probability of a global epidemic, an expression for  $\phi_{n-1,1}(\theta)$  (see (2.2)) for the SIS epidemic is required. Suppressing the explicit dependence on  $n$ , for  $k = 1, 2, \dots, n$ , let  $A_{[k]}$  denote the contribution to the severity  $A_n$  arising from a typical single sojourn of  $\{Y_n(t); t \geq 0\}$  at level  $\geq k$  and  $\phi_{[k]}(\theta) = \mathbb{E}(e^{-\theta A_{[k]}})$  be the moment generating function of  $A_{[k]}$ . Thus  $A_n = A_{[1]}$  and  $\phi_{n-1,1}(\theta) = \phi_{[1]}(\theta)$ . For  $k = 1, 2, \dots, n-1$  a typical sojourn of  $\{Y_n(t); t \geq 0\}$  at level  $\geq k$  comprises  $N_{k1} \sim \text{geom}(p_k)$  sojourns at level  $k$ , contributing  $U_1, U_2, \dots, U_{N_{k1}}$  to  $A_{[k]}$ , where  $U_i \sim \text{NE}(\gamma p_k^{-1})$ , and  $N_{k1} - 1$  sojourns at level  $\geq k+1$ , contributing  $V_1, V_2, \dots, V_{N_{k1}-1}$  to  $A_{[k]}$ , where  $V_i \sim A_{[k+1]}$ . Moreover, the Markov nature of  $\{Y_n(t); t \geq 0\}$  implies that  $N_{k1}, U_1, U_2, \dots, U_{N_{k1}}$  and  $V_1, V_2, \dots, V_{N_{k1}-1}$  are mutually independent. Now clearly  $\mathbb{E}(e^{-\theta U_i}) = \gamma p_k^{-1}/(\gamma p_k^{-1} + \theta)$ , so

$$(3.5) \quad \begin{aligned} \phi_{[k]}(\theta) &= \mathbb{E}[\exp(-\theta A_{[k]})] \\ &= \mathbb{E}\left[\exp\left\{-\theta\left(\sum_{i=1}^{N_{k1}} U_i + \sum_{i=1}^{N_{k1}-1} V_i\right)\right\}\right] \\ &= \sum_{j=1}^{\infty} (1-p_k)^{j-1} p_k \left(\frac{\gamma p_k^{-1}}{\gamma p_k^{-1} + \theta}\right)^j \phi_{[k+1]}(\theta)^{j-1} \\ (3.6) \quad &= \frac{\gamma}{\gamma + \theta + (n-k)\lambda_L(1 - \phi_{[k+1]}(\theta))} \quad (k = 1, 2, \dots, n-1), \end{aligned}$$

where the second sum in (3.5) is zero if vacuous. Finally, note that, since  $N_{n1} = 1$  and  $p_n = 1$ ,  $A_{[n]} = U_1$ , where  $U_1 \sim \text{NE}(\gamma^{-1})$ , so

$$(3.7) \quad \phi_{[n]}(\theta) = \frac{\gamma}{\gamma + \theta}.$$

Thus  $\phi_{n-1,1}(\theta) = \phi_{[1]}(\theta)$  can be computed recursively using (3.6) and (3.7), hence enabling the probability of a global epidemic to be computed. Note that if  $Y_n(0) = a > 1$  then the trajectory of  $\{Y_n(t); t \geq 0\}$  can be decomposed into successive sojourns at levels  $\geq a, \geq (a-1), \dots, \geq 1$ , making independent contributions to the severity of the epidemic. Thus, in the notation of Section 2.2,  $\phi_{n-a,a} = \prod_{i=1}^a \phi_{[i]}(\theta)$ . This is required to calculate the probability of a global epidemic when some households have more than one initial infective.

**3.2. Discussion of threshold parameters.** In this subsection, the threshold parameter  $R_*$  for the households SIR and SIS epidemic models when  $T_I \sim \text{NE}(\gamma)$  is discussed, where, without loss of generality it is assumed that  $\gamma = 1$ . For ease of exposition suppose that all the households

are of the same size  $n$  and the local infection rate is reparamaterised so that  $\lambda_L = n^{-1}\lambda'_L$ .

Figure 3 shows graphs of critical values of  $(\lambda'_L, \lambda_G)$  so that  $R_* = 1$  for both the SIR and SIS households epidemic for various household sizes  $n$ . First note that when  $n = 1$  the graphs for the SIR and SIS epidemics are the same, since in that case there is no internal spread within a household. However, when  $n > 1$  the SIS graph lies below the corresponding SIR graph since individuals can be infected more than once in the SIS epidemic, leading to greater internal spread within a household than in the corresponding SIR epidemic. As the household size  $n \rightarrow \infty$ , the SIS and SIR graphs approach the same limiting function given by  $\lambda_G = 1 - \lambda'_L$  if  $\lambda'_L < 1$  and  $\lambda_G = 0$  if  $\lambda'_L \geq 1$ . This is because as  $n \rightarrow \infty$  the processes of infectives in the early stages of single household SIR and SIS epidemics both converge to a birth-and-death process with birth and death rates  $\lambda'_L$  and 1, respectively (see Ball and Donnelly (1995)), having mean severity  $(1 - \lambda'_L)^{-1}$  if  $\lambda'_L < 1$  and  $\infty$  if  $\lambda'_L \geq 1$  (see Ball et al. (1997)). Note that in both the SIR and SIS epidemics, a typical infective makes local (global) contacts at rate  $\lambda'_L$  ( $\lambda_G$ ) throughout an infectious period having mean 1. Thus the reproductive ratio  $R_0 = \lambda'_L + \lambda_G$  for both epidemics. Note that  $R_0 = 1$  corresponds to the limit of  $R_* = 1$  as  $n \rightarrow \infty$ , so for small finite  $n$  the reproductive ratio does not give a good indicator as to whether global epidemics can occur. This point is returned to in Section 4. Note also the qualitative change in behaviour as  $\lambda'_L$  passes through its single household threshold of 1. Provided  $\lambda_G > 0$ , if  $\lambda'_L \geq 1$  then global epidemics can occur provided the household size  $n$  is sufficiently large, whilst if  $\lambda'_L < 1$ , global epidemics can only occur for sufficiently large  $n$  if  $\lambda_G > 1 - \lambda'_L$ . This is discussed further in Ball et al. (1997) and Ball (1999) for SIR and SIS epidemics, respectively. Finally, note the different limiting behaviour as  $\lambda'_L \rightarrow \infty$  in the two models when  $n \geq 2$ . As  $\lambda'_L \rightarrow \infty$ , the SIR model converges to the model of Becker and Dietz (1995) for highly infectious diseases, in which the whole household becomes infected once one individual in it is infected. Thus, as  $\lambda'_L \rightarrow \infty$ ,  $\mu_n \rightarrow n$  and, since  $\mathbb{E}(T_I) = 1$ , the critical value of  $\lambda_G$  tends to  $n^{-1}$ . However, in the SIS model, if  $\lambda'_L \rightarrow \infty$  then, for households of size  $n \geq 2$ , as soon as an infective is removed it is immediately re-infected. Thus  $\mu_n \rightarrow \infty$  as  $\lambda'_L \rightarrow \infty$  and consequently the critical value of  $\lambda_G$  tends to 0. Hence, for large values of  $\lambda'_L$ , the critical value of  $\lambda_G$  for the SIS epidemic can be several orders of magnitude smaller than that for the SIR epidemic.

**4. Meta-population models.** The models of the previous two sections have all been stochastic. In this section deterministic models of epidemics among households are briefly discussed. Consider first the Markov SIS households model of Section 3.1 and, for ease of exposition, suppose that the population comprises  $m$  households, labelled  $1, 2, \dots, m$ , each of size  $n$ . The classical deterministic model for this epidemic is given by

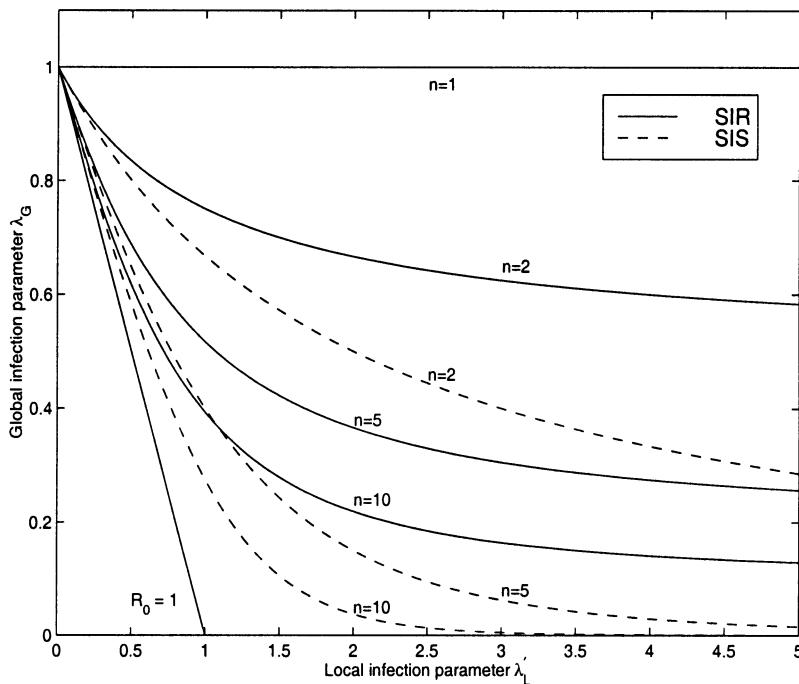


FIG. 3. Critical values of  $(\lambda'_L, \lambda_G)$  so that  $R_* = 1$  for SIR (solid line) and SIS (dashed line) epidemics.

$$(4.1) \quad \frac{dy_i}{dt} = \left( \lambda_L y_i + N^{-1} \lambda_G \sum_{j=1}^m y_j \right) (n - y_i) - \gamma y_i \quad (i = 1, 2, \dots, m),$$

where  $y_i(t)$  is the number of infectives in household  $i$  at time  $t$ . However, this model is generally inappropriate for the present households setting since, unless  $\lambda_L = 0$ , it only necessarily provides a good approximation to the more realistic stochastic model if the household size  $n$  is large. This is nicely illustrated by considering the ability of the reproductive ratio  $R_0$  to predict whether or not global epidemics can occur. Note that for the model (4.1),  $R_0 = (n\lambda_L + \lambda_G)/\gamma$  ( $= (\lambda'_L + \lambda_G)/\gamma$  in the notation of Section 3.2). If  $\lambda_L = 0$  then only global contacts occur and consequently the household structure becomes irrelevant. The model reduces to a homogeneously mixing one, so the population is locally, as well as globally, large. Further, it is easily checked that  $R_0 = R_*$ , so global epidemics can only occur if  $R_0 > 1$ . However, if  $\lambda_L > 0$  (so implicitly  $n \geq 2$ , since if  $n = 1$  there can be no local spread of infection) then the household structure is relevant and the population is not locally large. Further,  $R_0 \neq R_*$  and consequently  $R_0$  may not be a good indicator as to whether global epidemics can occur. Indeed, Figure 3 shows that  $R_0$  is a poor indicator, and hence the model (4.1) is inappropriate, unless either  $n$  is large (so the population is locally large)

or  $\lambda_L$  is small (so the population is approximately homogeneously mixing and hence also locally large).

An appropriate deterministic model can be derived by adopting a meta-population approach. For  $k = 0, 1, \dots, n$  and  $t \geq 0$ , let  $x_k(t)$  denote the proportion of households which have  $k$  infectives at time  $t$ . Then,

$$(4.2) \quad \frac{dx_k}{dt} = f_k(x_0, x_1, \dots, x_n) \quad (k = 0, 1, \dots, n),$$

where

$$\begin{aligned} f_0(x_0, x_1, \dots, x_n) &= -\frac{\lambda_G}{n} \sum_{i=1}^n ix_i + \gamma x_1 \\ f_k(x_0, x_1, \dots, x_n) &= -\lambda_L k(n-k)x_k - \frac{\lambda_G}{n}(n-k)x_k \sum_{i=1}^n ix_i - \gamma kx_k \\ &\quad + \lambda_L(k-1)(n-k+1)x_{k-1} \\ &\quad + \frac{\lambda_G}{n}(n-k+1)x_{k-1} \sum_{i=1}^n ix_i + \gamma(k+1)x_{k+1} \\ &\quad (k = 1, 2, \dots, n-1) \\ f_n(x_0, x_1, \dots, x_n) &= \lambda_L(n-1)x_{n-1} + \frac{\lambda_G}{n}x_{n-1} \sum_{i=1}^n ix_i - n\gamma x_n. \end{aligned}$$

See Ball (1999) for further details, where it is shown that the deterministic model (4.2) arises as the limit of the corresponding stochastic model as the number of households  $m \rightarrow \infty$ . The successive terms in the expression for  $f_k(x_0, x_1, \dots, x_n)$  given above correspond to the rate of change in the proportion of households with  $k$  infectives due to local infection within a household of size  $k$ , global infection of a susceptible in a household of size  $k$ , removal of an infective in a household of size  $k$  (all of which decrease  $x_k$ ), local infection within a household of size  $k-1$ , global infection of a susceptible in a household of size  $k-1$  and removal of an infective in a household of size  $k+1$  (all of which increase  $x_k$ ). The divisor  $n$ , rather than  $N$ , for  $\lambda_G$  arises from the fact that  $x_k(t)$  describes the evolution of the proportion, rather than the number, of households with  $k$  infectives.

The model (4.2) when  $n = 2$  is analysed in Ball (1999). The disease-free solution  $(1, 0, 0)$  is always an equilibrium point of the system of differential equations (4.2). If  $R_* \leq 1$  then it is the only equilibrium point and  $\Delta = \{(x_0, x_1, x_2) : \sum_{i=0}^2 x_i = 1, x_i \geq 0 (i = 0, 1, 2)\}$  is an ASR (asymptotic stability region) for this equilibrium point. If  $R_* > 1$  then there is a unique second (endemic) equilibrium point in  $\Delta$ , for which  $\Delta \setminus \{(1, 0, 0)\}$  is an ASR. Note that these results are qualitatively the same as those for the corresponding classical deterministic model (4.1) (Lajmanovich and Yorke (1976)), although the threshold parameter and proportion of individuals

infected at the endemic equilibrium are different. It does not seem straightforward to extend the above results to populations containing households of size  $n > 2$ , although it seems likely that similar results still hold. However, the local stability of the disease-free equilibrium point can be determined since linearising (4.2) about the disease-free point gives a system of differential equations which govern the mean size of a branching process having threshold parameter  $R_*$ .

It is also possible to write down an SIR equivalent of the meta-population model (4.2), although the details are more complicated since households now have to be classified according to both the numbers of infectives and susceptibles that they contain. A similar linearisation argument to that mentioned above shows that  $R_*$  is a threshold parameter for this deterministic model, whose final outcome when there is a trace of initial infection is described by results outlined in Section 2.3. The complexity of the expressions for  $R_*$  and the final outcome of the epidemic suggests that these results may be difficult to obtain directly from the deterministic model.

Note that although the deterministic and stochastic models have the same threshold parameter  $R_*$ , the interpretation of the threshold behaviour in the two models is quite different. For epidemics initiated by a trace of infection, in the deterministic formulation global epidemics never occur if  $R_* \leq 1$  and always occur if  $R_* > 1$ , whilst in the stochastic formulation, with sufficiently many households, global epidemics never occur if  $R_* \leq 1$  and occur with probability lying in  $(0, 1)$  if  $R_* > 1$ .

## 5. Vaccination strategies.

**5.1. Introduction.** One of the main purposes of constructing epidemic models is to use them to evaluate control measures, such as vaccination strategies. If all the households are of size 1 then our models reduce to standard homogeneously mixing models and the threshold parameter  $R_*$  reduces to the usual reproductive ratio  $R_0$ . For such models, if a proportion  $v$  of individuals are vaccinated and the vaccine is fully effective, in that it necessarily confers immunity, then  $R_0$  is reduced to  $R_{0v} = (1 - v)R_0$ , since a proportion  $v$  of contacts are with vaccinated individuals and hence do not result in the spread of infection. Thus, if  $R_0 > 1$ , in order to prevent global epidemics it is necessary to vaccinate a critical proportion  $v^*$  of individuals, given by

$$(5.1) \quad v^* = 1 - R_0^{-1}.$$

This formula is well known and goes back at least to Smith (1964).

The situation for epidemics among a community of households is less clear since (a) the models are more complicated and (b) there are now different ways of allocating vaccines to individuals in the population. For example, one could vaccinate all individuals in a proportion  $v$  of the households or one could vaccinate a proportion  $v$  of individuals chosen at random from the population. Thus the question of optimal vaccination strategies

now arises, which is considered for fully effective and partially effective vaccines in Sections 5.2 and 5.3, respectively. Comparison of vaccination strategies for households and homogeneously mixing models is studied in Section 5.4.

**5.2. Fully effective vaccines.** For both the SIR and Markov SIS models, analysed in Sections 2 and 3, respectively, the threshold parameter  $R_*$  takes the form

$$\begin{aligned} R_* &= \lambda_G \mathbb{E}(T_I) \sum_{n=1}^{\infty} \tilde{\alpha}_n \mu_n = \lambda_G \mathbb{E}(T_I) \sum_{n=1}^{\infty} (nm_n/N) \mu_n \\ &= N^{-1} \lambda_G \mathbb{E}(T_I) \sum_{k=1}^m n_k \mu_{n_k}, \end{aligned}$$

where the  $m$  households have been labelled  $1, 2, \dots, m$  and have respective sizes  $n_1, n_2, \dots, n_m$ . Suppose that a vaccination policy reduces  $n_k$  to  $n'_k$  ( $k = 1, 2, \dots, m$ ), so that the total number of individuals vaccinated is  $N_v = \sum_{k=1}^m (n_k - n'_k)$ , then  $R_*$  is reduced to

$$(5.2) \quad R_{*v} = N^{-1} \lambda_G \mathbb{E}(T_I) \sum_{k=1}^m n'_k \mu_{n'_k}.$$

Suppose that the sequence  $(n\mu_n)$  is convex, i.e. that  $D_{n+1} \geq D_n$  ( $n = 0, 1, \dots$ ), where  $D_n = (n+1)\mu_{n+1} - n\mu_n$  with  $\mu_0 = 0$ . Then it is easily seen (c.f. Ball et al. (1997)) that for fixed  $N_v$  the vaccination strategy that yields the greatest (least) reduction in  $R_*$  is one in which the  $N_v$  vaccines are allocated sequentially to individuals, always to an individual in a household containing the greatest (least) number of unvaccinated individuals. For populations comprised of households with equal size, the “best” (“worst”) strategy allocates the vaccines as equally (unequally) as possible among the households. However, as noted by Becker and Utev (1998), the worst strategy is likely to be closer to what happens in practice since parents are likely to decide to vaccinate either all or none of their children.

The above discussion rests on the assumption that the sequence  $(n\mu_n)$  is convex. Ball et al. (1997) conjecture that  $(n\mu_n)$  is convex for the SIR model of Section 2.1 and provide some numerical calculations which support that conjecture. They also note that  $(n\mu_n)$  is convex for the extension of the highly infectious disease model of Becker and Dietz (1995), in which if one person in a household is infected either the whole household is infected, with probability  $p$  say, or no-one else is infected with probability  $1 - p$ , so  $\mu_n = 1 - p + pn$ . It is easily verified, using (3.4) and a little algebra, that  $(n\mu_n)$  is convex for the Markov SIS model of Section 3.1.

Figure 4 shows the effect of four vaccination schemes on  $R_*$  for the SIR households epidemic model. The population used for this example consists of 1800 households, 300 of each size from 1 to 6 (so that 1/6 of

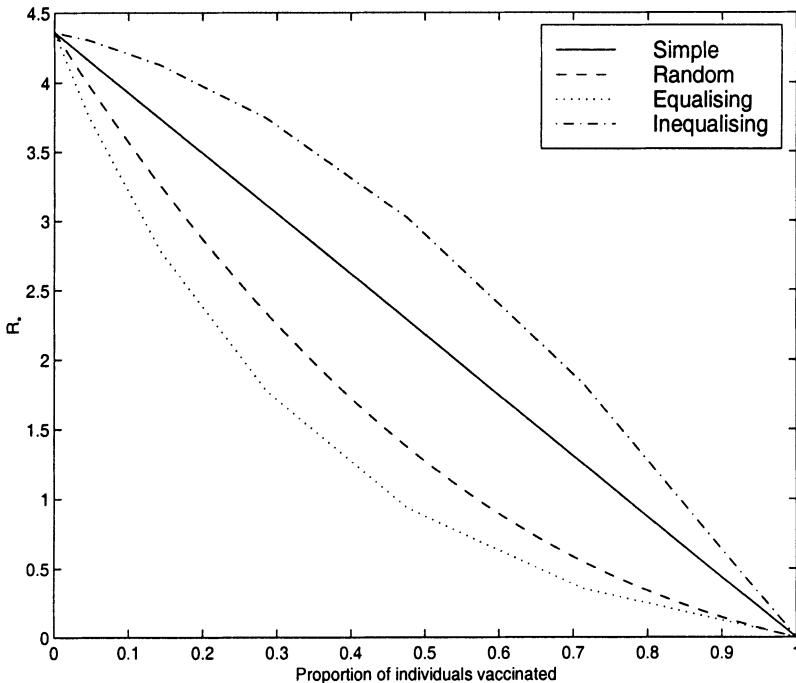


FIG. 4. Reduction of  $R_*$  by four vaccination strategies,  $T_I \sim \text{gamma}(2, 2.05)$  and  $(\lambda_L, \lambda_G) = (0.2, 0.3)$ , where 1/6 of the households are of each size from 1 to 6.

the households are of each size, it is in fact only these proportions that matter). The infectious period used was  $T_I \sim \text{gamma}(2, 2.05)$  and the parameters  $(\lambda_L, \lambda_G) = (0.2, 0.3)$ . The effect of each vaccination strategy was calculated using (5.2), except for the *Random* strategy which was computed using (5.3) below. The *Simple* scheme achieves pro-rata reduction in  $R_*$  by vaccinating entire households and selecting the same number of households of each size. The *Random* strategy selects individuals independently at random, taking no account of their household. The *Equalising* strategy (Ball et al. (1997)) concentrates the vaccinations in the largest households and the *Inequalising* strategy in the smallest. The choice of strategy clearly makes an enormous difference to the effectiveness of the vaccination campaign, whether measured horizontally on the graph, in terms of number of vaccines required to achieve a given  $R_*$ , or vertically, as  $R_*$  achieved for a given amount of vaccine. While different parameter values and household structures change the magnitude of the differences between the strategies (as  $\lambda_L \downarrow 0$  the model becomes homogeneous mixing and all strategies achieve pro-rata reduction in  $R_*$ ), the form of the curves and ordering of the four strategies remain essentially the same.

**5.3. Partially effective vaccines.** Suppose now that the vaccine is only partially effective, in that vaccinated individuals are rendered immune

independently with probability  $\epsilon$ . This situation is studied by Becker and Starczak (1997), whose approach is now briefly outlined.

For  $n = 1, 2, \dots$  and  $v = 0, 1, \dots, n$  let  $x_{nv}$  denote the proportion of households of size  $n$  that have had  $v$  members vaccinated. The probability that a global contact is with an individual residing in a household of size  $n$  having  $v$  members vaccinated is  $\bar{\alpha}_n x_{nv}$ . For  $k = n-v, n-v+1, \dots, n$ , such a household has  $k$  susceptibles if  $n-k$  of the vaccinations are successful, which happens with probability  $\binom{v}{n-k} \epsilon^{n-k} (1-\epsilon)^{v-n+k}$ , and given that there are  $k$  susceptibles, the conditional probability that a global contact with such a household is with a susceptible (and thus triggers a local household epidemic) is  $k/n$ . Hence  $R_*$  is reduced to

$$(5.3) \quad R_{*v} = \lambda_G \mathbb{E}(T_I) \sum_{n=1}^{\infty} \bar{\alpha}_n \sum_{v=0}^n x_{nv} \sum_{k=n-v}^n \binom{v}{n-k} \epsilon^{n-k} (1-\epsilon)^{v-n+k} \frac{k}{n} \mu_k.$$

The proportion of individuals vaccinated by the above policy,  $C_v$  say, equals the probability that a randomly chosen individual from the population is vaccinated, and hence is given by

$$(5.4) \quad C_v = \sum_{n=1}^{\infty} \bar{\alpha}_n \sum_{v=0}^n x_{nv} \frac{v}{n}.$$

Thus determination of the allocations of vaccines which (a) minimises  $R_{*v}$  for a given vaccine coverage  $C_v$  and (b) minimises  $C_v$  so that  $R_{*v} = 1$  are both linear programming problems, which can be solved numerically using standard computer software.

Becker and Starczak (1997) note that the above linear programming formulation does not provide insight into the form of the optimal allocation policies. They consider the approximation  $\mu_n = 1 - p + pn$  for SIR epidemics, referred to in Section 5.2, and show that for given vaccine coverage  $C_v$ ,  $R_{*v}$  is minimised by making  $\text{var}(N - \epsilon V)$  as small as possible, where  $(N, V)$  are the numbers of members and vaccinated members in a randomly selected household. They show further that  $\text{var}(N - \epsilon V) = 0$  if  $V = \mathbb{E}(V) + (N - \mathbb{E}(N))/\epsilon$  for every household (thus recovering the equalising strategy when  $\epsilon = 1$ ), although they note that this is unlikely to be achievable in practice since (i) the coverage  $C_v$  may be too low and (ii)  $V$  must be integer for every  $N$ . This latter point is also pertinent to the linear programming formulation (5.3) and (5.4), since for finite populations the possible values of the  $x_{nv}$ 's are constrained to a finite set, and strictly speaking an integer programming approach is required.

Note that if  $\epsilon = 1$  and  $x_{nv} = 0$  ( $v \neq n$ ) and  $x_{nn} = i$  ( $n = 1, 2, \dots$ ), so the vaccine is fully effective and a fixed proportion  $i$  of households of each size are completely immunised, then  $R_{*i} = (1-i)R_*$  and the critical coverage level is given by the classical formula (5.1), with  $R_0$  replaced by  $R_*$ .

This was pointed out by Becker and Dietz (1995), who also showed that, for highly infectious SIR epidemics, (5.1) holds when  $\epsilon < 1$  and  $x_{nn} = 1$  ( $n = 1, 2, \dots$ ), so everyone is vaccinated, provided  $R_0$  is replaced by the threshold parameter for the proliferation of infected individuals. However, this latter result does not hold generally when the highly infectious assumption is relaxed. The validity of (5.1) for SIR households models with  $T_I \sim \text{NE}(\gamma)$  is explored further in Becker and Dietz (1996).

**5.4. Comparison with homogeneous mixing epidemics.** In order to make a meaningful comparison between, for example, the critical vaccination coverages predicted by homogeneous mixing and households epidemic models, it is necessary to have a criterion for matching the classes of epidemic model. In practice, the parameter(s) governing an epidemic model are unknown and need to be estimated from various data. Suppose that an epidemic among an unvaccinated community is observed and a proportion  $z$  of individuals are ultimately infected. Then, for a homogeneously mixing population,  $R_0 = \lambda_G \mathbb{E}(T_I)$  can be estimated using (2.5) to yield  $\hat{R}_0 = -\log(1-z)/z$  and, using (5.1), the corresponding critical vaccination coverage level is  $v^* = 1 - \hat{R}_0^{-1}$ . Becker and Utev (1998) propose comparing the critical coverage levels for different epidemic models by examining the effects on the estimated coverage levels when the same data are used to estimate their parameters. The final outcome of the highly infectious disease model of Becker and Dietz (1995) depends on a single parameter,  $\lambda = \lambda_G \mathbb{E}(T_I)$ , which can be estimated from  $z$  using (2.3) and (2.4), thus facilitating a comparison of critical coverage levels (Becker and Utev (1998)). Such a calibration is not straightforward for the SIR households model of Section 2.1, since its final outcome depends on two parameters,  $\lambda = \lambda_G \mathbb{E}(T_I)$  and  $\lambda_L^* = \lambda_L \mathbb{E}(T_I)$ . However, any sensible estimation procedure for these parameters should be consistent, in the sense that these estimates converge to the unknown true values as the number of households in the population tends to infinity. This suggests calibrating the households and homogeneous mixing model by choosing the latter so that it has the same proportion of individuals ultimately infected as the former (c.f. Britton (1998)).

Use of this calibration is shown in Figures 5 and 6. Within each figure the global infection parameter  $\lambda_G$ , the vaccine efficiency  $\epsilon$ , the household structure and infectious period distribution  $T_I$  are kept fixed. For each value of  $\lambda_L$  the expected final proportion infected  $z$  was calculated using the SIR households model. Then the calibration was used to choose the homogeneous mixing model so that it also has an expected proportion  $z$  of individuals ultimately infected. This  $z$  yields  $R_0 = -\log(1-z)/z$  and the corresponding critical vaccination coverage level is  $v^* = \epsilon^{-1}(1 - R_0^{-1})$ . The graph is then plotted by applying  $v^*$  with different strategies. The *Homogeneous* line shows the threshold parameter of the homogeneous mixing epidemic, which is therefore identically equal to 1 until  $v^* > 1$  (that is,

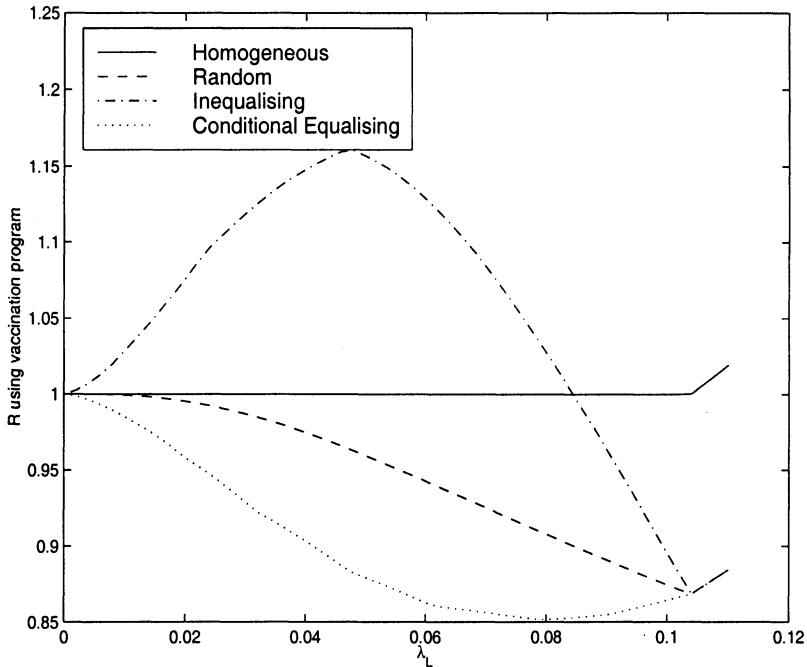


FIG. 5.  $\lambda_G = 0.25, \epsilon = 0.5, T_I \sim \text{gamma}(2, 2.05)$ , where 1/6 of the households are of each size from 1 to 6.

when  $R_0 > (1 - \epsilon)^{-1}$ ). When  $v^*$  passes through 1 it is no longer possible to reduce the homogeneous mixing epidemic below threshold, the graph is continued by assuming everybody is vaccinated. The other curves on the graph are the threshold parameter calculated from (5.3) for the SIR households epidemic model, where the  $v^*$  vaccinations are applied using three possible strategies. The *Random* strategy selects individuals independently at random, taking no account of their household. This may be the only strategy possible if the household information is unknown. In general this strategy does yield a below threshold population, as in Figure 5. However the example of Figure 6 shows that this is not always the case. The household structure used in Figure 6, where 20/21 of the households are of size 1 and 1/21 households are of size 5, was given by Becker and Utev (1998). They point out that the mean household size is 1.19 but the mean household size of a randomly selected individual is 1.80. The *Conditional Equalising* strategy concentrates the vaccinations in the households with the largest expected number of susceptibles, and thus is a natural generalisation of the *Equalising* strategy to partially effective vaccination. It is easily verified that, in the notation of Section 5.3, the *Conditional Equalising* strategy minimises  $\text{var}(N - \epsilon V)$ . Thus the *Conditional Equalising* strategy provides an explicit solution of the optimal allocation problem under the approximation  $\mu_n = 1 - p + pn$  (c.f. Becker and Starczak (1997) and Section 5.3).

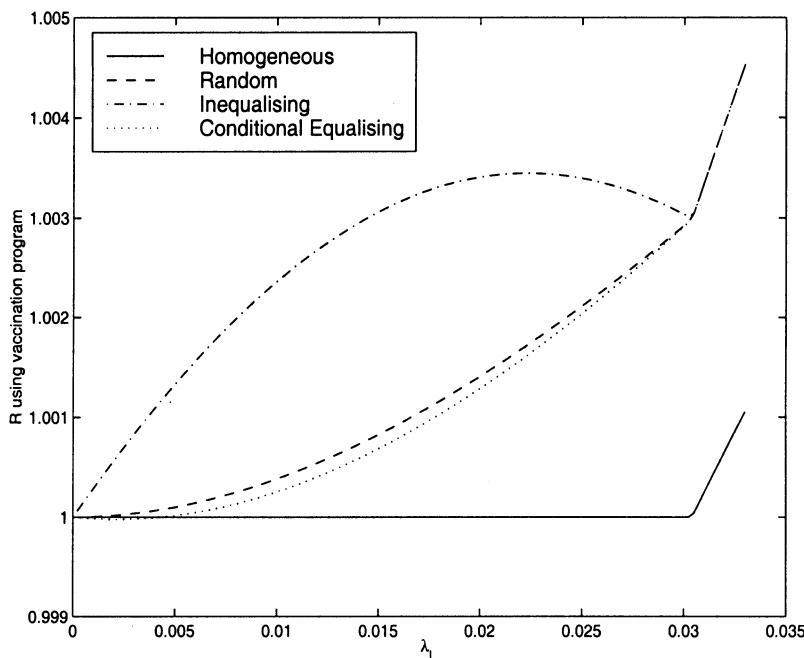


FIG. 6.  $\lambda_G = 1.2$ ,  $\epsilon = 0.8$ ,  $T_I \sim \text{gamma}(2, 2.05)$ , where 20/21 of the households are of size 1 and 1/21 households are of size 5.

For Figures 5 and 6 comparing the *Conditional Equalising* strategy with the optimal strategy (calculated from the linear programming formulation (5.3) and (5.4)) it is found that it is optimal or very close to optimal, to the extent that the optimal curve cannot be distinguished from the conditional equalising curve and is therefore not plotted. Further numerical testing also supports this conclusion, though the strategy is certainly not exactly optimal. As can be seen in both Figures 5 and 6 the *Conditional Equalising* strategy makes more effective use of the vaccine than the *Random* strategy, if the household structure is known. However, even this strategy is insufficient to keep the epidemic below threshold as  $\lambda_L$  increases in Figures 6. The *Inequalising* strategy concentrates the vaccinations in the households with the fewest susceptibles and is the worst possible use of the vaccines. In both Figures 5 and 6 these three strategies initially diverge from each other as  $\lambda_L$  increases from 0 but then converge when  $v^* = 1$ . This is because all strategies are the same for  $v^* = 0$  or 1 but these three differ from each other for  $v^* \in (0, 1)$ .

The most important point to note about the comparison of vaccination strategies using the calibration to homogeneous mixing models is whether or not the households model is below threshold. For Figure 5 the *Random* strategy, which requires no household information to implement, is sufficient to yield a below threshold population. With household information

more efficient strategies can also be used. On the other hand, there are cases, such as in Figure 6, when no strategy which only uses the amount of vaccine obtained from the homogeneous mixing calibration, is sufficient to yield a below threshold population. Thus in either case there is a gain to be made by using households based modelling and strategies — either to make a successful vaccination strategy more efficient, or to warn us our strategy will fail without further vaccination.

**6. Concluding comments.** Threshold parameters  $R_*$  have been determined for models for the spread of SIR and SIS epidemics among a population partitioned into a large number of households. For the stochastic models, the probability that a global epidemic occurs is non-zero if and only if  $R_* > 1$ . If  $R_* > 1$  the probability that a global epidemic occurs and the expected final size of SIR epidemics is determined. These results are illustrated with simulations. Vaccination strategies for households models are considered, for both fully and partially effective vaccines. While it is not possible to obtain a general form for an optimal vaccination policy much can be deduced and any specific allocation problem solved using linear programming techniques. A technique for calibrating the comparison of households based and homogeneously mixing models is given and examples demonstrate both potential inefficiency in applying calculations based on an incorrect homogeneous model and, in some cases, inadequacy of vaccination coverage. Thus it is valuable to take into account household structures when modelling public health problems.

The SIR households model of Section 2 can be extended in several directions, some of which are now briefly outlined. The model can be generalised to multitype epidemics (Becker and Hall (1996), Ball and Lyne (2001)), to incorporate important real-life heterogeneities, such as those owing to age, sex and response to vaccine. The process of infected households is approximated by a multitype branching process (Mode (1971)) and a threshold parameter for the epidemic process is given by the maximal eigenvalue of the mean matrix of the branching process. Other, more complex, global population structures can be included. One example, is a population partitioned into towns, which in turn are partitioned into households, with different infection rates for within-household, within-town and between-town infections, which leads to a multitype epidemic with type corresponding to town. Also, alternative local population structures can be considered. One example, motivated by the spread of infection between pigs in a line of stalls, is the “great-circle” model of Ball et al. (1997), in which individuals are located on a circle, with nearest-neighbour local infection and global infection as before. Other examples include populations that are partitioned in more than one way, such as according to household and school/workplace, (Becker and Dietz (1995), Andersson (1997)) and acquaintancy models (Andersson (1997), Diekmann et al. (1998)). It is also possible to consider models with more detailed individual behaviour (Becker et al. (1995)).

In principle, the SIS households model of Section 3 can be generalised along all of the above lines and associated threshold theorems developed. However, in practice, it is often not possible to explicitly compute the threshold. The extension of the households setting to other models displaying endemic behaviour, such as SIR models with vital dynamics, is an important area for future research. Once again, it is not too difficult to derive an invasion threshold theorem in general terms for such models. However, it is the analysis of their long-term behaviour that represents the real challenge.

Statistical inference for households epidemic models is an area of considerable practical importance. Statistical analysis for epidemic data assuming independent households has received considerable attention in the literature (see Becker (1989)). Some of these analyses incorporate the possibility of outside infection (see, for example, Addy et al. (1991)), though to date little has been done which incorporates dependence between households. For SIR models, hypothesis tests for the presence of enhanced local infection (so the entire population is not homogeneously mixing) have been developed by Britton (1997a-d) and Ball et al. (1997) give a method for estimating local and global infection parameters from final outcome data.

**Acknowledgement.** This research was supported in part by the UK Engineering and Physical Sciences Research Council, under research grant number GR/L56282.

## REFERENCES

- [1] C.L. ADDY, I.M. LONGINI, AND M. HABER, *A generalized stochastic model for the analysis of infectious disease final size data*, Biometrics **47** (1991), pp. 961–974.
- [2] H. ANDERSSON, *Epidemics in a population with social structures*, Math. Biosci. **140** (1997), pp. 79–84.
- [3] J.P. APARICIO, A.F. CAPURRO, AND C. CASTILLO-CHAVEZ, *Transmission and dynamics of tuberculosis on generalized households*, J. Theor. Biol. **206** (2000), pp. 327–341.
- [4] N.T.J. BAILEY, *The Mathematical Theory of Infectious Diseases and its Applications*, 2nd edn. Griffin, London, 1975.
- [5] F.G. BALL, *The threshold behaviour of epidemic models*, J. Appl. Prob. **20** (1983) pp. 227–241.
- [6] F.G. BALL, *A unified approach to the distribution of total size and total area under the trajectory of infectives in epidemic models*, Adv. Appl. Prob. **18** (1986), pp. 289–310.
- [7] F.G. BALL, *Threshold behaviour in stochastic epidemics among households*, in: Athens Conference on Applied Probability and Time Series, Volume I: Applied Probability (eds. C.C. Heyde, Y.V. Prohorov, R. Pyke, and S.T. Rachev), Lecture Notes in Statistics **114** (1996), pp. 253–266.
- [8] F.G. BALL, *Stochastic and deterministic models for SIS epidemics among a population partitioned into households*, Math. Biosci. **156** (1999), pp. 41–67.
- [9] F.G. BALL AND P. DONNELLY, *Strong approximations for epidemic models*, Stoch. Proc. Appl. **55** (1995), pp. 1–21.
- [10] F.G. BALL AND O.D. LYNE, *Stochastic multitype SIR epidemics among a population partitioned into households*, Adv. Appl. Prob. **33** (2001) pp. 99–123.

- [11] F.G. BALL, D. MOLLISON, AND G. SCALIA-TOMBA, *Epidemics with two levels of mixing*, Ann. Appl. Prob. **7** (1997), pp. 46–89.
- [12] R. BARTOSZYŃSKI, *On a certain model of an epidemic*, Appl. Math. **13** (1972), pp. 139–151.
- [13] N.G. BECKER, *Analysis of Infectious Disease Data*, Chapman and Hall, London (1989).
- [14] N.G. BECKER, A. BAHRAMPOUR, AND K. DIETZ, *Threshold parameters for epidemics in different community settings*, Math. Biosci. **129** (1995), pp. 189–208.
- [15] N.G. BECKER AND K. DIETZ, *The effect of the household distribution on transmission and control of highly infectious diseases*, Math. Biosci. **127** (1995), pp. 207–219.
- [16] N.G. BECKER AND K. DIETZ, *Reproduction numbers and critical immunity levels in epidemics in a community of households*, in: Athens Conference on Applied Probability and Time Series, Volume I: Applied Probability (eds. C.C. Heyde, Y.V. Prohorov, R. Pyke, and S.T. Rachev), Lecture Notes in Statistics **114** (1996), pp. 267–276.
- [17] N.G. BECKER AND R. HALL, *Immunization levels for preventing epidemics in a community of households made up of individuals of different types*, Math. Biosci. **132** (1996), pp. 205–216.
- [18] N.G. BECKER AND D.N. STARCZAK, *Optimal vaccination strategies for a community of households*, Math. Biosci. **139** (1997), pp. 117–132.
- [19] N.G. BECKER AND S. UTEV, *The effect of community structure on the immunity coverage required to prevent epidemics*, Math. Biosci. **147** (1998), pp. 23–39.
- [20] T. BRITTON, *Limit theorems and tests to detect within family clustering in epidemic models*, Commun. Statist. **26** (1997a), pp. 953–976.
- [21] T. BRITTON, *Tests to detect clustering of infected individuals within families*, Biometrika **53** (1997b), pp. 98–109.
- [22] T. BRITTON, *A test to detect within-family infectivity when the whole epidemic process is observed*, Scand. J. Statist. **24** (1997c), pp. 315–330.
- [23] T. BRITTON, *A test of homogeneity versus a specified heterogeneity in an epidemic model*, Math. Biosci. **141** (1997d), pp. 79–100.
- [24] T. BRITTON, *On critical vaccination coverage in multitype epidemics*, J. Appl. Prob. **35** (1998), pp. 1003–1006.
- [25] D.J. DALEY AND J. GANI, *A deterministic general epidemic model in a stratified population*, in: Probability, Statistics and Optimisation — A Tribute to Peter Whittle, (ed. F.P. Kelly), Wiley, Chichester (1994), pp. 117–132.
- [26] O. DIEKMANN, M.C.M. DE JONG AND J.A.J. METZ, *A deterministic epidemic model taking account of repeated contacts between the same individuals*, J. Appl. Prob. **35** (1998), pp. 448–462.
- [27] J.A.P. HEESTERBEEK AND K. DIETZ, *The concept of  $R_0$  in epidemic theory*, Statistica Neerlandica **50** (1996), pp. 89–110.
- [28] R.J. KRYSCIO AND C. LEFEVRE, *On the extinction of the SIS stochastic logistic epidemic*, J. Appl. Prob. **26** (1989), pp. 685–694.
- [29] T.G. KURTZ, *Solutions of ordinary differential equations as limits of pure jump Markov processes*, J. Appl. Prob. **7** (1970), pp. 49–58.
- [30] T.G. KURTZ, *Limit theorems for sequences of jump Markov processes approximating ordinary differential processes*, J. Appl. Prob. **8** (1971), pp. 344–356.
- [31] A. LAJMANOVICH AND J.A. YORKE, *A deterministic model for gonorrhea in a non-homogeneous population*, Math. Biosci. **28** (1976), pp. 221–236.
- [32] D. LUDWIG, *Final size distributions for epidemics*, Math. Biosci. **23** (1975), pp. 33–46.
- [33] C.J. MODE, *Multitype Branching Processes*, Elsevier, New York (1971).
- [34] D. MOLLISON, *Spatial contact models for ecological and epidemic spread*, J. R. Statist. Soc. B **39** (1977), pp. 283–326.
- [35] D. MOLLISON, V. ISHAM, AND B. GRENFELL *Epidemics: models and data*, J. R. Statist. Soc. A **157**, (1994), pp. 115–149.

- [36] D. MOLLISON AND S.A. LEVIN, *Spatial dynamics of parasitism*, in: Ecology of Infectious Diseases in Natural Populations (eds. Grenfell, B.T. and Dobson, A.), Cambridge Univ. Press (1995), pp. 384–398.
- [37] S. RUSHTON AND A.J. MAUTNER, *The deterministic model of a simple epidemic for more than one community*, Biometrika **42** (1955), pp. 126–132.
- [38] C.E.G. SMITH, *Factors in the transmission of virus infections from animals to man*, Sci. Basis Med. Annu. Rev. (1964), pp. 125–150.
- [39] R.K. WATSON, *On an epidemic in a stratified population*, J. Appl. Prob. **9** (1972), pp. 659–666.
- [40] P. WHITTLE, *The outcome of a stochastic epidemic — a note on Bailey's paper*, Biometrika **42** (1955), pp. 116–122.
- [41] T. WILLIAMS, *An algebraic proof of the threshold theorem for the general stochastic epidemic (abstract)*, Adv. Appl. Prob. **3** (1971), p. 223.

# INFECTION TRANSMISSION DYNAMICS AND VACCINATION PROGRAM EFFECTIVENESS AS A FUNCTION OF VACCINE EFFECTS IN INDIVIDUALS\*

CARL P. SIMON<sup>†</sup> AND JAMES S. KOOPMAN<sup>‡</sup>

**1. Introduction.** Ideal vaccine *effect* statistics should reflect biologically relevant parameters in an appropriate model of vaccine actions upon infection in the host. Ideal vaccine *effectiveness* statistics should reflect the effect of vaccination on the entire population or upon segments of that population such as vaccinated and unvaccinated individuals. The most commonly used vaccine effect statistic does not meet these ideals. It is one minus the risk in the vaccinated over the risk in the unvaccinated. These risks are sometimes calculated for disease and sometimes for infection. In this paper, we consider only infection. We label this statistic  $\alpha$  and the risks in the vaccinated and unvaccinated populations on which it is based as  $R_v$  and  $R_u$ , respectively:

$$(1) \quad \alpha = 1 - \frac{R_v}{R_u} = \frac{R_u - R_v}{R_u} = \text{PAR\%} .$$

In an article that outlines a set of more appropriate vaccine effectiveness statistics, Haber [2] points out that  $R_v$  and  $R_u$  can be affected by the indirect effects of vaccination in a population. But if the vaccine trial is conducted with low rates of vaccination in the overall population, indirect effects will be small enough to be ignored. In this case, we discuss below how  $\alpha$  can be interpreted both as a vaccine effect statistic and as a vaccine effectiveness statistic.

As a measure of population effectiveness, the statistic in Equation (1) can be interpreted as a relative risk difference or population attributable risk percent (PAR%) [7]. This interpretation requires treating the unvaccinated group as the high-risk or exposed group. One reason for the popularity of this statistic is that it estimates the fraction of unvaccinated individuals with infection that have their infection because they were not vaccinated. But this interpretation requires that the mode of biological action implied by this model is applicable, and this is unlikely to be the case.

The mechanistic model of vaccine effect in which this statistic estimates a stable parameter over any part of an epidemic has been clarified [10]. In this model, upon vaccination an individual is either fully protected

---

\*This work was supported by NIH Grant F002732.

<sup>†</sup>Departments of Mathematics, Economics and Public Policy, Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109-1109.

<sup>‡</sup>Department of Epidemiology, Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109-1109.

or receives no protection at all. If the individual receives vaccine protection, no level of exposure is capable of causing infection in that individual. An individual that fails to get this vaccine effect remains just like the unvaccinated individuals. We can derive the standard vaccine effect statistic from this model by dividing the vaccinated population into those who do and do not receive the effect and treating the total risk in the vaccinated population as the sum of risks in these two subpopulations. We do this in the following equation. Rearrangement of this equation yields Equation (1).

$$R_v = \alpha \cdot 0 + (1 - \alpha) \cdot R_u = (1 - \alpha) \cdot R_u .$$

Biologically the complete protection represented by the zero term in this equation corresponds to lowering the agent reproduction number in a host to below one. In that case, infection will always die out no matter how many times an individual may be exposed or how many agents may be in their exposure dose. This is called sterilizing immunity. The unrealistic assumption in this model is that individuals who do not get sterilizing immunity get no effect at all from vaccination. In the absence of sterilizing immunity, three additional functional effects of vaccine induced immunity can be defined giving four effects in total.

1. **Sterilizing effect ( $\alpha$ ):** complete protection against infection,
2. **Relative susceptibility effect ( $\lambda_S$ ):** decreased chance that an agent transmitted to a susceptible host finds a niche where it can proliferate,
3. **Contagiousness effect ( $\lambda_C$ ):** decreased production of infectious agent if proliferation begins, and therefore fewer agents available for transmission,
4. **Duration effect ( $\lambda_D$ ):** accelerated immune response that shortens the duration of infection.

The first effect is sometimes called an “all or none” effect. However, this seems inappropriate since if one fails to get the “all” effect, one is likely to get one of the three subsequent effects. The second effect has sometimes been labeled a “partial” effect or a “leaky” effect.

Heuristically,  $\alpha$  is the fraction of vaccinated individuals in which immunity prevents any agent growth;  $\lambda_S$  is the fraction of transmitted organisms that find a niche where they can start to grow. These two effects cannot be separated if either is very strong.

The last three effects are meaningful only if the first effect ( $\alpha$ ) is lacking. The last two effects can occur only when deficiencies in the first two permit some agent proliferation. The model underlying the standard vaccine effect statistic assumes that effects two, three, and four are not experienced by individuals in whom vaccine fails to stimulate sterilizing immunity. This assumption seems wholly unrealistic and potentially dangerous for both the choice of vaccine candidates and their field evaluation.

An example illustrates the potentially disastrous consequence of ignoring the effect on contagiousness in a vaccine trial. Imagine a trial of a

new influenza vaccine in a nursing home with two residents per room. To insure comparability of exposure in individuals receiving the old vaccine (A) to the new vaccine (B), investigators randomly assign one individual in each room to one vaccine, and the roommate to the other. Suppose both vaccines stimulate sterilizing immunity in the same proportions of individuals and generate the same level of relative susceptibility effects in those not getting a sterilizing effect. Suppose, however, that the new B vaccine also reduces contagiousness. In this case recipients of vaccine A would be indirectly protected by the B vaccine given to their roommates and would have lower infection rates. The standard analysis would attribute this difference to vaccine A's effects on susceptibility and would judge vaccine A to be the superior one, when exactly the opposite is true.

In this paper, we formulate a model with all four of the above vaccine effects and examine mathematical relationships that define threshold values and endemic equilibrium levels when a fraction  $f$  of the population is vaccinated. Eichner and Hadeler have examined a similar model with similar results [1]. We formulate the first three effects listed above exactly as they do. We formulate only the fourth effect differently. The difference is based upon mathematical convenience rather than a different conceptualization of causal mechanisms. Our results, however, are based upon a mathematical approach that differs from that of Eichner and Hadeler in that it can demonstrate global stability of the endemic equilibrium state in the model. Also, the way we formulate our results is more helpful for choosing between alternative vaccine formulations and between alternative vaccine trial designs.

To make our logic easier to follow, we begin with an examination of the underlying transmission model in the absence of vaccination. We then proceed to a model having only the first vaccine effect, followed by models that include the other three effects.

**2. The simplest model.** The simplest dynamic transmission model used to study vaccine efficacy is the standard SIR model in which, as stated earlier, individuals are either susceptible ( $X$ ), infected ( $Y$ ), or removed ( $Z$ ), i.e., immune, and there is a homogeneous population in which mixing is "random." The dependence of  $X$ ,  $Y$  and  $Z$  on time is given by the differential equations:

$$(2) \quad \begin{aligned} \dot{X} &= -\frac{c\beta XY}{N} + \mu N - \mu X \\ \dot{Y} &= +\frac{c\beta XY}{N} - \mu Y - kY \\ \dot{Z} &= +kY - \mu Z. \end{aligned}$$

In this system,

$c$  is the average number of possibly infection-transmitting contacts per person per unit time,

$\beta$  is the probability of infection given such a contact between an infected and a susceptible,

$\mu$  is the underlying birth and death rate in the population,

$k$  is the rate at which infected individuals recover and enter the immune class  $Z$ , and

$D = 1/(\mu+k)$  is the expected time to leave the infected population (by becoming immune or by dying).

The  $\mu N$  term in the first equation implies that all individuals are born susceptible. The term  $c\beta XY/N$ , often called the force of infection, is the number of contacts by susceptibles per unit time ( $cX$ ) times the probability that the contact is with an infective individual ( $Y/N$ ) times the probability that such a contact transmits the infection ( $\beta$ ).

Since  $\dot{N} = \dot{X} + \dot{Y} + \dot{Z} = 0$ ,  $N(t)$  is constant; one of the equations in (2) is redundant. Furthermore, the equations can be divided through by  $N$  to obtain the reduced system:

$$(3) \quad \begin{aligned}\dot{x} &= -c\beta xy + \mu(1-x) \\ \dot{y} &= +c\beta xy - (\mu + k)y,\end{aligned}$$

where  $x = X/N$  and  $y = Y/N$  are the fractions susceptible and infective. Write the second equation in (3) as

$$(4) \quad \dot{y} = c\beta y \left( x - \frac{\mu + k}{c\beta} \right) = c\beta y \left( x - \frac{1}{c\beta D} \right).$$

System (3) has two equilibria:

- i) the no-disease equilibrium  $(x_o, y_o) = (1, 0)$ , and
- ii) the “endemic” equilibrium  $(x_e, y_e) = \left( \frac{\mu+k}{c\beta}, \frac{\mu}{\mu+k} \left( 1 - \frac{\mu+k}{c\beta} \right) \right) = \left( \frac{1}{R_0}, \mu D \left( 1 - \frac{1}{R_0} \right) \right)$ , where the critical parameter is  $R_0 \equiv c\beta D$ , which can be interpreted to be the number of infections caused by a single infective in a population of susceptibles. For this reason,  $R_0$  is called the *basic reproduction number*. Rewriting (4) as

$$\dot{y} = c\beta y \left( x - \frac{1}{R_0} \right),$$

we see that, since  $x \leq 1$ , if  $R_0 \leq 1$  (so that  $1/R_0 \geq 1$ ),  $\dot{y}$  is always  $\leq 0$ . This implies that  $y(t)$  decreases monotonically to 0 and the no-disease equilibrium  $(x_o, y_o) = (1, 0)$  is a global attractor of system (2) and the only equilibrium. If  $R_0 > 1$ , then for  $x \in (1/R_0, 1]$ ,  $\dot{y} > 0$  and  $(x_o, y_o)$  is an unstable equilibrium. One can show that for  $R_0 > 1$  all solutions  $(x(t), y(t))$  of (2) starting with  $y(0) > 0$  tend to the endemic equilibrium as  $t \rightarrow \infty$ .

**3. Vaccines with only sterilizing immunity effects.** Despite our argument that a vaccine effect model with only an  $\alpha$  effect is biologically implausible, for heuristic purposes we first consider the use of a vaccine with only an  $\alpha$  effect. Let  $f$  be the fraction of population vaccinated and let  $\alpha$  equal the success rate of the vaccine, i.e., the fraction of those vaccinated that are completely protected by the sterilizing effect. Then,  $f\alpha$  is the fraction of susceptibles successfully vaccinated. We assume – throughout this paper – that everyone is born susceptible and that vaccination occurs soon “after birth,” that is, before infection can occur. Using the subscripts  $u$  and  $v$  for the unvaccinated and (not completely successfully) vaccinated subpopulations, respectively, system (3) becomes:

$$(5) \quad \begin{aligned} \dot{x}_u &= -c\beta x_u y_u - c\beta x_u y_v - \mu x_u + \mu - \mu f \\ \dot{x}_v &= -c\beta x_v y_u - c\beta x_v y_v - \mu x_v + \mu f - \mu f\alpha \\ \dot{y}_u &= +c\beta x_u y_u + c\beta x_u y_v - \mu y_u - k y_u \\ \dot{y}_v &= +c\beta x_v y_u + c\beta x_v y_v - \mu y_v - k y_v \\ \dot{z}_u &= +k y_u - \mu z_u \\ \dot{z}_v &= +k y_v - \mu z_v + \mu f\alpha. \end{aligned}$$

Analysis is simplified by combining the vaccinated and unvaccinated in each class:

$$(6) \quad x = x_u + x_v, \quad y = y_u + y_v, \quad z = z_u + z_v.$$

System (5) then becomes a generalization of system (3):

$$(7) \quad \begin{aligned} \dot{x} &= -c\beta x y - \mu x + \mu(1 - f\alpha) \\ \dot{y} &= +c\beta x y - \mu y - k y \\ \dot{z} &= k y - \mu z + \mu f\alpha. \end{aligned}$$

The no-disease equilibrium for (7) is:

$$(8) \quad (x_o, y_o, z_o) = (1 - f\alpha, 0, f\alpha).$$

The endemic equilibrium  $(x^*, y^*, z^*)$  is obtained by setting  $\dot{y} = 0$  and  $y^* \neq 0$  into the middle equation in (7), giving:

$$(9) \quad x^* = \frac{\mu + k}{c\beta} = \frac{1}{R_0}, \quad \text{where } R_0 \equiv \frac{c\beta}{\mu + k},$$

as above. Setting  $\dot{x} = 0$  in the first equation in (7) yields:

$$(10) \quad y^* = \frac{\mu[R_0(1 - f\alpha) - 1]}{c\beta}.$$

Finally, since  $y^* > 0$  at this endemic equilibrium,

$$R_0(1 - f\alpha) > 1 \quad \text{or} \quad 1 - \frac{1}{R_0} > f\alpha.$$

This leads to a necessary and sufficient condition of  $f$  and  $\alpha$  for an all-or-nothing vaccine to prevent an epidemic:

$$f\alpha > 1 - \frac{\mu + k}{c\beta} = 1 - \frac{1}{R_0}.$$

By carrying out the analysis just performed on system (7), one easily shows that the endemic equilibrium for system (5) is:

$$(11) \quad \begin{aligned} x_u^* &= \frac{1}{R_0} \frac{1-f}{1-f\alpha}, & x_v^* &= \frac{1}{R_0} \frac{f(1-\alpha)}{1-f\alpha}, \\ y_u^* &= y^* \frac{1-f}{1-f\alpha}, & y_v^* &= y^* \frac{f(1-\alpha)}{1-f\alpha}, \\ z_u^* &= \frac{1-f}{1-f\alpha} \cdot \frac{k}{c\beta} [R_0(1-f\alpha) - 1], \\ z_v^* &= \frac{f(1-\alpha)}{1-f\alpha} \cdot \frac{k}{c\beta} [R_0(1-f\alpha) - 1] + f\alpha, \end{aligned}$$

where  $y^*$  is as in (10).

**4. Vaccines having additional effects.** We now add realism to the previous model by allowing for additional effects such as contagiousness effects when complete sterilizing immunity is not experienced. Such effects seem highly likely since even if viral reproduction dynamics in the host are not lowered enough by the vaccine induced immune response to sterilize infection in the individual, they are likely to be lowered enough to slow viral replication and reduce the amount of virus available to be transmitted.

We continue to use  $f$  to denote the fraction of the population vaccinated and  $\alpha$  to denote the fraction of those vaccinated for whom the vaccination is completely successful, i.e., the vaccination drives the viral reproduction number below one. Let  $\beta$  (or  $\beta_{uu}$ ) denote the probability of transmitting infection when an unvaccinated infective contacts an unvaccinated susceptible. We assume that  $\beta$  is reduced to:

- $\beta_{uv}$  when an unvaccinated susceptible contacts a vaccinated infective,
- $\beta_{vu}$  when a vaccinated susceptible contacts an unvaccinated infective, and
- $\beta_{vv}$  when a vaccinated susceptible contacts a vaccinated infective.

Let  $\lambda_C$  denote the reduction of  $\beta$  to  $\beta_{uv}$  so that

$$\beta_{uv} = \lambda_C \beta_{uu}.$$

Then,  $\lambda_C \in [0, 1]$  captures the vaccine's effect in reducing the contagiousness of infective individuals. Let  $\lambda_S$  denote the reduction of  $\beta$  to  $\beta_{vu}$  so that

$$\beta_{vu} = \lambda_S \beta_{uu},$$

and  $\lambda_S$  captures the relative susceptibility effect. Assume that  $\beta_{vv} = \lambda_C \cdot \lambda_S \cdot \beta_{uu}$ .

Finally, in this more realistic model, we assume that the vaccine can reduce the average length of the infectious period, and so we define  $\lambda_D$  such that the vaccine reduces the average duration from  $1/(k+\mu)$  to  $\lambda_D/(k+\mu)$ . The parameter  $\lambda_D$  captures the duration effect of the vaccine. Now, the rate at which a vaccinated person recovers from the disease is  $[\mu(1-\lambda_D)+k]/\lambda_D$  instead of the usual  $k$  in system (5). The model becomes:

$$(12) \quad \begin{aligned}\dot{x}_u &= -c\beta x_u y_u - c\beta \lambda_C x_u y_v + \mu(1-f) - \mu x_u \\ \dot{x}_v &= -c\beta \lambda_S x_v y_u - c\beta \lambda_S \lambda_C x_v y_v + \mu f(1-\alpha) - \mu x_v \\ \dot{y}_u &= +c\beta x_u y_u + c\beta \lambda_C x_u y_v - \mu y_u - k y_u \\ \dot{y}_v &= +c\beta \lambda_S x_v y_u + c\beta \lambda_S \lambda_C x_v y_v - \frac{\mu+k}{\lambda_D} y_v \\ \dot{z}_u &= k y_u - \mu z_u \\ \dot{z}_v &= \frac{\mu(1-\lambda_D)+k}{\lambda_D} y_v - \mu z_v + \mu f \alpha.\end{aligned}$$

**Step 1. Effect of decreasing contagiousness:**  $\lambda_S = \lambda_D = 1$ .

Because system (12) is a bit more difficult to analyze than system (5), we first consider only the contagiousness effect  $\lambda_C$ , setting  $\lambda_S$  and  $\lambda_D$  equal to 1 and  $\alpha$  equal to 0. We can work without loss of generality with the top four equations of system (12). The new system is:

$$(13) \quad \begin{aligned}\dot{x}_u &= -c\beta x_u y_u - c\beta \lambda_C x_u y_v + \mu(1-f) - \mu x_u \\ \dot{x}_v &= -c\beta x_v y_u - c\beta \lambda_C x_v y_v + \mu f - \mu x_v \\ \dot{y}_u &= +c\beta x_u y_u + c\beta \lambda_C x_u y_v - (\mu+k) y_u \\ \dot{y}_v &= +c\beta x_v y_u + c\beta \lambda_C x_v y_v - (\mu+k) y_v.\end{aligned}$$

We now define  $\xi$ , an equivalent number of susceptibles, and  $\eta$ , an equivalent number of infectives, by:

$$(14) \quad \xi \equiv x_u + \lambda_C x_v \quad \text{and} \quad \eta \equiv y_u + \lambda_C y_v.$$

Upon adding and recombining, system (13) becomes:

$$(15) \quad \begin{aligned}\dot{\xi} &= -c\beta \xi \eta - \mu \xi + \mu(1-f(1-\lambda_C)) \\ \dot{\eta} &= +c\beta \xi \eta - (\mu+k)\eta = c\beta \eta \left( \xi - \frac{1}{R_0} \right),\end{aligned}$$

where  $R_0$  is still  $c\beta/(\mu+k)$ .

We proceed as in the analysis of the model with only the  $\alpha$  parameter effects of sterilizing immunity. When  $\xi - \frac{1}{R_0} < 1$ , then by (15)  $\dot{\eta} < 0$  and  $\eta(t)$  decreases toward 0. Now,

$$\xi = x_u + \lambda_C x_v \leq (1-f) + \lambda_C f.$$

We conclude that when

$$(16) \quad (1-f) + \lambda_C f < \frac{1}{R_0} \quad \text{or} \quad R_0(1-f + \lambda_C f) < 1,$$

the no-disease equilibrium

$$(17) \quad (\xi_o, \eta_o) = (1 - f(1 - \lambda_C), 0)$$

is a *globally asymptotically stable* equilibrium.

If  $R_0(1 - f(-\lambda_C)) > 1$ , or equivalently if

$$f > \frac{1 - \frac{1}{R_0}}{1 - \lambda_C},$$

orbits that start close to (17) move away from it — once again by (16). In this case, the no-disease equilibrium is unstable and all solutions  $(\xi(t), \eta(t))$  of (15) with  $\eta(0) > 0$  tend to the endemic equilibrium

$$(18) \quad (\xi^*, \eta^*) = \left( \frac{1}{R_0}, \frac{\mu}{c\beta} [R_0(1 - f(1 - \lambda_C)) - 1] \right).$$

Of course, for (18) to be the true endemic equilibrium we need  $\eta^* > 0$  in (18). But  $\eta^* > 0$  if and only if  $R_0(1 - f(1 - \lambda_C)) > 1$ , that is, if and only if inequality (16) is reversed, or in terms of the fraction  $f$  vaccinated, if and only if

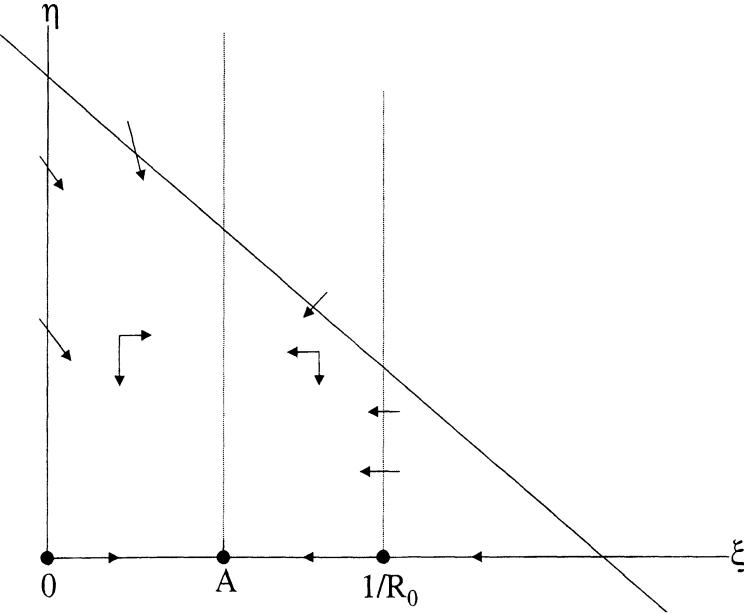
$$(19) \quad f < \frac{1 - \frac{1}{R_0}}{1 - \lambda_C}.$$

We sketch a proof to show what happens when (16) holds. Figure 1 shows a phase diagram for system (15), where  $A \equiv 1 - f(1 - \lambda_C)$  and  $A < \frac{1}{R_0}$  by hypothesis (16). To the left of the vertical line  $\{\xi = A\}$ ,  $\dot{\xi} < 0$  for all solutions  $(\xi(t), \eta(t))$  of (15). So, all orbits that start to the right of the line  $\xi = \frac{1}{R_0}$  eventually end up in this invariant region  $\{\xi \leq \frac{1}{R_0}\}$ . But in this region,  $\dot{\eta}/\eta < 0$ ; so  $\eta(t) \rightarrow 0$  for all solutions  $(\xi(t), \eta(t))$  of (15).

It is straightforward to decompose the combined variables (14) to obtain the endemic solution of system (13) from the endemic solution (18) of system (15):

$$(20) \quad \begin{aligned} x_u^* &= \frac{1}{R_0} \frac{1 - f}{1 - f + \lambda_C f} \\ x_v^* &= \frac{1}{R_0} \frac{f}{1 - f + \lambda_C f} \\ y_u^* &= \frac{\mu(1 - f)}{\mu + k} \left[ 1 - \frac{1}{R_0(1 - f + \lambda_C f)} \right] \\ y_v^* &= \frac{\mu f}{\mu + k} \left[ 1 - \frac{1}{R_0(1 - f + \lambda_C f)} \right]. \end{aligned}$$

As (16) indicates, the threshold  $R_0 < 1$  for system (1) without vaccines is replaced by threshold  $R_0(1 - f(1 - \lambda_C)) < 1$  for a system with a vaccine



with contagiousness effect. If everyone is vaccinated so that  $f = 1$ , the threshold for the disease to die out is simply  $R_0 \cdot \lambda_C < 1$ .

### Step 2. Effect of decreasing contagiousness, susceptibility, and duration.

We next remove the restrictions  $\lambda_S = 1$  and  $\lambda_D = 1$  and include the fraction getting sterilizing effects from vaccination  $\alpha$ , so that the effects of contagiousness, susceptibility and duration are all included. In this case, we work with the first four equations of system (12):

$$(21) \quad \begin{aligned} \dot{x}_u &= -c\beta x_u y_u - c\beta \lambda_C x_u y_v + \mu(1-f) - \mu x_u \\ \dot{x}_v &= -c\beta \lambda_S x_v y_u - c\beta \lambda_S \lambda_C x_v y_v + \mu f(1-\alpha) - \mu x_v \\ \dot{y}_u &= +c\beta x_u y_u + c\beta \lambda_C x_u y_v - (\mu+k)y_u \\ \dot{y}_v &= +c\beta \lambda_S x_v y_u + c\beta \lambda_S \lambda_C x_v y_v - \frac{(\mu+k)}{\lambda_D} y_v. \end{aligned}$$

The third and fourth equations in (21) can be combined to give:

$$(22) \quad \begin{aligned} \dot{y}_u + \lambda_C \dot{y}_v &= c\beta(x_u + \lambda_S \lambda_C \lambda_D x_v)(y_u + \lambda_C v_v) - (\mu+k)(y_u + \lambda_C y_v) \\ &= c\beta(y_u + \lambda_C y_v) \left[ (x_u + \lambda_S \lambda_C \lambda_D x_v) - \frac{1}{R_0} \right], \end{aligned}$$

where  $R_0 = c\beta/(\mu+k)$ . Since

$$x_u + \lambda_S \lambda_C \lambda_D x_v \leq (1-f) + \lambda_S \lambda_C \lambda_D f(1-\alpha),$$

if

$$(23) \quad (1-f) + \lambda_C \lambda_S \lambda_D f (1-\alpha) < \frac{1}{R_0}$$

or  $R_0 (1 - f(1 - \lambda_S \lambda_C \lambda_D (1 - \alpha))) < 1,$

the square bracketed term in (22) will always be negative, and so  $(y_u + \lambda_C y_v)(t) \rightarrow 0$  as  $t \rightarrow \infty$ , and thus *every* solution of system (21) tends to the no-disease equilibrium

$$(24) \quad (x_u, x_v, y_u, y_v, z_u, z_v) = (1-f, f(1-\alpha), 0, 0, 0, f\alpha),$$

when condition (23) holds. In terms of the fraction  $f$  vaccinated, condition (23) can be written as:

$$(25) \quad f > \frac{1 - \frac{1}{R_0}}{1 - \lambda_S \lambda_C \lambda_D (1 - \alpha)}.$$

Conversely, if (23) or (25) does not hold, i.e., if

$$(26) \quad 1 - f + \lambda_S \lambda_C \lambda_D (1 - \alpha) f > \frac{1}{R_0},$$

then solutions that start near (24) move away from (24); the no-disease equilibrium is unstable.

Once again, if everyone is vaccinated so that  $f = 1$ , the threshold condition (25) for the disease to die out is

$$(27) \quad R_0 \lambda_S \lambda_C \lambda_D (1 - \alpha) \leq 1.$$

In any case, (25) only makes sense when (27) holds. Furthermore, we see from (25) that the more potent the vaccine (i.e., the smaller is  $\lambda_S \lambda_C \lambda_D$ ), the smaller the fraction  $f$  needs to be to achieve population immunity.

The endemic equilibrium for system (21) is obtained by solving the nonlinear (quadratic) system of equations that arise by setting  $\dot{y}_u + \lambda_C \dot{y}_v = 0$ ,  $\dot{x}_u + \dot{y}_u = 0$ ,  $\dot{x}_v + \dot{y}_v = 0$ , and  $\dot{x}_v = 0$ , respectively, in system (21):

$$\begin{aligned} x_u + \lambda_S \lambda_C \lambda_D x_v &= \frac{1}{R_0} \\ \mu x_u + (\mu + k) y_u &= \mu(1-f) \\ \mu x_v + \frac{(\mu + k)}{\lambda_D} y_v &= \mu f(1-\alpha) \\ c\beta \lambda_S x_v (y_u + \lambda_C y_v) + \mu x_v &= \mu f(1-\alpha). \end{aligned}$$

The resulting endemic equilibrium of (21), the solution of these four equations, is:

$$\begin{aligned}
 x_u^* &= \frac{(1 - \lambda_S) - (\lambda_S R_0(1 - f(1 - \lambda_C \lambda_D(1 - \alpha))))}{2R_0(1 - \lambda_S)} \\
 &\quad + \frac{\sqrt{[\lambda_S R_0(1 - f + \lambda_C \lambda_D f(1 - \alpha)) - (1 - \lambda_S)]^2 + 4R_0(1 - f)(1 - \lambda_S)\lambda_S}}{2R_0(1 - \lambda_S)} \\
 (28) \quad x_v^* &= \frac{(1 - \lambda_S) + (\lambda_S R_0(1 - f(1 - \lambda_C \lambda_D(1 - \alpha))))}{2R_0(1 - \lambda_S)\lambda_S\lambda_C\lambda_D} \\
 &\quad - \frac{\sqrt{[\lambda_S R_0(1 - f + \lambda_C \lambda_D f(1 - \alpha)) - (1 - \lambda_S)]^2 + 4R_0(1 - f)(1 - \lambda_S)\lambda_S}}{2R_0(1 - \lambda_S)\lambda_S\lambda_C\lambda_D} \\
 y_u^* &= \frac{\mu}{\mu + k}(1 - f - x_u^*) \\
 y_v^* &= \frac{\mu}{\mu + k}\lambda_D(f(1 - \alpha) - x_v^*).
 \end{aligned}$$

Using a bit of calculus, one can verify that these solutions (28) reduce to the solutions (18) for the case  $\lambda_S = \lambda_D = 1$  and  $\alpha = 0$ .

**5. Significance of the critical vaccination fraction.** We have shown that the critical vaccination level in a randomly mixing population is:

$$f^* = \frac{1 - \frac{1}{R_0}}{1 - \lambda_S\lambda_C\lambda_D(1 - \alpha)}.$$

Our analysis demonstrates that the four vaccine effects act in mathematically the same way to reduce the critical vaccination fraction. That observation should provide guidance for both vaccine development and vaccine evaluation. But before we discuss what this guidance should be, let us be clear about what we have and have not demonstrated.

The equality of vaccine effects we have demonstrated is relevant only to critical vaccination fractions for elimination of infection from a population. The Liapunov function approach we used to derive our critical vaccination fraction does not allow us to draw conclusions about vaccine effects on *endemic* levels of infection. In fact, we believe that this equality of different vaccine effects does not hold for the extent to which different vaccine effects lower prevalence of infection. Neither does it hold when mixing is non-random.

We believe that when prevalence of infection is high, a sterilizing immunity ( $\alpha$ ) vaccine effect on susceptibility will cause a greater reduction in prevalence than a numerically equal relative susceptibility ( $\lambda_S$ ) vaccine effect. Likewise, we believe that either susceptibility effect will cause greater reductions in prevalence than numerically equal effects on contagiousness or on duration of infection. We leave the demonstration of these relationships for future work.

When mixing is non-random, no single vaccination fraction can define a threshold vaccination level. Different combinations of vaccination levels

in different population subgroups can put the population below threshold. The minimal number of vaccinations needed to eliminate infection in a population will be achieved by vaccinating different population subgroups at different levels. It may be that when high-risk subpopulations are difficult to vaccinate, the equality of vaccine effects on eliminating infection from a population may not hold. This is another topic for future work.

Given these caveats, let us now consider the implication of our findings for the vaccine developer who is considering how to alter vaccine formulations to maximize vaccine effects. Let us break down vaccine effects into two categories:

1. Vaccine effects that slow the reproduction dynamics of infectious agents in the host,
2. Vaccine effects that decrease the chances that infectious agents transferred from an infectious host to a susceptible host will survive to find fertile ground to proliferate in the susceptible host.

Consider step by step improvements in vaccine effects on infectious agent (e.g. virus) reproduction dynamics. A small slowing in virus reproduction dynamics will not get the viral reproduction number below one. It will, however, lower virus levels and thereby reduce contagiousness and the duration of infection. The reduction in duration will depend upon the relationship between the usual course of infection and the extent to which a secondary immune response gets the viral reproduction number below the critical level of one. As reproduction dynamics are further slowed, the viral reproduction number may be held below one right from the start of infection. In that case, complete elimination of susceptibility — the  $\alpha$  effect in our formulations — will be achieved. In this case, unlike relative susceptibility effects, no matter how many times someone is exposed and no matter how great the number of viruses in their exposures, no infection will result.

Vaccine effects on the survival of transmitted agents will be important only if vaccine effects on rate of agent reproduction have not lowered reproduction below the point where the reproduction of surviving agents is possible. But as long as the agent reproduction number is greater than one in the recipient of transferred agents, any vaccine effects on survivability can be overwhelmed by increasing the number of transferred agents until one such agent finally survives to begin reproducing. Pure mucosal immunity might result in pure relative susceptibility vaccine effects.

Previous classifications of vaccine effects have tended to classify them as “all or none” and “partial effects on susceptibility and contagiousness.” The biological considerations just presented suggest that we should classify them as “all or contagiousness” and “partial (or relative) susceptibility” effects.

What is the implication of this analysis for the design of vaccine trials? First, if the results of a vaccine trial are to indicate whether the biological effects a vaccine is generating are on survival of transmitted agents

or on agent reproduction dynamics, complete susceptibility effects should be distinguished from partial susceptibility effects. As Longini and Halloran [5] illustrate, these two different effects generate differing patterns of infection in vaccinated and nonvaccinated individuals as exposure levels increase, making this distinction theoretically possible. But this analysis indicates that an analysis of contagiousness effects is as much indicated as an analysis of susceptibility effects. Even if no susceptibility effects are experienced, our formulations indicate that contagiousness effects can just as feasibly eliminate infection as susceptibility effects. Moreover, the detection of contagiousness effects indicates that a step in vaccine development toward sterilizing effects has been taken. The detection of any susceptibility effect does not provide such an indication. For HIV vaccines where the main expectation from early trials is an indication as to how to proceed with further vaccine development, the distinction between different susceptibility effects and the detection of contagiousness effects thus become crucially important.

## REFERENCES

- [1] EICHNER M. AND K.P. HADELER, "Deterministic Models for the Eradication of Poliomyelitis: Vaccination with the Inactivated (IPV) and Attenuated (OPV) Polio Virus Vaccine." *Mathematical Biosciences* (1995), **127**: 149–166.
- [2] HABER M., "Estimation of the direct and indirect effects of vaccination." *Statistics in Medicine* (1999), **18**: 2101–2109.
- [3] KOOPMAN J.S., J.A. JACQUEZ, C.P. SIMON, ET AL., "The Role of Primary HIV Infection in the Spread of HIV Through Populations," *Journal of A.I.D.S.* (1997), **14**: 249–258.
- [4] KOOPMAN J.S. AND R.J. LITTLE, "Assessing HIV Vaccines Effects," *Amer. J. Epid.* (1995), **142**: 1113–1120.
- [5] LONGINI I.M. AND M.E. HALLORAN, "A Frailty Mixture Model for Estimating Vaccine Efficacy." *Appl Statist* (1996), **45**: 165–73.
- [6] METZ J.A.J. AND O. DIEKMANN (eds.), *The Dynamics of Physiologically Structured Populations*. Springer Verlag Lecture Notes in Biomathematics (1986), **68**.
- [7] ROTHMAN K. AND S. GREENLAND, *Modern Epidemiology*, Second Edition. Lippincott, Williams and Wilkins (1998).
- [8] SIMON C.P. AND J.A. JACQUEZ, "Reproduction Numbers and the Stability of Equilibria of SI Models for Heterogeneous Populations." *S.I.A.M. Journal of Applied Mathematics* (1992), **52**(2): 541—576.
- [9] SIMON C.P., J.A. JACQUEZ, AND J.S. KOOPMAN, "A Liapunov Function Approach to Computing  $R_0$ ." *Models for Infectious Diseases* (V. Isham & G. Medley, eds.) Cambridge University Press (1995).
- [10] SMITH P.G., RODRIGUEZ L.C., AND FINE P.E., "Assessment of the Protective Efficacy of Vaccines Against Common Diseases Using Case-Control and Cohort Studies." *International Journal of Epidemiology* (1984), **13**(1): 87–93.

# THE INFLUENCE OF DIFFERENT FORMS OF CROSS-PROTECTIVE IMMUNITY ON THE POPULATION DYNAMICS OF ANTIGENICALLY DIVERSE PATHOGENS

NEIL FERGUSON\* AND VIGGO ANDREASEN†

**Abstract.** We develop simple epidemic models of co-circulating strains of an infectious disease in which the strains interact immunologically via cross-protective acquired immune responses. Two limiting forms of cross-protective immunity are explored: reduction of infectivity on infection with a strain that against which a degree of cross-protective immunity exists from prior exposure to a heterologous strain, and reduction of susceptibility to infection after exposure to the second strain. After developing a generic model framework capable of representing both forms of action, we show that model formulation can be simplified for some simple cross-immunity structures in the case of infectivity reduction. We then discuss equilibria and stability properties of the generic model, before investigating in detail the special case of allele-based cross-immunity, where antigenic relatedness depends on the number of alleles shared between two strains of a haploid pathogen. For this system, we present conditions for the stability of the symmetric and boundary equilibria in the case of purely infectivity-mediated cross-immunity, and illustrate numerically the wide range of complex limit cycle or chaotic dynamics that dominate a large region of parameter space. Finally, we describe the similarities between the dynamics exhibited by systems with each form of immunity action, and discuss biological applications of such models.

**Key words.** strains, cross-protection, epidemic model, limit cycles, stability analysis.

**1. Introduction.** Many pathogens occur in different variants or strains that elicit a partial cross-reaction in the sense that antibodies produced in response to infection by one strain give some but not a full immune reaction when the host is challenged by antigens from a related strain. A variety of past work [8, 7, 3] has shown that such cross-protective responses can induce structuring of the pathogen population into a collection of discrete strains, in a manner akin to simple competition models of niche adaptation [10].

The exact manner in which cross-protective immune responses act remains unknown however, and may depend on the host-pathogen system in question. Ackerman et al [1] suggested several ways in which cross-immunity may affect disease transmission, the most important of which are via a reduction in the severity and infectivity of a second infection [8, 7] or reduced susceptibility to secondary challenges [4, 2, 9]. The key difference between these two alternatives is that in the case of infectivity

---

\*Institute of Genetics, University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, UK. The work of the first author was supported by a research fellowship from the Royal Society.

†Department of Mathematics and Physics, Roskilde University, DK-4000 Roskilde, Denmark. The work of the second author was supported by Danish Natural Science Research Council grant 97-0141-2.

reduction, the affected individual still experiences the same force of infection as all other individuals susceptible to that strain and on infection gains complete immunity to the strain, while for susceptibility reduction the force of infection experienced is reduced, effectively meaning that the host is sometimes able to clear infection purely with a pre-existing cross-protective response and without generation of strain-specific immunity or significant infectiousness. In this paper we explore how these two types of cross-immunity influence pathogen transmission dynamics.

The rest of this paper is organized as follows. First we develop a model that allows us to include both types of cross-immunity in the description of multiple-strain transmission dynamics and show how this formulation relates to previous models. We next provide some analytical results comparing the effects of the two types of cross-immunity. Finally we focus on the special symmetric situations that arise if cross-immunity is determined by a few alleles in a haploid pathogen. Here we use numerical methods to compare the dynamics caused by the two types of cross-immunity.

**2. Model formulation.** In standard epidemic theory, the population is divided into three groups, the susceptibles, the infected and infectious, and the immune and recovered. To simplify the description we shall use a slightly different notation, where  $S$  denotes the number of susceptible hosts while  $R$  denotes the number of host who have some time in their life been infected, in other words  $R$  gives the number of hosts who are immune to the infection or who are currently infected. Assuming a constant population size  $N = S + R$ , disease transmission can be described by the following model

$$\begin{aligned}\dot{S} &= \mu N - \mu S - \Lambda S \\ \dot{R} &= \Lambda S - \mu R \\ \dot{\Lambda} &= \beta \Lambda S - (\mu + \sigma) \Lambda.\end{aligned}$$

Here  $\Lambda$  gives the force of infection, i.e. the rate at which susceptibles get infected. The transmission coefficient  $\beta$  gives the proportionality between the number of infected hosts and  $\Lambda$ . Note that  $\Lambda = \beta Y$ , where  $Y$  is the number of infected hosts, assumed to be a subset of the compartment  $R$ . Finally  $\mu$  and  $\sigma$  denote the death rate and the rate of recovery respectively.

To account for the interaction between cocirculating, cross-reacting strains, we shall need a more detailed division of the host population. Andreasen *et al.* [3] used an index set notation where  $S_{\mathcal{J}}$  denotes the number of hosts who have acquired immunity to the strains in the set  $\mathcal{J}$ , where  $\mathcal{J}$  is a subset of the set of all strains  $\mathcal{K}$ .

To simplify the description, we will assume that the immunity to an infecting strain is obtained immediately upon infection and that multiple concurring infections with different strains (superinfection) is possible. The definition of  $S_{\mathcal{J}}$  thus implies that  $S_{\mathcal{J}}$  denotes the number of host who have

been or are currently infected by the strains in  $\mathcal{J}$  and who are not and have never been infected with the strains in  $\mathcal{K} \setminus \mathcal{J}$ .

Following Andreasen *et al.* [3], it is now straightforward to determine the transmission dynamics of the cocirculating strains

$$(2.1) \quad \begin{aligned} \dot{S}_\emptyset &= \mu N - \mu S_\emptyset - S_\emptyset \sum_{i \in \mathcal{K}} \Lambda^i \\ \dot{S}_{\mathcal{J}} &= \sum_{i \in \mathcal{J}} \gamma_{\mathcal{J} \setminus i}^i \Lambda^i S_{\mathcal{J} \setminus i} - \mu S_{\mathcal{J}} - S_{\mathcal{J}} \sum_{j \notin \mathcal{J}} \gamma_{\mathcal{J}}^j \Lambda^j \\ \dot{\Lambda}^j &= \Lambda^j \left( \beta_j \sum_{\mathcal{J} \subset \mathcal{K}} \gamma_{\mathcal{J}}^j \tau_{\mathcal{J}}^j S_{\mathcal{J}} - (\mu + \sigma) \right). \end{aligned}$$

The two types of cross-immunity are both included in this formulation. To account for the reduction in infectivity, the transmission coefficient  $\beta_j$  is reduced by a factor  $\tau_{\mathcal{J}}^j$  for infectious individuals with immune history  $\mathcal{J}$ . Susceptibility reduction has two effects. Firstly it reduces the force of infection for strain  $j$  by a factor  $\gamma_{\mathcal{J}}^j$ ,  $0 \leq \gamma_{\mathcal{J}}^j \leq 1$  for susceptibles  $S_{\mathcal{J}}$  who previously have been infected with the strains in  $\mathcal{J}$ . Secondly reduction in susceptibility also reduces the growth rate of  $\Lambda^i$  resulting from infection of  $S_{\mathcal{J}}$ -individuals, since disease prevalence among these hosts is reduced.

The cross-immunity factors  $\tau_{\mathcal{J}}^i$  and  $\gamma_{\mathcal{J}}^i$  reflect the relatedness between the challenging strain  $i$  and the immunizing strains  $\mathcal{J}$  and in general we would expect that  $\tau$  and  $\gamma$  are decreasing functions in the sense that  $\mathcal{J} \subset \mathcal{L} \Rightarrow \gamma_{\mathcal{J}}^i \geq \gamma_{\mathcal{L}}^i$  and  $\tau_{\mathcal{J}}^i \geq \tau_{\mathcal{L}}^i$ . This is valid for the majority of pathogens for which cross-reactive immunity generates a protective response. Some pathogens, most notably dengue virus [5, 6], appear to induce enhancement of susceptibility or infectivity on secondary infection, and while this case can be expressed naturally within the model formulation above by assigning  $\tau_{\mathcal{J}}^i$  and  $\gamma_{\mathcal{J}}^i$  to be over 1, we neglect this interesting special case for the remainder of this paper.

If cross-immunity affects only the transmissibility, the description can be simplified substantially provided that the cross-immunity structure is somewhat simple. To develop this more compact representation, we focus on strain  $i$  and assume that the strength of cross-immunity is determined only by the cross-reaction of the most related previous infection. Thus for each strain  $i \in \mathcal{K}$  we define a nested sequence of neighborhoods  $\{i\} = \mathcal{N}_0(i) \subseteq \mathcal{N}_1(i) \subseteq \dots \subseteq \mathcal{N}_m(i) = \mathcal{K}$  describing collections of strains that elicit less and less cross-reaction to strain  $i$ . Hosts whose most related infection is at level  $k$  (*i.e.* in set  $\mathcal{N}_k(i) \setminus \mathcal{N}_{k-1}(i)$ ) have disease transmissibility reduced by a factor of  $\tau_k$ . Since cross-reaction grows with the degree of relatedness, we have  $0 = \tau_0 \leq \tau_1 \leq \tau_2 \leq \dots \leq 1$ .

From the viewpoint of strain  $i$ , the population based immunity can now be represented by listing the number of hosts  $T_k^i$  who have immunity to one or more strains in the  $\mathcal{N}_k(i)$ -neighborhood of  $i$ , for all  $k = 0, \dots, m$ .

The size of these compartments can easily be expressed in terms of our previously used variables in that

$$T_k^i = \sum_{\mathcal{N}_k(i) \cap \mathcal{J} \neq \emptyset} S_{\mathcal{J}}.$$

The dynamic equations of these quantities can be found by summing appropriately over the previous differential equations for the  $S_{\mathcal{J}}$ . However it is simpler to observe that the number of individuals who do not have  $\mathcal{N}_k(i)$ -immunity is  $N - T_k^i$ . Since the rate at which individuals acquire infection by a given strain is assumed to be independent of the individual's immune history, the rate of flow into  $\mathcal{N}_k(i)$  is determined by the combined force of infection exerted by all strains in  $\mathcal{N}_k(i)$  while individuals can leave the  $T_k^i$ -class only through deaths. This yields

$$(2.2) \quad \dot{T}_k^i = \left( \sum_{j \in \mathcal{N}_k(i)} \Lambda_j \right) (N - T_k^i) - \mu T_k^i.$$

It remains to determine the force of infection of strain  $i$ ,

$$\Lambda_i = \beta_i \sum_{\mathcal{J}} \tau_{\mathcal{J}}^i I_{\mathcal{J}}^i,$$

where  $\tau_{\mathcal{J}}^i = \tau_k$  if  $\mathcal{J} \cap (\mathcal{N}_k \setminus \mathcal{N}_{k-1})$ . Making flow considerations as the above ones or summing over the appropriate dynamic equations we obtain a dynamic equation for  $\Lambda_i$ ,

$$(2.3) \quad \dot{\Lambda}_i = \beta_i \Lambda_i \sum_{k=1}^m \tau_k (T_k^i - T_{k-1}^i) - (\mu + \sigma) \Lambda_i,$$

where  $T_m^i = N$  by definition.

The resulting dynamic description of the system is compact compared with the case of susceptibility reduction since the total number of dynamic equations for  $n$  strains each of which have  $m$  immunity neighborhoods is reduced from  $2^n + n$  to  $n \times (m + 1)$ . This was the model formulation used by Gupta and Ferguson [7]. However, note that in the latter paper,  $\gamma$ , the degree of cross-protection, was defined in an opposite sense to here (*i.e.*  $\gamma = 1$  represented 100% cross-protection, and  $\gamma = 0$  no cross-protection).

A similar structure can be obtained if cross-immunity takes the form that some individuals obtain a broader immune protection than others. However we have not been able to obtain a similar simplification of the system for the situation where cross-immunity affects susceptibility.

**3. Equilibria and invasion conditions.** To compare equilibria and their stability for the two models of cross-immunity, we return to the original model (2.1). Equilibria of this model can be determined implicitly in

terms of the forces of infection  $(\Lambda^1, \dots, \Lambda^n) = \underline{\Lambda}$ . First observe that  $S_{\mathcal{J}}(\underline{\Lambda})$  can be determined recursively on the cardinality of the index set  $\mathcal{J}$  in that

$$S_{\emptyset}(\underline{\Lambda}) = \frac{\mu N}{\mu + \sum \Lambda^i}, \quad S_{\mathcal{J}}(\underline{\Lambda}) = \frac{\sum_{i \in \mathcal{J}} \gamma_{\mathcal{J} \setminus i}^i S_{\mathcal{J} \setminus i}(\underline{\Lambda})}{\mu + \sum_{i \notin \mathcal{J}} \Lambda^i}.$$

Equilibria of model (2.1) are now characterized as the non-negative solutions to the implicit equations

$$\Lambda^j = \Lambda^j \frac{\beta_i}{\mu + \sigma} \sum_{\mathcal{J}} \gamma_{\mathcal{J}}^j \tau_{\mathcal{J}}^i S_{\mathcal{J}}(\underline{\Lambda}), \quad j \in \mathcal{K}.$$

Model (2.1) has multiple boundary equilibria and for certain cross-immunity structures there may in fact exist multiple internal equilibria as well, see [3, 9] for details. We shall not characterize the full set of equilibria, but turn our attention to the stability of selected equilibria.

The disease-free equilibrium when  $\underline{\Lambda} = \underline{0}$  and  $S_{\emptyset} = N, S_{\mathcal{J}} = \emptyset$  is independent of  $\gamma_{\mathcal{J}}^i$  and  $\tau_{\mathcal{J}}^i$  and from the dynamic equation for  $\Lambda$  it is clear that the equilibrium is stable if  $R_i = \beta_i / (\mu + \sigma) < 1$  for all  $i \in \mathcal{J}$  and is unstable if  $R_i > 1$  for at least one strain. Thus the stability of the disease free equilibrium is independent of the exact nature of the cross-immunity.

The ability of a novel strain to invade an equilibrium  $S_{\mathcal{J}}^{\dagger}$  where the strain is not present can be determined from the dynamic equation for  $\Lambda^i$  and we find that  $\Lambda^i$  will increase provided that

$$(3.1) \quad R_i \sum_{\mathcal{J}} \tau_{\mathcal{J}}^i \gamma_{\mathcal{J}}^i S_{\mathcal{J}}^{\dagger} > 1.$$

The condition (3.1) can be seen as a generalization of the standard epidemic threshold condition, for details see [3].

In general the value of the left hand side of (3.1) will depend on  $\tau$  and  $\gamma$  in a rather complicated way since cross-immunity will influence the equilibrium values of  $S_{\mathcal{J}}^{\dagger}$ . However in the special situation where there is no cross-immunity between the resident strains, the value of  $S_{\mathcal{J}}^{\dagger}$  is independent of the cross-immunity parameters. Thus invasion by a non-resident strain  $j$  into such a community will depend on the cross-immunity only through the products

$$\tau_{\mathcal{J}}^j \gamma_{\mathcal{J}}^j, \quad \mathcal{J} \subset \mathcal{K}.$$

In this situation, the invasion condition for a strain that confers cross-immunity to some resident strains is independent of the nature of the cross-immunity and susceptibility- and infectivity-reduction enters multiplicatively into the threshold expression.

Consideration of the stability of internal equilibria where all strains coexist is considerably more complex, however. In the special case of symmetric strains and simple cross-immunity structures, it is possible to derive stability criteria analytically for the completely symmetric equilibrium (see below), but only in the case of infectivity mediated immunity.

#### 4. Allele-based cross-immunity.

**4.1. Model definition.** To obtain a more detailed description of the dynamical aspects of the interacting strains, we for the remainder of this paper focus on the cross-immunity structure that arises if immunity is determined by a few loci in a haploid pathogen. In their studies of antigenically diverse pathogens, Gupta et al [8, 7] assume that cross-immunity is not present if the host has had no prior infection with pathogens that share any allele with the challenging strain. If the challenging strain shares at least one allele with a strain that has already immunized the host, then the susceptibility/infectivity is reduced by a factor of  $c$ . Infection with a strain confers complete immunity to that strain.

This framework assumes a simple one-to-one mapping from pathogen genotype onto antigenic phenotype, such that each locus codes for a single antigenically dominant (protein) epitope. As such, it is a suitable model for, say *Neisseria meningitidis* B, but not for the A and C forms in which the polysaccharide capsid is antigenically dominant.

For a two locus  $A$  and  $B$  two allele system, the cross-immunity experienced by a challenging  $AB$ -pathogen in our notation becomes

$$\begin{aligned} AB \in \mathcal{J} &\Rightarrow \varrho_{\mathcal{J}}^{AB} = 0 \\ \{AB\} \cap \mathcal{J} = \emptyset \wedge \{Ab, aB\} \cap \mathcal{J} \neq \emptyset &\Rightarrow \varrho_{\mathcal{J}}^{AB} = c \\ \{AB, Ab, aB\} \cap \mathcal{J} = \emptyset &\Rightarrow \varrho_{\mathcal{J}}^{AB} = 1, \end{aligned}$$

with similar expressions for the other strains. Here  $\varrho$  denotes either  $\gamma$  or  $\tau$ . It is straightforward to construct similar expressions for the general  $n$  locus,  $m$  allele system. Each of the  $m^n$  genotypes in such a system shares alleles with  $r_{n,m} = m^n - (m-1)^n - 1$  others.

We note that the assumptions of Gupta et al, satisfy the rule that cross-immunity is determined solely by the most related of the prior infections. Thus in the case of reduction in infectivity, the model reduction scheme (2.2–2.3) from section 2 applies.

**4.2. Equilibria and stability.** The cross-immunity structure in these allele-systems is sufficiently simple to allow us to obtain some general observations.

Let us define a discordant set of strains to be a maximal set such that no two members have any alleles in common. In the two locus, two allele system there are two such sets ( $AB, ab$ ) and ( $Ab, aB$ ) while in the two locus, three allele system four such sets exist, again each with 2 members. More generally, for the  $n$  locus,  $m$  allele system there are  $(m!)^{n-1}$  discordant sets each with  $m$  members. Boundary equilibria consisting only of a discordant set of strains play a central role in the dynamics of these systems since for a region of parameter space in which cross-protective responses are large (small  $c$ ) they appear to be the only stable equilibria. In essence, when cross-protection is high, competitive exclusion (akin to niche selection

[10]) acts to eliminate immunologically overlapping strains. The particular discordant set selected is determined by initial conditions.

Our invasion result from the previous section shows that the stability of any boundary equilibrium consisting of a discordant set will depend only on the magnitude of  $c = \tau\gamma$  and thus is independent of the way cross-immunity acts. In the case of a completely symmetric system in which all strains have the same  $R_0$ , it is straightforward to show in the general  $n$  locus,  $m$  allele case that the boundary equilibria of discordant sets of strains are stable to invasion so long as  $\tau\gamma < \frac{1}{2R_0}$ .

If all strains are identical in their infectivity  $\beta_i$ , duration of infection  $\sigma$ , and strength of the cross-immunity response  $c$ , then the cross-immunity structure is completely symmetrical in all strains and a result by Andreasen *et al.* [3] shows that there exists a unique symmetric equilibrium if and only if  $R = R_i > 1$ . This result holds irrespective of whether cross-immunity reduces susceptibility or infectivity. However, the stability of this equilibrium depends on the parameter values. We have not succeeded in calculating the stability of this completely symmetric equilibrium for the general case of mixed susceptibility and infectivity mediated cross-immunity, but for the infectivity-only case (*i.e.* all  $\tau_J^i = 1$ ) the equilibrium is stable provided that

$$x(1-x)^2 R_0 (\sigma/\mu + 1) \left( 1 - \left[ (m-1)^{(n-2)} + 1 \right] (1-c) \right) + \\ ([m^n - (m-1)^n - 1] x + 2) ([m^n - (m-1)^n - 1] x + 1) > 0$$

and

$$c ([m^n - (m-1)^n - 1] x + 1)^2 - 1 + c > 0$$

where  $R_0 = \beta/(\sigma + \mu)$  and  $x$  is the proportion of the population that have ever been exposed to any particular strain, given by the larger root of:

$$c [m^n - (m-1)^n - 1] x^2 - ([m^n - (m-1)^n - 1] [c - 1/R_0] - 1) x - \\ (1 - 1/R_0) = 0$$

We do not show the derivations of these results here; while lengthy, much of the detail is mundane.

Figure 1 shows threshold values of  $c$  (derived from the inequalities above) as a function of  $R_0$ ,  $(\sigma + \mu)/\mu$  and the number of loci and alleles modelled. Note that the number of loci and number of alleles affect stability in approximately the same manner, so we only show results for cases where the number of loci and alleles are the same. As both the number of loci and alleles increases the range of values of  $c$  for which the symmetric equilibrium is stable narrows, and structuring of the pathogen population into discordant sets becomes more likely.

The additional symmetry of the 2 locus 2 allele system (the fact a discordant set is made up of exactly half of all genotypes) means that, for

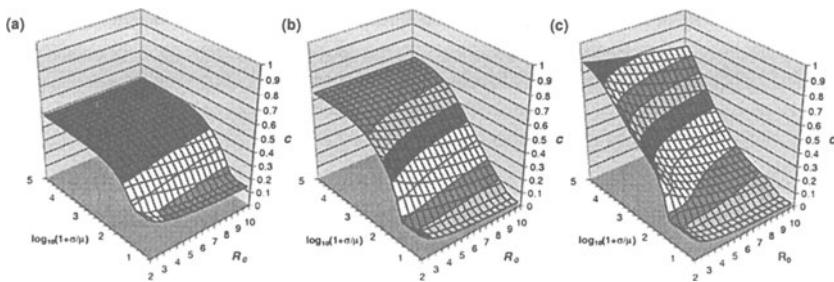
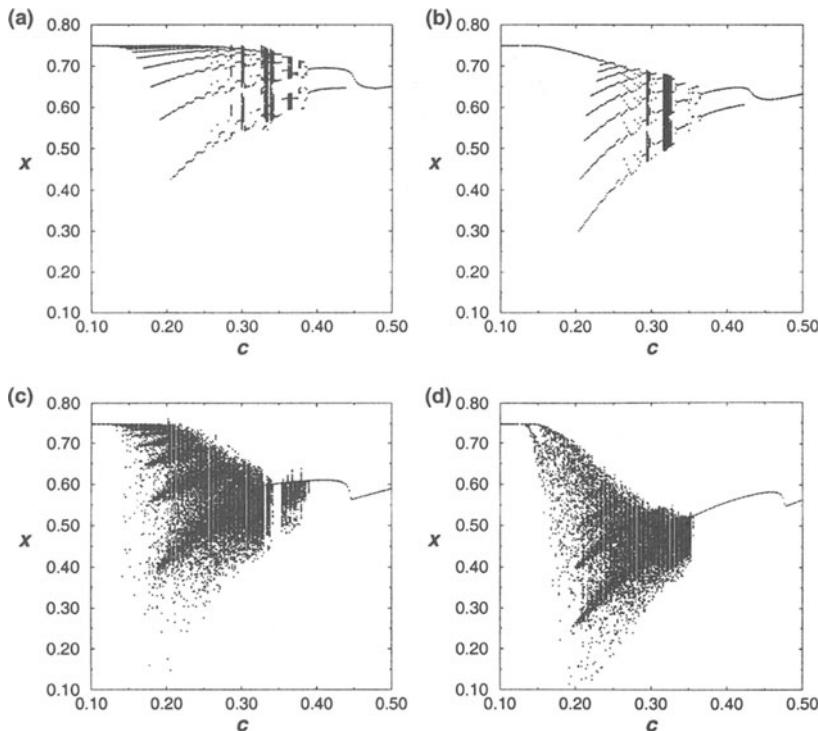


FIG. 1. Values of cross-protection,  $c$ , above which the symmetric internal equilibrium (all strains coexisting) is stable, as a function of the key dimensionless parameters  $R_0$  and  $(\sigma + \mu)/\mu$ , for systems with (a) 2 loci and 2 alleles, (b) 3 loci and 3 alleles, and (c) 4 loci and 4 alleles; cross-immunity is mediated through reduction in infectivity.

values of  $c$  just above the  $\frac{1}{2R_0}$  threshold at which the boundary equilibria become stable (but below the threshold of stability for the symmetric internal equilibrium), additional asymmetric internal equilibria exist. Numerical exploration indicates that where such equilibria exist, they appear to be locally stable, and no stable limit cycles or chaotic solutions exist. At  $c = \frac{1}{2R_0}$  these asymmetric equilibria become identical with the boundary equilibria.

Figure 1 also shows that in some cases (low  $R_0$  and low  $\sigma/\mu$ ), the symmetric equilibrium can be locally stable below the threshold  $c = \frac{1}{2R_0}$  where the symmetric boundary equilibria (and several asymmetric such equilibria) are also locally stable. For these cases, initial conditions determine the final state, though the basins of attraction of the two types of equilibria can differ markedly in size (with the boundary equilibria being the more common final states).

**4.3. Limit cycle behaviour.** For parameter regions where neither the discordant set boundary equilibria nor the fully symmetric equilibrium is stable, limit cycles and chaotic attractors dominate system dynamics, except where asymmetric internal equilibria exist (see above). Analysis of the dynamical behaviour in this regime is largely limited to numerical investigation, and Figure 2 shows bifurcation plots for the 2 locus, 2 allele and 2 locus, 3 allele systems for both infectivity and susceptibility mediated cross-protection. While the dynamics are very similar, infectivity and susceptibility cases differ in the precise details of the locations at which bifurcations occur. Note that while these figures show changes in system dynamics as a function of  $c$ , for fixed  $R_0$  and  $\sigma/\mu$ , a full characterisation of system behaviour would require exploration of the full 3 dimensional parameter space. This is, however, currently prevented by computational limitations - particularly for the susceptibility inhibition case where all unique immune histories need to be tracked.



**FIG. 2.** Bifurcation diagrams showing limit cycle and chaotic behaviour of genetic cross-immunity model as a function of  $c = \gamma\tau$ , the degree of cross-protection between strains sharing alleles. Diagrams were constructed by plotting all maxima of  $x$ , the proportion exposed to genotype  $i$  (where  $i$  is fixed but arbitrary), over a 8000 year period, with an initial 12000 years of each solution being discarded to allow equilibration. Solutions were integrated using Burlisch-Stoer methods. Parameter values:  $\mu = 0.02/\text{year}$ ,  $\sigma = 10/\text{year}$ ,  $R_0 = 4$ . 4 different model variants are shown: (a) 2 locus, 2 allele system with infectivity mediated immunity,  $\tau = 1$ , (b) 2 locus, 2 allele system with susceptibility mediated immunity,  $\gamma = 1$ , (c) 3 locus, 2 allele system with infectivity mediated immunity,  $\tau = 1$ , (d) 2 locus, 2 allele system with susceptibility mediated immunity,  $\gamma = 1$ .

These systems exhibit a wide range of complex behaviours in the parameter regime where no equilibrium is stable, including simple limit cycles, deterministic intermittency (in the 3 locus 2 allele case, the regions around  $c = 0.3$  to  $0.4$  immediately before the simple limit cycle becomes stable), and large-amplitude chaos (low  $c$  for the 3 locus, 2 allele case). It is interesting that chaotic behaviour dominates far more for the 3 allele systems than for the 2 allele systems, but that the 3 allele system shows a chaotic echo of the same limit cycle structure exhibited by the 2 allele system. Furthermore, simple bifurcation plots such as in Figure 2 cannot capture the complexities of the different types of population structuring that are

seen as cross-protection is increased. In particular, these systems exhibit a number of symmetry breaking transformations, such that for regions of moderate  $c$  where limit cycles dominate, all discordant set members have coherent dynamics. For smaller  $c$ , the symmetry between discordant set members is gradually lost, but is then regained as the chaotic oscillations become very large in amplitude and (quasi-)period. These behaviours are illustrated in Figure 3, which shows example time-series from the 2 locus, 2 allele model for both infectivity- and susceptibility-mediated cross immunity and for 3 values of  $c$ :  $c = 0.4, 0.28$  and  $0.15$ . The coherence between members of discordant sets that is seen to exist for the  $c = 0.4$  and  $c = 0.15$  cases is partially lost for  $c = 0.28$ . As the number of alleles is increased, this decoherence becomes more pronounced over a wider range of  $c$  values.

**5. Conclusions.** We have shown how the selective pressures imposed by cross-reactive immunity between antigenically related variants of a pathogen can induce a rich and complex variety of epidemic dynamics at a population level. This paper has concentrated on exploring how the precise mode of action of cross-reactive immunity modulates these dynamics by looking at two extreme cases: where immunity acts to reduce susceptibility to infection, and where it acts to reduce the infectivity of the infected individual. Biological reality for many pathogens will typically lie between these extremes. We have demonstrated that while the mathematical formulation and analysis of systems with susceptibility-mediated immunity is many times more difficult than for infectivity-mediated immunity, system dynamics are - paradoxically but perhaps unsurprisingly - relatively insensitive to the type of immunity action selected. We were disappointingly unable to reduce the dimensionality of the  $O(2^N)$  state variables required to track all immune histories in the susceptibility-mediated immunity case in a manner comparable to what can be achieved for the infectivity-mediated immunity system for simple cross-immunity structures. However, the paper by Gomes & Medley also in this publication offers insight on how to overcome this obstacle for a variety of special cases. The elegant analysis therein should greatly facilitate both the analysis and numerical exploration of the dynamics of these complex systems, particularly when the number of strains becomes large.

Better understanding the differences caused by different modes of action of cross-protective immunity is further motivated by the observation that the only type of cross-reactive immunity that can be easily detected from population-level epidemiological data, most notably strain-specific serology, is that which modifies susceptibility to infection [6]. The reason is clear - when immunity acts to reduce susceptibility, individuals within a population will experience different forces of infection to a particular strain depending on their previous immune history. Where only infectivity is reduced, all individuals experience the same force of infection, and estimation of the degree of cross-reactive immunity through stratification of the host

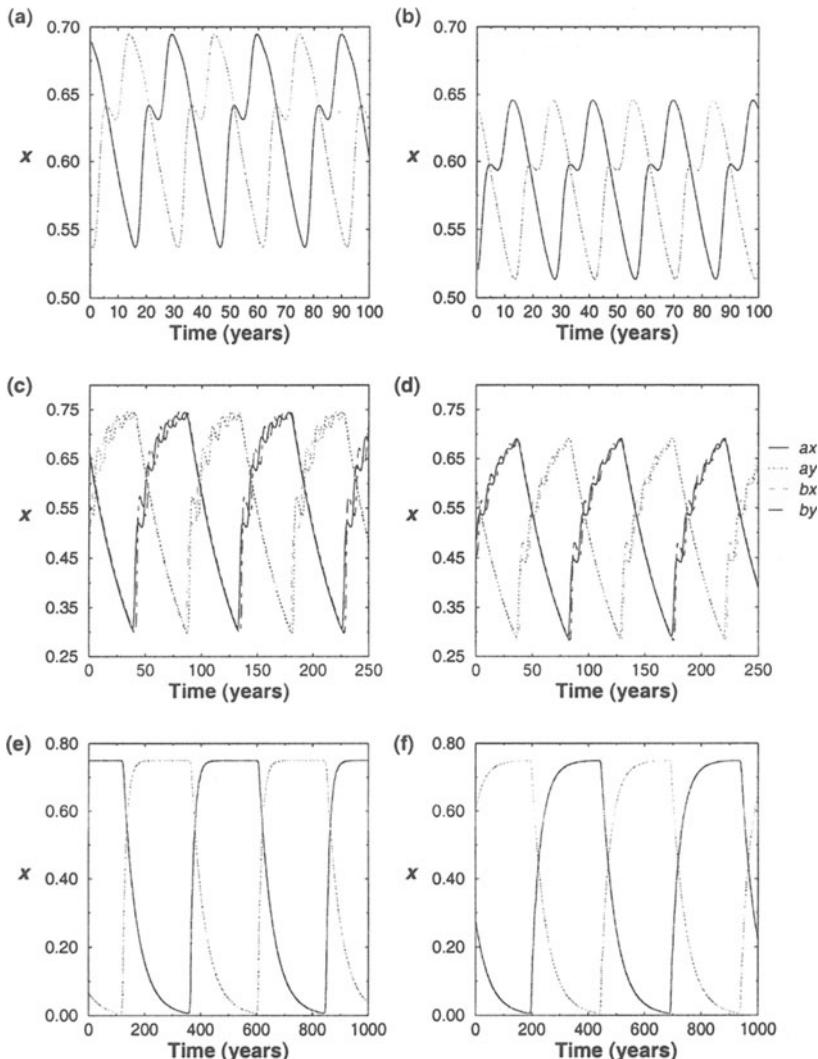


FIG. 3. Example limit-cycle solutions of the 2 locus, 2 allele model, plotting the proportion exposed to each genotype,  $x$ , against time. Graphs (a), (c) and (e) refer to infectivity-mediated cross-immunity, and (b), (d) and (f) to susceptibility-mediated cross-immunity. Note how the dynamics are remarkably similar for the two cases. Graphs (a) and (b) have  $c = 0.4$ , (c) and (d) have  $c = 0.28$ , and (e) and (f) have  $c = 0.15$ . Also note that in graphs (a), (b), (e) and (f), the  $ax$  and  $by$  curves are coincident, as are the  $ay$  and  $bx$  curves. In graphs (c) and (d), this symmetry is broken, if only slightly.

population by immune history becomes impossible. Of course, the optimal way to test the qualitative predictions made by the models explored above is through longitudinal studies collecting strain-specific infection incidence data. Arguably the most scientifically robust means to this goal is to collect and sequence pathogen isolates from infected individuals, since serological data may not reflect the immunologically dominant antigenic structure. However, such studies are open to serious selection biases arising from the fact that most pathogen isolates are collected from hospitalised individuals with serious disease. This is particularly a problem for diseases such as dengue [5, 6], where only a small minority of infections result in serious morbidity and there is evidence that pathogenicity is correlated with secondary infections and strain type.

With the increasing availability of molecular epidemiological data, it is also becoming more important to realistically capture the relationship between pathogen genotype and antigenic phenotype withing transmission models. In this paper we have concentrated on exploring a cross-immunity structure in which a few polymorphic loci code (in an entirely 1 to 1 manner) for a corresponding number of antigenically dominant epitopes. We have further simplified this structure by making all strains identical, and imposing a very simple antigenic distance metric on the system - namely that if 2 strains share any alleles they experience the same degree of cross-immunity irrespective of the number of shared alleles. Only strains sharing no alleles are considered antigenically distinct. This formulation is probably sufficient for cases where a few protein epitopes are antigenically dominant (*e.g.* Neisseria meningitidis B), but less satisfactory where the relationship between genotype and phenotype is more complex - such as when polysaccharide antigens are dominant (*e.g.* Neisseria meningitidis A & C) or when many loci determine antigenic properties (*e.g.* Influenza) and expression variation may exist (*e.g.* Neisseria gonorrhoea). For such systems, it may be more appropriate to model strain structure in a purely phenotypic manner, such as assuming antigenic variants can be ordered by similarity in some simple 1 or 2 dimensional space. This has been the approach adopted by Andreasen *et al.* [2, 3] to explore the determinants of antigenic drift in Influenza. However, in many cases, the dynamics of such systems are still very similar to those of the 'genetic' model used here. Much scope therefore remains for powerful general analyses to give insight into the transmission and evolutionary dynamics of a wide range of antigenically variable pathogens.

## REFERENCES

- [1] E. ACKERMAN, J.I.M. LONGINI, S. SEAHOLM, AND A.S. HEDIN, *Simulation model for viral interference in influenza*, Int. J. Epidemiol., 19 (1990), pp. 444-454.
- [2] V. ANDREASEN, S.A. LEVIN, AND J.LIN, *A model of influenza a drift evolution*, Z. Angew. Math. Mech., 76 (1996), Suppl. 2, pp. 421-424.

- [3] V. ANDREASEN, J. LIN, AND S.A. LEVIN, *The dynamics of cocirculating influenza strains conferring partial cross-immunity*, J. Math. Biol., 35 (1997), pp. 825–842.
- [4] C. CASTILLO-CHAVEZ, H.W. HETHCOTE, V. ANDREASEN, S.A. LEVIN, AND W.M. LIU, *Epidemiological models with age structure, proportionate mixing, and cross-immunity*, J. Math. Biol., 27 (1989), pp. 233–258.
- [5] N. FERGUSON, R. ANDERSON, AND S. GUPTA, *The effect of antibody dependent enhancement on the transmission dynamics and persistence of multiple strain pathogenesis.*, Proc. Natl. Acad. Sci. USA, 96 (1999), pp. 790–794.
- [6] N. FERGUSON, C. DONNELLY, AND R. ANDERSON, *Transmission dynamics and epidemiology of dengue: insights from age-stratified serological surveys*, Phil. Trans. Roy. Soc. Lond. B, 280 (1998), pp. 912–915.
- [7] S. GUPTA, N. FERGUSON, AND R. ANDERSON, *Chaos, persistence and evolution of strain structure in antigenically diverse infectious agents*, Science, 280 (1998), pp. 912–915.
- [8] S. GUPTA, M. MAIDEN, I. FEAVERS, S. NEE, R. MAY, AND R. ANDERSON, *The maintenance of strain structure in populations of recombining infectious agents.*, Nature Medicine, 2 (1996), pp. 437–442.
- [9] J. LIN, V. ANDREASEN, AND S.A. LEVIN, *Dynamics of Influenza A drift: the linear three-strain model*, Math. Biosci., 162 (1999), pp. 33–51.
- [10] R. MAY, *Stability and complexity in model ecosystems*, Princeton University Press, Princeton, 1973.

# DYNAMICS OF MULTIPLE STRAINS OF INFECTIOUS AGENTS COUPLED BY CROSS-IMMUNITY: A COMPARISON OF MODELS

M. GABRIELA M. GOMES\* AND GRAHAM F. MEDLEY\*

**Abstract.** The aim of this paper is to provide an overview of existing models where multiple strains are coupled by cross-immunity. We discuss their differences and similarities, and propose a method to abstract some universal properties intrinsic to the coupling structure. More precisely, the coupling structure of a multiple-strain system can be organized as a matrix that is often invariant under many symmetry operations. Symmetries are known to constrain the behaviour of dynamical systems in many ways. Some symmetry effects are intuitive, but sometimes they can be rather subtle. Given that the assumptions and mechanisms of coupling strains are expected to have a major influence in determining the behaviour of the system, methods and techniques for abstracting their effects are valuable.

**1. Introduction.** Systems of differential equations have become standard tools in epidemiology. Models are made mathematically tractable by making various homogeneity assumptions. There is now a very good agreement amongst researchers about basic modelling approaches to infections where homogeneity can be assumed safely. An overview of the basic model types and their application to specific diseases was given by Anderson and May [1]. Increasing interest is being placed on the effects that population heterogeneities may have on the basic models. In particular, there is evidence that many pathogen populations are structured into a discrete set of strains that interact by conferring partial cross-immunity, and this structure is expected to have important effects on the disease epidemiology. Interaction by enhancement is also known to exist, but here we focus on cross-protection interactions. In order to stress the effects of the strain structure, we assume throughout that the host population is homogeneous.

Several models for multiple strain infections have been proposed in the literature. The pioneering work of Dietz [5] and Castillo-Chavez *et al.* [3] has been greatly extended in recent years. The complexity of multiple-strain models allows a great variability in modelling strategies. In particular, this paper will focus on the deterministic frameworks proposed by Andreasen *et al.* [2, 12], Gupta *et al.* [9, 10], and May and Nowak [15, 16, 19]. A novel framework, which we will not discuss further, is being proposed by Gog and Swinton [8]. Modelling approaches compared in this paper have the unifying feature of describing the dynamical interactions of strains that confer partial cross-protection to infection with other strains. Differences between the models can be due to differences in biological assumptions, differences in which aspects are emphasized, or differences in modelling techniques. One aspect that varies wildly between the referred

---

\*Ecology & Epidemiology Group, Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom. Email: M.G.M.Gomes@warwick.ac.uk.

models is the degree to which details about the structure of the pathogen population are included. In particular, any symmetries in that structure are expected to exert a strong influence in the behaviour of the system and deserve special attention.

Ferguson and Andreasen [7] compared the dynamics of two models under two different cross-immunity actions: in one cross-immunity reduces susceptibility to secondary challenges [2], and in the other cross-immunity reduces the infectivity of a secondary infection [9]. These two different approaches have also been compared by White *et al.* [20]. The main conclusion was that the behaviour governed by the system is rather insensitive to the type of cross-immunity action. To allow a direct comparison to be made, the authors wrote the two systems in the same coordinates and imposed the same symmetries on the strain structure. They succeeded to compare the dynamics with the use of numerical methods. In the adopted coordinates, one model turned out to have a much more complicated form than the other, and analytical comparison was of forbidding complexity.

In this paper we introduce a different system of coordinates, and show how this alternative framework opens the possibility for further analysis. The primary aim of this exercise is to provide tools to investigate the effects that different biological interactions can have on the model dynamics. Comparisons of the actual dynamics are beyond the scope of this paper. Although the formalism used in this paper is very general, some key operations will be illustrated for the particular case of four circulating strains. In all the models throughout this paper the birth rate is defined such that the total population size is constant, and the host population variables are normalized to the total population. Maternal protection is not included in any of the models described here.

Other aspects of multiple-strain systems have been investigated theoretically. Investigations on the effects of vaccination include the work of McLean [17], White *et al.* [20], and Lipsitch [13, 14]. Interactions where infection with one strain enhances the probability of future infection with other strains were also investigated by White *et al.* [20], and Ferguson *et al.* [6]. These different aspects will not be considered further in this paper.

**2. Andreasen, Lin, and Levin model.** In this section we describe a model developed by Andreasen, Lin, and Levin (hereafter referred to as ALL) and reported in [2, 12]. The model provides a detailed description of the dynamics of a host population that is exposed to an arbitrary number,  $n$ , of strains that potentially co-circulate and interact by conferring partial cross-immunity among strains.

### 2.1. The original variables, $(S, I)$ .

**2.1.1. Definition of the  $(S, I)$  coordinates.** The notation of [2, 12] will be used here as much as possible. The set of all strains that may be present is represented by

$$(2.1) \quad \mathcal{N} = \{1, 2, \dots, n\}.$$

The dynamical variables in the model are

$S_{\mathcal{J}}$ : proportion of currently uninfected individuals that have previously been infected by strains in the subset  $\mathcal{J} \subseteq \mathcal{N}$ . There are  $2^n$  such classes.

$I_{\mathcal{J}}^i$ : proportion of individuals who are currently infected by strain  $i \in \mathcal{N}$  and have previously recovered from infections with strains in the subset  $\mathcal{J} \subseteq \mathcal{N} \setminus i$ . There are  $n2^{n-1}$  such classes.

Throughout this paper, we use the same notation for the “proportion of individuals” in a certain class and the “set of individuals” in that class. It will be clear from the context whether we are referring to proportions or sets. Using this convention, the set structure of the uninfected classes is illustrated in the diagram of Fig. 1(a) for the particular case  $n = 4$ . The diagram for classes infected with strain 1 is shown in Fig. 1(b). The structure of classes infected with other strains can be similarly represented. All sets are disjoint, which makes the modelling simple and intuitive.

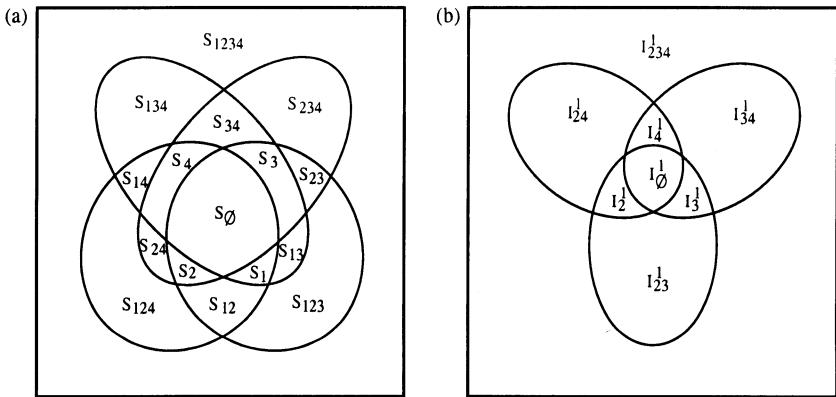


FIG. 1. Diagrams representing the structure of a host population exposed to four co-circulating strains in the coordinates  $(S, I)$ . The uninfected classes,  $S$ , are illustrated in (a), and the classes  $I^1$ , infected with strain 1, are illustrated in (b). Classes infected with other strains have a structure analogous to that shown in (b). All sets are disjoint.

**2.1.2. ALL model in  $(S, I)$  coordinates.** The general form of the model proposed and analysed by ALL consists of a system of ordinary differential equations as

$$(2.2) \quad \dot{S}_{\emptyset} = \mu - \sum_{i \in \mathcal{N}} \sigma_{\emptyset}^i \Lambda^i S_{\emptyset} - \mu S_{\emptyset}$$

$$(2.3) \quad \dot{S}_{\mathcal{J}} = (1 - \mu) \sum_{i \in \mathcal{J}} I_{\mathcal{J} \setminus i}^i - \sum_{i \notin \mathcal{J}} \sigma_{\mathcal{J}}^i \Lambda^i S_{\mathcal{J}} - \mu S_{\mathcal{J}}$$

$$(2.4) \quad \dot{I}_{\mathcal{J}}^i = \sigma_{\mathcal{J}}^i \Lambda^i S_{\mathcal{J}} - (\nu + \mu) I_{\mathcal{J}}^i,$$

where  $\Lambda^i$  is the force of infection of strain  $i$  defined as

$$(2.5) \quad \Lambda^i = \beta_i \sum_{\mathcal{J} \in \mathcal{N} \setminus i} I_{\mathcal{J}}^i.$$

The parameters in equations (2.2-2.5) are

$\mu$ : death rate. The assumption of constant population size is met in this model when the birth rate equals the death rate.

$\nu$ : rate of loss of infectiousness.

$\beta_i$ : transmission coefficient for strain  $i$ , for  $i \in \mathcal{N}$ .

$\sigma_{\mathcal{J}}^i$ : cross-immunity factor. For individuals with immune history  $\mathcal{J} \subseteq \mathcal{N}$ , cross-immunity reduces susceptibility to a strain  $i \in \mathcal{N} \setminus \mathcal{J}$  to a factor  $\sigma_{\mathcal{J}}^i$ . The effect of cross-immunity is total when this factor is 0, and there is no cross-immunity when this factor is 1. Otherwise it will take values between these two extremes.

Note the three terms in the differential equation for the uninfected classes: one additive and two subtractive. The two subtractions correspond to infection and death, respectively. The additive term corresponds to births in the totally susceptible class,  $S_0$ , and recovery from infection in the case of classes  $S_{\mathcal{J}}$ , for  $\mathcal{J} \neq \emptyset$ . The differential equation for the infected classes consists of two terms: one additive and one subtractive. The addition corresponds to new infections, and the subtraction corresponds to loss of infected individuals due to death or recovery from infection. Built into the model is also the assumption that hosts are all born totally susceptible and immunity acquired upon infection with a given strain is permanent. Recovery from infection, which here is assumed to coincide with infectivity, occurs at the rate  $\nu$  and is strain independent .

## 2.2. New variables, $(Z, Y)$ .

**2.2.1. Definition of the  $(Z, Y)$  coordinates.** To facilitate comparison of the ALL model with models proposed and analysed by other authors, we introduce the variables

$Z_{\mathcal{J}}$ : proportion of currently uninfected individuals that have *not* been infected by strains in  $\mathcal{J} \subseteq \mathcal{N}$ . There are  $2^n$  such classes.

$Y_{\mathcal{J}}^i$ : proportion of individuals who are currently infected by strain  $i$  and have *not* been infected by strains in  $\mathcal{J} \subseteq \mathcal{N} \setminus i$ . There are  $n2^{n-1}$  such classes.

The parameters remain unchanged. The set structure of the uninfected host population satisfies the inclusive relations

$$(2.6) \quad Z_{\mathcal{J}_2} \supseteq Z_{\mathcal{J}_1} \quad \text{if} \quad \mathcal{J}_2 \subset \mathcal{J}_1,$$

and similarly for the infected classes

$$(2.7) \quad Y_{\mathcal{J}_2}^i \supseteq Y_{\mathcal{J}_1}^i \quad \text{if} \quad \mathcal{J}_2 \subset \mathcal{J}_1.$$

This structure contrasts with the original coordinates  $(S, I)$ , where all classes are disjoint. The overlapping nature of the host population in the new coordinates  $(Z, Y)$ , makes the model more difficult to describe in the first instance. However, it will facilitate comparison with models proposed by other authors. Furthermore, as will become clear in Sec. 3, the new variables are more convenient when different biological assumptions are made. Uninfected individuals are structured with no overlap with infecteds, as before. Fig. 2(a) illustrates the set structure of the uninfected classes for the case  $n = 4$ . Classes infected with strain 1 are illustrated in Fig. 2(b).

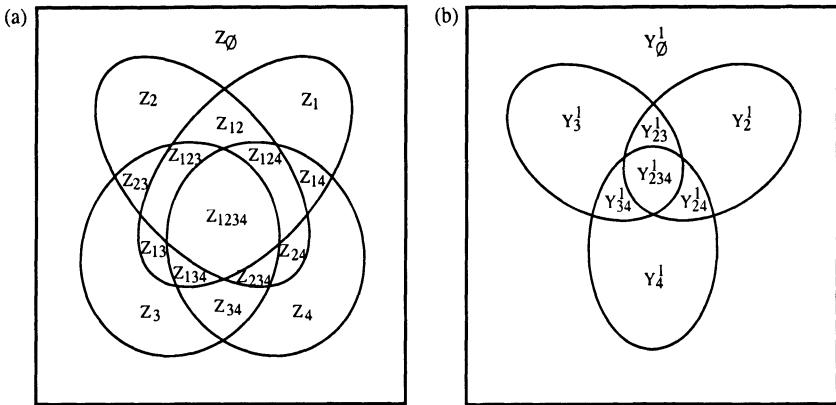


FIG. 2. *Diagrams representing the structure of a host population (equivalent to that in Fig. 1) exposed to four co-circulating strains in the coordinates  $(Z, Y)$ . The uninfected classes,  $Z$ , are illustrated in (a), and the classes  $Y^1$ , infected with strain 1, are illustrated in (b). Classes infected with other strains have a structure analogous to that shown in (b). There is no overlap between  $Z$  and  $Y^i$  classes. However, there are inclusive relations within the uninfected classes, and within the infected classes. Namely,  $Z_{\mathcal{J}_2} \supseteq Z_{\mathcal{J}_1}$  if  $\mathcal{J}_2 \subset \mathcal{J}_1$ , and similarly  $Y_{\mathcal{J}_2}^i \supseteq Y_{\mathcal{J}_1}^i$  if  $\mathcal{J}_2 \subset \mathcal{J}_1$ .*

The new variables are obtained from the original ones by the equations

$$(2.8) \quad Z_{\mathcal{J}} = \sum_{\mathcal{L} \subseteq N \setminus \mathcal{J}} S_{\mathcal{L}}$$

$$(2.9) \quad Y_{\mathcal{J}}^i = \sum_{\mathcal{L} \subseteq N \setminus \{\mathcal{J}, i\}} I_{\mathcal{L}}^i,$$

which define a linear change of coordinates. Some technical details associated with this change of coordinates are described in the appendix. In the following section, we rewrite the system of ordinary differential equations (2.2-2.5) using the new variables.

**2.2.2. ALL model in  $(Z, Y)$  coordinates.** The change of coordinates introduced in (2.8, 2.9) transforms the system (2.2-2.5) into

$$(2.10) \quad \dot{Z}_{\mathcal{J}} = \mu + (1 - \mu) \sum_{i \notin \mathcal{J}} Y_{\mathcal{J}}^i - \sum_{i \notin \mathcal{J}} \Lambda^i C_{\mathcal{J} \cup i}^i(Z) - \sum_{i \in \mathcal{J}} \Lambda^i C_{\mathcal{J}}^i(Z) - \mu Z_{\mathcal{J}}$$

$$(2.11) \quad \dot{Y}_{\mathcal{J}}^i = \Lambda^i C_{\mathcal{J} \cup i}^i(Z) - (\nu + \mu) Y_{\mathcal{J}}^i,$$

where the force of infection,  $\Lambda^i$ , is now written in the simpler form

$$(2.12) \quad \Lambda^i = \beta_i Y_{\emptyset}^i,$$

and the  $C_{\mathcal{J}}^i$  are linear functions of  $Z$  that contain all the information about the cross-protection interactions. These terms have a rather complicated form due to the fact that the new variables characterize the host population in terms of classes that are no longer disjoint. In order to define  $C_{\mathcal{J}}^i$  we introduce some auxiliary notations. Let  $\mathcal{L}$  be any subset of  $\mathcal{N}$ , let  $\ell$  be the order (here meaning number of elements) of  $\mathcal{L}$ , and denote

$\mathcal{O}_m(\mathcal{L})$ : subsets of  $\mathcal{L}$  which have order  $m$ , where  $m$  is an integer between 0 and  $\ell$ .

For concreteness we have, for example, that when  $\mathcal{L} = \{1, 2, 4\}$  then

$$(2.13) \quad \mathcal{O}_2(\mathcal{L}) = \{\{1, 2\}, \{1, 4\}, \{2, 4\}\}.$$

Now letting  $\mathcal{L}$  be a subset of  $\mathcal{N} \setminus \mathcal{J}$  and  $\ell$  the order of  $\mathcal{L}$  as before, we define, for each strain  $i$ , a linear combination of the cross-immunity coefficients as

$$(2.14) \quad \rho_{\mathcal{L}}^i = \sum_{0 \leq m \leq \ell} (-1)^m \sum_{\mathcal{L}_m \subset \mathcal{O}_m(\mathcal{L})} \sigma_{\mathcal{L}_m}^i.$$

Similarly, we define, for each set of strains  $\mathcal{J}$ , a linear combination of the uninfected variables as

$$(2.15) \quad W_{\mathcal{J}, \mathcal{L}}(Z) = \sum_{0 \leq m \leq \ell} (-1)^m \sum_{\mathcal{L}_m \subset \mathcal{O}_m(\mathcal{L})} Z_{\mathcal{J} \cup \mathcal{L}_m}.$$

Finally, we let  $j$  be the order of the set  $\mathcal{J}$ , and write the expressions  $C_{\mathcal{J}}^i$  that appear in the differential equations (2.10, 2.11) as

$$(2.16) \quad C_{\mathcal{J}}^i(Z) = \sum_{0 \leq m \leq n-j} (-1)^m \sum_{\mathcal{L}_m \subset \mathcal{O}_m(\mathcal{N} \setminus \mathcal{J})} \rho_{\mathcal{L}_m}^i W_{\mathcal{J}, \mathcal{L}_m}(Z).$$

**3. Introducing coinfection in ALL model.** So far we have not changed any of the biological assumptions in the ALL model, but have simply changed coordinates. We proceed by making our first change in the model assumptions. ALL assumed that individuals that are presently infected with a certain strain are not susceptible to coinfection by a different strain. For this reason it was essential for the model to keep infected

and uninfected individuals in separate compartments. We change this assumption by allowing coinfection (using the terminology of May and Nowak [16]). More precisely, we assume that individuals presently infected with a certain strain can become coinfected by a different strain to which they have not been previous exposed. The new system is more conveniently modelled by replacing the uninfected variables  $Z$  by the combinations

$$(3.1) \quad X_{\mathcal{J}} = Z_{\mathcal{J}} + \sum_{i \notin \mathcal{J}} Y_{\mathcal{J}}^i,$$

which in words corresponds to

$X_{\mathcal{J}}$ : proportion of *all* individuals that have not been infected by strains in  $\mathcal{J} \subseteq \mathcal{N}$ , no matter if they are currently infected or not. There are  $2^n$  such classes.

As a result we obtain the system of equations

$$(3.2) \quad \dot{X}_{\mathcal{J}} = \mu - \sum_{i \in \mathcal{J}} \Lambda^i C_{\mathcal{J}}^i(X) - \mu X_{\mathcal{J}}$$

$$(3.3) \quad \dot{Y}_{\emptyset}^i = \Lambda^i C_i^i(X) - (\nu + \mu) Y_{\emptyset}^i,$$

where the force of infection,  $\Lambda^i$ , is as in (2.12), and the  $C_{\mathcal{J}}^i$  have the same expressions as described in (2.14-2.16), but are now functions of  $X$  (rather than  $Z$ ). Since now  $X_{\mathcal{J}}$  represents the proportion of all individuals susceptible to all strains in  $\mathcal{J}$  (including uninfected and infected), the terms corresponding to recovery from infection, and infection with strains in  $\mathcal{K} \setminus \mathcal{J}$  are no longer explicit in the equations for  $X_{\mathcal{J}}$ .

Since the force of infection of strain  $i$  is simply  $Y_{\emptyset}^i$  multiplied by the transmission coefficient,  $\beta_i$ , and recovery from infected classes is no longer explicit in the  $X$  equations, the remaining infected classes,  $Y_{\mathcal{J}}^i$ , decouple from the rest of the system and can be omitted. Consequently, we can delete the subscript of the infected variables and, without ambiguity, let  $Y^i$  represent  $Y_{\emptyset}^i$ . This notation will be used from now onwards. The total number of equations is reduced from  $2^n + n2^{n-1}$  to  $2^n + n$ . Recall that this reduction is due to the assumption of coinfection. As will be observed in later sections, further reductions were obtained by other authors by imposing certain symmetries on the cross-immunity matrix, or by using a probabilistic approach to estimate the proportion,  $X_{\mathcal{J}}$ , of individuals susceptible to all strains in  $\mathcal{J}$  as the product  $\prod_{j \in \mathcal{J}} X_j$ .

**4. Cross-immunity matrix.** It is commonly assumed, in particular by ALL, that individuals that have not experienced infection are totally susceptible to all strains. More formally, it is assumed that

$$(4.1) \quad \sigma_{\emptyset}^i = 1,$$

for any  $i \in \mathcal{N}$ . The factors of cross-immunity between any two strains,  $\sigma_j^i$ , for  $i, j \in \mathcal{N}$ , can be organized as the  $n \times n$  matrix

$$(4.2) \quad \Sigma = \begin{pmatrix} \sigma_1^1 & \sigma_1^2 & \cdots & \sigma_1^n \\ \sigma_2^1 & \sigma_2^2 & \cdots & \sigma_2^n \\ \vdots & \vdots & & \vdots \\ \sigma_n^1 & \sigma_n^2 & \cdots & \sigma_n^n \end{pmatrix}.$$

Another common assumption, which will be made here, is that infection with a given strain provides total immunity to further infections with that same strain. More formally, we assume that

$$(4.3) \quad \sigma_i^i = 0,$$

for any  $i \in \mathcal{N}$ . As for the cross-immunity that a set of strains confers to infection with a new strain, ALL make the general assumption that additional immunity cannot decrease cross-protection. However, most of their analytical and numerical results were obtained for the particular case where additional immunity does not alter cross-protection. This is to say that

$$(4.4) \quad \sigma_{\mathcal{J}}^i = \min_{j \in \mathcal{J}} \sigma_j^i.$$

This simplifying assumption is very common in multiple-strain models and will be maintained throughout this paper.

**5. Gupta, Ferguson, and Anderson model.** Much theoretical work on multiple-strain infectious diseases was also carried out by Gupta *et al.* [9, 10]. This section will be mostly concerned with the more recent model studied by Gupta, Ferguson, and Anderson (hereafter referred to as GFA) and reported in Ref. [9]. In our view, one of the most important contributions of the model is to express the multiple-strain pathogen population in terms of their genetic constitution. This approach was first introduced in the earlier work of Gupta *et al.* [10], which will be briefly described at the end of this section for perspective.

**5.1. Pathogen genetics as constraints on the cross-immunity matrix.** Without further assumptions on the structure of the pathogen population, the cross-immunity matrix,  $\Sigma$ , presents many degrees of freedom. Any constraints realistically imposed on this matrix are expected to have a major impact in determining the extent of analytical work. By expressing the various circulating strains in terms of variant alleles at a number of “relevant” loci, Gupta *et al.* [10] reduced the matrix  $\Sigma$  to a much simpler form. The same structure was used later by GFA and further analytical results were obtained. As an illustration, we describe the case where there are two loci, each with two alleles. In this case, four combinations (each defining a strain) can be generated:

- Strain 1: encoded by  $a_1b_1$ ;
- Strain 2: encoded by  $a_1b_2$ ;
- Strain 3: encoded by  $a_2b_2$ ;
- Strain 4: encoded by  $a_2b_1$ .

The assumptions that the cross-protection factor between any strains that share alleles is  $\sigma$ , and that there is no cross-protection (the factor is 1) between strains that do not share alleles, leads to the highly symmetric matrix

$$(5.1) \quad \Sigma = \begin{pmatrix} 0 & \sigma & 1 & \sigma \\ \sigma & 0 & \sigma & 1 \\ 1 & \sigma & 0 & \sigma \\ \sigma & 1 & \sigma & 0 \end{pmatrix},$$

which has only one degree of freedom.

Coincidentally, the same matrix was used in ALL's study of their four-strain model due to the assumption that strains were organized in a ring and cross-protection extended to nearest neighbours only.

Using the strain structure of GFA, a "symmetric" matrix  $\Sigma$  with only one degree of freedom can be uniquely written given just the number of relevant loci and the number of alleles at each loci.

**5.2. GFA model in  $(X, Y)$  coordinates.** The model proposed and analysed by GFA is obtained from our coinfection model (3.2, 3.3) by making the following two alterations:

1. Specify the cross-protection factors as in Sec. 4, where the matrix  $\Sigma$  is deduced from the number of relevant loci and the number of alleles at each loci as illustrated above.
2. Assume that immunity to a given strain does not affect susceptibility to another strain, but only reduces its transmission probability to a factor  $\sigma$ .

The first specifies the configuration of the strain-coupling network and is expressed by the cross-immunity matrix  $\Sigma$ . The second refers to the actual coupling mechanism between each pair of strains (ALL implemented a reduced susceptibility mechanism as described in previous sections, and GFA work with a reduced infectivity mechanism that will be described in this section.)

Implementation of 1 involves complicated combinatorics associated with the action of the cross-immunity matrix. In broad terms, it introduces symmetries in the cross-immunity configuration, and this results in a drastic simplification of the cross-protection terms,  $C_{\mathcal{J}}^i$ , in equations (3.2, 3.3). As these are the terms responsible for coupling the dynamics of the various strains, condition 1 leads to a dimensionality reduction by decoupling many equations from the basic system.

Alteration 2 reflects whether cross-immunity acts by reducing host susceptibility or infectivity to subsequent challenges. These alternatives

were investigated numerically by Ferguson and Andreasen [7], who observed no major effect on the behaviour of the system. Analytical comparisons were then problematic because simplifications induced by 1 were highly dependent on which of the assumptions in 2 had been made. Using the system of coordinates introduced here, the effect of 1 on the  $C_J^i$  terms is independent on the implementation of 2. This separation of effects allows direct analytical comparisons to go much further if desired.

As an illustration in the case of four circulating strains, the general form of  $C_1^1$  according to (2.14-2.16) is

$$\begin{aligned} C_1^1(X) = & \sigma_\emptyset^1 X_1 - (\sigma_\emptyset^1 - \sigma_2^1)(X_1 - X_{12}) - (\sigma_\emptyset^1 - \sigma_3^1)(X_1 - X_{13}) \\ & - (\sigma_\emptyset^1 - \sigma_4^1)(X_1 - X_{14}) \\ & + (\sigma_\emptyset^1 - \sigma_2^1 - \sigma_3^1 + \sigma_{23}^1)(X_1 - X_{12} - X_{13} + X_{123}) \\ & + (\sigma_\emptyset^1 - \sigma_2^1 - \sigma_4^1 + \sigma_{24}^1)(X_1 - X_{12} - X_{14} + X_{124}) \\ & + (\sigma_\emptyset^1 - \sigma_3^1 - \sigma_4^1 + \sigma_{34}^1)(X_1 - X_{13} - X_{14} + X_{134}) \\ & + (\sigma_\emptyset^1 - \sigma_2^1 - \sigma_3^1 - \sigma_4^1 + \sigma_{23}^1 + \sigma_{24}^1 + \sigma_{34}^1 - \sigma_{234}^1) \\ & \times (X_1 - X_{12} - X_{13} - X_{14} + X_{123} + X_{124} + X_{134} - X_{1234}). \end{aligned}$$

By imposing the GFA structure (assumption 1 expressed above) on the cross-immunity matrix, the expression for  $C_1^1$  simplifies to

$$(5.2) \quad C_1^1(X) = X_1 - (1 - \sigma)(X_1 - X_{124}).$$

As suggested by GFA, this simplification generalizes to say that when this genetic structure is assumed, the originally complicated cross-protection terms,  $C_i^i$ , take the simpler form

$$(5.3) \quad C_i^i(X) = X_i - (1 - \sigma)(X_i - X_{\mathcal{I}}),$$

where  $\mathcal{I}$  is the set of all strains that share alleles with  $i$  (including  $i$  itself). Replacing in equation (3.3), we get the new equation for the infected individuals as

$$(5.4) \quad \dot{Y}^i = \Lambda^i [X_i - (1 - \sigma)(X_i - X_{\mathcal{I}})] - (\nu + \mu)Y^i,$$

or equivalently,

$$(5.5) \quad \dot{Y}^i = \Lambda^i [X_{\mathcal{I}} + \sigma(X_i - X_{\mathcal{I}})] - (\nu + \mu)Y^i.$$

The force of infection, remains  $\Lambda^i = \beta_i Y^i$  as in (2.12) throughout this section.

We do not wish to elaborate much further here on the effect of the GFA structure on the more general terms  $C_J^i$ , where  $i \in \mathcal{J} \subset \mathcal{N}$ , because as will be realized below, many such term are removed from the model by

the second assumption of GFA. However, since other modellers may wish to keep these terms we give a brief illustration based on

$$\begin{aligned} C_{12}^1(X) = & \sigma_0^1 X_{12} - (\sigma_0^1 - \sigma_3^1)(X_{12} - X_{123}) - (\sigma_0^1 - \sigma_4^1)(X_{12} - X_{124}) \\ & + (\sigma_0^1 - \sigma_3^1 - \sigma_4^1 + \sigma_{34}^1)(X_{12} - X_{123} - X_{124} + X_{1234}). \end{aligned}$$

By applying the GFA structure (assumption 1 above) the complexity of this term reduces to

$$C_{12}^1(X) = X_{12} - (1 - \sigma)(X_{12} - X_{124}).$$

Implementation of the second assumption corresponds to removal of the cross-protection terms from the equations for the  $X$  variables. This simply transforms equation (3.2) into

$$(5.6) \quad \dot{X}_{\mathcal{J}} = \mu - \sum_{i \in \mathcal{J}} \Lambda^i X_{\mathcal{J}} - \mu X_{\mathcal{J}}.$$

Note that different  $X$  variables are no longer directly coupled. Now coupling depends entirely on equations (5.4) for the infected variables  $Y^i$ .

The system of equations (5.4, 5.6) is that proposed by GFA written in a system of coordinates that we find more convenient for the comparative purposes of this paper. Their set of equations maps into ours by a linear change of coordinates. In order to fully describe the dynamics of this system we only need the equations for the classes  $X_i$ ,  $X_{\mathcal{I}}$  and  $Y^i$ . This makes a total of at most  $3n$  equations. This is considerably less than the  $2^n \times n 2^{n-1}$  equations of the analogous model of ALL, or even the  $2^n + n$  equations of the coinfection model without incorporation of the genetic structure.

Finally, the system originally proposed by Gupta *et al.* [10] is even simpler. It corresponds to replacing the differential equations for the classes  $X_{\mathcal{I}}$  by the products  $\prod_{i \in \mathcal{I}} X_i$ , reducing the number of equations to  $2n$ . To our knowledge, there is no explicit comparison of the two approaches in the literature. It would be valuable to understand to what extent using the products  $\prod_{i \in \mathcal{I}} X_i$  (instead of the differential equations for  $X_{\mathcal{I}}$ ) alters the behaviour of the system.

**6. May, and Nowak model.** Models of multiple-strain infections were also developed by May and Nowak (hereafter referred to as MN) and reported in [15, 16, 19]. As the primary aim of the referred series of papers is to investigate the evolution of pathogen virulence, special attention was given to the death rate,  $v_i$ , induced by each individual strain on its hosts. The authors ordered the  $n$  strains by increasing virulence ( $v_i < v_{i+1}$ , for  $1 \leq i \leq n$ ), and developed two different models based on different rules for the interaction of different strains. As it has been the case throughout this paper, it is assumed that deaths are balanced by births so the total population size is constant and all variables normalized to the total population. An assumption made by MN that has not been met before in this paper

is the absence of recovery from infection with acquired immunity. Infected hosts are infectious until their death. Since deaths are balanced by births of totally susceptible individuals, a subtle consequence of this assumption is to generate a larger pool of totally susceptible individuals (comparatively to models that include recovery into immune classes).

In the interest of consistency of notation, the MN models will be written here in notation that is slightly different from their original versions. The variables in the equations will be the following:

$Z$ : proportion of uninfected individuals. Due to the absence of recovery from infection into immune classes, this is the same as the proportion of totally susceptible individuals.

$Y^i$ : proportion of individuals who are currently infected by strain  $i$ , for  $1 \leq i \leq n$ .

**6.1. MN superinfection model.** In this first model developed and analysed by MN [15, 19], it is assumed that any strain,  $i$ , can infect hosts that are either totally susceptible or infected by a less virulent strain,  $j < i$ . When this happens, it is assumed that infection with the less virulent strain ends. Using the terminology of MN, a more virulent strain can “take over” hosts that are already infected with a less virulent strain, and this process is called “superinfection”. It is assumed that superinfection occurs at a rate  $\sigma$  relatively to infection of uninfected hosts. The equations modelling the dynamics of the  $n$  different classes of infected individuals are

$$(6.1) \quad \dot{Y}^i = \Lambda^i \left( Z + \sigma \sum_{j=1}^{i-1} Y^j \right) - \sum_{j=i+1}^n \Lambda^j Y^j - (v_i + \mu) Y^i,$$

where  $\Lambda^i$  is the force of infection of strain  $i$  defined as before

$$(6.2) \quad \Lambda^i = \beta_i Y^i,$$

and the parameters are the death rate,  $\mu$ , the cross-protection factor between strains,  $\sigma$ , the transmission coefficients,  $\beta_i$ , and virulence,  $v_i$ , associated with individual strains. The number of uninfected individuals is simply obtained as  $Z = 1 - \sum_{j=1}^n Y^j$ .

In this case, interactions between strains can be represented by the “asymmetric” cross-immunity matrix

$$(6.3) \quad \Sigma = \begin{pmatrix} 0 & \sigma & \sigma & \cdots & \sigma \\ 0 & 0 & \sigma & \cdots & \sigma \\ 0 & 0 & 0 & \cdots & \sigma \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

This particular matrix translates the assumptions that hosts cannot be infected by the same strain more than once, and hosts infected with some

strain,  $j$ , can be infected by a different strain  $i$  if and only if  $i > j$ . This secondary infection happens with a probability reduced to a factor  $\sigma$ .

**6.2. MN coinfection model.** A second model proposed by MN [16] is built under the assumption that individual hosts can be simultaneously infected by any number of different strains. This process is called “coinfection”. In this case, the model equations for the classes of individuals infected with each of the  $n$  strains (including individuals coinfecte with other strains) are

$$(6.4) \quad \dot{Y}^i = \Lambda^i [Z + \sigma(Y - Y^i)] - (\bar{v}_i + \mu)Y^i,$$

where  $Y = 1 - Z$  is the total number of infected hosts, and the proportion of uninfected individuals is now defined as the product

$$(6.5) \quad Z = \prod_{j=1}^n (1 - Y^j).$$

Note that here, in contrast with the superinfection model of section 6.1, there is overlap between classes infected with different strains and the proportion of infected individuals is generally less than the sum over all infected classes. The forces of infection,  $\Lambda^i$ , and the parameters  $\mu$ ,  $\sigma$ , and  $\beta_i$  are as those defined for model (6.1). The average disease induced death rate (or virulence), denoted by  $\bar{v}_i$ , is now more complicated. Given that hosts may simultaneously harbour several pathogen strains, MN assume that the death rate of an infected host is determined by the most virulent strain present. In order to estimate the average virulence, they estimate the probability that a host is not infected with a strain more virulent than  $i$  as the product

$$(6.6) \quad P_i = \prod_{j=i+1}^n (1 - Y^j),$$

leading to the average virulence of strain  $i$  as

$$(6.7) \quad \bar{v}_i = v_i P_i + \sum_{j=i+1}^n v_j Y^j P_j.$$

This approach of multiplying probabilities is, in essence, similar to that used by Gupta *et al* [10] and briefly referred at the end of Sec. 5. Again we stress our concern that, to our knowledge, there is no comparative study of this probabilistic approach with one that would incorporate extra differential equations for classes of individuals infected with different numbers of strains. The similarity between equations (5.5) of GFA (and [10]) and equations (6.4) of MN is also worth noting.

The model studied by MN was restricted to the case of no cross-protection between different strains ( $\sigma = 1$ ). We introduce this term to facilitate comparison with other models, and to highlight the cross-immunity matrix, which in this case takes the “symmetric” form

$$(6.8) \quad \Sigma = \begin{pmatrix} 0 & \sigma & \sigma & \cdots & \sigma \\ \sigma & 0 & \sigma & \cdots & \sigma \\ \sigma & \sigma & 0 & \cdots & \sigma \\ \vdots & \vdots & \vdots & & \vdots \\ \sigma & \sigma & \sigma & \cdots & 0 \end{pmatrix}.$$

This particular matrix translates the assumptions that hosts cannot be infected by the same strain more than once, and hosts infected with some strain,  $j$ , can be infected by any different strain,  $i$ , with a probability reduced to a factor  $\sigma$ . Here the cross-protection factor is the same between any pair of different strains.

**7. Discussion.** Different models describing the dynamics of  $n$  distinct strains of infectious agents coinfecting a host population are compared in this paper. Strains interact due to the assumption that previous infection with a given strain may confer partial cross-immunity to infection with other strains. To ease comparison, all models were written in the same system of coordinates, and notation was made uniform. In all models considered, the birth rate is defined such that the total population size is constant, and all hosts are born totally susceptible. As a result of this comparison, six points were identified as the main differences between models proposed by different authors. We proceed by describing these points and raising some questions.

1. *Coinfection of a host by different strains.* In the ALL model, it is assumed that an infected host cannot be coinfecte by a different strain. Because of this assumption, it was necessary for the model to keep records of the proportions of uninfected and infected individuals with all possible immune histories. Such model consists of  $2^n + n2^{n-1}$  equations. We introduce coinfection in the ALL model, in the sense that a host can be simultaneously infected by any number of strains. As a result, proportions of uninfected and infected individuals can be recorded together, as long as they have the same history of past infections. Due to this assumption, the number of equations reduced to  $2^n + n$ . The model of GFA, and one of the models of MN, also reflect the assumption of coinfection. Can we make this simplifying assumption safely?

2. *Details of a single cross-immunity interaction.* The ALL model is built under the assumption that immunity to a given strain,  $j$ , reduces susceptibility to another strain,  $i$ , to a factor  $\sigma_j^i$ , and as a result it reduces its transmission probability to the same factor. In contrast, GFA assume

that immunity to strain  $j$  reduces the transmission probability of strain  $i$  to a factor  $\sigma_j^i$ , without affecting its susceptibility. The resulting model is slightly simpler. The immediate consequence of this difference is that the GFA model is building less immunity in the host population than the ALL model. How do the long term consequences of the two assumptions compare?

3. *Symmetries of the cross-immunity matrix.* The cross-immunity factors between different pairs of strains can be organized as a  $n \times n$  matrix,  $\Sigma$ , which we call “cross-immunity matrix”. In the model systems, this matrix has the important role of coupling several differential equations. In its general form, the matrix  $\Sigma$  presents many degrees of freedom, which makes a thorough mathematical analysis very difficult (and the generality of the result would make it basically unapplicable). As has been emphasized throughout this paper, researchers deal with this problem by imposing symmetry constraints on  $\Sigma$ . To a certain extent, there is a biological motivation behind the choice of these symmetries, but there is also an unavoidable degree of arbitrariness. Can we reduce this degree of arbitrariness by finding good agreements between epidemiological data and the behaviour of symmetric systems?

4. *Estimate of the proportion of hosts with a given immune history.* Different models use different estimates for the proportion of hosts susceptible to all strains in a set  $\mathcal{J}$ . For example, ALL and GFA include differential equations describing the dynamics of proportions of hosts with all the relevant immune histories. In contrast, some authors (for example, Gupta *et al.* [10] and MN [16]) estimate the proportion of individuals susceptible to all strains in  $\mathcal{J}$  as the product  $\prod_{j \in \mathcal{J}} X_j$ , which makes the model considerably simpler. To what extent does this simplifying assumption alter the behaviour of the system?

5. *Loss of infectiousness.* Excluding natural mortality, there are two ways by which hosts leave the infectious classes: they either recover and become immune (as ALL and GFA), or they die due to infection (as MN). The dynamical equations for the infectious classes do not explicitly distinguish between the two cases. The distinction is made in the susceptible classes. In the case of recovery with acquired immunity, hosts leave the infectious classes to enter classes that have some immunity. In contrast, deaths due to infection have the effect of increasing the birth rate of totally susceptible individuals due to the assumption of constant population size. What are the long term effects of this increased pool of totally susceptible individuals?

6. *Distributions of transmission rates, and loss of infectiousness rates.* The symmetries of the cross-immunity matrix translate into symmetries of the dynamical system if and only if they are not broken by differences in transmission rates or differences in loss of infectiousness rates. What is a realistic distribution for these rates, and how does it relate to the structure of the cross-immunity matrix?

A detailed comparison of the actual models is a problem on its own, and at this stage we do not wish to make formal statements about the effects of model differences in the associated behaviour. What we can do safely at this stage, is to discuss general results that can be extrapolated from the extensively studied fields of bifurcation theory and dynamical systems to the particular systems reported in this paper. In particular, we would like to speculate about potential consequences of symmetries that are often imposed.

*A. Strain formation as a symmetry-breaking bifurcation.* Multiple-strain models are usually invariant under many symmetry operations as we described in this paper. Symmetry assumptions have the effect of greatly reducing the model complexity and increase the extent of analytical treatment. Different biological justifications for the imposed symmetries have been proposed by different authors (in particular, Gupta et al. [10, 9] and Andreasen *et al* [2]). Proposed models fit in the general theme of coupled cell systems for which general theoretical results have been developed. Using a different approach but similar motivations, Cohen and Stewart [4] modelled the process of speciation as symmetry-breaking bifurcation in a rather general context. Strain formation is an analogous process, and these symmetry and symmetry-breaking approaches are expected to suggest answers to fundamental questions. For example, results of Ref. [4] suggest that the number of species (strains) that stably cocirculate in a population can be related to the symmetry of the speciation (strain formation) process.

*B. Heteroclinic cycles and intermittent behaviour.* In symmetric systems, heteroclinic cycles can be robust for symmetry preserving perturbations and stable (see Krupa [11] for a review). From [11]: “A stable cycle defines a mechanism of intermittency – a solution approaching it spends long periods near equilibria and makes fast transitions from one equilibrium to the next. In a perfectly symmetric system the return times increase monotonically and rapidly approach infinity, thus making the intermittent behaviour uninteresting. However, under small, symmetry breaking perturbations, the cycling behaviour persists (even if there is no longer a cycle) and the transition times no longer converge to infinity. ... Hence, in applications, the existence of a stable heteroclinic cycle in the idealized model can be linked to the occurrence of intermittency.” Heteroclinic cycles and periodic orbits were found by ALL [2, 12] after imposing some symmetries in their multiple-strain model, and some analogy was made with observed cyclic epidemics of influenza following the appearance of the first pandemic of a new subtype.

*C. Regular and irregular burst.* It has been shown (Moehlis and Knobloch [18]) that a perturbation of certain symmetric systems can lead to regular and irregular bursts of large dynamic range. In these situations, each burst

is followed by oscillations of decreasing amplitude until another burst occurs. It is tempting to make analogy between this process and the pattern of influenza epidemiology described above.

To our knowledge, it is not known how realistic are the symmetries of the cross-immunity matrix or the symmetries in distributions of transmission rates and loss of infectiousness rates, and in our view, they should be imposed with caution. The effects of symmetry in the behaviour of multiple-strain systems deserve careful investigation.

**8. Conclusions.** A variety of models have been proposed to investigate the behaviour of multiple-strain pathogen populations where “distinct” strains interact by cross-reactivity of the host immune response. Comparison of existing models has been largely restricted to numerical methods due to technical difficulties with mathematical analysis. This paper introduces a framework that facilitates model comparisons to be made analytically.

**Acknowledgements.** MGMG acknowledges the support of the Wellcome Trust, and the hospitality of the IMA, University of Minnesota, where this work was initiated. GFM acknowledges the support of the British Biotechnology & Biological Science Research Council (E08527).

## APPENDIX

**A. Change of coordinates from  $(S, I)$  to  $(Z, Y)$ .** Here we investigate some technicalities associated with the change of coordinates (2.8, 2.9). First we describe in detail the restriction to the  $2^n$  uninfected classes. We then generalize the results to each of the  $n$  groups of  $2^{n-1}$  classes of individuals currently infected with strain  $i$ , for any  $i \in \mathcal{N}$ .

**A.1. Ordering the uninfected variables.** A linear change of coordinates can be represented by a matrix,  $P_n$ , with nonzero determinant, and such matrix depends on the way that the variables are ordered. In this section we order the variables in a way that will lead to a convenient matrix.

Let  $\{0, 1\}^n$  represent the set of vectors with  $n$  components whose values are 0 or 1, and let the vector  $A = (A_1, A_2, \dots, A_n)$  represent a generic element of  $\{0, 1\}^n$ . A one-to-one correspondence between elements of  $\{0, 1\}^n$  and subsets of  $\mathcal{N}$  is established by defining  $\mathcal{J}(A)$ , such that strain  $i$  belongs to  $\mathcal{J}(A)$  if and only if  $A_i = 1$ . For example,

$$(A.1) \quad \mathcal{J}(1, 0, 1, 1) = \{1, 3, 4\}.$$

We proceed by ordering the original uninfected classes as

$$(A.2) \quad S_{\mathcal{J}(A^{2^n})}, S_{\mathcal{J}(A^{2^{n-1}})}, \dots, S_{\mathcal{J}(A^2)}, S_{\mathcal{J}(A^1)},$$

and the new uninfected classes in the reversed order

$$(A.3) \quad Z_{\mathcal{J}(A^1)}, Z_{\mathcal{J}(A^2)}, \dots, Z_{\mathcal{J}(A^{2^n-1})}, S_{\mathcal{J}(A^{2^n})},$$

where  $A^j$  are the vectors in  $\{0, 1\}^n$  ordered as follows. Let  $A^1$  be the zero vector

$$A^1 = (\overbrace{0, 0, \dots, 0}^n),$$

and generate  $A^2$  by introducing a 1 in the first component

$$A^2 = (1, \overbrace{0, \dots, 0}^{n-1}).$$

Then complete the sequence of vectors by induction

$$(A.4) \quad A^{j+1} = s(A^j),$$

with the aim of generating vectors in an order that increases the number of 1s and their position in the vector until the all 1 vector

$$A^{2^n} = (\overbrace{1, 1, \dots, 1}^n)$$

is reached. More precisely,  $s$  operates on a generic vector

$$A = (A_1, A_2, \dots, A_n) \in \{0, 1\}^n$$

as follows. Let  $k < n$  be the largest nonnegative integer such that  $A_i = 0$  for all  $i$  such that  $n - k < i \leq n$ . This is to say that the last  $k$  components of the vector  $A$  have value 0 and the preceding component has value 1. The definition of  $s$  depends on whether  $k$  is strictly positive or zero.

1. If  $0 < k < n$ , then move the last 1 one position forward. More formally, in this case we have

$$\begin{aligned} s(A_1, \dots, A_{n-k-1}, 1, \overbrace{0, 0, \dots, 0}^k) \\ = (A_1, \dots, A_{n-k-1}, 0, 1, \overbrace{0, \dots, 0}^{k-1}). \end{aligned}$$

2. If  $k = 0$ , then the last position of the vector  $A$  has value 1, and there is no forward position to move it to. In this case, let  $\ell \leq n$  be the largest positive integer such that  $A_i = 1$  for all  $i$  such that  $n - \ell < i \leq n$ . Furthermore, let  $m \leq n - \ell$  be the largest positive integer such that  $A_i = 0$  for all  $i$  such that  $n - \ell - m < i \leq n - \ell$ . This is to say that the last  $\ell$  components of the vector  $A$  have value 1, and the preceding  $m$  components have value 0. The definition of  $s$  depends on  $\ell$  and  $m$  as follows:

- 2.1. If  $0 < \ell < n$  and  $0 < m < n - \ell$ , then  $A_{n-\ell-m} = 1$ . In this case, move this 1 one position forward and the last  $\ell$ -block of 1s back as

$$\begin{aligned} s(A_1, \dots, A_{n-\ell-m-1}, 1, \overbrace{0, \dots, 0}^m, \overbrace{1, \dots, 1}^\ell) \\ = (A_1, \dots, A_{n-\ell-m-1}, 0, \overbrace{1, \dots, 1}^{\ell+1}, \overbrace{0, \dots, 0}^{m-1}). \end{aligned}$$

- 2.2. If  $0 < \ell < n$  and  $m = n - \ell$ , then introduce a new 1 in the first component,  $A_1$ , and move the last  $\ell$ -block of 1s back as

$$s(\overbrace{0, \dots, 0}^{n-\ell}, \overbrace{1, \dots, 1}^\ell) = (\overbrace{1, \dots, 1}^{\ell+1}, \overbrace{0, \dots, 0}^{n-\ell-1}).$$

- 2.3. If  $\ell = n$ , then the  $2^n$  elements of the set  $\{0, 1\}^n$  have been generated and the construction ends.

The construction just described is illustrated for  $n = 4$  as follows:

$A^1$	$A^2$	$A^3$	$A^4$	$A^5$	$A^6$	$A^7$	$A^8$	$A^9$	$A^{10}$	$A^{11}$	$A^{12}$	$A^{13}$	$A^{14}$	$A^{15}$	$A^{16}$
$\downarrow$															
0	1	0	0	0	1	1	1	0	0	0	1	1	1	0	1
0	0	1	0	0	1	0	0	1	1	0	1	1	0	1	1
0	0	0	1	0	0	1	0	1	0	1	1	0	1	1	1
0	0	0	0	1	0	0	1	0	1	1	0	1	1	1	1

Recall that the order of this set of vectors determines the order of the variables  $S_J$  and  $Z_J$ , as in (A.2) and (A.3), respectively. This order is a key ingredient to uncover certain features of the matrix that changes the coordinates  $S_J$  into  $Z_J$ .

**A.2. Change of coordinates.** There is a change of coordinates that transforms the original uninfected classes,  $S_J$  ordered as in (A.2), into the new uninfected classes,  $Z_J$  ordered as in (A.3). Given a number  $n$  of strains, we can determine the matrix  $P_n$  of dimensions  $2^n \times 2^n$  such that

$$(A.5) \quad \begin{pmatrix} Z_{J(A^1)} \\ Z_{J(A^2)} \\ \vdots \\ Z_{J(A^{2^n-1})} \\ Z_{J(A^{2^n})} \end{pmatrix} = P_n \begin{pmatrix} S_{J(A^{2^n})} \\ S_{J(A^{2^n-1})} \\ \vdots \\ S_{J(A^2)} \\ S_{J(A^1)} \end{pmatrix}.$$

Some important properties shared by the matrices  $P_n$  for all  $n$  are

1.  $P_n$  is an upper triangular matrix.
2. All diagonal entries on  $P_n$  are 1.

Consequently,  $P_n$  has determinant 1. Having a nonzero determinant is essential in the definition of a change of coordinates, and determinant 1 is important to insure that the population size is not being altered. The two properties above can be formally proved by tedious combinatorial arguments, which we omit. A good intuition can be obtained by inspection on the matrix  $P_4$  shown in the change of coordinates below.

$$\left( \begin{array}{c} Z_\emptyset \\ Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Z_{12} \\ Z_{13} \\ Z_{14} \\ Z_{23} \\ Z_{24} \\ Z_{34} \\ Z_{123} \\ Z_{124} \\ Z_{134} \\ Z_{234} \\ Z_{1234} \end{array} \right) = \left( \begin{array}{cccccccccccccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{array} \right) \left( \begin{array}{c} S_{1234} \\ S_{234} \\ S_{134} \\ S_{124} \\ S_{123} \\ S_{34} \\ S_{24} \\ S_{23} \\ S_{14} \\ S_{13} \\ S_{12} \\ S_4 \\ S_3 \\ S_2 \\ S_1 \\ S_\emptyset \end{array} \right)$$

The lower triangle of the matrix  $P_4$  is empty meaning that all entries are zero in that region.

By analogy, we proceed by constructing a change of coordinates for the  $2^{n-1}$  classes of individuals infected with a given strain. After appropriate ordering (by analogy with Sec. A.1), classes,  $I^1$ , of individuals infected with a given strain are transformed into the new coordinates,  $Y^1$ , by the  $2^{n-1} \times 2^{n-1}$  the matrix  $P_{n-1}$ . Following a general trend of this paper, we give the case  $n = 4$  as example. In this case we show how the matrix  $P_3$  transforms the classes infected with strain 1. This is,

$$\left( \begin{array}{c} Y_\emptyset^1 \\ Y_1^1 \\ Y_2^1 \\ Y_3^1 \\ Y_4^1 \\ Y_{23}^1 \\ Y_{24}^1 \\ Y_{34}^1 \\ Y_{234}^1 \end{array} \right) = \left( \begin{array}{cccccccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right) \left( \begin{array}{c} I_{234}^1 \\ I_{34}^1 \\ I_{24}^1 \\ I_{23}^1 \\ I_4^1 \\ I_3^1 \\ I_2^1 \\ I_0^1 \end{array} \right).$$

Classes infected with other strains are transformed by the same matrix  $P_3$  as long as the variables are appropriately ordered.

## REFERENCES

- [1] R.M. Anderson and R.M. May (1991). *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press.
- [2] V. Andreasen, J. Lin, and S.A. Levin (1997). The dynamics of cocirculating influenza strains conferring partial cross-immunity. *J. Math. Biol.* **35**, 825–842.
- [3] C. Castillo-Chavez, 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.
- [4] J. Cohen, and I. Stewart (1999). Polymorphism viewed as a phenotypic symmetry-breaking. University of Warwick preprint.
- [5] K. Dietz (1979). Epidemiologic interference of virus populations. *J. Math. Biol.* **8**, 291–300.
- [6] N. Ferguson, R.M. Anderson, and S. Gupta (1999). The effect of antibody-dependent enhancement on the transmission dynamics and persistence of multiple-strain pathogens. *PNAS* **96**, 790–794.
- [7] N. Ferguson and V. Andreasen (1999). The influence of different forms of cross-protective immunity on the population dynamics of antigenically diverse pathogens. Submitted to this volume.
- [8] J. Gog and J. Swinton (2000). A status-based approach to multiple strain dynamics. Submitted to *J. Math. Biol.*
- [9] S. Gupta, N. Ferguson, and R.M. Anderson (1998). Chaos, persistence and evolution of strain structure in antigenically diverse infectious agents. *Science* **280**, 912–915.
- [10] S. Gupta, M.C.J. Maiden, I.M. Feavers, S. Nee, R.M. May, and R.M. Anderson (1996). The maintenance of strain structure in populations of recombining infectious agents. *Nature Med.* **2**, 437–442.
- [11] M. Krupa (1997). Robust heteroclinic cycles. *J. of Nonl. Sci.* **7**, 129–176.
- [12] J. Lin, V. Andreasen, and S.A. Levin (1999). Dynamics of influenza A drift: The linear three-strain model. *Math. Biosci.* **162**, 33–51.
- [13] M. Lipsitch (1997). Vaccination against colonizing bacteria with multiple serotypes. *Proc. Natl. Acad. Sci.* **94**, 6571–6576.
- [14] M. Lipsitch (1999). Bacterial vaccines and serotype replacement: Lessons from *Haemophilus influenzae* and prospects for *Streptococcus pneumoniae*. *Emerging Infectious Diseases* **5**, 336–345.
- [15] R.M. May and M.A. Nowak (1994). Superinfection, metapopulation dynamics, and the evolution of diversity. *J. Theor. Biol.* **170**, 95–114.
- [16] R.M. May and M.A. Nowak (1995). Coinfection and the evolution of parasite virulence. *Proc. R. Soc. Lond. B* **261**, 209–215.
- [17] A.R. McLean (1995). Vaccination, evolution and changes in the efficacy of vaccines: A theoretical framework. *Proc. R. Soc. Lond. B* **261**, 389–393.
- [18] J. Moehlis and E. Knobloch (1998). Forced symmetry-breaking as a mechanism for bursting. *Phys. Rev. Lett.* **80**, 5329–5332.
- [19] M.A. Nowak and R.M. May (1994). Superinfection and the evolution of parasite virulence. *Proc. R. Soc. Lond. B* **255**, 81–89.
- [20] L.J. White, M.J. Cox, and G.F. Medley (1998). Cross immunity and vaccination against multiple microparasite strains. *IMA J. Math. App. Med. Biol.* **15**, 211–233.

# VIRULENCE EVOLUTION IN MACRO-PARASITES

ANDREA PUGLIESE\*

**1. Introduction.** Evolutionary issues are very relevant in the comprehension of emerging or re-emerging disease (Ewald 1994): it seems likely that benign diseases may become, by evolutionary changes in virulence or in the capability of evading immune response, a new threat for the health of humans and animals, and thus be considered ‘emerging diseases’ (Dieckmann et al. 2000).

Starting from the seminal papers by Levin and Pimentel (Levin & Pimentel 1981), and Anderson and May (Anderson & May 1982), it has been widely accepted that (in contrast to old evolutionary wisdom) parasite virulence may be adaptive, if it increases the transmission of parasites to new hosts. Most models (Anderson & May 1982, Bremermann & Thieme 1989, Pugliese 2000b) show that, under reasonable assumptions, an intermediate level of virulence should be sustained. A better understanding of these issues may provide advice in the design of therapy and vaccines.

The basic method in these analyses has been formalized by Metz and co-authors (Metz et al. 1996, Dieckmann 1997, Geritz et al. 1998) as adaptive dynamics, although implicitly used by many previous authors. In this method, a phenotypic trait  $x$  is considered to be evolving under the effect of selection and mutation. However, the stochastic interaction of selection and mutation is treated only implicitly; instead, relying on the assumption that the ecological time scale is faster than the rate at which new favorable mutations arise, one first finds the ecological attractor  $E_x$  of the system with the population monomorphic at type  $x$ ; then, one computes the invasion coefficient  $s_x(y)$  of any other feasible type  $y$  in a population close to  $E_x$ . Only types  $y$  such that  $s_x(y) > 0$  may invade into the population, although there is a positive probability that any mutant will go extinct because of demographic stochasticity. If for a type  $x$ , there exists some other type  $y$  such that  $s_x(y) > 0$ , it is certain that, over evolutionary times, some mutant will successfully invade. On the other hand, a state  $\bar{x}$  such that  $s_{\bar{x}}(y) < 0$  for all  $y \neq \bar{x}$  is often called an ESS, evolutionary stable state, although other terms may be more appropriate (Metz et al. 1996). Not necessarily the final state of evolution will be an ESS (Eshel 1996), but it is often possible to understand evolutionary dynamics by considering ‘pairwise invasibility plots’.

Many diseases, with important effects on human health (Anderson & May 1991), are caused by the so-called (Anderson & May 1978) macro-parasites, mainly helminths. Evolutionary models for these interactions are

---

\*Dipartimento di Matematica, Università di Trento, Via Sommarive 14, 38050 Povo (TN), Italy.

however missing, mainly because the models used for describing the dynamics of macro-parasitic infections are more complex than for micro-parasites. The original model presented by Kostizin (Kostizin 1934) consists of an infinite system of differential equations. On the other hand, the approximations introduced by Anderson and May (Anderson & May 1978) can not be extended in a straightforward way to multi-species or strains interactions: the two-species approximations so far available (Dobson 1985, Roberts & Dobson 1995, Gatto & De Leo 1998) are not suitable for evolutionary considerations; in fact, in these models any type  $x$  can be invaded by all types  $y$  close to  $x$ , making it impossible the use of the adaptive dynamics framework.

In this paper, I present a setting for the analysis of virulence evolution in macro-parasites, together with some preliminary results. These methods and results may prove useful to better understand the pathology induced by this kind of parasites. Furthermore, within this setting it becomes possible to study competition among parasites at two different levels: competition for new hosts to infect, and competition within each host. This important theoretical issue has been addressed for micro-parasites by using simplifying assumptions: instantaneous replacement of strains in a host (Nowak & May 1994, Castillo-Chavez & Velasco-Hernandez 1998, Mena-Lorca et al. 1999); or a maximum of two different strains at the same time in one host (van Baalen & Sabelis 1995, Mosquera & Adler 1998). In the present setting, it would be possible to assume any kind of interaction among the parasites within a host, so that, at least in principle, the problem can be studied in depth.

In Section 2, I present the main model, an infinite system of differential equations describing the interaction of a host species with one (or two) species of macro-parasites. I then summarize the computation (Pugliese 2000a) of the “invasibility” condition, i.e. the condition under which parasite species 2 increases in number when a small number of parasites are introduced in a host population at equilibrium with parasite species 1. It is remarked that mutual invasibility occurs in a narrow parameter region, so that coexistence of two parasite species is possible, but unlikely. In Section 3, I introduce the methods of adaptive dynamics into this setting, by choosing an explicit function modeling the trade-off between parasite fertility and virulence; it is shown, partly analytically, partly numerically, that there always exists a unique stable state, to which virulence will converge over evolutionary times. The method of adaptive dynamics implicitly assumes asexual reproduction (see however Metz et al. 1996); in Section 4, I consider instead two simple models of sexual reproduction in parasites: in the first, it is assumed that parasites are hermaphrodite; in the second, they are assumed dioecious (males and females). I show there that virulence at ESS is lower when reproduction is sexual than when it is asexual, especially when parasites are dioecious. In Section 5, I assume that infections occurs in clumps of parasite larvae, not one by one as

assumed in the main model. It is shown that the larger the mean size of the clump, the lower will virulence be at the ESS. Finally, in Section 6, I discuss some limitations of the present models and results.

**2. Models for two macro-parasite species.** As mentioned above, the starting point is the model first presented by Kostizin (Kostizin 1934), in which the main variables are  $p_i(t)$ , the number of hosts carrying  $i$  parasites. Parasites in one host may increase from  $i - 1$  to  $i$  because of new infections at rate  $\varphi(t)$ ; may decrease from  $i + 1$  to  $i$  because of the death of one parasite: it is assumed that each parasite dies (independently of the number of parasites in the same host) at rate  $\sigma$ . One must also keep track of hosts' births and deaths: it is assumed that new hosts are born at rate  $b$ , while hosts die at a natural rate  $\mu$  plus an additional rate  $\alpha$  for each parasite harbored: it is thus assumed that parasites increase hosts' mortality. Here I assume that parasites have no effects on hosts' fertility; it is known (Diekmann & Kretzschmar 1991, Kretzschmar 1993) that introducing this possibility may induce a more complex behavior in the model, which is probably extraneous to the problem considered here. Performing all the bookkeeping (for instance, when a new infection occurs in a host with  $i - 1$  parasite,  $p_i$  will increase, and  $p_{i-1}$  will decrease), one arrives at the following system of equations:

$$(1) \quad \begin{aligned} \frac{d}{dt} p_0(t) &= -(\mu + \varphi)p_0(t) + \sigma p_1(t) + bN(t) \\ \frac{d}{dt} p_i(t) &= -(\mu + \varphi + i(\alpha + \sigma))p_i(t) + \sigma(i+1)p_{i+1}(t) + \varphi(t)p_{i-1}(t) \quad i \geq 1. \end{aligned}$$

where  $N(t) = \sum_{i=0}^{\infty} p_i(t)$  is the total number of hosts and  $P(t) = \sum_{i=0}^{\infty} ip_i(t)$  is the total number of adult parasites.

A partial analysis of this system is possible, finding for instance conditions for an endemic equilibrium (Kretzschmar 1993). For this purpose, it is convenient to introduce in the model host age, even though no parameters will be assumed to depend on that. The advantage of this introduction is that (at equilibrium) the parasites on hosts of a given age  $a$  will be distributed according to a simple law (in the basic model, a Poisson), while the overall parasite distribution will be a mixture, not easy to interpret by itself.

The main variables  $p_i$  will then be assumed to depend on age  $a$  and time  $t$ . As well known, the total derivative in time corresponds to  $\frac{\partial}{\partial t} + \frac{\partial}{\partial a}$ . Hence, one arrives at the following system of equations:

$$(2) \quad \begin{cases} Dp_i(a, t) = -(\mu + \varphi(t) + i(\alpha + \sigma))p_i(a, t) + \sigma(i+1)p_{i+1}(a, t) \\ \quad + \varphi(t)p_{i-1}(a, t) & i = 0, 1, \dots \\ p_0(0, t) = bN(t) \\ p_i(0, t) = 0 & i \geq 1. \end{cases}$$

where now

$$\begin{aligned} N(t) &= \sum_{i=0}^{\infty} \int_0^{\infty} p_i(a, t) da \quad (\text{total population size}) \\ P(t) &= \sum_{i=0}^{\infty} \int_0^{\infty} i p_i(a, t) da \quad (\text{total number of parasites}) \\ D &= \frac{\partial}{\partial t} + \frac{\partial}{\partial a} \quad \text{and} \quad p_{-1} \equiv 0. \end{aligned}$$

The infection rate  $\varphi(t)$  has still to be specified. In a more complete model (Anderson & May 1978) infections occur through encounters between free-living parasite larvae and hosts, according to a mass action law. If the life span of free-living larvae is short relative to the other processes, then time-scale arguments yield a reduced model where

$$(3) \quad \varphi(t) = \frac{hP(t)}{c + N(t)}.$$

This simplifying assumption will be used in the rest of the paper.

The analysis of evolutionary dynamics is based, first of all, on identifying the attractors of the system with only one type present. It can be shown that there exists a unique positive equilibrium of (2); some of its properties have been obtained in (Pugliese 2000a); I will summarize quickly the main results and the method.

It appears from extensive simulations that this equilibrium is always globally attractive, when it exists. Unfortunately, there is no formal proof of its stability.

Assume that an equilibrium  $\bar{p}_i(a)$  of (2) exists. At equilibrium the infection rate  $\frac{h\bar{P}}{c + \bar{N}}$  is a constant  $\bar{\varphi}$ . Then  $\bar{p}_i(a)$  follow (as functions of  $a$ ) an immigration-and-death process; it is known (Bailey 1964, Isham 1995) that their distribution is Poisson at each  $a$ . In other words, a stationary solution for (2) has the form

$$(4) \quad \bar{p}_i(a) = N(a) \frac{(x(a))^i}{i!} e^{-x(a)}.$$

Through some computations, one finds that the mean number of parasites per host of age  $a$ ,  $x(a)$ , is

$$(5) \quad x(a) = \frac{\bar{\varphi}}{\sigma + \alpha} (1 - e^{-(\sigma + \alpha)a})$$

where  $\bar{\varphi}$  is found as the unique root of a nonlinear equation.

The number of hosts of age  $a$ ,  $N(a)$ , can be written as  $K\pi(a)$  where

$$(6) \quad \pi(a) = e^{-\mu a} \exp\left\{-\alpha \int_0^a x(s) ds\right\} = \exp\left\{-\mu a - \frac{\alpha \bar{\varphi}}{\sigma + \alpha} \left(a - \frac{1 - e^{-(\sigma + \alpha)a}}{\sigma + \alpha}\right)\right\}$$

represents the probability of surviving to age  $a$  (at equilibrium), and  $K$  can be obtained from  $\bar{\varphi}$  through

$$(7) \quad \bar{\varphi} = \frac{h\bar{P}}{c + \bar{N}} = \frac{hK \int_0^\infty x(a)\pi(a) da}{c + K \int_0^\infty \pi(a) da}.$$

Let us now consider the introduction of a second parasite species; as in all the other models considered, it will be assumed that parasites interact only by causing the death of a common host (Anderson & May 1978, Dobson 1985, Kretzschmar 1993, Gatto & De Leo 1998).

It is not necessary to write explicitly the system with both parasite species present. Instead, as stated in the Introduction, it is sufficient to compute the invasion coefficient of a second parasite species introduced (in small numbers) into a host population at equilibrium with the first parasite; this equilibrium has been described above and will be named here  $E_1$ . As discussed in the Introduction, computation of this invasion coefficient is relevant to adaptive dynamics only if  $E_1$  is an attractor for system (2); although a proof is missing, this will be assumed to be true, as it seems from numerical simulation.

In this context, it is convenient to compute  $R_0$ , the basic reproductive ratio of a second parasite near  $E_1$ , instead of the invasion coefficient. Although more formal definitions are possible (Diekmann et al. 1990),  $R_0$  can be defined as the average number of adult 2-parasites generated over its life by one adult 2-parasite when the host population is at  $E_1$ . It is intuitive that, if  $R_0 < 1$ , the species 2 will not be able to invade an equilibrium where species 1 coexists with their host; vice versa, such an invasion will occur if  $R_0 > 1$ . In fact, it can be proved (Pugliese 2000a, Diekmann & Heesterbeek 2000) that  $R_0 > 1$  is equivalent to  $s_1(2) > 0$  (in the notation of the Introduction), and implies that  $E_1$  is unstable for the system with both parasite species present.

When both  $E_1$  and  $E_2$  are unstable for that system, the two parasites species will be deemed to coexist. This can be justified through persistence theory (Hutson & Schmitt 1992). Moreover, if a bifurcation diagram is drawn for the system with both parasite species present, using (for instance)  $h_2$  as bifurcation parameter, it is easy to see that, at the value of  $h_2$  at which  $E_1$  becomes unstable (i.e. expression (11) is equal to 1), an equilibrium  $E^*$  with a positive number of both parasite species bifurcates transcritically from  $E_1$ .

Going back to the computation of  $R_0$ , it is clear from the form of system (2) that  $R_0$  has the form

$$(8) \quad R_0 = \frac{h_2 \bar{N}_1}{c_2 + \bar{N}_1} T$$

where  $T$  represents the expected lifetime of a 2 parasite that has just infected an average host in a population at equilibrium  $E_1$ , and the subscripts (1 or 2) specify that the parameters or the equilibrium value belong to the corresponding species.

The only difficulty lies in computing  $T$ . While the details are given in (Pugliese 2000a), I here introduce the ingredients that will be needed later. Since the parasites do not interact (as long as their common host is alive), the probability that a 2-parasite will be alive, and in a host containing  $l$  1-parasites, time  $s$  after having infected a host is (Pugliese 2000a)

$$(9) \quad p(l, s) = \frac{\int_0^\infty e^{-(\sigma_2 + \alpha_2)s} \bar{p}_l(a+s) da}{\int_0^\infty N_1(a) da}.$$

where  $\bar{p}_l(a)$  is given in (4).

Hence, the probability  $p(s)$  of being alive is the sum over  $l$  from 0 to  $\infty$  of the previous expression, and

$$(10) \quad T = \int_0^\infty p(s) ds = \int_0^\infty \sum_{l=0}^\infty p(l, s) ds = \frac{\int_0^\infty \int_0^\infty \pi_1(a+s) e^{-(\sigma_2 + \alpha_2)s} da ds}{\int_0^\infty \pi_1(a) da} \\ = \frac{\int_0^\infty \pi_1(a)(1 - e^{-(\sigma_2 + \alpha_2)a}) da}{(\sigma_2 + \alpha_2) \int_0^\infty \pi_1(a) da}.$$

From this expression one obtains immediately  $R_0$  through (8). It is convenient to write  $R_0$ , using also (3), in a more abstract form, valid when  $c_1 = c_2$ , as

$$(11) \quad R_0 = \frac{h_2}{h_1} F(\alpha_2 + \sigma_2, \alpha_1 + \sigma_1)$$

where

$$(12) \quad F(y_2, y_1) = \frac{\varphi_1}{y_2} \frac{\int_0^\infty \pi_1(a)(1 - e^{-y_2 a}) da}{\int_0^\infty \pi_1(a)x_1(a)} \\ = \frac{\alpha_1 \varphi_1}{b - \mu y_2} \int_0^\infty \pi_1(a)(1 - e^{-y_2 a}) da.$$

The function  $F$  satisfies

$$(13) \quad F(y_2, y_1) \begin{cases} < 1 & \text{if } y_1 < y_2; \\ = 1 & \text{if } y_1 = y_2; \\ > 1 & \text{if } y_1 > y_2. \end{cases}$$

Hence, if  $h_2 \leq h_1$ , and  $\alpha_2 + \sigma_2 \geq \alpha_1 + \sigma_1$ , species 2 will not be able to invade the state where hosts are at equilibrium with parasite species 1, and vice versa species 1 will invade a host-species 2 equilibrium.

This result can be stated as a *principle of partial competitive exclusion*: competitors that are inferior in one or more respects and superior in none are doomed to extinction. This is in agreement with intuition (Armstrong & McGehee 1980), but in contrast with the results of Dobson (1985).

On the other hand, if there is a trade-off between virulence  $\alpha$  and parasite fertility  $h$ , coexistence may occur, provided the parameters satisfy a certain condition. Precisely, for each value of the parameters  $\alpha$  and  $\sigma$  such that  $\alpha_1 + \sigma_1 > \alpha_2 + \sigma_2$ , there exist  $\rho_{\min}$  and  $\rho_{\max}$  with  $1 < \rho_{\min} < \rho_{\max}$  such that if  $\rho_{\min} < h_1/h_2 < \rho_{\max}$ , the two parasite species coexist.

In all the cases examined  $\rho_{\min}$  and  $\rho_{\max}$  are actually very close to each other, so that the parameters must fall in a very narrow region for coexistence to occur.

**3. Evolutionary dynamics of virulence in simple macroparasite models.** In the previous section, we found the condition under which the population of one parasite type can increase, when that parasite is introduced, in small numbers, in a host population at equilibrium with another parasite type. On the basis of this “invasibility” condition, we found the conditions for the coexistence of two parasite types. In this section we turn to evolutionary considerations: we assume that infinitely many parasite types may arise by mutation, and analyze, through the adaptive dynamics paradigm, the expected evolutionary path of virulence.

We will denote parasite types by their virulence  $\alpha$ , and assume we can write explicitly the trade-off between fertility,  $h$ , and virulence  $\alpha$ , as a function  $h(\alpha)$ . We will assume (in agreement with previous authors (Anderson & May 1982))  $h(\alpha) \geq 0$ ,  $h'(\alpha) \geq 0$  and  $h''(\alpha) \leq 0$ . Specific examples may be  $h(\alpha) = C\alpha^\beta$  or  $h(\alpha) = \frac{C\alpha}{\alpha + k}$ .

Letting  $y = \alpha + \sigma$ , we use the invasibility condition  $F(y_2, y_1) \cdot h_2/h_1 > 1$ , and look for a value  $\bar{\alpha}$  such that a parasite species of type  $\alpha$ , different from  $\bar{\alpha}$ , can not invade into a host population at equilibrium with parasite type  $\bar{\alpha}$ . Thus we fix  $\alpha_1$  at such a putative value and consider (11) as a function of  $\alpha = \alpha_2$ . Thus we define

$$(14) \quad g(y) = \frac{b}{y} \int_0^\infty \bar{\pi}(a)(1 - e^{-ya}) da = \int_0^\infty \bar{P}(a)e^{-ya} da$$

through an integration by parts, where  $\bar{\pi}(a)$  is the survival function (6) for  $\bar{y}$  and

$$(15) \quad \bar{P}(a) = b \int_a^\infty \bar{\pi}(a) da.$$

Then the value  $\bar{\alpha}$  must be such that the maximum of the function

$$(16) \quad R(\alpha) = h(\alpha)g(\alpha + \sigma)$$

is at  $\alpha = \bar{\alpha}$ .

To find the maximum, we need to compute the first and second derivatives of  $R(\alpha)$ . It is convenient to differentiate (14) obtaining

$$(17) \quad g^{(n)}(y) = (-1)^n \int_0^\infty a^n \bar{P}(a) e^{-ya} da.$$

Hence, the condition  $R'(\bar{\alpha}) = 0$  translates into

$$(18) \quad \frac{\int_0^\infty a \bar{P}(a) e^{-\bar{y}a} da}{\int_0^\infty \bar{P}(a) e^{-\bar{y}a} da} = \frac{h'(\bar{\alpha})}{h(\bar{\alpha})}.$$

From numerical computations, it seems that in most cases (18) has a unique solution  $\bar{\alpha} > 0$ . I could not find a proof for this, however.

(18) is a necessary and sufficient condition for  $\bar{\alpha}$  to be what (Geritz et al. 1998) called an evolutionary singular state. It will really be uninivable, at least locally, if  $R''(\alpha) < 0$  at such a point. Through straightforward computations one sees that the sign of  $R''(\bar{\alpha})$  is the same as that of

$$(19) \quad S(\bar{\alpha}) = h''(\bar{\alpha}) \int_0^\infty \bar{P}(a) e^{-\bar{y}a} da - 2h'(\bar{\alpha}) \int_0^\infty a \bar{P}(a) e^{-\bar{y}a} da + h(\bar{\alpha}) \int_0^\infty a^2 \bar{P}(a) e^{-\bar{y}a} da.$$

The only positive term is the rightmost one. In order to estimate the sign of  $S$ , the following lemma, obtained from (Hadeler & Dietz 1983) with an obvious change in the assumptions, is useful:

**LEMMA 3.1.** *Let  $g, u, v$  be locally summable real functions on the (finite or infinite) interval  $(a, b)$ , with  $u(x), v(x) \geq 0$ , such that*

$$(20) \quad x \leq y \implies g(x) \geq g(y) \text{ and } v(x)u(y) \leq u(x)v(y).$$

*Then*

$$(21) \quad \int_a^b g(x)u(x) dx \cdot \int_a^b v(x) dx \geq \int_a^b g(x)v(x) dx \cdot \int_a^b u(x) dx$$

*provided the integrals exist.*

*Moreover, when  $u, v$  and  $g$  are positive, and  $u(x)/v(x)$  is strictly decreasing, then (21) holds with strict inequality.*

We use the lemma with  $g(a) = \bar{P}(a)$  (a decreasing function),  $v(a) = a^2 e^{-\bar{y}a}$  and  $u(a) = ae^{-\bar{y}a}$ . Then using also the fact that

$$(22) \quad \int_0^\infty v(a) da = \int_0^\infty a^2 e^{-\bar{y}a} da = \frac{2}{\bar{y}} \int_0^\infty a e^{-\bar{y}a} da = \frac{2}{\bar{y}} \int_0^\infty u(a) da$$

we have

$$(23) \quad \int_0^\infty a^2 \bar{P}(a) e^{-\bar{y}a} da < \frac{2}{\bar{y}} \int_0^\infty a \bar{P}(a) e^{-\bar{y}a} da.$$

Hence

$$(24) \quad \begin{aligned} S(\bar{\alpha}) &< h''(\bar{\alpha}) \int_0^\infty \bar{P}(a) e^{-\bar{y}a} da \\ &\quad - 2h'(\bar{\alpha}) \int_0^\infty a \bar{P}(a) e^{-\bar{y}a} da + \frac{2}{\bar{y}} h(\bar{\alpha}) \int_0^\infty a \bar{P}(a) e^{-\bar{y}a} da. \end{aligned}$$

Inserting now (18) in (24) yields

$$(25) \quad S(\bar{\alpha}) < \int_0^\infty a \bar{P}(a) e^{-\bar{y}a} da \left( \frac{h(\bar{\alpha}) h''(\bar{\alpha})}{h'(\bar{\alpha})} - 2h'(\bar{\alpha}) + \frac{2}{\bar{y}} h(\bar{\alpha}) \right).$$

For the function  $h(\alpha) = \frac{C\alpha}{\alpha+k}$ , one sees that the RHS of (25) is negative for all values of  $\alpha$  and  $\sigma > 0$ ; hence  $R''(\bar{\alpha}) < 0$ . We have then proved that, when  $h(\alpha) = \frac{C\alpha}{\alpha+k}$ , any solution of (18) is actually an evolutionary stable state. It appears from numerical simulations that the same is true for the function  $h(\alpha) = C\alpha^\beta$ .

This shows that there is indeed a unique evolutionary stable strategy; this will be of intermediate virulence, in the sense that it will be larger than 0 and lower than other values that would still yield a viable parasite. However, it appears (see Fig. 1) in reasonable numerical examples, that the virulence at the evolutionary stable strategy, though intermediate in this sense, is rather high compared with hosts' natural death rate and optimal strategies for micro-parasites. This result is not unexpected, since macro-parasites in a host are, in these models, unrelated to each other and thus place relatively little value in host's survival.

**4. The effect of sexual reproduction.** The previous model is certainly too simple to account for all the complexities in host-parasite interactions. Moreover, the virulence predicted by this model seems definitely too large relatively to what is seen in actual observations (Dobson & Hudson 1992, Dobson & Merenlender 1991). Two complications will be explored here: the effect of sexual reproduction and of multiple infections resulting from clumped dispersal. The resulting models will be extremely simple, and will miss important features of real interactions, but have the only aim of exploring the effect of sex and clumped dispersal on the evolutionarily stable level of virulence. This section is devoted to the presentation and analysis of two simple models of parasite sexual reproduction, while multiple infections will be examined in the next section;

Sexual reproduction must definitely play a very important role in the biology of parasitic worms, since it is ubiquitous despite the obvious advantages clonal reproduction would seem to have in such a system. It

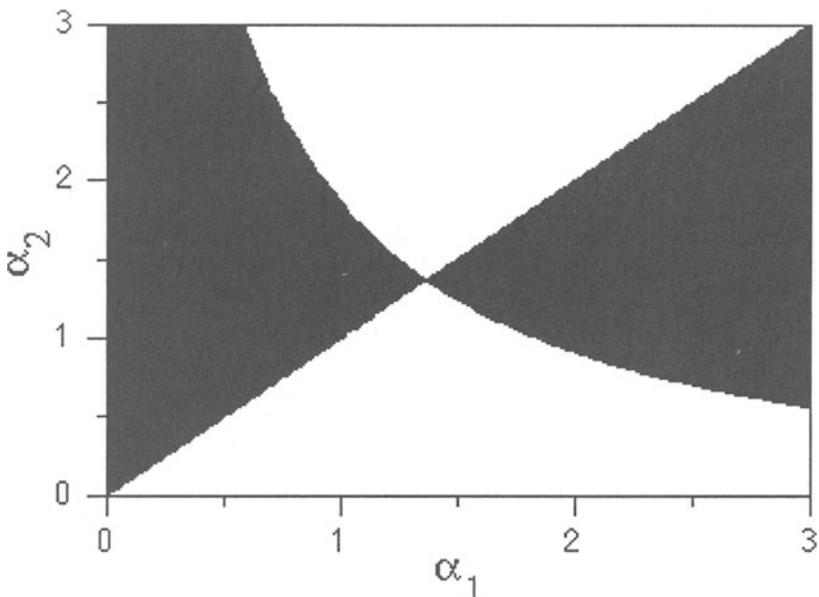


FIG. 1. The pairwise invasibility plot for the basic model, found through the condition (11). A strain with virulence  $\alpha_2$  can invade into a population at equilibrium with a strain with virulence  $\alpha_1$  when  $(\alpha_1, \alpha_2)$  is in the black region. The value of  $\alpha$  (around 1.38) where the two curves cross is an evolutionarily stable strategy. Parasite fertility  $h$  is obtained as  $h(\alpha) = C\alpha^\beta$ . Parameter values are  $b = 1$ ,  $\mu = 0.5$ ,  $\sigma = 2$ ,  $\beta = 0.3064$ ; the value of  $C$  does not matter for the invasibility plot.

seems likely (Hamilton 1980, Ebert & Hamilton 1996) that sex is related to the immune response mounted by vertebrate hosts; understanding its role certainly requires the consideration of coevolutionary models.

Here, I simply take the existence of sexual reproduction for granted; therefore, in the model, sexual contacts among parasites within a host will be necessary for the production of eggs, and hence of new infections. In order to keep the models as simple as possible, I consider two cases: in the first case, adult parasites are assumed to be hermaphroditic (a widespread condition among Trematoda) but outside fertilization is assumed to be necessary for the production of eggs (in reality, it seems that often eggs of lower quality can be produced through self-fertilization). In the second case (widespread among nematodes) adult parasites will be males or females; it will be assumed that the overall sex ratio is 1:1, and that one male is enough to fertilize all females in a host.

**4.1. Hermaphroditic parasites.** The basic model is still (2) except that the infection rate  $\varphi(t)$  will not be given by  $\varphi(t) = \frac{hP(t)}{c+N(t)}$  as in (3); in fact the numerator in (3) is the rate of production of new parasite eggs,

that now will not be proportional to the total number of parasites, but only to those present in a host carrying at least 2 parasites (those alone in a host will not be able to reproduce). Hence now

$$(26) \quad \varphi(t) = \frac{h \sum_{i=2}^{\infty} i \int_0^{\infty} p_i(a, t) da}{c + N(t)}$$

The procedure used in the previous sections can be repeated to a large extent. First of all, one considers the equilibrium with only one parasite type present. Formula (4) still holds, as well as the expressions (5) for  $x(a)$  and (6) for  $\pi(a)$ . The only difference is in the fact that from (26) one obtains

$$(27) \quad \begin{aligned} \bar{\varphi} &= \frac{hK \int_0^{\infty} \pi(a) \sum_{i=2}^{\infty} i \frac{(x(a))^i}{i!} e^{-x(a)} da}{c + K \int_0^{\infty} \pi(a) da} \\ &= \frac{hK \int_0^{\infty} \pi(a) x(a) (1 - e^{-x(a)}) da}{c + K \int_0^{\infty} \pi(a) da} \end{aligned}$$

from which the value of  $K$  can be obtained.

Assume now we introduce a mutant gene that influences the value of  $h$ ,  $\alpha$  and  $\sigma$ . As long as the number of mutant genes is small, it will almost never find itself in homozygosity; hence, in order to study whether such a mutant will invade into the population, we need to know its effect in heterozygosity with the resident gene (Charlesworth 1981). Assume then that a heterozygous will have parameter values  $h_2$ ,  $\alpha_2$  and  $\sigma_2$ , and let us compute the appropriate value of  $R_0$ . Assuming equal effects on the male and female function (see the next subsection for a different assumption), half of the time the mutant gene will spread through the female function, half of the time through the male. Thus, we simply compute the number of successful infecting larvae coming from eggs produced over its life by a heterozygous. Since only half of these larvae will contain the mutant gene, this quantity will be the reproductive ratio  $R_0$  for the mutant gene.

The number of infecting larvae produced by a heterozygous female over its life time will be equal to the product of  $\frac{h_2 \bar{N}_1}{c_2 + \bar{N}_1}$  with the time spent in a host with at least another parasite. Using the same notation as in Section 2, the latter can be written as

$$\begin{aligned}
(28) \quad & \int_0^\infty \sum_{l=1}^\infty p(l, s) ds = \frac{\int_0^\infty \int_0^\infty \pi_1(a+s)(1-e^{-x_1(a+s)})e^{-(\sigma_2+\alpha_2)s} da ds}{\int_0^\infty \pi_1(a) da} \\
& = \frac{\int_0^\infty \pi_1(a)(1-e^{-x_1(a)})(1-e^{-(\sigma_2+\alpha_2)a}) da}{(\sigma_2 + \alpha_2) \int_0^\infty \pi_1(a) da}.
\end{aligned}$$

From this expression one can write  $R_0$  in an abstract form, valid when  $c_1 = c_2$ , as

$$(29) \quad R_0 = \frac{h_2}{h_1} F(\alpha_2 + \sigma_2, \alpha_1 + \sigma_1)$$

where, using (27),

$$(30) \quad F(y_2, y_1) = \frac{\varphi_1}{y_2} \frac{\int_0^\infty \pi_1(a)(1-e^{-x_1(a)})(1-e^{-y_2 a}) da}{\int_0^\infty \pi_1(a)x_1(a)(1-e^{-x_1(a)}) da}.$$

Using this new definition of  $F$ , one can proceed as in Section 3 and find the conditions for the existence of an evolutionary stable state. Here, I study this problem only numerically. Using the same parameter values as in Fig. 1, I find the neutrality curve shown in Fig. 2. The curve is similar to that obtained in the previous section, but it is shifted down, so that virulence at the evolutionary stable state will be somewhat lower than the virulence predicted from Fig. 1.

**4.2. Dioecious parasites.** The basic model is still (2), but we have to compute the infection rate  $\varphi(t)$  in this case. As stated above, I will assume that the overall sex ratio is fixed 1:1 and that one male is enough to fertilize all females within a host.

We start by considering a population with only one parasite type. When  $l$  parasites are together in a host, their total rate of reproductive output will be equal to  $hk$ , where  $k$  is the number of females, as long as  $k < l$ . Having assumed a 1:1 sex ratio, the probability  $p_k^l$  that exactly  $k$  of the  $l$  parasites are females is simply  $\binom{l}{k} \left(\frac{1}{2}\right)^l$ ; hence the average reproductive output of  $l$  parasites together in a host is

$$(31) \quad h \sum_{k=0}^{l-1} k \binom{l}{k} \left(\frac{1}{2}\right)^l = h \frac{l}{2} \left(1 - \left(\frac{1}{2}\right)^{l-1}\right).$$

Hence we obtain that, always considering negligible the time spent as infecting larva, the rate of new infections is

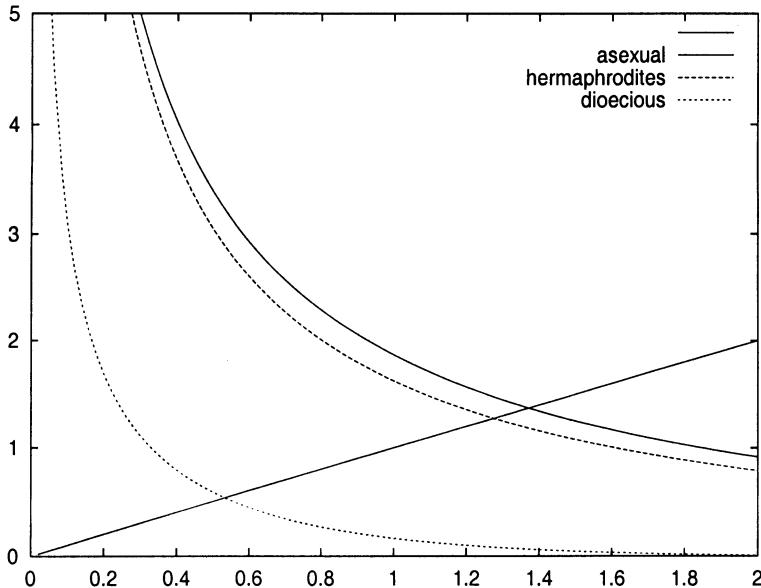


FIG. 2. For each value of  $\alpha_1$ , the value of  $\alpha_2$  ( $= N(\alpha_1)$ ) is shown such that the equilibrium with only parasites 1 is neutrally stable relatively to the invasion of parasites 2 (the 'neutrality' curve). In these models, parasites with values of  $\alpha_2$  intermediate between  $\alpha_1$  and  $N(\alpha_1)$  will be able to invade (see Fig. 1). The point where the neutrality curve crosses the line  $\{y = x\}$  is an ESS. The solid line refers to the model without sex (the same curve as in Fig. 1); the dashed line to the model for hermaphroditic parasites; the dotted line to the model for dioecious parasites. Parameter values as in Fig. 1.

$$(32) \quad \varphi(t) = \frac{h \sum_{l=2}^{\infty} \frac{l}{2} \left(1 - \left(\frac{1}{2}\right)^{l-1}\right) \int_0^{\infty} p_l(a, t) da}{c + N(t)}$$

Using again (at a potential equilibrium) (4), (5) and (6), one obtains from (32)

$$(33) \quad \begin{aligned} \bar{\varphi} &= \frac{hK \int_0^{\infty} \pi(a) \sum_{l=2}^{\infty} l \left(1 - \left(\frac{1}{2}\right)^{l-1}\right) \frac{(x(a))^l}{l!} e^{-x(a)} da}{2(c + K \int_0^{\infty} \pi(a) da)} \\ &= \frac{hK \int_0^{\infty} \pi(a) x(a) (1 - e^{-x(a)/2}) da}{2(c + K \int_0^{\infty} \pi(a) da)} \end{aligned}$$

from which the value of  $K$  can be obtained.

Let us consider now the question of whether a mutant gene can invade into a homogeneous population. I will assume that a mutant gene exists

that influences the value of  $h$ ,  $\alpha$  and  $\sigma$ ; its action is such that the effect of the gene on  $\alpha$  and  $\sigma$  is the same whether the individual parasite containing that gene is male or female; on the other hand, the effect on  $h$  occurs only in females; it is assumed that all males in a host contribute equally to the fertilization of the females in that host. This assumption is made to avoid addressing very complex issues such as sexual selection and sperm competition. Clearly, the assumption is such that the cost of a highly virulent gene is borne both by males and females, while the advantage (high fecundity) occurs only in females.

Having made this assumption, we can compute the invasion coefficient of the mutant gene; we remind, as before, that, when the mutant genes are rare, parasites homozygous for the mutant gene will occur very rarely; hence, in the linearization, we need to consider the effect of the mutant gene only when it is in heterozygosity with the resident gene. When the heterozygous parasite is a male, and is present in a host with  $l$  other parasites (all of which can be considered to be homozygous with the resident gene) the rate of reproductive output of the females fertilized by him is

$$(34) \quad \sum_{k=0}^l \frac{hk}{l-k+1} p_k^l = h \sum_{k=0}^l \frac{k}{l-k+1} \binom{l}{k} \left(\frac{1}{2}\right)^l = h \left(1 - \left(\frac{1}{2}\right)^l\right),$$

where  $p_k^l = \binom{l}{k} \left(\frac{1}{2}\right)^l$  is the probability that, of the  $l$  other parasites,  $k$  will be females, and  $\frac{1}{l-k+1}$  is the probability that, in that situation, a random female is fertilized by the heterozygous male.

When the heterozygous parasite is a female, she will reproduce at rate  $h_2$  as long as there is at least one male in the same host; hence, when she is present in a host with  $l$  other parasites, her rate of reproductive output will be

$$(35) \quad h_2 \left(1 - \left(\frac{1}{2}\right)^l\right).$$

In order to compute the invasion coefficient for the mutant gene, we have to consider that half of the time it will be in a male, half of the time in a female; moreover, the mutant gene will be passed to only half of the descendants. Hence, its reproductive ratio will be  $1/4$  of the lifetime reproductive output of a heterozygous female plus  $1/4$  of the total reproductive output of the females fertilized by a heterozygous male, each multiplied by the probability that the egg will develop into a larva that will successfully infect another host. Thus, letting the parameters of a heterozygous having subscript 2, while those of the resident homozygous subscript 1, we obtain

$$\begin{aligned}
(36) \quad R_0 &= \frac{(h_2 + h_1)}{4} \frac{\bar{N}_1}{c + \bar{N}_1} \int_0^\infty \sum_{l=1}^\infty \left( 1 - \left( \frac{1}{2} \right)^l \right) p(l, s) ds \\
&= \frac{(h_2 + h_1)}{4} \frac{\bar{N}_1}{c + \bar{N}_1} \frac{\int_0^\infty \int_0^\infty \pi_1(a+s)(1-e^{-x_1(a+s)/2})e^{-(\sigma_2+\alpha_2)s} da ds}{\int_0^\infty \pi_1(a) da} \\
&= \frac{(h_2 + h_1)}{4} \frac{K}{c + \bar{N}_1} \frac{\int_0^\infty \pi_1(a)(1-e^{-x_1(a)/2})(1-e^{-(\sigma_2+\alpha_2)a}) da}{(\sigma_2 + \alpha_2)} \\
&= \frac{(h_2 + h_1)}{2h_1} \frac{\varphi_1}{(\sigma_2 + \alpha_2)} \frac{\int_0^\infty \pi_1(a)(1-e^{-x_1(a)/2})(1-e^{-y_2 a}) da}{\int_0^\infty \pi_1(a)x_1(a)(1-e^{-x_1(a)/2})}.
\end{aligned}$$

Using this expression, I computed numerically the ‘dioecious’ curve shown in Fig. 2. It can be seen that virulence at the evolutionarily stable strategy is much lower in this case than that found in the previous ones.

**5. Clumped infections.** It has been observed that often parasite infections occur in clumps; this has been surmised as a possible explanation of observed aggregation of parasite distributions (Isham 1995, Pugliese et al. 1998). While Isham (1995) has considered the effect of several distributions for clump size on the distribution of adult parasites in a cohort of hosts, Pugliese et al. (1998) have considered only a (truncated) Poisson distribution for clump size, but integrated this into a dynamical model of infection in a host population.

Building on this, I (Pugliese 2000a) have found the invasion condition of a second type of parasites into a population at equilibrium with a type of parasites. Since infections occur in ‘parcels’, one needs to compute the average number of infecting ‘parcels’ produced by a newly established ‘parcel’. Here I only report the final result

$$\begin{aligned}
(37) \quad R_0 &= \frac{h_2}{h_1} \frac{b}{b - \mu} \frac{\alpha_1 \varphi_1}{\alpha_2 (1 - e^{-\lambda_2})} \\
&\times \int_0^\infty \pi(a) \left( 1 - \exp \left\{ -\frac{\alpha_2 \lambda_2}{\sigma_2 + \alpha_2} (1 - e^{-(\sigma_2 + \alpha_2)a}) \right\} \right) da
\end{aligned}$$

where  $\lambda$  is the parameter of the (truncated) Poisson distribution of clump size: because of the truncation, the average size of a clump is actually  $\lambda/(1 - e^{-\lambda})$ .

Using this formula, one can compute numerically the types that can invade into the resident type. Assuming again  $h = h(\alpha)$ , one can visualize the evolutionary dynamics. In this case too, there appears to be a unique evolutionarily stable state. Hence, for each value of the average clump

size  $\lambda$ , the level of virulence  $\alpha$  at the evolutionarily stable state can be computed: in Fig. 3, I show the evolutionarily stable virulence  $\alpha$  as a function of  $\lambda$ . It appears that, already for low values of  $\lambda$ , the optimal virulence is much lower than what found in the basic model (Section 3).

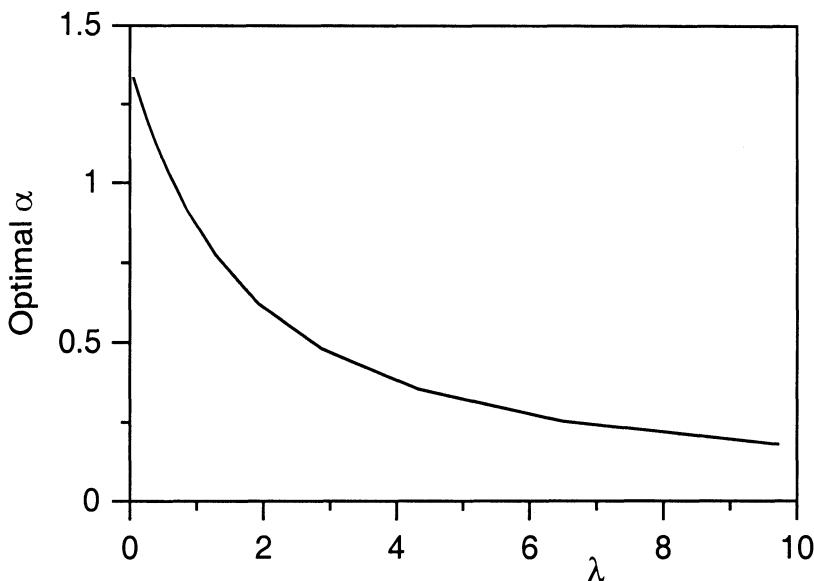


FIG. 3. The value of virulence  $\alpha$  at the evolutionarily stable state is shown for each value of  $\lambda$ . Parameter values as in Fig. 1.

One may note that, by choosing the average number of infecting ‘parcels’ produced by one ‘parcel’ of larvae as a substitute of the “invasibility” condition, it is implicitly assumed that all parasite larvae in an infecting ‘parcel’ are genetically identical. Then, under this scenario, adult parasites in a host will be related at least to a part of the other parasites in the same host, and will place a greater value in host’s survival. We may then say that kin selection is operating in this system to decrease optimal virulence (see Eshel (1977) for a verbal description of similar phenomena).

**6. Discussion.** Different species of parasitic worms coexisting on the same host species are often rather similar as for habitat and resources used (see however Holmes 1973); one may wonder what is the reason for their coexistence in spite of the principle of competitive exclusion (Armstrong & McGehee 1980). Dobson (Dobson 1985) has argued that complex explanations are not needed for the coexistence of macroparasites; in fact, similar species can easily coexist in the models he analyzed, as long as the parasite distributions were highly aggregated, as it is the rule in empirical data (Shaw & Dobson 1995).

However, an analysis of the original infinite model (Kostizin 1934, Kretzschmar 1993) shows that Dobson’s results depended on the approxi-

mations he used to build the low-dimensional model he analyzed. Assuming that parasites interact only through hosts' deaths, parasite coexistence is indeed possible, but the parameter region where coexistence occurs is tiny (Pugliese 2000a). Moreover, the size of the coexistence region has nothing to do with aggregation. This leaves the question open of why macroparasites do coexist on the same host species; more complex models including interactions among parasites seem to be necessary for an explanation.

The use of the infinite model makes it possible to study evolutionary dynamics using the methods of adaptive dynamics. As discussed in the Introduction, the main tool in this method is the computation of the invasion coefficient  $s_x(y)$  of feasible types  $y$  into a population fixed at some state  $x$  (this could also be a polymorphism, but the analysis would be more complex); the state  $x$  will move (stochastically) on evolutionary time-scales towards types  $y$  such that  $s_x(y) > 0$ . In the approximate model used by Dobson (Dobson 1985) all types  $y$  close to any type  $x$  are such that  $s_x(y) > 0$ ; thus, the adaptive dynamics approach is pointless.

On the other hand, in the infinite model discussed here, the invasibility condition (11) is such that invasion is possible at most from one direction (see Fig. 1), and one can use the adaptive dynamics approach. Beyond numerically computed 'pairwise invasibility plots', some analytical results on the existence of an evolutionarily stable state have been presented in Section 3.

The value of virulence  $\alpha$  at the evolutionarily stable state is rather high, compared to what is obtained for microparasites and what is known empirically (Dobson & Hudson 1992, Dobson & Merenlender 1991). I investigated here two mechanisms that may reduce the optimal level of virulence: clumped infections, and sexual reproduction. It is shown that both factors will reduce the optimal level of virulence relative to the basic model. Moreover, they do not exclude each other, so the overall reduction may result from a sum of reasons.

Sexual reproduction has been taken here for granted, and only its consequences on the level of virulence have been examined. Understanding the role of sex in parasites certainly requires considering coevolutionary interactions (Ebert & Hamilton 1996). These have not been examined here, because it would be much more complex. This is a serious limitation of the present models: in fact, parasite virulence may be low because of evolution of resistance in hosts (Frank 1996). The argument often raised that parasite evolution will be on a faster time-scale than host evolution, while reasonable when parasites are bacteria or viruses (see however Andreasen & Christiansen 1993), is not tenable for macro-parasites. Therefore, the present analysis represents a first technical step in describing phenotypic evolution of virulence in macroparasites, and its results, while suggestive, are certainly not conclusive.

The models used here for sexual reproduction are meant only to illustrate the important effect sex can have on virulence. The large difference

found between optimal virulence between hermaphrodite and dioecious parasites (Fig. 2) certainly depends on the lower probability of finding a mate for dioecious than for hermaphrodites; but it is also caused by the different assumptions used in the respective models: in the dioecious model, mutant genes were assumed to affect fertility in females only; hence, a gene coding for high virulence would yield a cost (high death rate) and a benefit (high fertility) to females carrying it, but only a cost to males; on the other hand, for hermaphrodites, it was assumed a benefit both for male and female functions. Many other assumptions can be made; the results would probably be between these two extremes, but the analysis is certainly more complex. This model is distant from reality in many other respects, for instance it is known that sex ratio is generally far from 1:1 (Poulin 1997), but the conclusion seems to be robust.

As for the role of clumped larval dispersal on virulence, the strong reduction in the evolutionarily stable level of virulence shown in Fig. 3 is due, at least to some extent, to the implicit assumption that all parasites in a clump are genetically identical. A more complete model would involve at the same time sexual reproduction and clumped dispersal; depending on the details of sexual reproduction, and of the mechanism of clump formation, it is likely that some parasites in an infecting ‘parcel’ would still be related. The present model yields the expectation that parasite virulence should decrease with mean clump size, while quantitative predictions would depend on details of complex models, for which analytical results seem unlikely.

This result is apparently in contrast with the result by Sasaki & Iwasa (1991) that multiple infections are expected to increase parasite virulence; however, their model considers micro-parasites, and assumes that multiple infections are from genetically unrelated parasites. In the present model the rate at which individual parasites is acquired is kept constant, while increasing the mean size of each infection event; they considered individual infections, and studied the effect of increasing the infection rate.

It seems likely, but it is difficult to investigate analytically, that limited larval dispersal and territorial hosts, may give rise, even if the individual size of each infecting clump is small, to a similar reduction of expected virulence. In fact, kin selection could operate under those circumstances as well.

The final consideration concerns the mathematical structure of the models studied. They consist of infinite systems of first-order partial differential equations. Despite their complexities, their equilibria have a very simple structure, understandable with the help of biological intuition. Invasion conditions into these equilibria have been presented here using only heuristic arguments; a more formal derivation of the invasion condition for the basic model is presented in (Pugliese 2000a); it seems likely that similar derivations could be obtained for the models with sexual reproduction, or clumped infections. What is totally missing, however, is a proof that

the positive equilibrium  $E_1$  [or  $E_2$ ] is indeed an attractor of the system (2) with only one parasite species; this has been implicitly assumed in the computation of invasion conditions; if  $E_1$  were unstable for system (2) for some parameter values, the present analysis would not make sense for those parameter values. An advancement in the analysis of the asymptotic behavior of this kind of systems would therefore be very important for the progress of this area.

**Acknowledgments.** I thank Carlos Castillo-Chavez, and two anonymous referees for their suggestions on the first draft of this paper. I also thank Mike Boots, Yoh Iwasa, and Akira Sasaki for useful comments. I thank CNR for partial support under Grant n. 98.03639.ST74 "Metodi e modelli matematici nello studio dei fenomeni biologici".

## REFERENCES

- ANDERSON, R.M. & MAY, R.M. (1978). Regulation and stability of host-parasite populations interactions 1-2, *J. Anim. Ecol.* **47**: 219–247, 249–267.
- ANDERSON, R.M. & MAY, R.M. (1982). Coevolution of hosts and parasites, *Parasitology* **85**: 411–426.
- ANDERSON, R.M. & MAY, R.M. (1991). *Infectious diseases of humans: dynamics and control*, Oxford Univ. Press, Oxford.
- ANDREASEN, V. & CHRISTIANSEN, F.B. (1993). Disease-induced natural selection in a diploid host, *Theor. Pop. Biol.* **44**: 261–298.
- ARMSTRONG, R.A. & MCGEHEE, R. (1980). Competitive exclusion, *Amer. Nat.* **115**: 151–170.
- BAILEY, N.T.J. (1964). *The elements of stochastic processes*, Wiley, New York.
- BREMERMANN, H.J. & THIEME, H.R. (1989). A competitive exclusion principle for pathogen virulence, *J. Math. Biol.* **27**: 179–190.
- CASTILLO-CHAVEZ, C. & VELASCO-HERNANDEZ, J.X. (1998). On the relationship between evolution of virulence and host demography, *J. Theor. Biol.* **192**: p. 437.
- CHARLESWORTH, B. (1981). *Evolution in age-structured populations*, Cambridge Univ. Press, Cambridge.
- DIECKMANN, U. (1997). Can adaptive dynamics invade?, *Trends Ecol. Evol.* **12**: 128–131.
- DIECKMANN, U., METZ, J., SABELIS, M. & SIGMUND, K. (eds) (2000). *The adaptive dynamics of infectious diseases: in pursuit of virulence management*, Cambridge Univ. Press.
- DIEKMANN, O. & HEESTERBEEK, J.A.P. (2000). *Mathematical Epidemiology of Infectious Diseases*, Wiley, New York.
- DIEKMANN, O., HEESTERBEEK, J.A.P. & METZ, J.A.J. (1990). On the definition and the computation of the basic reproduction ratio  $r_0$  in models for infectious diseases in heterogeneous populations, *J. Math. Biol.* **28**: 365–382.
- DIEKMANN, O. & KRETZSCHMAR, M. (1991). Patterns in the effects of infectious diseases on population growth, *J. Math. Biol.* **29**: 539–570.
- DOBSON, A.P. (1985). The population dynamics of competition between parasites, *Parasitology* **91**: 317–347.
- DOBSON, A.P. & HUDSON, P.J. (1992). Regulation and stability of a free-living host-parasite system. II. population models, *J. Anim. Ecol.* **61**: 487–498.
- DOBSON, A.P. & MERENLENDER, A. (1991). Coevolution of macroparasites and their hosts, in C. Toft, A. Aeschlimann & L. Bolis (eds), *Parasite-Host Association. Coexistence or Conflict?*, Oxford Scientific Publications, pp. 83–101.
- EBERT, D. & HAMILTON, W.D. (1996). Sex against virulence : the coevolution of parasitic diseases, *Trends Ecol. Evol.* **11**: 79–82.

- ESHEL, I. (1977). The founder effect - an ecogenetical approach, *Theor. Pop. Biol.* **3**: 410–24.
- ESHEL, I. (1996). On the changing concept of evolutionary population stability as a reflection of a changing point of view in the quantitative theory of evolution, *J. Math. Biol.* **34**: 485–510.
- EWALD, P.W. (1994). *The evolution of Infectious Disease*, Oxford University Press.
- FRANK, S.A. (1996). Models of parasite virulence, *Quarterly Review of Biology* **71**: 37–78.
- GATTO, M. & DE LEO, G. (1998). Interspecific competition among macroparasites in a density-dependent host population, *J. Math. Biol.* **37**: 467–490.
- GERITZ, S.A.H., KISDI, E., MESZÉNA, G. & METZ, J.A.J. (1998). Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree, *Evol. Ecol.* **12**: 35–.
- HADELER, K.P. & DIETZ, K. (1983). Nonlinear hyperbolic partial differential equations for the dynamics of parasite populations, *Comp. Math. Appl.* **9**: 415–430.
- HAMILTON, W.D. (1980). Sex versus non-sex versus parasite, *Oikos* **35**: 282–290.
- HOLMES, J.C. (1973). Site segregation by parasitic helminths: interspecific interactions, site segregation, and their importance to the development of helminth communities, *Can. J. Zool.* **51**: 333–347.
- HUTSON, V. & SCHMITT, K. (1992). Permanence and the dynamics of biological systems, *Math. Biosci.* **111**: 1–71.
- ISHAM, V. (1995). Stochastic models of host-macroparasite interaction, *Annals of Applied Probability* **5**: 720–740.
- KOSTIZIN, V.A. (1934). *Symbiose, parasitisme et évolution (étude mathématique)*, Hermann, Paris. Translated in F. Scudo and J. Ziegler (eds.) (1978), *The Golden Age of Theoretical Ecology*, pp. 369–408, Lecture Notes in Biomathematics 22, Springer-Verlag, Berlin.
- KRETZSCHMAR, M. (1993). Comparison of an infinite dimensional model for parasitic diseases with a related 2-dimensional system, *J. Math. Anal. Appl.* **176**: 235–260.
- LEVIN, S.A. & PIMENTEL, D. (1981). Selection for intermediate rates of increase in parasite-host systems, *Amer. Nat.* **117**: 308–315.
- MENA-LORCA, J., VELASCO-HERNANDEZ, J. & CASTILLO-CHAVEZ, C. (1999). Density-dependent dynamics and superinfection in an epidemic model, *IMA J. of Math. Applied to Medicine and Biology* **16**: 307–317.
- METZ, J., GERITZ, S., MESZENA, G., JACOBS, F. & VAN HEERWAARDEN, J. (1996). Adaptive dynamics: A geometrical study of the consequences of nearly faithful reproduction, *Stochastic and Spatial Structures of Dynamical Systems, van Strien SJ, Verduyn Lunel SM (eds.)*, Proceedings of the Royal Dutch Academy of Science (KNAW Verhandelingen), North Holland, Amsterdam, pp. 183–231.
- MOSQUERA, J. & ADLER, F. (1998). Evolution of virulence: a unified framework for coinfection and superinfection, *J. theor. Biol.* **195**: 293–313.
- NOWAK, M.A. & MAY, R.M. (1994). Superinfection and the evolution of parasite virulence, *Proc. R. Soc. London B* **255**: 81–89.
- POULIN, R. (1997). Population abundance and sex ratio in dioecious helminth parasites, *Oecologia* **111**: 375–380.
- PUGLIESE, A. (2000a). Coexistence of macroparasites without direct interactions, *Theor. Pop. Biol.* **57**: 145–165.
- PUGLIESE, A. (2000b). Evolutionary dynamics of virulence, in U. Dieckmann & J. Metz (eds), *Elements of adaptive dynamics*, Cambridge Univ. Press. in press.
- PUGLIESE, A., ROSÀ, R. & DAMAGGIO, M.L. (1998). Analysis of a model for macroparasitic infection with variable aggregation and clumped infections, *J. Math. Biol.* **36**: 419–447.
- ROBERTS, M.G. & DOBSON, A.P. (1995). The population dynamics of communities of parasitic helminths, *Math. Biosci.* **126**: 191–214.

- SASAKI, A. & IWASA, Y. (1991). Optimal growth schedule of pathogens within a host: switching between lytic and latent cycles, *Theor. Pop. Biol.* **39**: 201–239.
- SHAW, D.J. & DOBSON, A.P. (1995). Patterns of macroparsite abundance and aggregation in wildlife populations: a quantitative review, *Parasitology* **111 Suppl.**: 111–133.
- VAN BAALEN, M. & SABELIS, M.W. (1995). The dynamics of multiple infection and the evolution of virulence, *Amer. Nat.* **146**: 881–910.

# MATHEMATICAL MODELS FOR SCHISTOSOMIASIS WITH DELAYS AND MULTIPLE DEFINITIVE HOSTS

JIANHONG WU\* AND ZHILAN FENG†

**Abstract.** A mathematical model for the transmission dynamics of *Schistosomiasis japonicum* is derived. The model consists of a system of retarded functional differential equations to take into account two important factors of the transmission process of this disease, i.e., the transit-time distribution and multiple definitive hosts (both human and non-human). The strong monotonicity principle recently established by Wu is used to show that the solution of our model equations defines an eventually strongly monotone semiflow which allows us to give a rather complete qualitative description of the global dynamics of the model.

**1. Introduction.** Schistosomiasis is a parasitic infection which has been estimated to afflict about 200 million people [16]. During the last two decades, many mathematical models and analyses have been devoted to the study of the transmission dynamics of such a parasitic infection. However, most of them ignore the fact that the transmission process requires a definite length of time and focus on *Schistosomiasis mansoni* and *Schistosomiasis haematobium* in which man is the only significant vertebrate host. For details, we refer to [16], [22], and [32]. The importance of alternative mammalian definitive hosts has been well illustrated for *S. japonicum* [2]. *S. japonicum* causes high prevalence of schistosomiasis in the lake regions of China, where the re-infection of human patients has been reported in greater than 12% of cases in which the patient was previously treated with Praziquantal [34]. It is found that in China *Schistosoma japonicum* infects cattle, horses, sheep, dogs, cats and many other mammals (see e.g. [32]). Pigs are also “important in the zoonotic transmission of *S. japonicum*” [28], and both cattle (*Bos taurus*) and water buffalo (*Bubalus bubalis*) commonly act as domestic animal hosts for this parasite as well [5]. Workers have realized that reservoirs of infection warrant control if re-infection of humans is to be stopped [34], and to this end vaccines for both cattle [2] and pigs [28] have been developed. Accordingly, epidemiological and ecological characteristics which affect transmission of human schistosomes are beginning to be included as components of field studies [3]. Both a more realistic understanding of schistosome population dynamics and greater success for disease control will be attained if this trend continues.

The life cycle of the schistosome is described as follows. There are two adult forms of schistosomes, male and female which mate heterosexually. Typically, the infected definitive host contains worms of both genders

---

\*Department of Mathematics and Statistics, York University, North York, Ontario, M3J 1P3, Canada. J. Wu is partially supported by Natural Sciences and Engineering Research Council of Canada.

†Department of Mathematics, Purdue University, W. Lafayette, IN 47907. This author's research was partially supported by NSF grant DMS 9974389.

which seem to have little difficulty in finding each other in the liver of the definitive host, where mating occurs. Some of the fertilized eggs succeed in passing through the wall of the blood vessel in which they are laid, hence through the wall of the intestine or bladder (depending on the species of schistosome) until they fall into the lumen of the organ and are voided with either feces or urine. Some of the eggs ultimately may be deposited in fresh water streams or lakes, where small ciliated larvae (miracidia) emerge. They swim actively and if one comes into contact with an appropriate molluscan host (snail), it rapidly penetrates the snail tissue. By a peculiar process of repeated asexual multiplication within the snail, thousands of a second larval form (cercariae) are produced. The cercariae resulting from a single miracidium have the same sex. When mature, the cercariae are shed periodically by the snail throughout the lifetime of the infection, which coincides, more or less, with the lifetime of the snail. They then enter a free-swimming state designed for invasion of the definitive host. On coming into contact with a definitive host a cercaria attaches itself to the skin and quite rapidly penetrates it, while at the same time sloughing larval structures to become a juvenile schistosome. This is followed by migration to the liver, maturation, mating, migration to the permanent vascular abode, copulation, and oviposition, which begins the cycle all over again. So viewed schematically (see Figure 1) the transmission dynamics of schistosomiasis can be seen to depend on two interrelated flows: one a draft of eggs from definitive hosts to snail, the other a stream of cercariae from snails to definitive hosts.

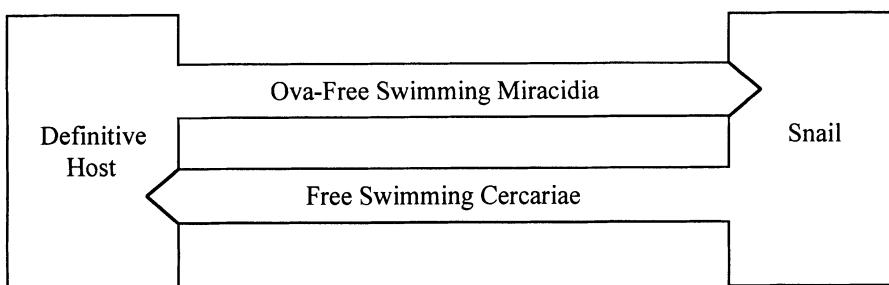


FIG. 1. A transmission diagram of schistosomiasis with one group of definitive hosts.

In this paper, we propose a model describing the transit-time distribution of the transmission process and the evolution of the parasitic infection of *Schistosomiasis japonicum*. The consideration of multiple definitive hosts leads to a system of retarded functional differential equations. By using the strong monotonicity principle recently established by Wu [33] for retarded functional differential equations with infinite delay, it is shown

that the solution of our model equations defines an eventually strongly monotone semiflow so that the powerful decomposition theory of solution operators due to Hale and Kato [4], the spectral theory of linear systems due to Naito [17–19], Nussbaum [24–27], and the monotone dynamical systems theory due to Hirsch [6–8], Matano [13–15], Smith [29–31] can be applied to give a rather complete picture of the global behavior of solutions. Particularly, in the case where zero solution is the unique nonnegative equilibrium point, we show that the zero solution is globally asymptotically stable, and hence the infection can not maintain itself; in the case where multiple equilibria exist, we show that the equilibrium point set together with the invariant curves associated with each unstable equilibrium point forms a tree-like structure with the minimum equilibrium at the base, and that the infection maintains itself in stable configurations.

This paper is organized as follows. In Section 2, after stating the necessary parasitological background of *Schistosomiasis japonicum* (following Nasell and Hirsch [23]), we develop in detail the model equations of the transmission dynamics of *Schistosomiasis japonicum*. The qualitative description of the overall dynamics of the model equations is presented in Section 3. This section also provides several schematic pictures of the global dynamics. Section 4 contains a brief biological interpretation of our qualitative results.

**2. Formulation of the model equations.** Since *Schistosoma japonicum* infects not only man but a wide range of animals as well, the life cycle corresponding to each definitive host is related by the snail through the miracidia flow from definitive hosts to snails and the cercariae flow from snails to definitive hosts. Figure 2 gives a schematic description of the transmission of *Schistosoma japonicum* in definitive host A (such as man) and definitive host B (such as cattle).

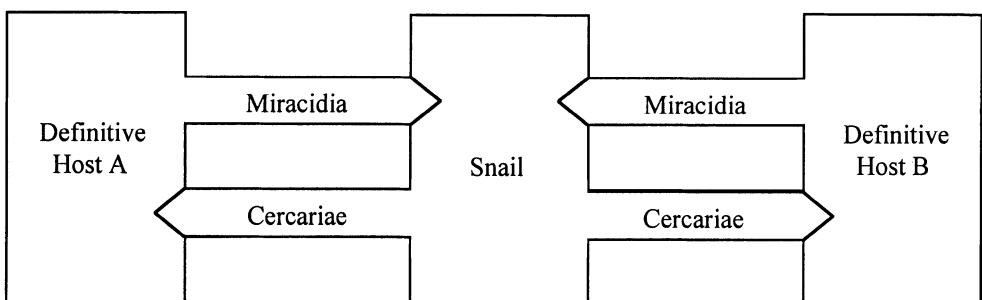


FIG. 2. A transmission diagram of schistosomiasis with two groups of definitive hosts.

In [32] a mathematical model was established for the transmission dynamics of *Schistosoma japonicum* among humans and cattle. A detailed

analysis indicates the possibility of the coexistence of infection in human beings and animals. Particularly, it is shown that in certain situations, although the initial average load of the schistosome in man is zero, schistosomiasis will spread in man and animals and the average load of the schistosome in man will rise and stay at a definite level if the initial average load in animals reaches a definite level.

Our model is built on that of [32]. To describe the model equation in [32], we consider an idealized focus of infection — a relatively isolated community where each group of definitive hosts are equally likely exposed to the risk of infection and are not subject to the processes of birth, death, immigration or emigration. Births and deaths, but not immigration or emigration, will be assumed to occur in the intermediate host population under the simplifying hypotheses that at the instant a snail dies an uninfected snail is born. For ease of reference, we denote in the sequel by  $P_1, \dots, P_n$  the definitive hosts which may be infected by *Schistosoma japonicum*, denote by  $u_i(t)$  the average load of mated *Schistosoma japonicum* in mature form in each individual of  $P_i$  at time  $t$ , by  $N_i$  the total population of  $P_i$  and by  $N_0$  the total number of snails.

Since the infectivity of the definitive host depends on its paired worm burden, we introduce  $m_i(b_i)$  to denote the mating probability of the worm in  $P_i$  when the average load of *Schistosoma japonicum* is  $b_i$ . In the work by MacDonald [11, 12] and its subsequent refinements and elaborations (e.g., Lee and Lewis [9], Nasell [20, 21] and Nasell and Hirsch [23]), adult schistosomes are assumed to be independently randomly distributed among the definitive hosts. This gives a Poisson distribution for the number of schistosomes per individual and yields the formula for  $m_i$ :  $m_i(b_i) = 1 - e^{-b_i} [I_0(b_i) + I_1(b_i)]$  where  $I_0(b_i)$  and  $I_1(b_i)$  are Bessel functions. However, some theoretical argument and empirical evidence suggest that schistosomes are not independently randomly distributed, but rather are distributed in a clumped or contagious fashion. Such clumping of worms has been described by May [16] and Bradley and May [1], where they derived various formulas for  $m_i$  by using the negative binomial. In this paper, we do not require an explicit expression for  $m_i$ . Instead, we follow Nasell and Hirsch to assume that the function  $g_i(b_i) = b_i m_i(b_i)$  (which gives the average load of paired schistosomes when the average load of schistosomes is  $b_i$ ) satisfies (see [23]):

$$(H) \quad g_i(b_i) \geq 0, \quad g'_i(b_i) \geq 0, \quad g''_i(b_i) > 0, \quad b_i g'_i(b_i) - g_i(b_i) \geq 0, \quad \text{if } b_i \geq 0,$$

$$g_i(0) = 0, \quad g'_i(0) = 0, \quad g_i(b_i) \rightarrow \infty, \quad g'_i(b_i) \rightarrow \infty, \quad \text{as } b_i \rightarrow \infty,$$

and all the inequalities are strict for  $b_i > 0$ . Denote by  $E_i$  the frequency of contact with polluted water of each individual in  $P_i$  per unit time, by  $v_i$  the oviposition rate of a paired female schistosome in each individual of  $P_i$  per unit time, by  $K_{1i}$  the probability that the living eggs from each individual of  $P_i$  will enter into the reservoir that contains snails, and by  $\sigma$

the probability that living eggs will penetrate the snail tissue. Then the infection rate of a snail by miracidia is

$$\sum_{i=1}^n N_i E_i u_i v_i K_{1i} \sigma / N_0 = \sum_{i=1}^n K_i u_i,$$

where

$$K_i = \frac{N_i E_i v_i K_{1i} \sigma}{N_0}$$

represents the ability of a paired female schistosome to deliver viable miracidia to a given uninfected snail.

Let  $p$  be the probability that a snail will survive in a unit time. Then in that unit time, the ratio of the number of infected snails to the total number of snails is

$$\frac{\sum_{i=1}^n K_i u_i}{\sum_{i=1}^n K_i u_i - \ln p}.$$

Denote by  $\beta_i$  the death rate of worms in each individual of  $P_i$ , by  $w$  the total number of cercariae shed by each infected snail per unit of time, and by  $K_{2i}$  the probability of each cercaria to penetrate a definitive host of  $P_i$ . Then

$$L_i = N_0 w K_{2i} E_i$$

is a measure of the potential of the intermediate host population to deliver schistosomes to a given definitive host. Noting that when the average load of paired schistosomes is  $u_i$ , the average load of schistosomes is  $g_i^{-1}(u_i)$ , we obtain the following expression

$$(2.1) \quad \begin{aligned} u_i(t + \Delta t) &= g_i(g_i^{-1}(u_i(t)) e^{-\beta_i \Delta t}) \\ &+ \Delta t g_i \left( L_i \frac{\sum_{j=1}^n K_j u_j(t)}{\sum_{j=1}^n K_j u_j(t) - \ln p} \right). \end{aligned}$$

Letting  $\Delta t \rightarrow 0$ , we get the following ordinary differential equation

$$(2.2) \quad \begin{aligned} \frac{d}{dt} u_i(t) &= -\beta_i g'_i(g_i^{-1}(u_i(t))) g_i^{-1}(u_i(t)) \\ &+ g_i \left( L_i \frac{\sum_{j=1}^n K_j u_j(t)}{\sum_{j=1}^n K_j u_j(t) - \ln p} \right). \end{aligned}$$

In the model equation (2.2) for *Schistosoma japonicum* and most models for *Schistosoma mansoni* and *Schistosoma haemotobium*, it is tacitly assumed that the infection, i.e., the miracidia-cercariae flow between different hosts, requires no time. However, the actual situation in nature is

that the time required for the mutual infection among definitive hosts can not be neglected. Various factors such as the latent periods during which cercariae and schistosomes develop to their mature forms, the time required for miracidia to find a snail, the latent period between the moment when a snail is infected by a miracidium and the moment it begins shedding cercariae and the time required for cercariae to find an appropriate definitive host, cause a time lag for the transmission of infection. Therefore the model should be a system of differential equations with retarded arguments as opposed to the aforementioned case where the model is a system of ordinary differential equations.

For the formulation of the delay phenomenon in the process of transmission we start with an idealized case. Suppose we have only one group of definitive hosts  $P_1$  and one (imaginary) pipe through which miracidia and cercariae pass. We assume that the transit time required for the whole transmission process — a mated female schistosome produces eggs, a snail is infected by the miracidia developed from the eggs, and the juvenile schistosome in definitive hosts, resulting from the contact of the host and the cercariae shed from the infected snail, becomes mature — is  $\tau$ . Then relation (2.1) is modified as

$$u_1(t + \Delta t) = g_1(g_1^{-1}(u_1(t))e^{-\beta_1 \Delta t}) + \Delta t g_1 \left( \frac{L_1 K_1 u_1(t - \tau)}{K_1 u_1(t - \tau) - \ln p} \right).$$

By letting  $\Delta t \rightarrow 0$ , we obtain the following delay equation

$$\frac{d}{dt} u_1(t) = -\beta_1 g'_1(g_1^{-1}(u_1(t)))g_1^{-1}(u_1(t)) + g_1 \left( \frac{L_1 K_1 u_1(t - \tau)}{K_1 u_1(t - \tau) - \ln p} \right).$$

Note that the function  $I_1(t) = g_1 \left( \frac{L_1 K_1 u_1(t - \tau)}{K_1 u_1(t - \tau) - \ln p} \right)$  can be expressed by

$$I_1(t) = g_1 \left( \frac{\int_0^\infty K_1 u_1(t - s) dF_0(s)}{\int_0^\infty K_1 u_1(t - s) dF_0(s) - \ln p} \right),$$

where

$$F_0(s) = e(s - \tau) \text{ and } e(u) = \begin{cases} 0, & u \leq 0 \\ 1, & u > 0. \end{cases}$$

The function  $F_0(s)$  is a probability distribution function on  $[0, \infty)$ .

Because the transmission stages from definitive hosts to snails and from snails to definitive hosts involve all manners of sociological and geographical complications (see, e.g. [16]), the transmission process is not uniform. For example, the time required for a juvenile schistosome to become mature and for an infected snail to begin shedding cercariae are different for different

schistosomes and snails. Therefore, it is reasonable to decompose the pipe into  $N$  distinct parallel subpipes with transit time  $\tau_i$ ,  $i = 1, \dots, N$ . Suppose a proportion  $\alpha_i$  of the miracidia-cercariae flow passes through the  $i$ -th subpipe, where  $\alpha_i \geq 0$  and  $\sum_{i=1}^N \alpha_i = 1$ , then we have

$$u_1(t + \Delta t) = g_1(g_1^{-1}(u_1(t))e^{-\beta_1 \Delta t}) + \Delta t \cdot g_1 \left( \frac{L_1 \sum_{i=1}^N \alpha_i K_1 u_1(t - \tau_i)}{\sum_{i=1}^N \alpha_i K_1 u_1(t - \tau_i) - \ln p} \right)$$

and thus

$$\frac{d}{dt} u_1(t) = -\beta_1 g'_1(g_1^{-1}(u_1(t)))g_1^{-1}(u_1(t)) + g_1 \left( \frac{L_1 \sum_{i=1}^N \alpha_i K_1 u_1(t - \tau_i)}{\sum_{i=1}^N \alpha_i K_1 u_1(t - \tau_i) - \ln p} \right).$$

Again we note that the function

$$I_2(t) = g_1 \left( \frac{L_1 \sum_{i=1}^N \alpha_i K_1 u_1(t - \tau_i)}{\sum_{i=1}^N \alpha_i K_1 u_1(t - \tau_i) - \ln p} \right)$$

can be expressed by

$$I_2(t) = g_1 \left( \frac{L_1 \int_0^\infty K_1 u_1(t - s) dF_N(s)}{\int_0^\infty K_1 u_1(t - s) dF_N(s) - \ln p} \right),$$

where

$$F_N(s) = \sum_{i=1}^N \alpha_i e(s - \tau_i) \quad \text{for } s \geq 0.$$

The function  $F_N(s)$  characterizes the transit-time distribution of the pipe and has a demonstrative meaning similar to that given for the function  $F_0(s)$ .

From real analysis theory it is known that any distribution function  $F(s)$  can be approximated by step functions like  $F_N(s)$ . On this base, we introduce the following definition.

**DEFINITION 2.1.** *The function  $F : [0, \infty) \rightarrow [0, \infty)$  is called the transit-time distribution function of the pipe  $T$ , if*

- (i)  *$F$  is monotone nondecreasing and continuous from the left;*
- (ii)  *$F(0) = 0$  and  $\lim_{s \rightarrow \infty} F(s) = 1$ ;*
- (iii) *the increasing rate of new paired schistosomes is*

$$g_1 \left( \frac{L_1 \int_0^\infty K_1 u_1(t - s) dF(s)}{\int_0^\infty K_1 u_1(t - s) dF(s) - \ln p} \right).$$

Therefore we obtain the following retarded equation

$$\begin{aligned} \frac{d}{dt} u_1(t) &= -\beta_1 g'_1(g_1^{-1}(u_1(t)))g_1^{-1}(u_1(t)) \\ &\quad + g_1 \left( L_1 \frac{\int_0^\infty K_1 u_1(t - s) dF(s)}{\int_0^\infty K_1 u_1(t - s) dF(s) - \ln p} \right). \end{aligned}$$

Returning to our general problem where there are  $n$  groups of definitive hosts, we assume that there are  $n^2$  pipes  $T_{ij}$ ,  $1 \leq i, j \leq n$ , coming out from host  $P_j$  and coming into host  $P_i$ , and that the transit-time distribution function of  $T_{ij}$  is  $F_{ij}(s)$ . Then the rate at which new paired schistosomes in  $P_i$  are produced is

$$g_i \left( \frac{L_i \sum_{j=1}^n \int_0^\infty K_j u_j(t-s) dF_{ij}(s)}{\sum_{j=1}^n \int_0^\infty K_j u_j(t-s) dF_{ij}(s) - \ln p} \right)$$

and the model equation becomes

$$\begin{aligned} \frac{d}{dt} u_i(t) = & -\beta_i g'_i(g_i^{-1}(u_i(t))) g_i^{-1}(u_i(t)) \\ & + g_i \left( \frac{L_i \sum_{j=1}^n \int_0^\infty K_j u_j(t-s) dF_{ij}(s)}{\sum_{j=1}^n \int_0^\infty K_j u_j(t-s) dF_{ij}(s) - \ln p} \right). \end{aligned}$$

A combination of two particular cases will be investigated in the following section. The first case is when  $F_{ij}(s) = e(s - \tau_{ij})$  for constants  $\tau_{ij} \geq 0$ ,  $i, j = 1, \dots, n$ . In this case the model equation becomes functional differential equation with finite delay

$$\begin{aligned} \frac{d}{dt} u_i(t) = & -\beta_i g'_i(g_i^{-1}(u_i(t))) g_i^{-1}(u_i(t)) \\ (2.3) \quad & + g_i \left( \frac{L_i \sum_{j=1}^n K_j u_j(t - \tau_{ij})}{\sum_{j=1}^n K_j u_j(t - \tau_{ij}) - \ln p} \right). \end{aligned}$$

When  $\tau_{ij} \equiv 0$ , (2.3) reduces to the ordinary differential equation (2.2). The second case is when  $F_{ij}(s) = \delta_{ij} \int_0^\infty u^{k_{ij}} e^{-\alpha_{ij} u} du$ , where  $k_{ij} \geq 0$  are integers,  $\alpha_{ij} \geq 0$  are constants, and  $\delta_{ij} > 0$  are constants such that  $\lim_{s \rightarrow \infty} F_{ij}(s) = 1$ . In this case the model equation becomes an integrod-differential equation

$$\begin{aligned} \frac{d}{dt} u_i(t) = & -\beta_i g'_i(g_i^{-1}(u_i(t))) g_i^{-1}(u_i(t)) \\ & + g_i \left( \frac{L_i \sum_{j=1}^n \int_0^\infty \delta_{ij} K_j u_j(t-s) s^{k_{ij}} e^{-\alpha_{ij} s} ds}{\sum_{j=1}^n \int_0^\infty \delta_{ij} K_j u_j(t-s) s^{k_{ij}} e^{-\alpha_{ij} s} ds - \ln p} \right). \end{aligned}$$

**3. Global dynamics of model equations.** Motivated by the discussion in the last section, we now consider the following model equations

$$\begin{aligned} (3.1) \quad & \frac{d}{dt} u_i(t) = -h_i(u_i(t)) \\ & + g_i \left( \frac{L_i \sum_{j=1}^n K_j \left( \eta_{ij} u_j(t - \tau_{ij}) + \int_{-\infty}^t \delta_{ij}(t-s)^{k_{ij}} e^{-\alpha_{ij}(t-s)} u_j(s) ds \right)}{\sum_{j=1}^n K_j \left( \eta_{ij} u_j(t - \tau_{ij}) + \int_{-\infty}^t \delta_{ij}(t-s)^{k_{ij}} e^{-\alpha_{ij}(t-s)} u_j(s) ds \right) - \ln p} \right) \end{aligned}$$

where  $\beta_i$ ,  $L_i$ ,  $K_j$ ,  $\delta_{ij}$ ,  $\eta_{ij}$ ,  $\alpha_{ij}$  are positive numbers,  $k_{ij}$  are nonnegative integers,  $0 < p < 1$ ,  $\eta_{ij} + \delta_{ij} \int_0^\infty s^{k_{ij}} e^{-\alpha_{ij}s} ds = 1$ ,  $g_i$  satisfies the assumption (H) and

$$h_i(u_i) = \beta_i g'_i(g_i^{-1}(u_i)) g_i^{-1}(u_i), \quad 1 \leq i, j \leq n.$$

Therefore,  $h_i$  is continuously differentiable and increasing.

Note that the general term involving integration is of the form

$$\begin{aligned} & \int_{-\infty}^t (t-s)^{k_{ij}} e^{-\alpha_{ij}(t-s)} u_j(s) ds \\ &= \sum_{\ell=0}^{k_{ij}} \binom{\ell}{k_{ij}} \left( \int_{-\infty}^0 (-s)^{k_{ij}-\ell} e^{\alpha_{ij}s} u_j(s) ds \right) t^\ell e^{-\alpha_{ij}t} \\ &+ \int_0^t (t-s)^{k_{ij}} e^{\alpha_{ij}(s-t)} u_j(s) ds. \end{aligned}$$

This leads us to select the state space  $C_{r,\alpha,k}$  which is defined as follows. Let  $R_- = (-\infty, 0]$  and  $R_+ = [0, +\infty)$ , and let  $\tilde{X}$  be the set of all bounded and continuous functions from  $R_-$  to  $R^n$ . Suppose  $\alpha_{ij} > 0$  and  $k_{ij}$  are given nonnegative integers,  $i, j = 1, \dots, n$ . We define  $p : \tilde{X} \rightarrow R_+$  by

$$p(\tilde{\varphi}) = \max_{1 \leq i \leq n} \sup_{-r_i \leq \theta_i \leq 0} |\tilde{\varphi}_i(\theta_i)| + \sup_{1 \leq i, j \leq n} \sup_{0 \leq m \leq k_{ij}} \left| \int_{-\infty}^0 (-s)^{k_{ij}-m} e^{\alpha_{ij}s} \tilde{\varphi}_j(s) ds \right|,$$

where  $r = (r_1, \dots, r_n) \in R_n$  with  $r_j = \max_{1 \leq i \leq n} \tau_{ij}$ ,  $\alpha = (\alpha_{ij})$  and  $k = (k_{ij})$ .

Evidently,  $p$  is a seminorm and the quotient space  $C_{r,\alpha,k} \triangleq \tilde{X}/p$  is a Banach space with the norm

$$|\varphi| = \max_{1 \leq i \leq n} \sup_{-r_i \leq \theta_i \leq 0} |\varphi_i(\theta_i)| + \sup_{1 \leq i, j \leq n} \sup_{0 \leq m \leq k_{ij}} \left| \int_{-\infty}^0 (-s)^{k_{ij}-m} e^{\alpha_{ij}s} \varphi_j(s) ds \right|.$$

It can be shown that  $C_{r,\alpha,k}$  satisfies Axioms 1–16 in [33].

In what follows for any given  $v = (v_1, \dots, v_n)^T$ ,  $u = (u_1, \dots, u_n)^T \in R^n$ ,  $v \leq u$  means  $v_i \leq u_i$  for  $1 \leq i \leq n$ ,  $v < u$  means  $v \leq u$  but  $v \neq u$ , and if  $\varphi \in C_{r,\alpha,k}$  then  $\varphi \geq 0$  means  $\varphi(0) \geq 0$ .

For any  $x, y, z \in R^n$ , define

$$\begin{aligned} f_i \left( \bigvee_{1 \leq j \leq n} x_j, \bigvee_{1 \leq j \leq n} y_j, \bigvee_{1 \leq j \leq n} z_j \right) = \\ - h(x_i) + g_i \left( \frac{L_i \sum_{j=1}^n K_j (\eta_{ij} y_j + \delta_{ij} z_j)}{\sum_{j=1}^n K_j (\eta_{ij} y_j + \delta_{ij} z_j) - \ln p} \right). \end{aligned}$$

Then (3.1) can be rewritten as

$$\begin{aligned} \frac{d}{dt}u_i(t) = \\ f_i\left(\bigvee_{1 \leq j \leq n} u_j(t), \bigvee_{1 \leq j \leq n} u_j(t - \tau_{ij}), \bigvee_{1 \leq j \leq n} \int_{-\infty}^t (t-s)^{k_{ij}} e^{\alpha_{ij}(s-t)} u_j(s) ds\right). \end{aligned}$$

Because of the increasing property of the function  $\frac{s}{s - \ln p}$  for  $s \geq 0$ , we can easily verify that  $f_i$  satisfies the assumptions (H1–H4) in [33]. Therefore, we know that Equation (3.1) defines an eventually strongly monotone semiflow  $T(t) : C_{r,\alpha,k} \rightarrow C_{r,\alpha,k}$  ( $T(t)\varphi \triangleq u_t(\varphi)$ ,  $t \geq 0$ ), i.e.,  $\varphi < \psi$  implies that  $T(t)\varphi \leq T(t)\psi$  for  $t$  sufficiently large. Using Theorems 2.1 and 2.6 in [33] we can show that if  $\varphi \in C_{r,\alpha,k}$  with  $\varphi \geq 0$ , then  $u(t; 0, \varphi) \geq 0$  and  $T(t)\varphi \geq 0$  for all  $t \geq 0$ . Moreover, system (3.1) is point dissipative. Therefore, all the solutions with nonnegative initial data remains nonnegative and bounded. By Theorem 2.4 in [33], there exists a global attractor. The structure of the attractor is described in the following results.

Let  $x = (x_1, \dots, x_n) \in R^n$  with  $x \geq 0$ . After some algebra we know that  $x$  is an equilibrium of (3.1) if and only if  $x_1$  solves the following equation

$$h_1(x_1) = g_1\left(\frac{L_1 \sum_{j=1}^n K_j p_j(x_1)}{\sum_{j=1}^n K_j p_j(x_1) - \ln p}\right).$$

Assume that

$$(3.2) \quad 0 \text{ is a regular value of } h_1(x_1) - g_1\left(\frac{L_1 \sum_{j=1}^n K_j p_j(x_1)}{\sum_{j=1}^n K_j p_j(x_1) - \ln p}\right).$$

Then it can be shown that there exists at least one equilibrium  $(0, \dots, 0)$ , and that the number of equilibria is finite. In fact, we have the following:

**Result 1.** *If the assumption (3.2) is satisfied, then the number of equilibria is odd, and the equilibria are totally ordered*

$$x^1 \leq x^2 \leq \dots \leq x^{2^m+1}, \quad x^1 = (0, \dots, 0).$$

It can be shown that there exists a unique equilibrium, which is the infection-free equilibrium  $x^1 = 0$ , if the following inequality holds

$$(3.3) \quad v(x) \sum_{j=1}^n \frac{L_j K_j g'_j(L_j q(x))}{h'_j(x_j)} < 1 \quad \text{for all } x \geq 0,$$

where

$$v(x) = \frac{-\ln p}{\left(\sum_{j=1}^n K_j x_j - \ln p\right)^2}, \quad q(x) = \frac{\sum_{j=1}^n K_j x_j}{\sum_{j=1}^n K_j x_j - \ln p}.$$

In this case, by Corollary 3.1 in [33] we have the global stability of the infection-free equilibrium:

**Result 2.** *If (3.3) holds, then for any  $\varphi \in C_{r,\alpha,k}$  with  $\varphi \geq 0$ ,  $\lim_{t \rightarrow \infty} T(t)\varphi = 0$ .*

Therefore, every nonnegative solution is convergent to zero as  $t \rightarrow \infty$ , and thus the infection can not maintain itself.

The case of multi-equilibria is more complicated but cannot be chaotic. For example, by Theorem 3.7 in [33], we know that System (3.1) has no attracting periodic orbits. To give more information about global dynamics, we consider the stability of each equilibrium point. It is easy to calculate that at equilibrium  $x$ , the linear variational equation is

$$\begin{aligned} \frac{d}{dt} u_i(t) &= -h'_i(x_i)u_i(t) + L_i v(x)g'_i(L_i q(x)) \sum_{j=1}^n K_j \left( \eta_{ij} u_j(t - \tau_{ij}) \right. \\ &\quad \left. + \delta_{ij} \int_{-\infty}^t (t-s)^{k_{ij}} e^{-\alpha_{ij}(t-s)} u_j(s) ds \right). \end{aligned}$$

One can verify that the condition (3.2) in [33] holds with  $\lambda_0 = -\min_{1 \leq i,j \leq n} \alpha_{ij}$ . Therefore, by Theorem 3.2 in [33], the stability of  $x$  is determined by the stability of the corresponding ordinary differential equation

$$\frac{d}{dt} u_i(t) = -h'_i(x_i)u_i(t) + L_i v(x)g'_i(L_i q(x)) \sum_{j=1}^n K_j u_j(t).$$

Let  $A_i(x) = L_i v(x)g'_i(L_i q(x))$ . Then, for any  $\ell$ ,  $1 \leq \ell \leq n$ , we consider

$$\begin{aligned} &(-1)^\ell \det \begin{pmatrix} -h'_1(x_1) + A_1(x)K_1 & A_1(x)K_2 & \dots & A_1(x)K_\ell \\ A_2(x)K_1 & -h'_2(x_2) + A_2(x)K_2 & \dots & A_2(x)K_\ell \\ A_\ell(x)K_1 & A_\ell(x)K_2 & \dots & -h'_\ell(x_\ell) + A_\ell(x)K_\ell \end{pmatrix} \\ &= (-1)^\ell \det \begin{pmatrix} -h'_1(x_1) + A_1(x)K_1 & A_1(x)K_2 & \dots & A_1(x)K_\ell \\ h'_1(x_1) \frac{A_2(x)}{A_1(x)} & -h'_2(x_2) & \dots & 0 \\ h'_1(x_1) \frac{A_\ell(x)}{A_1(x)} & 0 & \dots & -h'_\ell(x_\ell) \end{pmatrix} \\ &= (-1)^\ell (-1)^{\ell-1} h'_2(x_2) \dots h'_\ell(x_\ell) \left[ -h'_1(x_1) + A_1(x)K_1 + \sum_{j=2}^\ell \frac{A_j(x)K_j h'_1(x_1)}{h'_j(x_j)} \right] \\ &= h'_1(x_1) \dots h'_\ell(x_\ell) \left[ 1 - \sum_{j=1}^\ell \frac{L_j K_j g'_j(L_j q(x))}{h'_j(x_j)} v(x) \right]. \end{aligned}$$

Note that

$$v(x) \sum_{j=1}^n \frac{L_j K_j g'_j(L_j q(x))}{h'_j(x_j)} < 1$$

at  $x^1, x^3, \dots, \dots, x^{2m+1}$ , and

$$v(x) \sum_{j=1}^n \frac{L_j K_j g'_j(L_j q(x))}{h'_j(x_j)} > 1$$

at  $x^2, x^4, \dots, x^{2m}$ . By Theorems 3.2 and 3.3 in [33] we obtain

**Result 3.** *If the inequality in (3.3) is reversed for some  $x \geq 0$ , then there exist multiple equilibria which have the following properties: (a)  $x^1, x^3, \dots, x^{2m+1}$  are (locally) asymptotically stable and  $x^2, x^4, \dots, x^{2m}$  are unstable; (b) there exists a monotone increasing orbit connecting  $x^1$  to  $x^2$  and a monotone decreasing orbit connecting  $x^3$  to  $x^2$ , and the identical assertion holds for the other equilibria  $x_3, x_4, \dots$ . Moreover,*

$$\begin{aligned}\lim_{t \rightarrow \infty} T(t)\varphi &= x^{2k-1} && \text{for } x^{2k-1} \leq \varphi < x^{2k}, k = 1, \dots, m, \\ \lim_{t \rightarrow \infty} T(t)\varphi &= x^{2k+1} && \text{for } x^{2k} < \varphi \leq x^{2k+1}, k = 1, \dots, m, \\ \lim_{t \rightarrow \infty} T(t)\varphi &= x^{2m+1} && \text{for } \varphi \geq x^{2m+1}.\end{aligned}$$

Result 3 can be illustrated by the schematic pictures of the global dynamics shown in Figure 3. Biologically, Result 3 indicates that in the case where multiple equilibria exist, if the initial parasite load is low, then the infection can not maintain itself; however if the initial parasite load is high, then the infection will maintain at a stable configuration. We are unable to show that every solution of the model equation is convergent to an equilibrium, but we do know from Theorems 3.4 and 3.6 in [33] that in most situations (for almost every initial condition) the infection maintain itself in stable configurations as the following result indicates.

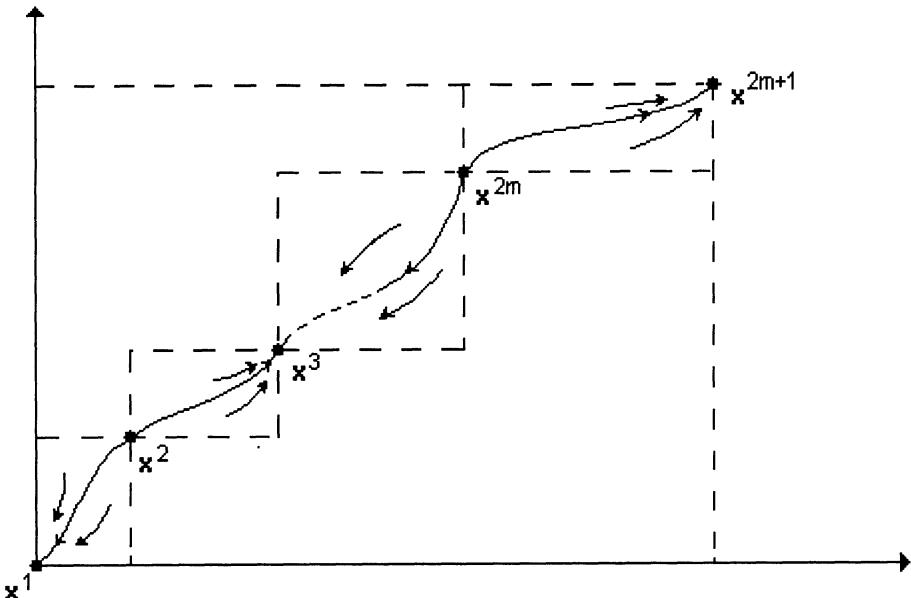


FIG. 3. The graph illustrates Result 3 for the global dynamics of (3.1).

**Result 4.** *The union of the basins of attractions of  $x^1, x^3, \dots, x^{2m+1}$  is open and dense, and the global attractor is contained in the closed interval  $[[0, x^{2m+1}]] = \{\varphi; 0 \leq \varphi \leq x^{2m+1}\}$ .*

**4. Discussion.** In this paper we propose a system of delay differential equations as a model for the transmission dynamics of *schistosomiasis japonicum*. The model takes into account multiple definitive hosts as well as the time lags for the transmission of infection. It is shown that the system has either only the infection-free equilibrium  $x^1 = 0$  or an odd number of multiple equilibria  $x^1 \leq x^2 \leq x^3 \leq \dots \leq x^{2m+1}$ . We give the condition for the global asymptotical stability of the infection-free equilibrium (see (3.3)), and we show that, if multiple non-trivial equilibria exist, then the disease will either die out or persist in the population at a constant level depending on the initial distribution of the parasites among different definitive host populations.

The condition (3.3) allows us to gain some insights into the transmission of infection and to determine the role of various parameters in the disease dynamics. Notice that  $g'_i(L_i q(x))$  and  $h'_i(x_i)$  correspond to the "birth" and "death" rates of worms, respectively, in the host population  $P_i$ . Also note that  $L_i$  and  $K_i$  represent the forces of infection associated with the two transmission processes respectively: from the intermediate hosts to the definitive hosts and from the definitive hosts to the intermediate hosts. Hence the expression on the left hand side of (3.3) provides a measure of the reproductive number of the parasites. Since all the environmental factors are included in  $L_i$  and  $K_i$ , according to the results of our model, one may try to reduce the values of  $L_i$  and  $K_i$  such that the reproductive number of the parasites (the LHS of (3.3)) is below one. In this case the parasite population will go to extinction.

Our results indicate that the qualitative behavior of our model equations is not sensitive to the delay. This coincides with the theoretical analysis of Lee and Lewis [9], Lewis [10] and Nasell [21] by introducing a transition rate to include the latency in snail infections in a MacDonalds like model, with the qualitative study of May [16] by introducing a delay in the MacDonald-Nasell-Hirsch model, and with the computer simulation by Lee and Lewis [9]. Our results also show that non-human definitive hosts play an important role in the persistence of the disease. Consider a situation in which  $n = 2$  with  $P_1$  and  $P_2$  denoting the human and non-human hosts, respectively, and the disease is endemic in both of the populations. If a control program concerns only the human hosts, then even in the case when  $L_1$  and  $K_1$  can be reduced to be very small, the disease may still persist in both of the host populations as long as  $L_2$  and/or  $K_2$  are large enough so that (3.3) does not hold.

## REFERENCES

- [1] BRADLEY, J. AND MAY, R.M., Consequences of helminth aggregation for the dynamics of schistosomiasis, *Trans. Roy. Soc. Trop. Med. Hyg.* **72**: 262–273, 1978.
- [2] BOWLES, J., BLAIR, D., AND McMANUS, D.P., A molecular phylogeny of the human schistosomes, *Molec. Phylogen. and Evolut.* **4**(2): 103–109, 1995.

- [3] HAGAN, P. AND GRYSEELS, B., Schistosomiasis research and the European Community, *Tropic. and Geogr. Medic.* **46**(4):259–268, 1994.
- [4] HALE, J.K. AND KATO, J., Phase space for retarded equations with infinite delay, *Funkcialaj Ekvac.* **21**: 11–41, 1978.
- [5] HE, Y., XU, S., SHI, F., SHEN, W., HUS, S., AND HSU, H., Comparative studies on the infection and maturation of *Schistosoma japonicum* in cattle and buffaloes, *Acta Zool. Sinica* **38**(3): 266–271, 1992.
- [6] HIRSCHM M.W., The dynamical systems approach to differential equations, *Bull. Amer. Math. Soc.* **11**: 1–64, 1984.
- [7] ———, Systems of differential equations that are competitive or cooperative II: Convergence almost everywhere, *SIAM J. Math. Anal.* **16**: 423–439, 1985.
- [8] ———, Stability and convergence in strongly monotone dynamical systems, *J. Reine Angew Math.* **383**: 1–53, 1988.
- [9] LEE, K.L. AND LEWIS, E.R., Delay time models of population dynamics with application to schistosomiasis control, *IEEE Trans. Biomed. Eng.* **BME-23**: 225–233, 1976.
- [10] LEWIS, T., A model for the parasitic disease bilharziasis, *Adv. Appl. Probab.* **7**: 673–704, 1975.
- [11] MACDONALD, G., The dynamics of helminth infections, with special reference to schistosomes, *Trans. R. Soc. Trop. Med. Hyg.* **59**: 489–506, 1965.
- [12] ———, On the Scientific basis of tropical hygiene, *Trans. R. Soc. Trop. Med. Hyg.* **59**: 611–620, 1965.
- [13] MATANO, H., Asymptotic behavior and stability of solutions of semilinear diffusion equations, *Publ. Res. Inst. Math. Sci.*, **15**: 401–454, 1979.
- [14] ———, Existence of nontrivial unstable sets for equilibrium of strongly order-preserving systems, *J. Fac. Sci. Univ. Tokyo, Sec. t-1A Math.* **30**: 645–673, 1983.
- [15] ———, Strongly order-preserving local semi-dynamical systems — theory and applications, *Proceedings of Autumn Course on Semigroups*, Pitman, New York, 1985.
- [16] MAY, R.M., Togetherness among schistosomes, its effects on the dynamics of the infection, *Math. Biosci.* **35**: 301–343, 1977.
- [17] NAITO, T., Integral manifolds for linear functional differential equations on some Banach space, *Funkcialaj Ekvac.* **13**: 199–213.
- [18] ———, On linear autonomous retarded equations with an abstract space for infinite delay, *J. Differential Equations* **33**: 74–91, 1979.
- [19] ———, On autonomous linear functional differential equations with infinite retardations, *J. Differential Equations* **2**: 297–315, 1976.
- [20] NASELL, I., A hybrid model of schistosomiasis with snail latency, *Theor. Popul. Biol.* **10**: 47–69, 1976.
- [21] ———, On eradication of schistosomiasis, *Theor. Popul. Biol.* **10**: 133–144, 1976.
- [22] NASELL, I. AND HIRSCH, W.M., A mathematical model of some helminthic infections, *Comm. Pure Appl. Math.* **25**: 459–477, 1972.
- [23] ———, The transmission dynamics of schistosomiasis, *Comm. Pure Appl. Math.* **26**: 395–453, 1973.
- [24] NUSSBAUM, R., The radius of the essential spectrum, *Duke Math. J.* **37** : 473–488, 1970.
- [25] ———, Positive operators and elliptic eigenvalue problems, *Math. Zeit.* **186**: 247–269, 1984.
- [26] ———, A folk theorem in the spectral theory of  $c_0$ -semigroups, *Pac. J. Math.* **113**: 433–449, 1984.
- [27] ———, Hilbert's Projective Metric and Iterated Nonlinear Maps, *Mem. Amer. Math. Soc.* **391**, 1988.
- [28] SHI, Y., JIANG, C., HAN, J., LI, Y., AND RUPPEL, A., Immunization of pigs against infection with *Schistosoma japonicum* using ultraviolet-attenuated cercaria, *Parasitology* **106**(5): 459–462, 1993.

- [29] SMITH, H., Periodic solutions of periodic competitive and cooperative systems, *SIAM J. Math. Anal.* **17**: 1289–1318, 1986.
- [30] ———, Invariant curves for mappings, *SIAM J. Math. Anal.* **17**: 1053–1067, 1986.
- [31] ———, Monotone semiflows generated by functional differential equations, *J. Differential Equations* **66**: 420–442, 1987.
- [32] WU, J., LIU, N., AND ZUO, S., The qualitative analysis of model of the transmission dynamics of Japanese schistosomiasis, *Applied Mathematics, a Journal of Chinese Universities* **2**: 352–362, 1987.
- [33] WU, J., Strong monotonicity principles and applications to Volterra integrodifferential equations, *Differential Equations: Stability and Control*, Saber, ed., Marcel Dekker, 1990.
- [34] WU, Z., BU, K., YUAN, L., YANG, G., ZHU, Z., AND LIU, Q., Factors contributing to reinfection with schistosomiasis japonica after treatment in the lake region of China, *Acta Tropica* **52**(2): 83–88, 1993.

# INFECTIOUS DISEASE MODELS WITH CHRONOLOGICAL AGE STRUCTURE AND EPIDEMIOLOGICAL AGE STRUCTURE\*

FRED BRAUER†

**Abstract.** Disease models with recovery rates depending on disease-age and with exponentially distributed natural life spans have been studied by H. Thieme, C. Castillo-Chavez, and others. Here, we formulate S-I-R models in which both disease recovery and natural life spans have arbitrary distributions. We focus on the relation between the basic reproductive number, the mean life span, and the mean age at infection. The S-I model, with no recovery, is analyzed completely and partial stability results are obtained for models with mean infective period much shorter than mean life span.

**Key words.** epidemic models, disease-age dependence.

**AMS(MOS) subject classifications.** 92D 30.

**1. Introduction.** The interaction between demographic and epidemiological effects is an important aspect of infectious disease models. It is common to assume a natural death rate proportional to population size, or equivalently, an exponential distribution of life spans. The purpose of this paper is to explore the implications of assuming an arbitrary distribution of natural life spans, without going to a fully age-structured model. The obvious topics to examine are the basic reproductive number, stability of the disease-free equilibrium, and existence and stability of endemic equilibria. In addition, we shall formulate a relation between the basic reproductive number, the mean life span, and the mean age at infection which generalizes the simple relation for the exponentially distributed life span case originally derived by K. Dietz [8,9].

Most past work which deals with age of infection has assumed an exponentially distributed infective period [13] or an arbitrary infective period distribution with exponentially distributed life spans [3,12,14,15,16]. We are unable to analyze the stability of the endemic equilibrium in general, but there are two special cases which we can analyze completely, namely the case of exponentially distributed life spans and the case of an S-I model, describing a disease which lasts for a lifetime but does not cause death. Another special case of interest is where the mean infective period is very short compared to the mean life span, and we are able to obtain some partial stability results in this case.

**2. The basic model and its equilibria.** We consider an S-I-R model under the following assumptions:

---

\*This research is supported by NSERC grant OGPIN 203901-99.

†Department of Mathematics, University of British Columbia, Vancouver, B.C. V6T 1Z2, Canada; brauer@math.ubc.ca.

- (i) There is a constant number  $\Lambda$  of births in unit time and a fraction  $Q(a)$  of members survive to age  $a$ , so that the mean life span is

$$L = \int_0^\infty Q(a)da.$$

The function  $Q(a)$  has  $Q(0) = 1$  and is non-negative and monotone non-increasing, and the mean lifespan  $L$  is finite.

- (ii) On the average, an infective makes  $\beta N$  contacts sufficient to transmit infection in unit time, where  $N$  is the total population size. Since a fraction  $S/N$  of these contacts is with a susceptible and thus produces a new infection, the number of new infectives in unit time is

$$I \cdot \beta N \cdot \frac{S}{N} = \beta SI.$$

- (iii) The fraction of infectives who remain infective a time  $u$  (infection age) after having become infective (if still alive) is  $P(u)$ . The function  $P(u)$  has  $P(0) = 1$  and is non-negative and monotone non-increasing.

One consequence of these assumptions is that the mean infective period, corrected for natural mortality, for members of the population who were infected at (chronological) age  $a$  is

$$\int_0^\infty \frac{Q(a+u)}{Q(a)} P(u)du.$$

At equilibrium, the fraction of the population which is of age  $a$  is

$$\frac{Q(a)}{\int_0^\infty Q(a)du}$$

and therefore the mean infective period for the entire population is

$$(2.1) \quad \begin{aligned} \tau &= \frac{\int_0^\infty Q(a)[\int_0^\infty \frac{Q(a+u)}{Q(a)} P(u)du]du}{\int_0^\infty Q(a)da} \\ &= \frac{\int_0^\infty P(u)[\int_0^\infty Q(a+u)da]dy}{\int_0^\infty Q(a)du}. \end{aligned}$$

Our model formulation will be a single equation for the number of infectives  $I(t)$  at time  $t$ , but we will also be able to give an expression for the number of susceptibles  $S(t)$  at time  $t$  in terms of  $I(t)$ .

In order to generalize our model to situations where the constant contact rate  $\beta$  or the constant birth rate  $\Lambda$  are replaced by functions of total population size, we would also need an equation for the total population

size  $N(t)$  at time  $t$ . Later, we will discuss how this may be done under various assumptions about recovery and disease deaths.

In order to formulate our model, we follow members of the population who were born at time  $(t - a - u)$  and who become infective at time  $(t - u)$  while of chronological age  $a$ . Those who survive and remain infective until time  $t$  have chronological age  $(a + u)$  and infection age  $u$  at time  $t$ . We may then obtain  $I(t)$  by integrating with respect to both  $a$  and  $u$ .

Since a fraction  $\beta I$  of susceptibles become infective in unit time, of the  $\Lambda$  members born at time  $(t - a - u)$ , with  $\Lambda Q(a)$  survivors at time  $(t - u)$ , the number of survivors who have not been infected and thus remain susceptible is

$$\Lambda Q(a) e^{-\int_{t-a-u}^{t-u} \beta I(\alpha) d\alpha}$$

The rate of new infections of age  $a$  at time  $(t - u)$  is

$$\beta I(t-u) \Lambda Q(a) e^{-\int_{t-a-u}^{t-u} \beta I(\alpha) d\alpha}$$

and of these a fraction  $\frac{Q(a+u)}{Q(a)} P(u)$  survive and remain infective to time  $t$  with chronological age  $(a + u)$  and infection age  $u$ . Thus the number of infectives at time  $t$  with chronological age  $(a + u)$  and infection age  $u$  is

$$\beta I(t-u) \Lambda Q(a+u) P(u) e^{-\int_{t-a-u}^{t-u} \beta I(\alpha) d\alpha}$$

and this leads to the integral equation

$$(2.2) \quad I(t) = \int_0^\infty \beta \Lambda I(t-u) P(u) \left[ \int_0^\infty Q(a+u) e^{-\int_{t-a-u}^{t-u} \beta I(\alpha) d\alpha} da \right] du$$

as our basic model. Since the number of members born at time  $(t - a)$  who remain alive and susceptible until time  $t$  is

$$\Lambda Q(a) e^{-\int_{t-u}^t \beta I(\alpha) d\alpha}$$

we also obtain the representation

$$(2.3) \quad S(t) = \int_0^\infty \Lambda Q(a) e^{-\int_{t-a}^t \beta I(\alpha) d\alpha} da.$$

Equilibria of (2.2) are given by  $I = 0$  (disease-free) or

$$(2.4) \quad \beta \Lambda \int_0^\infty P(u) \left[ \int_0^\infty Q(a+u) e^{-\beta I a} da \right] du = 1$$

(endemic).

Since the carrying capacity of the population is

$$K = \Lambda \int_0^\infty Q(a)da$$

and the mean infective period  $\tau$  is given by (2.1), the basic reproductive number  $R_0$ , defined as the mean number of secondary infections caused by a single infective introduced into a susceptible population over the course of the infection, is given by

$$(2.5) \quad R_0 = \beta \Lambda \int_0^\infty P(u) \left[ \int_0^\infty Q(a+u)da \right] du.$$

Combination of (2.4) and (2.5) gives the expression

$$(2.6) \quad R_0 = \frac{\int_0^\infty P(u) [\int_0^\infty Q(a+u)da] du}{\int_0^\infty P(u) [\int_0^\infty Q(a+u)e^{-\beta I_a}da] du}.$$

When the population is at an endemic equilibrium  $I$ , susceptibles who become infected at age  $a$  do so at a rate  $e^{-\beta I_a}$ ; thus the mean age  $A$  at infection of those who do become infected is given by

$$A = \frac{1}{\beta I}.$$

The relation (2.6) gives an expression for  $R_0$  in terms of the mean life span  $L = \int_0^\infty Q(a)da$  and the mean age at infection  $A$ . This expression depends on the distributions  $Q$  and  $P$ .

In the special case  $Q(a) = e^{-\mu a}$ , for which  $L = \frac{1}{\mu}$ , simplification of the integrals in (2.6) gives

$$(2.7) \quad R_0 = \frac{\int_0^\infty P(u)e^{-\mu u} [\int_0^\infty e^{-\mu a}da] du}{\int_0^\infty P(u)e^{-\mu u} [\int_0^\infty e^{-(\mu+\beta I)a}da] du} = \frac{\int_0^\infty e^{-\mu a}da}{\int_0^\infty e^{-(\mu+\beta I)a}da} \\ = \frac{\mu + \beta I}{\mu} = 1 + \frac{\beta I}{\mu} = 1 + \frac{L}{A}.$$

This relation, useful for estimating  $R_0$  when  $\beta$  can not be measured effectively, was first established by K. Dietz [8,9] for the ordinary differential equation model

$$(2.8) \quad \begin{aligned} S' &= \Lambda - \beta SI - \mu s \\ I' &= \beta SI - \mu I - \alpha I, \end{aligned}$$

which is the special case of (2) with  $Q(a) = e^{-\mu a}$ ,  $P(u) = e^{-\alpha u}$ . It holds also for the somewhat more general case of (2.2) with  $Q(a) = e^{-\mu a}$ ,  $P(u)$  arbitrary

$$(2.9) \quad \begin{aligned} S' &= \Lambda - \beta SI - \mu S \\ I(t) &= \int_{-\infty}^t \beta S(v)I(v)P(t-v)e^{-\mu(t-v)}dv \end{aligned}$$

studied in [5,6]. The equation for  $S$  in the models (2.8) and (2.9) is obtained by differentiation under the integral sign in (2.3). If  $Q(a)$  is not an exponential function, the relation between  $R_0$ ,  $L$ , and  $A$  takes a form different from (2.7), as has been pointed out by May [11] for populations with a fixed life span. In general, this relation will also depend on the form of  $P(u)$ .

**3. Models for diseases with short infective period.** For many diseases the disease time scale is much faster than the demographic time scale. This is true, for example, of the classical “childhood” diseases such as measles, chicken pox, mumps, and rubella. It is also true for fatal diseases which lead to rapid death from disease. Such diseases can be modelled as limiting cases by a perturbation analysis, as has been done by V. Andreasen [1,2]. We shall model short infective periods in a less precise manner by assuming that  $Q(a+u)$  may be replaced by  $Q(a)$  for all  $a \geq 0$  and all  $u$  such that  $P(u)$  is not negligible. Our analyses will also apply if  $Q(a+u)$  may be approximated by  $Q(a)p(u)$  with  $p(u)$  an arbitrary function of  $u$ , and this includes the case  $Q(a) = e^{-\mu a}$  treated in the preceding section. With the replacement of  $Q(a+u)$  by  $Q(a)$ , (6) becomes

$$(3.1) \quad R_0 \approx \frac{\int_0^\infty P(u) [\int_0^\infty Q(a)da] du}{\int_0^\infty P(u) [\int_0^\infty Q(a)e^{-\beta I a}da] du} = \frac{\int_0^\infty Q(a)da}{\int_0^\infty Q(a)e^{-\beta I a}da}.$$

This expression agrees with the one given by May [11] for a population with a fixed life span. We believe, however, that (3.1) is only an approximation and not an exact expression. If  $Q(a) = \begin{cases} 1, & 0 \leq a \leq L \\ 0, & a > L \end{cases}$ , (10) reduces to

$$(3.2) \quad R_0 \approx \frac{L}{\int_0^L e^{-\beta I a}da} = \frac{\beta I L}{1 - e^{-\beta I L}} = \frac{L/A}{1 - e^{-L/A}}.$$

The relations (2.7), (3.1), and (3.2) between  $R_0$ ,  $L$ , and  $A$  depend on the assumption that the system is at an asymptotically stable endemic equilibrium. In order to validate these relations, we must carry out an equilibrium stability analysis.

**4. Analysis of the characteristic equation.** We have observed that equilibria of (2.2) are given by  $I = 0$  and (2.4). We may consider the left side of (2.4) as a function of  $I$ ,

$$F(I) = \beta \Lambda \int_0^\infty P(u) \left[ \int_0^\infty Q(a+u)e^{-\beta I a}da \right] du.$$

Then  $F(0) = R_0$ ,  $\lim_{I \rightarrow \infty} F(I) = 0$ , and  $F'(I) \leq 0$  for  $0 \leq I < \infty$ . Thus if  $R_0 < 1$  there is no endemic equilibrium (solution of  $F(I) = 1$ ), while if  $R_0 > 1$  there is a unique endemic equilibrium.

The linearization of the equation (2.2) at an equilibrium  $I_0$  obtained by letting  $I(t) = I + x(t)$ , expanding with the aid of Taylor's theorem, and retaining only linear terms is

$$\begin{aligned} x(t) &= \int_0^\infty \beta \Lambda P(u) \left[ \int_0^\infty Q(a+u) e^{-\beta I a} da \right] x(t-u) dy \\ &\quad - \int_0^\infty \beta \Lambda \beta I P(u) \left[ \int_0^\infty Q(a+u) e^{-\beta I a} \left\{ \int_{t-a-u}^{t-u} x(\alpha) dx \right\} da \right] du. \end{aligned}$$

If we define

$$(4.1) \quad \Phi_I(u) = P(u) \int_0^\infty Q(a+u) e^{-\beta I a} da$$

and let  $\hat{\Phi}_I(\lambda)$  denote the Laplace transform of  $\Phi_I(u)$ ,

$$(4.2) \quad \hat{\Phi}_I(\lambda) = \int_0^\infty e^{-\lambda u} \Phi_I(u) du,$$

the characteristic equation at the equilibrium  $I$  is

$$(4.3) \quad \begin{aligned} \beta \Lambda \left\{ \hat{\Phi}_I(\lambda) - \beta I \int_0^\infty P(u) \right. \\ \left. \left[ \int_0^\infty Q(a+u) e^{-\beta I a} \left( \frac{1-e^{-\lambda a}}{\lambda} \right) da \right] e^{-\lambda u} du \right\} = 1 \end{aligned}$$

and the equilibrium condition (2.4) is

$$(4.4) \quad \beta \Lambda \int_0^\infty \Phi_I(u) du = \beta \Lambda \hat{\Phi}_I(0) = 1.$$

At the disease-free equilibrium  $I = 0$ , the characteristic equation (4.3) becomes

$$\beta \Lambda \hat{\Phi}_0(\lambda) = 1$$

and it is known [4] that all roots are in the left half-plane if and only if

$$R_0 = \beta \Lambda \hat{\Phi}_0(0) < 1.$$

Thus the disease-free equilibrium is asymptotically stable if and only if  $R_0 < 1$ . If  $R_0 > 1$  there is a unique endemic equilibrium  $I$  determined by (4.4). Using (4.4), we may write the characteristic equation (4.3) as

$$(4.5) \quad \begin{aligned} \frac{\hat{\Phi}_I(\lambda)}{\hat{\Phi}_I(0)} &= 1 + \frac{\beta I}{\hat{\Phi}_I(0)} \int_0^\infty P(u) \\ &\quad \left[ \int_0^\infty Q(a+u) e^{-\beta I a} \left( \frac{1-e^{-\lambda a}}{\lambda} \right) da \right] e^{-\lambda u} du. \end{aligned}$$

We are not able to analyze this characteristic equation in general, but we will show that in some special cases all roots are in the left half plane and thus the endemic equilibrium is (locally) asymptotically stable.

In the special case  $Q(a) = e^{-\mu a}$ , the model (2.2) reduces to the system (2.9) for which it has been shown that the endemic equilibrium is asymptotically stable if  $R_0 > 1$  [5,6].

This may also be shown by analyzing the characteristic equation (4.5). We may show that if  $Q(a) = e^{-\mu a}$ , then

$$\hat{\Phi}_I(\lambda) = \frac{\hat{P}(\lambda + \mu)}{\mu + \beta I}$$

and (4.5) becomes

$$\frac{\hat{P}(\lambda + \mu)}{\hat{P}(\mu)} = 1 + \frac{\beta I}{\lambda + \mu + \beta I} \frac{\hat{P}(\lambda + \mu)}{\hat{P}(\mu)}$$

or

$$\frac{\hat{P}(\lambda + \mu)}{\hat{P}(\mu)} \frac{\lambda + \mu}{\lambda + \mu + \beta I} = 1.$$

Since  $\left| \frac{\hat{P}(\lambda + \mu)}{\hat{P}(\mu)} \right| \leq 1$  and  $\left| \frac{\lambda + \mu}{\lambda + \mu + \beta I} \right| \leq 1$  for  $R\lambda \geq 0$ , there can be no roots in the right half plane (except for an extraneous root  $\lambda = 0$ ), and thus the endemic equilibrium is asymptotically stable.

One way to approach the analysis of the characteristic equation (4.5) is to consider

$$(4.6) \quad \begin{aligned} \frac{\hat{\Phi}_I(\lambda)}{\hat{\Phi}_I(0)} &= 1 + \frac{r\beta I}{\hat{\Phi}_I(0)} \int_0^\infty P(u) \\ &\quad \left[ \int_0^\infty Q(a+u)e^{-\beta I a} \left( \frac{1 - e^{-\lambda a}}{\lambda} \right) da \right] e^{-\lambda u} du. \end{aligned}$$

It is known [4] that for  $r = 0$  all roots of (4.6) are in the left half plane except for a simple root  $\lambda = 0$  which moves into the left half plane when  $r$  increases. In order to prove that all roots of (4.5) are in the left half plane, it suffices to show that as  $r$  increases from 0 to 1 no root can enter the right half plane. There are three ways in which a root could appear in the right half plane:

- (i) A root could appear “at infinity”, but this possibility may be ruled out since elementary properties of the Laplace transform imply that as  $|\lambda| \rightarrow \infty$  with  $R\lambda \geq 0$  the left side of (4.6) approaches 0 while the right side of (4.6) approaches 1.
- (ii) There could be a root  $\lambda = 0$  for some value of  $r$ ,  $0 < r \leq 1$ , but this possibility may be ruled out since as  $\lambda \rightarrow 0$ , the left side of (4.6) approaches 1 while the right side approaches

$$1 + \frac{r\beta I}{\hat{\Phi}_I(0)} \int_0^\infty P(u) \left[ \int_0^\infty Q(a+u)ae^{-\beta I a} da \right] du > 1.$$

- (iii) There could be a pair of pure imaginary roots  $\lambda = \pm iy$  for some  $y > 0$  and some  $r, 0 < r \leq 1$ .

The third possibility cannot be ruled out in general, but in order to establish asymptotic stability of the endemic equilibrium in a particular case we need only rule out the possibility of a pair of pure imaginary roots. Thus we have obtained the following theorem (which is not of much use since its central hypothesis is difficult to verify):

**THEOREM 4.1.** *If  $R_0 > 1$ , the endemic equilibrium of the model (2.2) is (locally) asymptotically stable provided*

$$(4.7) \quad \begin{aligned} \frac{\hat{\Phi}_I(iy)}{\hat{\Phi}_I(0)} \neq 1 + \frac{r\beta I}{\hat{\Phi}_I(0)} \int_0^\infty P(u) \\ \left[ \int_0^\infty Q(a+u) \left( \frac{1 - e^{-iya}}{iy} \right) da \right] e^{-iya} du \end{aligned}$$

for  $0 < y < \infty$ ,  $0 < r \leq 1$ .

However, there is a corollary which extends a result in [13] to the effect that for small disease prevalence ( $\beta I \rightarrow 0$ ) the endemic equilibrium is asymptotically stable.

Another special case of interest is when the mean infective period is very short compared to the mean life span. We have described this in the preceding section by assuming that  $Q(a+u) \approx Q(a)$  for  $0 \leq a < \infty$  and those values of  $u$  for which  $P(u)$  is not negligible. With this assumption, the condition (4.7) becomes

$$(4.8) \quad \begin{aligned} \frac{\hat{P}(iy)}{\hat{P}(0)} \neq 1 + \frac{r\beta I}{\hat{Q}(\beta I)} \frac{\hat{P}(iy)}{\hat{P}(0)} \int_0^\infty Q(a) \frac{1 - e^{-iya}}{iy} da, \\ \frac{\hat{P}(iy)}{\hat{P}(0)} \left[ 1 - \frac{r\beta I}{\hat{Q}(\beta I)} \int_0^\infty Q(a) \frac{1 - e^{-iya}}{iy} da \right] \neq 1, \\ 1 - \frac{r\beta I}{\hat{Q}(\beta I)} \int_0^\infty Q(a) \frac{1 - e^{-iyu}}{iy} da \neq \frac{\hat{P}(0)}{\hat{P}(iy)}. \end{aligned}$$

If we separate (4.8) into real and imaginary parts we obtain two relations

$$(4.9) \quad \begin{aligned} 1 - \frac{r\beta I}{\hat{Q}(\beta I)} \int_0^\infty Q(a) \frac{\sin ya}{y} da &= \frac{\hat{P}(0) \int_0^\infty P(u) \cos yudu}{|\hat{P}(iy)|^2} \\ \frac{r\beta I}{\hat{Q}(\beta I)} \int_0^\infty Q(a) \frac{1 - \cos ya}{y} da &= \frac{\hat{P}(0) \int_0^\infty P(u) \sin yudu}{|\hat{P}(iy)|^2} \end{aligned}$$

and the stability condition is that these are not both satisfied for any  $y$ ,  $0 < y < \infty$ , and  $r$ ,  $0 < r \leq 1$ . Since

$$\int_0^\infty Q(a) \frac{\sin yu}{y} da \geq 0$$

for  $0 \leq y < \infty$ , a sufficient condition for stability is

$$\frac{\hat{P}(0) \int_0^\infty P(u) \cos yu du}{|\hat{P}(iy)|^2} \geq 1, \quad (0 < y < \infty)$$

or

$$(4.10) \quad \int_0^\infty P(u) du \int_0^\infty P(u) \cos yu du \geq \left| \int_0^\infty P(u) e^{iyu} du \right|^2, \quad (0 < y < \infty).$$

It is possible to verify that (4.10) is satisfied for the particular choices  $P(u) = e^{-\alpha u}$  and  $P(u) = \begin{cases} 1, & 0 \leq u \leq \tau \\ 0, & u > \tau \end{cases}$ . We note, however, that a necessary condition that (4.10) be satisfied is  $\int_0^\infty P(u) \cos yu du \geq 0$ , which is true if  $P$  is a convex function ( $P(u) \geq 0$ ,  $P'(u) \leq 0$ ,  $P''(u) \geq 0$  for  $0 \leq u < \infty$ ), but is false if  $\int_0^\infty P(u) \cos yu du < 0$  for any  $y$ . Thus for certain choices of  $P$ , all roots of the approximate characteristic equation obtained when  $Q(a+u)$  is replaced by  $Q(a)$  in (4.5) have negative real part. Because the roots of the characteristic equations depend continuously on the functions in the equation, this is also true for the “true” characteristic equation provided  $Q(a+u)$  is a sufficiently good approximation to  $Q(a)$ . Thus, for these choices of  $P$ , the endemic equilibrium is asymptotically stable provided the infection time scale is sufficiently fast relative to the demographic time scale. It is natural to conjecture that the endemic equilibrium is always asymptotically stable when the infection time scale is fast relative to the demographic time scale.

**5. The S-I model.** The opposite extreme to models with very short infective periods is the S-I model, describing a disease in which infectivity lasts for life but does not cause death. Some sexually transmitted diseases, such as herpes, are of this type. It is possible to give a direct formulation for an S-I model in a population with age-dependent mortality, but we may also derive it from our general model (2.2) by taking  $P(u) = 1$  ( $0 \leq u < \infty$ ). Then (2.2) becomes

$$(5.1) \quad \begin{aligned} I(t) &= \int_0^\infty \beta \Lambda I(t-u) \left[ \int_0^\infty Q(a+u) e^{-\int_{t-a-u}^{t-u} \beta I(\alpha) d\alpha} da \right] du \\ &= \int_0^\infty \beta \Lambda I(t-u) \left[ \int_0^\infty Q(b) e^{-\int_{t-b}^{t-u} \beta I(\alpha) d\alpha} db \right] du \\ &= \int_0^\infty \Lambda Q(b) \left[ \int_0^b \beta I(t-u) e^{-\int_{t-b}^{t-u} \beta I(\alpha) d\alpha} du \right] db \end{aligned}$$

$$= \int_0^\infty \Lambda Q(b) \left[ 1 - e^{-\int_{t-b}^t \beta I(\alpha) d\alpha} \right] db.$$

If we add the equations (5.1) and (2.3) and write  $N(t) = S(t) + I(t)$ , the total population size, we have

$$N(t) = \int_0^\infty \Lambda Q(a) da = K.$$

Thus total population size is constant, and this indicates that our model is actually the limit equation of a correct model. We have started our models at  $t = -\infty$  rather than at time  $t = 0$  with prescribed initial data, and this leads us to the limit equation. The theory of [10] shows that the asymptotic behaviour of a model with assigned initial data is the same as that of the limit equation and thus we have not lost any generality.

Equilibria of (5.1) are given by  $I = 0$  or

$$(5.2) \quad \beta \Lambda \int_0^\infty Q(a) \left( \frac{1 - e^{-\beta I a}}{\beta I} \right) da = 1.$$

As a function of  $I$ , the left side of (5.2) has limit  $\beta \Lambda \int_0^\infty a Q(a) da$  as  $I \rightarrow 0+$ , is monotone decreasing, and approaches zero as  $I \rightarrow \infty$ . Thus (5.1) has only the disease free equilibrium if

$$(5.3) \quad R_0 = \beta \Lambda \int_0^\infty a Q(a) da < 1.$$

If  $R_0 > 1$  there is a unique endemic equilibrium determined by (5.2).

We now show that the disease-free equilibrium is asymptotically stable if and only if  $R_0 < 1$ . The characteristic equation at the disease-free equilibrium is

$$(5.4) \quad \beta \Lambda \int_0^\infty Q(a) \frac{1 - e^{-\lambda a}}{\lambda} da = 1.$$

Since

$$\left| \frac{1 - e^{-\lambda a}}{\lambda} \right| < a$$

for  $R\lambda \geq 0$ , the roots of this characteristic equation are in the left half plane if  $R_0 > 1$ . Thus the disease-free equilibrium is asymptotically stable if  $R_0 < 1$ . If  $R_0 > 1$ , there is a positive real root of (5.4) and thus the disease-free equilibrium is unstable.

If  $R_0 > 1$ , there is an endemic equilibrium  $I > 0$  given by (5.2) and the linearization of (5.1) at this equilibrium is

$$x(t) = \int_0^\infty \Lambda Q(a) e^{-\beta I a} \left[ \int_{t-a}^t \beta x(\alpha) d\alpha \right] da.$$

The characteristic equation is

$$(5.5) \quad \beta\Lambda \int_0^\infty Q(a)e^{-\beta Ia} \left( \frac{1 - e^{-\lambda a}}{\lambda} \right) da = 1.$$

For  $R\lambda \geq 0$

$$e^{-\beta Ia} \left( \frac{1 - e^{-\lambda a}}{\lambda} \right) \leq ae^{-\beta Ia}$$

and

$$\frac{1 - e^{-\beta Ia}}{\beta I} = e^{-\beta Ia} \frac{e^{\beta Ia} - 1}{\beta I} \geq ae^{-\beta Ia}.$$

Thus

$$e^{-\beta Ia} \left( \frac{1 - e^{-\lambda a}}{\lambda} \right) \leq \frac{1 - e^{-\beta Ia}}{\beta I}$$

and it follows from this estimate and (5.2) that (5.5) has no roots in  $R\lambda \geq 0$ . Thus the endemic equilibrium is asymptotically stable, and we have established the following result.

**THEOREM 5.1.** *The S-I model (5.1) has a unique asymptotically stable equilibrium, the disease-free equilibrium if  $R_0$ , given by (5.3), is less than 1 and the endemic equilibrium given by (5.2) if  $R_0 > 1$ .*

For the S-I model, the relation (2.6) takes the form

$$R_0 = \frac{\beta I \int_0^\infty aQ(a)da}{\int_0^\infty (1 - e^{-\beta Ia})Q(a)da} = \frac{1}{A} \frac{\int_0^\infty aQ(a)da}{L - \int_0^\infty e^{-\frac{a}{A}}Q(a)da}$$

with  $L = \int_0^\infty Q(a)da$ ,  $A = \frac{1}{\beta I}$ . This may not be at all close to the simple classical estimate  $1 + \frac{L}{A}$  for  $R_0$ . For example, if  $Q(a) = 1$  ( $0 \leq a \leq L$ ),  $Q(a) = 0$  ( $a > L$ ), we have

$$R_0 = \frac{1}{2} \frac{L/A}{1 - \frac{1 - e^{-L/A}}{L/A}}.$$

As  $L/A$  grows, we see that  $R_0$  is approximated by  $\frac{1}{2} \frac{L}{A}$ .

**6. Extensions.** Natural extensions of our basic model (2.2) would include replacement of the constant birth rate  $\Lambda$  by a birth rate  $g(N(t-a-u))$  depending on total population size and replacement of the constant contact rate  $\beta$  by a saturating contact rate  $\hat{C}(N(t-u))$  which is a non-increasing function  $\hat{C}(N)$  of total population size. This would require expansion of the model to at least two variables. If there are no disease deaths, we may use  $N$  as a second variable and use the equation

$$(6.1) \quad N(t) = \int_0^\infty g(N(t-a))Q(a)da$$

together with the analogue of (2.2),

$$(6.2) \quad I(t) = \int_0^\infty \hat{C}(N(t-u)) I(t-u) P(u) \left[ \int_0^\infty g(N(t-a-u)) Q(a+u) e^{-\int_{t-a-u}^{t-u} \beta I(\alpha) d\alpha} da \right] du$$

to form an S-I-S model. For an S-I-R model, we could add the analogue of (2.3),

$$(6.3) \quad S(t) = \int_0^\infty g(N(t-a)) e^{-\int_{t-a}^t \beta I(\alpha) d\alpha} da$$

to these two equations. Because (6.1) involves only the variable  $N$ , this equation uncouples in the stability analysis and reduces the stability analysis to the single equation (6.2) for the S-I-S model and the pair of equations (6.2) and (6.3) for the S-I-R model. The theory of asymptotically autonomous systems [7] allows us to drop the equation (6.1) from the model and set the total population size equal to the carrying capacity.

For a universally fatal disease, we could use the equation (6.2) together with the sum of (6.2) and (6.3) as an equation for  $N = S + I$ . If some but not all infectives die of disease, the situation is more complicated. We would have to differentiate the equation (6.2), identify the term containing  $P'(u)$  which describes departure from the infective class, and then make an assumption to separate this term into recoveries and disease deaths to form an equation for either recovered members  $R$  which tracks their ages or for total population size  $N$  which tracks disease mortality. The analysis of this model would be a formidable task.

Other extensions which have been undertaken include models with several exponentially distributed stages in the infective period [14] and models with age-dependent contact rates [17]. In addition, we have left several questions about the stability of the endemic equilibrium unanswered. The only case we have analyzed fully, with an arbitrary life span distribution, is the S-I model. In addition, we have seen that when the life span has an exponential distribution the incorporation of infection age dependence in a model, does not affect the conclusions.

## REFERENCES

- [1] V. ANDREASEN, *Multiple time scales in the dynamics of infectious diseases*, Mathematical Approaches to Problems in Resource Management and Epidemiology (C. Castillo-Chavez, S.A. Levin, C.A. Shoemaker, eds.), Lecture Notes in Biomathematics 81, Springer-Verlag, Berlin (1989), 142–151.
- [2] ———, *The effect of age-dependent host mortality on the dynamics of an endemic disease*, Math. Biosc. 114 (1993), 29–58.
- [3] S.P. BLYTHE AND R.M. ANDERSON, *Variable infectiousness in HIV transmission models*, IMA J. Math. Appl. Med. Biol. 5 (1988), 181–200.

- [4] F. BRAUER, *A class of Volterra integral equations arising in delayed-recruitment population models*, Nat. Resource Modeling 2 (1987), 259–278.
- [5] F. BRAUER, *Models for the spread of universally fatal diseases*, J. Math. Biol. 28 (1990), 451–462.
- [6] C. CASTILLO-CHAVEZ, K.L. COOKE, W. HUANG, AND S.A. LEVIN, *On the role of long periods of infectiousness in the dynamics of acquired immunodeficiency syndrome (AIDS)*, Mathematical Approaches to Problems in Resource Management and Epidemiology (C. Castillo-Chavez, S.A. Levin, C.A. Shoemaker, eds.), Lecture Notes in Biomathematics 81, Springer-Verlag, Berlin (1989), 177–189.
- [7] C. CASTILLO-CHAVEZ AND H.R. THIEME, *Asymptotically autonomous epidemic models*, Mathematical Population Dynamics: Analysis of Heterogeneity, I. Theory of Epidemics (O. Arino, D. Axelrod, M. Kimmel, M. Langlais, eds.) Wuerz, Winnipeg (1995), 33–50.
- [8] K. DIETZ, *Transmission and control of arbovirus diseases*, Epidemiology (D. Ludwig and K.L. Cooke, eds.) Society for Industrial & Applied Mathematics, Philadelphia (1975), 104–121.
- [9] ———, *The incidence of infectious diseases under the influence of seasonal fluctuations*, Mathematical Models in Medicine (J. Berger, W. Buhlen, R. Rogges, P. Tautu, eds.) Lecture Notes in Biomathematics 11, Springer-Verlag, Berlin (1976), 1–15.
- [10] J.J. LEVIN AND D.F. SHEA, *On the asymptotic behaviour of the bounded solutions of some integral equations I, II, III*, J. Math. Anal. & Appl. 37 (1972), 42–82, 288–326, 537–575.
- [11] R.M. MAY, *Population biology of microparasitic infections*, Mathematical Ecology (T.G. Hallam and S.A. Levin, eds.), Biomathematics, Vol. 17, Springer-Verlag, Berlin (1986), 405–442.
- [12] R.M. MAY AND R.M. ANDERSON, *The transmission dynamics of human immunodeficiency virus*, Phil. Trans. Roy. Soc. London, Series B, 321 (1989), 565–607.
- [13] H.R. THIEME, *Stability change of the endemic equilibrium in age-structured models for the spread of S-I-R type infectious diseases*, Differential Equations Models in Biology, Epidemiology, and Ecology (S. Busenberg and M. Martelli, eds.), Lecture Notes in Biomathematics 92, Springer-Verlag, Berlin (1991), 139–158.
- [14] H.R. THIEME, *Endemic models with arbitrarily distributed periods of infection*, Mathematical Approach for Emerging and Reemerging Infectious Diseases: Models, Methods and Theory (Carlos Castillo-Chavez with Sally Blower, Pauline van den Driessche, Denise Kirschner, and Abdul-Aziz Yakubu, eds.), Springer-Verlag, Berlin (2001), to appear.
- [15] H.R. THIEME AND C. CASTILLO-CHAVEZ, *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.), Lecture Notes in Biomathematics 83, Springer-Verlag, Berlin (1989), 157–176.
- [16] ———, *How may infection-age-dependent infectivity affect the dynamics of HIV/AIDS?*, SIAM. J. App. Math. 53 (1993), 1447–1479.
- [17] Y. ZHOU AND Z. MA, *Analysis of an epidemic model with age and infection age structure*, Mathematical Approach for Emerging and Reemerging Infectious Diseases: Models, Methods and Theory (Carlos Castillo-Chavez with Sally Blower, Pauline van den Driessche, Denise Kirschner, and Abdul-Aziz Yakubu, eds.), Springer-Verlag, Berlin (2001), to appear.

# EFFECTS OF GENETIC HETEROGENEITY ON HIV TRANSMISSION IN HOMOSEXUAL POPULATIONS

SHU-FANG HSU SCHMITZ\*

**Abstract.** Several AIDS cohort studies observe that the incubation period between HIV infection and AIDS onset can be shorter than 3 years in about 10% seropositive individuals, or longer than 10 years in about 10–15% individuals. On the other hand, many individuals remain seronegative even after multiple exposures to HIV. These distinct outcomes have recently been correlated with some mutant genes in HIV co-receptors (e.g., CCR5, CCR2 and CXCR4). For instance, the mutant alleles  $\Delta 32$  and  $m303$  of CCR5 may provide full protection against HIV infection in homozygotes and partial protection in heterozygotes; moreover, infected heterozygotes may progress more slowly than individuals who have no mutant alleles. Frequencies of these mutant alleles are not very low in Caucasian populations, therefore their effects may not be insignificant. To investigate the impact of such heterogeneity on the spread of HIV, Hsu Schmitz (2000a,b), based on available data, proposes a specific type of models where susceptibles are classified as having no, partial or full natural resistance to HIV infection and infecteds as rapid, normal or slow progressors. She also applies the models to CCR5- $\Delta 32$  mutation in San Francisco gay men. This manuscript sketches her models with focus on the basic model without treatment and an extended model with treatment in certain proportion of newly infected individuals. The same example of CCR5- $\Delta 32$  in San Francisco gay men is used, but some parameters are estimated in different ways. The results are very similar to those in Hsu Schmitz (2000a,b) with the following two main conclusions: 1) without any intervention, HIV infection will continue to spread in this population and the epidemic is mainly driven by the normal progressors; 2) treating only a certain proportion of newly infected individuals with currently available therapies is unlikely to eradicate the disease. Additional interventions are thus necessary for disease control.

**Key words.** HIV, AIDS, mathematical model,  $R_0$ , homosexual, mutation, CCR5, treatment.

**1. Introduction.** Several recent genetic studies have demonstrated the protective effects of certain mutant genes on HIV infection or/and AIDS pathogenesis. For instance, Samson et al. (1996) and Dean et al. (1996) report a mutant allele,  $\Delta 32$ , of CCR5 chemokine receptor gene at a high allele frequency ( $\sim 10\%$ ) in Caucasian populations. Homozygotes with two  $\Delta 32$  alleles may escape from HIV-1 infection and heterozygous infecteds may progress more slowly than other infecteds who have no mutant alleles. Quillent et al. (1998) characterize another CCR5 gene mutation,  $m303$ , among Europeans at an allele frequency of under 1%. Individuals with genotype  $m303/m303$  or  $\Delta 32/m303$  acquire resistance to HIV-1 infection. Similarly, the  $m303$  heterozygosity may give partial protection against infection and slow down the progression once infected. On the other hand, Martin et al. (1998) identify a promotor allele, CCR5P1, which accelerates AIDS progression. The frequency of recessive homozygote CCR5P1/P1 is

---

\*Institut für Mathematische Statistik und Versicherungslehre (IMSV), Universität Bern, 3012 Bern, Switzerland; E-mail: shu-fang.hsu@stat.unibe.ch.

about 7–13% in general population. Infected individuals with this genotype progress to AIDS more rapidly than those with other genotypes.

In another chemokine, CCR2, Smith et al. (1997) describe a mutation, 64I, at an allele frequency of 10–15% among Caucasians and African Americans. This mutant gene indicates a 2–4 years delay of progression among infecteds and this effect is genetically independent of that of CCR5-Δ32. In stromal-derived factor (SDF-1, the principal ligand for CXCR4), Winkler et al. (1998) identify a gene variant, 3'A, that shows recessive restriction on AIDS pathogenesis. HIV infected individuals with SDF1-3'A/3'A (homozygous recessive) genotype have a significantly lower relative hazard to AIDS onset and the protection is approximately twice that seen with CCR2 or CCR5 protection. Moreover, CCR and SDF1 protection seem to be additive. Very recently Faure et al. (2000) discover a variant haplotype, I249M280, in CX<sub>3</sub>CR1 coreceptor among Caucasians with an allele frequency of 13.5%. Like CCR5P1, infected individuals homozygous for this haplotype progress to AIDS more rapidly than those with other haplotypes.

All the mutant genes mentioned above confirm the existence of genetic heterogeneity with respect to susceptibility to HIV infection and to rate of AIDS progression in general populations. Such heterogeneity gives a partial explanation of why some HIV infected individuals progress very fast while some very slowly (Sheppard et al. 1993 and Phair 1994), and why some individuals remain seronegative even after multiple exposures to HIV from infected partners (Detels et al. 1994, Paxton et al. 1996, Fowke et al. 1996).

Dushoff and Levin (1995) and Coutinho et al. (1999) have investigated the effects of heterogeneity on disease invasion in more general ways. To investigate the impact of above mentioned genetic heterogeneity on the spread of HIV, Hsu Schmitz (2000a,b) proposes, as driven by available data, a specific type of one-sex models with susceptibles classified by genotype as having no, partial or full resistance to HIV infection and infecteds classified as rapid, normal or slow progressors. Compared with the general model in Dushoff and Levin (1995) where transmission rate does not depend on the infectious individual, our models allow such dependence. In contrast to the model in Coutinho et al. (1999) where heterogeneity was described by frailty, i.e., a random variable with certain probability distribution for each individual, our models treat heterogeneity as a fixed characteristic within each group as in Dushoff and Levin (1995), i.e., within-group variation is ignored.

This manuscript provides a more concise representation of earlier work of Hsu Schmitz (2000a,b), with focus on selected models without and with treatment. The same application to homosexual population in San Francisco with CCR5-Δ32 mutation is presented, but for some parameters different values are used. The manuscript is organized as follows. Section 2 describes the basic model without treatment. The extended model incorporating treatment is presented in Section 3. Application of our models is illustrated in Section 4. Estimates of some parameters used in this

manuscript are different from those in Hsu Schmitz (2000a,b). Comparisons with previous results are presented. Finally some concluding remarks are provided in Section 5.

**2. The basic model.** Based on the classical one-sex *S-I* model, Hsu Schmitz (2000a) derives the basic model by stratifying susceptibles into three groups according to their genotype related to natural resistance to HIV: no resistance ( $S_1$ ), partial resistance ( $S_2$ ) and complete resistance ( $S_3$ ). The  $S_3$ -individuals never become infected. Similarly, infecteds are classified into three groups according to progression rate: rapid progressor ( $I_1$ ), normal progressor ( $I_2$ ) and slow progressor ( $I_3$ ). AIDS patients are assumed sexually inactive, thus they do not play a role in HIV transmission and are not included in the model.

A constant recruitment rate,  $\pi$ , is assumed to replenish the three susceptible groups with respective fractions,  $g_i$  ( $i = 1, 2, 3$  and  $\sum_i g_i = 1$ ), which are related to frequencies of relevant genotypes. Genotype frequencies might fluctuate with time. However, the actual dynamic of  $g_i$  is usually unknown. Here  $g_i$  are taken as constant for convenience. All individuals are subject to the common per-capita natural removal rate,  $\mu$ . The per-capita progression rates for  $I_j$  individuals are denoted by  $\gamma_j$  ( $j = 1, 2, 3$ ).

The infectiousness of  $I_j$  individuals is reflected by the per-partnership transmission rates,  $\beta_j$  ( $j = 1, 2, 3$ ). It has been shown that viral load and infectiousness may change dramatically during the incubation period. However, for the sake of simplicity,  $\beta_j$  are assumed constant here as in Anderson, Gupta and May (1991) and McLean and Blower (1993). Among the three infected groups, rapid progressors ( $I_1$ ) are assumed to have the highest viral load, thus are most infectious; and slow progressors ( $I_3$ ) are assumed to have the lowest viral load, thus are least infectious. More specifically, the following relation is hypothesized:

$$(1) \quad \beta_1 \geq \beta_2 \geq \beta_3.$$

During the partnership between an  $S_2$ -individual and an  $I_j$ -individual, the transmission rate  $\beta_j$  of the infected partner is reduced to  $x\beta_j$ , with  $0 < x < 1$  to account for partial resistance to HIV in  $S_2$ -individuals. Estimate of transmission rate obtained from all infected individuals is readily available, but the corresponding estimates stratified by progression rate ( $\beta_j$ ) are not. It is assumed that the transmission rate of normal progressors ( $I_2$ , who are the majority among infecteds) can be approximated by the estimate obtained from the pooled infected population. The transmission rates are then reparameterized as follows:

$$(2) \quad \beta_1 = b_1\beta, \quad \beta_2 = b_2\beta, \quad \beta_3 = b_3\beta,$$

with  $b_2 \equiv 1$ . Relation (1) implies

$$(3) \quad b_1 \geq 1 \quad \text{and} \quad 0 \leq b_3 \leq 1.$$

Newly infected  $S_i$ -individuals ( $i = 1, 2$ ) join the three infected groups with respective proportions  $f_{ij}$ , which satisfy

$$(4) \quad 0 \leq f_{ij} \leq 1 \quad \text{and} \quad \sum_{j=1}^3 f_{ij} = 1 \quad \text{for} \quad i = 1, 2.$$

The new infecteds who come from  $S_1$  are expected to generate a larger fraction of rapid progressors ( $I_1$ ) and a smaller fraction of slow progressors ( $I_3$ ) than those coming from  $S_2$ ; that is,

$$(5) \quad f_{11} > f_{21} \quad \text{and} \quad f_{13} < f_{23}.$$

Currently most individuals do not know their genotypes at loci related to HIV susceptibility or/and AIDS pathogenesis, hence, it is reasonable to assume that genetic heterogeneity has no influence on pairing behavior. However, disease status might have some influence on pairing behavior. To keep our model simple, such influence is neglected as in Anderson, Gupta and May (1991) and McLean and Blower (1993). Thus, a common average number of partners per unit time, denoted by  $c$ , is applicable to all individuals. Because we are looking at a homosexually active population, processes of pair formation and dissolution are not followed explicitly; instead, a proportional mixing pattern is assumed. The total number of sexual partnerships is defined as

$$(6) \quad \Lambda = c \Phi_n = c \left( \sum_{i=1}^3 S_i + \sum_{j=1}^3 I_j \right),$$

with subscript  $n$  indicating the null, or the basic model, which also applies to the notation that follows. For a susceptible, given he pairs (i.e., forms a pair with an individual), the chance of pairing with an  $I_j$ -individual is  $c I_j / \Lambda = I_j / \Phi_n$  (Busenberg and Castillo-Chavez 1991). Thus, the force of infection for  $S_1$ -individuals is

$$(7) \quad \sigma_{1n} = c \sum_{j=1}^3 \beta_j I_j / \Lambda = \beta \sum_{j=1}^3 b_j I_j / \Phi_n,$$

and for  $S_2$ -individuals is

$$(8) \quad \sigma_{2n} = c x \sum_{j=1}^3 \beta_j I_j / \Lambda = x \beta \sum_{j=1}^3 b_j I_j / \Phi_n = x \sigma_{1n}.$$

The incidences from  $S_i$  ( $i = 1, 2$ ) are then

$$(9) \quad \delta_{1n} = c S_1 \sigma_{1n},$$

$$(10) \quad \delta_{2n} = c S_2 \sigma_{2n} = x c S_2 \sigma_{1n}.$$

These newly infected individuals enter the class  $I_j$  at the rates

$$(11) \quad \rho_{jn} = f_{1j} \delta_{1n} + f_{2j} \delta_{2n} = c \sigma_{1n} (f_{1j} S_1 + x f_{2j} S_2), \quad j = 1, 2, 3,$$

called the incidence of  $I_j$ . The basic model can now be summarized by the following system of equations:

$$(12) \quad \begin{aligned} \dot{S}_1 &= g_1 \pi - \mu S_1 - \delta_{1n} \\ \dot{S}_2 &= g_2 \pi - \mu S_2 - \delta_{2n} \\ \dot{S}_3 &= g_3 \pi - \mu S_3 \\ \dot{I}_1 &= \rho_{1n} - (\mu + \gamma_1) I_1 \\ \dot{I}_2 &= \rho_{2n} - (\mu + \gamma_2) I_2 \\ \dot{I}_3 &= \rho_{3n} - (\mu + \gamma_3) I_3 \end{aligned}$$

To study the potential of disease spreading, we shall compute the basic reproductive number (Diekmann et al. 1990), which indicates whether a disease may invade a population in demographic steady state when there is no disease present. The basic reproductive number for the basic model (12), denoted by  $R_{0n}$ , is

$$(13) \quad R_{0n} = c \beta E_n$$

$$(14) \quad = K \{Q_1 + 1 + Q_3\}$$

with

$$(15) \quad E_n = \frac{b_1 \tau_1}{\mu + \gamma_1} + \frac{\tau_2}{\mu + \gamma_2} + \frac{b_3 \tau_3}{\mu + \gamma_3},$$

$$(16) \quad K = c \beta \frac{\tau_2}{\mu + \gamma_2},$$

$$(17) \quad Q_1 = b_1 \left( \frac{\mu + \gamma_2}{\mu + \gamma_1} \right) \left( \frac{\tau_1}{\tau_2} \right),$$

$$(18) \quad Q_3 = b_3 \left( \frac{\mu + \gamma_2}{\mu + \gamma_3} \right) \left( \frac{\tau_3}{\tau_2} \right),$$

$$(19) \quad \tau_j = f_{1j} g_1 + x f_{2j} g_2.$$

If  $R_{0n} > 1$ , then the disease will successfully invade. Compared with normal progressors, the relative contributions of rapid and slow progressors to  $R_{0n}$  are reflected by  $Q_1$  and  $Q_3$ , which are determined by  $b_j$ ,  $\tau_j$  and  $\gamma_j$ . If  $Q_1 < 1$  and  $Q_3 < 1$ , then the group of normal progressors,  $I_2$ , contributes the major part,  $K$ , to  $R_{0n}$ . If  $K > 1$ , then certainly  $R_{0n} > 1$  and the disease will spread. If  $K < 1$ , it may still be possible to have  $R_{0n} > 1$  if  $K$ ,  $Q_1$  and  $Q_3$  are not too small.

**3. The extended model.** Treatments for HIV infected individuals are readily available in many developed countries. It has been shown that timing of treatment plays an important role on outcome and the best timing is at the early disease stages (Rosenberg et al. 2000). Following Hsu Schmitz (2000b), we thus consider an ideal situation in which a common proportion ( $m$ ) of newly infected individuals are effectively treated. Theoretically it might be more efficient to treat a certain group with a higher proportion, e.g., a higher treatment proportion for rapid progressors. However, upon infection it is difficult to predict with high accuracy if the individual will be a rapid, normal or slow progressor even if we know his genotype. Therefore, for practical consideration we use a common treatment proportion for all newly infected individuals. We assume that the transmission rate of effectively treated individuals ( $T_j$ ) is reduced from  $\beta_j$  to  $z\beta_j$  with  $0 \leq z < 1$ , and the progression rate is reduced from  $\gamma_j$  to  $y\gamma_j$  with  $0 \leq y < 1$ . The value of 1 is not included in the ranges of  $z$  and  $y$  because it means that treatment does not reduce infectiousness and progression rate at all, which does not seem to be reasonable based on current knowledge. Although treatment might change people's pairing behavior (e.g., treated infecteds have more/less sexual partners per unit of time than those do not receive treatment), for simplicity we do not pursue this issue here, but treat it extensively elsewhere (Hsu Schmitz 2001). Thus, the common pairing activity  $c$  and the proportional mixing pattern in the basic model are still in effect. Let

$$(20) \quad \Phi_t = \sum_{i=1}^3 S_i + \sum_{j=1}^3 (I_j + T_j)$$

with subscript  $t$  indicating the extended model with treatment, which also applies to the notation that follows. The forces of infection for  $S_1$ - and  $S_2$ -individuals are then

$$(21) \quad \sigma_{1t} = \beta \left( \sum_{j=1}^3 b_j I_j + z \sum_{j=1}^3 b_j T_j \right) / \Phi_t,$$

$$(22) \quad \sigma_{2t} = x \sigma_{1t}.$$

The incidences from  $S_i$ -individuals ( $i = 1, 2$ ) are

$$(23) \quad \delta_{1t} = c S_1 \sigma_{1t},$$

$$(24) \quad \delta_{2t} = c S_2 \sigma_{2t} = x c S_2 \sigma_{1t}.$$

These newly infected individuals enter  $j$ th infected group ( $I_j$  and  $T_j$ ) at the incidence rates

$$(25) \quad \rho_{jt} = f_{1j} \delta_{1t} + f_{2j} \delta_{2t} = c \sigma_{1t} (f_{1j} S_1 + x f_{2j} S_2), \quad j = 1, 2, 3.$$

The extended model is described by the following system of equations:

$$(26) \quad \begin{aligned} \dot{S}_1 &= g_1 \pi - \mu S_1 - \delta_{1t} \\ \dot{S}_2 &= g_2 \pi - \mu S_2 - \delta_{2t} \\ \dot{S}_3 &= g_3 \pi - \mu S_3 \\ \dot{I}_1 &= (1-m)\rho_{1t} - (\mu + \gamma_1)I_1 \\ \dot{I}_2 &= (1-m)\rho_{2t} - (\mu + \gamma_2)I_2 \\ \dot{I}_3 &= (1-m)\rho_{3t} - (\mu + \gamma_3)I_3 \\ \dot{T}_1 &= m \rho_{1t} - (\mu + y \gamma_1)T_1 \\ \dot{T}_2 &= m \rho_{2t} - (\mu + y \gamma_2)T_2 \\ \dot{T}_3 &= m \rho_{3t} - (\mu + y \gamma_3)T_3 \end{aligned}$$

The basic reproductive number for the extended model (26) is

$$(27) \quad R_{0t} = c\beta F_\tau,$$

where

$$(28) \quad \begin{aligned} F_\tau &= \frac{(1-m)b_1\tau_1}{\mu + \gamma_1} + \frac{(1-m)\tau_2}{\mu + \gamma_2} + \frac{(1-m)b_3\tau_3}{\mu + \gamma_3} \\ &\quad + \frac{mz b_1 \tau_1}{\mu + y \gamma_1} + \frac{mz \tau_2}{\mu + y \gamma_2} + \frac{mz b_3 \tau_3}{\mu + y \gamma_3}. \end{aligned}$$

The difference between  $R_{0n}$  and  $R_{0t}$  is denoted by

$$(29) \quad \begin{aligned} \Delta_t &= R_{0n} - R_{0t} \\ &= c\beta m(b_1\tau_1 H_1 + \tau_2 H_2 + b_3\tau_3 H_3), \end{aligned}$$

where

$$(30) \quad H_j = \frac{1}{\mu + \gamma_j} - \frac{z}{\mu + y \gamma_j}.$$

If we require that the treatment is effective not only in the individual level, but also in the population level to slow down the spread of the disease, then we expect  $\Delta_t > 0$ , which implies not all  $H_j$  can be negative. For given  $\mu$  and  $\gamma_j$ ,  $H_j$  can become negative when  $y \rightarrow 0$  and  $z \rightarrow 1$ , i.e., treatment significantly prolongs the incubation period, but does not reduce the infectiousness. This is in agreement with Anderson et al. (1991). Fortunately, recent clinical trials with highly active antiretroviral therapy (HAART) have shown that such treatments reduce the plasma HIV-RNA to undetectable levels (e.g., see Lillo et al. 1999 and Ibanez et al. 1999). Due to the reduction of viral load, the infectiousness may as well be reduced.

Based on (27) or (29), the critical value of each treatment related parameter (i.e.,  $m$ ,  $y$ ,  $z$ ) to have  $R_{0t} < 1$  can be calculated. Because there

is only one equation, only one unknown parameter can be solved at a time. For example, the common treatment proportion  $m$  with given  $R_{0n}$  has to be

$$(31) \quad m > m^* = \frac{R_{0n} - 1}{c\beta U},$$

where

$$(32) \quad \begin{aligned} U &= b_1 \tau_1 H_1 + \tau_2 H_2 + b_3 \tau_3 H_3 \\ &= E_n - z E_y \end{aligned}$$

with

$$(33) \quad E_y = \frac{b_1 \tau_1}{\mu + y \gamma_1} + \frac{\tau_2}{\mu + y \gamma_2} + \frac{b_3 \tau_3}{\mu + y \gamma_3}.$$

Because  $0 \leq y < 1$ , all three terms in  $E_y$  are larger than the corresponding terms in  $E_n$ ; thus,  $E_y > E_n$ . If  $(100 \times m)\%$  newly infected individuals are effectively treated and  $m > m^*$ , then the epidemic will eventually die out. As a proportion,  $m^*$  is constrained in the range of  $(0,1]$ . The value 0 is not included here because it means no treatment. The condition  $m^* > 0$  requires  $(R_{0n} - 1)U > 0$ . If  $R_{0n} > 1$ , which is the usual case for the current HIV/AIDS epidemic, it requires

$$(34) \quad U > 0.$$

If  $U < 0$ , which is equivalent to  $H_j < 0$  for all  $j$ , then this treatment is not effective in the population level and should not be applied, otherwise we get negative  $m^*$  and  $\Delta_t$ . On the other hand, the condition  $m^* \leq 1$  requires

$$(35) \quad U \geq U_m = \frac{R_{0n} - 1}{c\beta},$$

or equivalently

$$(36) \quad z \leq z_m = \frac{1}{c\beta E_y}.$$

If  $R_{0n} > 1$ , the  $U$ 's satisfying (35) also satisfy (34). If condition (35) or (36) is not satisfied, we obtain  $m^* > 1$ , which is not realistic. Thus either of these two conditions can be used to check the possibility of disease eradication with the given treatment. If it proves possible, then the required common treatment proportion can be calculated by (31).

Alternatively for given  $m$  and  $y$ , the reduction factor for infectiousness has to be

$$(37) \quad z < z^* = \frac{1 + R_{0n}(m - 1)}{c\beta m E_y}$$

to eliminate the disease by the treatment. Note that when  $m = 1$ ,  $z^*$  is equivalent to  $z_m$  in (36). Again, as a proportion,  $z^*$  is constrained in the range of  $[0,1]$ . The value 1 is not included here because it means that the treatment has no effect in reducing infectiousness. The condition  $z^* \geq 0$  requires

$$(38) \quad m \geq 1 - \frac{1}{R_{0n}}.$$

If condition (38) is not satisfied, we get negative  $z^*$ . The other condition  $z^* < 1$  requires

$$(39) \quad m > \frac{R_{0n} - 1}{R_{0n} - c\beta E_y} = \frac{R_{0n} - 1}{c\beta(E_n - E_y)}.$$

If  $R_{0n} > 1$ , then the right hand side is negative and any  $m \in (0, 1]$  will satisfy (39). If  $R_{0n} < 1$ , then the disease will die out by itself and treatment is not necessary for disease control in the population level. In brief, only condition (38) has to be satisfied for having a realistic  $z^*$  when  $R_{0n} > 1$ .

To obtain the critical value of reduction factor for progression rate,  $y^*$ , with given  $m$  and  $z$ , one needs to solve the following polynomial for  $y$ :

$$(40) \quad \begin{aligned} 0 = & \gamma_1 \gamma_2 \gamma_3 W y^3 \\ & + \{\gamma_1 \gamma_2 (\mu W - \tau_1 b_1) + \gamma_1 \gamma_3 (\mu W - \tau_2) + \gamma_2 \gamma_3 (\mu W - \tau_3 b_3)\} y^2 \\ & + \{\gamma_1 (\mu W - \tau_2 - \tau_3 b_3) + \gamma_2 (\mu W - \tau_1 b_1 - \tau_3 b_3) + \gamma_3 (\mu W - \tau_1 b_1 - \tau_2)\} \mu y \\ & + (\mu W - \tau_1 b_1 - \tau_2 - \tau_3 b_3) \mu^2, \end{aligned}$$

where

$$(41) \quad W = \frac{1 - (1 - m) R_{0n}}{c \beta m z}.$$

There are at most three solutions to this polynomial, but only the solution satisfying  $0 \leq y < 1$  should be chosen. The value 1 is not included here because it means that the treatment has no effect in reducing progression rate.

#### 4. Example.

**4.1. Estimates of parameters.** We use the CCR5- $\Delta 32$  mutation in the population of gay men in San Francisco, as in Hsu Schmitz (2000a,b), to illustrate the application of formulas and conditions derived above. Whenever possible, we take parameter values specific to this population; otherwise we take estimates from other Caucasian AIDS cohorts.

As in McLean and Blower (1993), we take natural removal rate  $\mu = 1/32$  year $^{-1}$ , overall recruitment rate  $\pi = 2000/\text{year}$ , and product of overall transmission rate and average number of partners per unit time  $\beta c =$

0.62/year. Because normal progressors are the majority among infected individuals, we assume that  $\beta_2 c$  can be approximated by the overall estimate obtained from the pooled infected population, i.e.,  $\beta_2 c = \beta c = 0.62/\text{year}$ . No information about  $\beta_1$  and  $\beta_3$ , or equivalently about  $b_1$  and  $b_3$ , is available, thus they are considered as free parameters within respective ranges:  $1 \leq b_1 \leq 4$  (this upper bound is chosen for illustrative purpose) and  $0 \leq b_3 \leq 1$ .

Based on pooled data of Caucasians in five AIDS cohorts including homosexually active men in the San Francisco City Clinic Cohort (SFCC), Smith et al. (1997, Figure 3) present number of individuals and genotype frequencies stratified by incubation time into 6 categories: < 3.5 years, 3.5 to < 7 years, 7 to < 10 years, 10 to < 13 years, 13 to < 16 years, and  $\geq 16$  years. As defined in Hsu Schmitz (2000a,b), rapid progressors have an incubation time of less than 3.5 years (the first category), slow progressors of more than 13 years (the last two categories), and normal progressors of in between 3.5 and 13 years. This implies  $1/\gamma_1 < 3.5$  years,  $3.5 \leq 1/\gamma_2 \leq 13$  years and  $1/\gamma_3 > 13$  years. Assuming that the first category has a lower bound of 1 year and the last category has an upper bound of 19 years, we obtain the average progression rates  $\gamma_1 = 1/2.25$ ,  $\gamma_2 = 1/8.93$  and  $\gamma_3 = 1/15$  by using the midpoint as the average within each category. These  $\gamma$  values are similar to those guessed values (1/2, 1/8 and 1/16) used in Hsu Schmitz (2000a,b), with a difference of -0.056, -0.013 and 0.004, respectively.

Using data of SFCC presented in Dean et al. (1996, Table 2), the genotype frequencies are estimated as  $g_1 = 0.75$ ,  $g_2 = 0.23$  and  $g_3 = 0.02$ . The fractions  $f_{ij}$  may be approximated from survival curves on incubation time stratified by genotype. Although Dean et al. (1996) also present such graph specifically for SFCC individuals whose seroconversion was well documented (their Figure 3C), the sample size is relatively small and thus the results might not be representative. We decide to take the survival curves from a larger sample including SFCC (their Figure 3B) to estimate  $f_{ij}$ . This figure shows that among individuals without  $\Delta 32$  alleles, about 10% progress at 3.5 years and about 26% remain AIDS-free at 13 years; thus  $f_{11} = 0.1$ ,  $f_{12} = 0.64$  and  $f_{13} = 0.26$ . Among individuals with one  $\Delta 32$  allele, about 6% progress at 3.5 years and about 46% remain AIDS-free at 13 years; thus  $f_{21} = 0.06$ ,  $f_{22} = 0.48$  and  $f_{23} = 0.46$ . These  $f_{ij}$  also satisfy the hypothesized condition (5). The  $f_{11}$ ,  $f_{12}$ ,  $f_{13}$  and  $f_{21}$  are very similar to those estimated and used in Hsu Schmitz (2000a,b), with absolute differences  $\leq 0.043$ ; while the  $f_{22}$  is smaller and the  $f_{23}$  is larger than the counterpart by a difference of -0.125 and 0.128, respectively. This proves that the rough estimating procedure for  $f_{ij}$  proposed by Hsu Schmitz (2000a) does provide reasonable estimates.

As indicated by Samson et al. (1996), the frequency of heterozygotes in cohorts of seropositive Caucasians is 35% lower than in the general populations. Hence, we take  $\Delta 32$ -induced reduction factor for transmission rate

$x = 1 - 0.35 = 0.65$ . Although the actual relation between the observed difference in genotype frequency and the reduction factor  $x$  might not be so simple, it has been shown that value of  $x$  has only minor effects on the epidemic (Hsu Schmitz 2000a).

**4.2. Without treatment.** Without treatment the magnitude of  $R_{0n}$  for different values of  $b_1$  and  $b_3$  can be calculated using (14) as follows:

$$(42) \quad R_{0n} = 2.388 (0.109 b_1 + 1 + 1.670 b_3),$$

which is also plotted in Figure 1. Because the coefficient of  $b_3$  is about 15.3 times of that of  $b_1$ ,  $R_{0n}$  is much more sensitive to value of  $b_3$  than to  $b_1$ . With  $b_1 \in [1, 4]$  and  $b_3 \in (0, 1]$ , we obtain  $0.046 \leq Q_1 \leq 0.183$  and  $0 < Q_3 \leq 0.699$ , which are clearly less than unity. Hence, the normal progressors give the major contribution to  $R_{0n}$  with  $K = 2.388 > 1$ . The resulting range of  $R_{0n}$  is between 2.498 and 4.496. Even under the “best” situation with the minimum values of  $b_1 = 1$  and  $b_3 = 0$ ,  $R_{0n}$  is still larger than unity. Hence, the disease will continue spreading out in this population if no intervention is given. If  $b_1 = 1$  and  $b_3 = 1$ , we obtain  $R_{0n} = 4.168$  with  $Q_1 = 0.046 < Q_3 = 0.699$ ; while if  $b_1 = 4$  and  $b_3 = 0.25$ ,  $R_{0n} = 3.244$  with  $Q_1 = 0.183 > Q_3 = 0.175$ . Thus, depending on the values of  $b_1$  and  $b_3$ , the relative contribution of rapid progressors to  $R_{0n}$  may be smaller or larger than that of slow progressors.

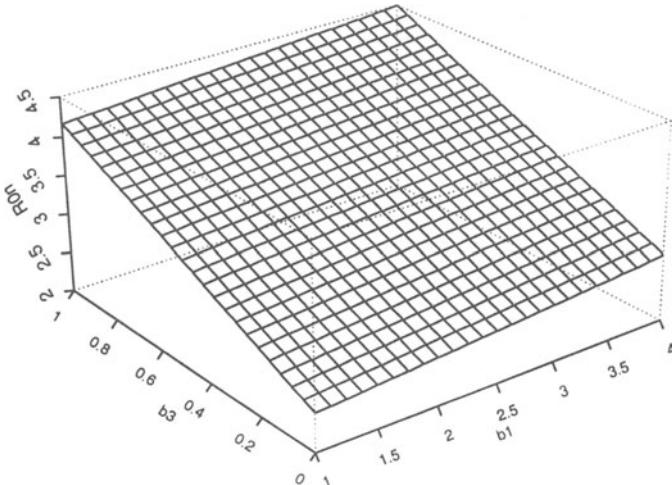


FIG. 1. *Basic model without treatment. Basic reproductive number ( $R_{0n}$ ) as a function of the multipliers of reference transmission rate,  $b_1$  and  $b_3$ , respectively for rapid and slow progressors.*

Compared with the counterparts in Hsu Schmitz (2000a) where different values of  $\gamma_j$  and  $f_{ij}$  are used, the  $K$  and coefficients of  $b_j$  here are larger, thus the  $R_{0n}$  is also larger. This is probably due to larger  $f_{13}$  and  $f_{23}$ , which yield more slow progressors. Anyway, the qualitative conclusions are similar.

**4.3. With treatment.** When certain proportion ( $m$ ) of newly infected individuals are effectively treated, the basic reproductive number will be changed to  $R_{0t}$ , as defined in (27). Based on (36), Figure 2 presents the allowable upper bounds of the treatment-induced reduction factor for transmission rate  $z$ , i.e.,  $z_m$ , for the treatment-induced reduction factor for progression rate  $y \in [0, 1]$  under four different  $(b_1, b_3)$  situations: “best” with  $b_1 = 1$  and  $b_3 = 0$ , “intermediate” with  $b_1 = 2$  and  $b_3 = 0.5$ , “neutral” with  $b_1 = b_3 = 1$ , and “worst” with  $b_1 = 4$  and  $b_3 = 1$ . Under all four situations, the  $z_m$  increases nearly linearly with  $y$ ; that is, the larger the  $y$  is, the larger the upper bound for  $z$  is allowed; or equivalently, the smaller reduction in progression rate is, the smaller reduction in transmission rate is required. Also the  $z$  has to be relatively smaller than  $y$ , which suggests more efforts should be given to reduction of transmission rate than to reduction of progression rate induced by the treatment. Moreover, the difference between situations also increases with  $y$ . For any given value of  $y$ , the order of the  $z_m$  is “best” > “intermediate” > “neutral” > “worst”; that is, the better the situation is, the larger the upper bound for  $z$  is allowed, thus the less strength in reduction of transmission rate induced by the treatment is required. For instance, when  $y = 0.7$ , the  $z_m$ ’s of these four situations are 0.306, 0.224, 0.186 and 0.172, respectively. Compared with the “neutral” situation where the three infected groups have the same transmission rate, the consideration of heterogeneity in transmission rate among infected groups may make the same treatment more effective under the “best” and “intermediate” situations or less effective under the “worst” situation. If treatment does not reduce progression rate at all (i.e.,  $y = 1$ ), which is unlikely to be true based on available clinical data, a  $\geq 60\%$  reduction in transmission rate (i.e.,  $z \leq 0.4$ ) is required even under the “best” situation. This requirement seems to be difficult to fulfill.

If the treatment does yield  $y = 0.7$ , i.e., a 30% reduction in progression rate or a 43% increase in incubation time, then the  $z$  has to be less than or equal to 0.172 in order to eliminate the disease under the “worst” situation. Figure 3 presents the critical values of the common treatment proportion,  $m^*$ , for  $b_1 \in [1, 4]$  and  $b_3 \in [0, 1]$  with  $y = 0.7$  and  $z = 0.17$ . The value of  $m^*$  increases with  $b_1$  as well as with  $b_3$ , but is much more sensitive to  $b_3$ . The required treatment proportion ranges between 77% under the “best” situation and 99.7% under the “worst” situation. With currently available treatments, it might be possible to reduce the progression rate by 30% (i.e.,  $y = 0.7$ ), but there is little information about reduction of transmission rate. It might be difficult to reduce this rate by 83% (i.e.,  $z = 0.17$ ).

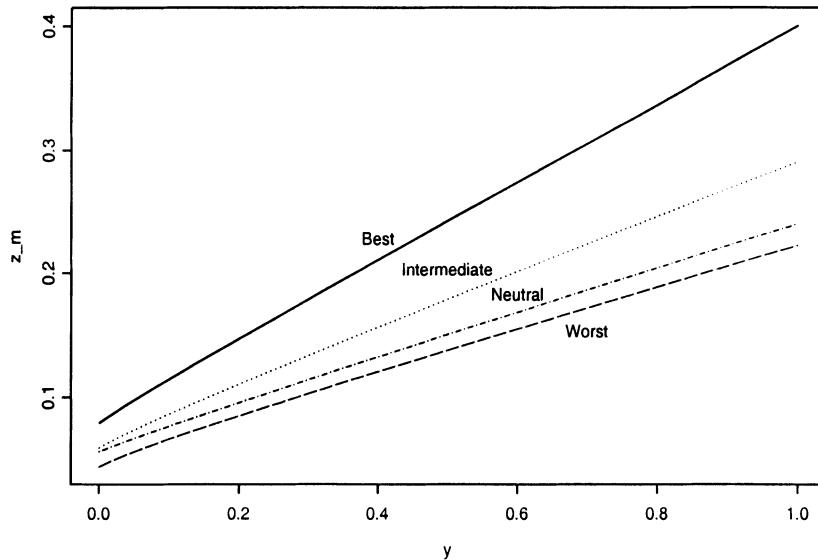


FIG. 2. *Extended model with treatment.* Allowable upper bounds of treatment-induced reduction factor for per-partnership transmission rate ( $z_m$ ) as a function of treatment-induced reduction factor for progression rate ( $y$ ) under four different situations: “best” with  $b_1 = 1$  and  $b_3 = 0$ , “intermediate” with  $b_1 = 2$  and  $b_3 = 0.5$ , “neutral” with  $b_1 = b_3 = 1$ , and “worst” with  $b_1 = 4$  and  $b_3 = 1$  ( $b_1$  and  $b_3$  are as defined in figure 1 legend and text).

as required. Therefore, additional interventions, e.g., using condoms and vaccination, are necessary to help in this context. Overall, it is unlikely to eliminate the disease in this population with treatment alone.

Compared with the counterparts in Hsu Schmitz (2000b) where different values of  $\gamma_j$  and  $f_{ij}$  are used, the qualitative conclusions are similar. However, the specific value of  $y$  here is 0.7, different from 0.5 used in Hsu Schmitz (2000b); thus quantitative results are not comparable.

**5. Concluding remarks.** We have presented a specific type of models to incorporate genetic heterogeneity into HIV/AIDS epidemiology. In our models susceptibles with different genotypes may enter the same infected group. It might be more informative, and probably also more realistic, to have three infected groups specifically for  $S_1$ -individuals and another three infected groups for  $S_2$ -individuals. However, this requires additional parameters to be estimated from more detailed data, which are not readily available. Therefore, such extention was not considered.

For the example of CCR5- $\Delta 32$  mutation in San Francisco gay men, we have used some parameter values a bit different from those in Hsu Schmitz (2000a,b) and obtained similar results:

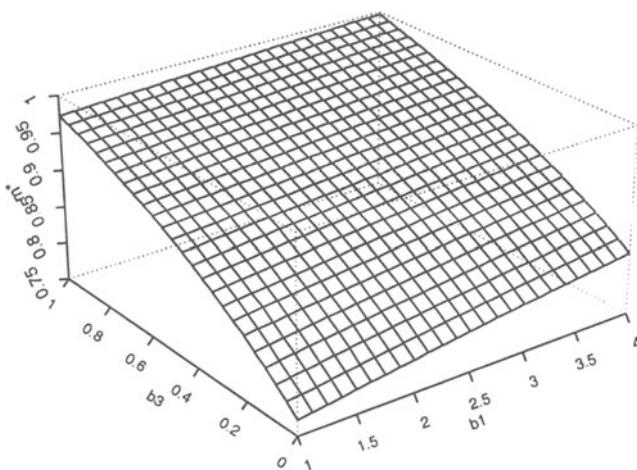


FIG. 3. Extended model with treatment. Minimum common treatment proportion ( $m^*$ ) required to have the basic reproductive number  $R_{0t} \leq 1$  as a function of  $b_1$  and  $b_3$  (as defined in figure 1 legend and text) with treatment-induced reduction factor for progression rate  $y = 0.7$  and treatment-induced reduction factor for per-partnership transmission rate  $z = 0.17$ .

### Without any intervention

1. HIV infection will continue to spread in this population.
2. The major contribution to the epidemic is from the normal progressors, who are the majority among infecteds.
3. Depending on the group transmission rates ( $\beta_j$  or  $b_j$ ), the rapid progressors may contribute more or less than the slow progressors to disease spread.
4. The magnitude of the basic reproductive number  $R_{0n}$  is much more sensitive to transmission rate of slow progressors ( $\beta_3$  or  $b_3$ ) than to that of rapid progressors ( $\beta_1$  or  $b_1$ ). Thus, data should be collected to allow an accurate estimate of  $\beta_3$  or  $b_3$ .

### With treatment in newly infected individuals

1. The limiting factor for the treatment to be effective in the population level is the treatment-induced reduction factor for the transmission rate ( $z$ ). The requirement may be difficult to fulfill with currently available treatments. We therefore suggest medical researchers to focus more on reducing transmission rates than on reducing progression rates when developing new treatments.
2. The required treatment proportion ( $m$ ) is more sensitive to the transmission rate of slow progressors ( $\beta_3$  or  $b_3$ ) than to that of rapid progressors ( $\beta_1$  or  $b_1$ ). This again suggests that data should be collected to allow an accurate estimate of  $\beta_3$  or  $b_3$ .

3. Compared with the “neutral” situation with homogeneous transmission rate among infected groups, the consideration of heterogeneity in transmission rate may make the same treatment more effective under the “best” and “intermediate” situations or less effective under the “worst” situation.
4. Treatment alone is unlikely to eliminate the disease in this population. Additional interventions are necessary to help in this context.

As mentioned in the Introduction, there are several mutant alleles at different loci related to susceptibility to HIV or/and rate of progression to AIDS. Thus, it may be more appropriate to consider a combined locus accommodating several relevant loci, instead of focusing on a single locus. For instance, Smith et al. (1997) combine the CCR5 locus and the CCR2 locus into a compound locus. More data should be collected to cover this aspect. Overall, data for various parameters stratified by relevant genotypes or progression rate are still limited. Clinical researchers are recommended to take genetic heterogeneity into account and record data in an according way.

Effects of vaccination on HIV transmission in homosexual populations with genetic heterogeneity have been investigated in Hsu Schmitz (2000b). In addition, the impact of change in pairing behavior induced by treatment and vaccination has been explored (Hsu Schmitz 2001). Other possible extension to our models may take into account heterogeneous sensitivity to treatments, incomplete treatment due to patient refusal, drug resistance, additional treatments for patients at later disease stages, heterogeneous transmission rates during incubation period, etc.

**Acknowledgement.** This research was supported by the Swiss NSF grant No. 20-49319.96.

## REFERENCES

- [1] ANDERSON, R.M., GUPTA, S., AND MAY, R.M. Potential of community-wide chemotherapy or immunotherapy to control the spread of HIV-1. *Nature* **350**:356-359 (1991).
- [2] BUSENBERG, S. AND CASTILLO-CHAVEZ, C. 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 Journal of Mathematics Applied in Medicine & Biology* **8**:1-29 (1991).
- [3] COUTINHO, F.A.B., MASSAD, E., LOPEZ, L.F., ET AL. Modelling heterogeneities in individual frailties in epidemic models. *Mathematical and Computer Modelling* **30**:97-115 (1999).
- [4] DEAN, M., CARRINGTON, M., WINKLER, C., ET AL. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CCR5 structural gene. *Science* **273**:1856-1862 (1996).
- [5] DETELS, R., LIU Z., HENNESSEY, K., ET AL. Resistance to HIV-1 infection. Multicenter AIDS Cohort Study. *J Acquir Immune Defic Syndr Hum Retrovirol* **7**:1263-1269 (1994).

- [6] DIEKMANN, O., HEESTERBEEK, J.A.P., AND METZ, J.A.J. On the definition and the computation of the basic reproduction ratio  $R_0$  in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology* **28**:365–382 (1990).
- [7] DUSHOFF, J AND LEVIN, S. The effects of population heterogeneity on disease invasion. *Mathematical Biosciences* **128**:25–40 (1995).
- [8] FAURE, S., MEYER, L., COSTAGLIOLA, D., ET AL. Rapid progression to AIDS in HIV<sup>+</sup> individuals with a structural variant of the chemokine receptor CX<sub>3</sub>CR1. *Science* **287**:2274–2277 (2000).
- [9] FOWKE, K.R., NAGELKERKE, N.J.D., KIMANI, J., ET AL. Resistance to HIV-1 infection among persistently seronegative prostitutes in Nairobi, Kenya. *Lancet* **348**:1347–1351 (1996).
- [10] HSU SCHMITZ, S.-F. A mathematical model of HIV transmission in homosexuals with genetic heterogeneity. *Journal of Theoretical Medicine* **2**:285–296 (2000a).
- [11] HSU SCHMITZ, S.-F. Treatment and vaccination against HIV/AIDS in homosexuals with genetic heterogeneity. *Mathematical Biosciences* **167**:1–18 (2000b).
- [12] HSU SCHMITZ, S.-F. Impact of changes in contact rates on HIV transmission under treatment and vaccination in homosexual populations with genetic heterogeneity (submitted) (2001).
- [13] IBANEZ, A., PUIG, T., ELIAS, J., ET AL. Quantification of integrated and total HIV-1 DNA after long-term highly active antiretroviral therapy in HIV-1-infected patients. *AIDS* **13**(9):1045–1049 (1999).
- [14] LILLO, F.B., CIUFFREDA, D., VEGLIA, F., ET AL. Viral load and burden modification following early antiretroviral therapy of primary HIV-1 infection. *AIDS* **13**(7):791–796 (1999).
- [15] MARTIN, M.P., DEAN, M., SMITH, M.W., ET AL. Genetic acceleration of AIDS progression by a promoter variant of CCR5. *Science* **282**:1907–1911 (1998).
- [16] MCLEAN, A.R., BLOWER, S.M. Imperfect vaccines and herd immunity to HIV. *Proc. R. Soc. Lond. B* **253**:9–13 (1993).
- [17] PAXTON, W.A., MARTIN, S.R., TSE, D., ET AL. Relative resistance to HIV-1 infection of CD4 lymphocytes from persons who remain uninfected despite multiple high-risk sexual exposure. *Nature Medicine* **2**:412–417 (1996).
- [18] PHAIR, J.P. Keynote address: variations in the natural history of HIV infection. *AIDS REs Hum RETrovir* **10**:883–885 (1994).
- [19] QUILLENT, C., OBERLIN, E., BRAUN, J., ET AL. HIV-1-resistance phenotype conferred by combination of two separate inherited mutations of CCR5 gene. *The Lancet* **351**:14–18 (1998).
- [20] ROSENBERG, E.S., ALTFELD, M., POON, S.H., ET AL. Immune control of HIV-1 after early treatment of acute infection. *Nature* **407**:523–526 (2000).
- [21] SAMSON, M., LIBERT, F., DORANZ, B.J., ET AL. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* **382**:722–725 (1996).
- [22] SHEPPARD, H.W., LANG, W., ASCHER, M.S. The characterization of non-progressors: long-term HIV-1 infection with stable CD4+ T-cell levels. *AIDS* **7**:1159–1166 (1993).
- [23] SMITH, M.W., DEAN, M., CARRINGTON, M., ET AL. Contrasting genetic influence of CCR2 and CCR5 variants on HIV-1 infection and disease progression. *Science* **277**:959–965 (1997).
- [24] WINKLER, C., MODI, W., SMITH, M.W., ET AL. Genetic restriction of AIDS pathogenesis by an SDF-1 chemokine gene variant. *Science* **279**:389–393 (1998).

# AGE-STRUCTURED CORE GROUP MODEL AND ITS IMPACT ON STD DYNAMICS

CARLOS CASTILLO-CHAVEZ\* AND WENZHANG HUANG†

**Abstract.** The recruitment of new susceptibles into a core group of sexually-active individuals may depend on the current levels of infection within a population. We extend the formalism of Hadeler and Castillo-Chavez (1995), that includes prevalence dependent recruitment rates, to include age structure within core and noncore populations. Some mathematical results are stated but only a couple of proofs are included since our objectives are to highlight the modeling process and the dynamic possibilities. This paper concludes with an example where endemic distributions can be supported when the basic reproductive number  $R_0$  is less than one. Systems that are capable of supporting multiple attractors are more likely to support disease re-emergence. This model is likely to support stable multiple attractors when  $R_0 < 1$ .

**Introduction.** Core groups are generally small subpopulations capable of supporting strong transmission rates and high disease prevalence. Effective disease management strategies not only cannot ignore core groups but in fact must focus on them (Hethcote and Yorke 1984). The role of behavior on the dynamics of sexually transmitted disease (STDs) gained additional importance with the emergence of HIV, in western societies, nearly two decades ago (Castillo-Chavez 1989; Hethcote and Van Ark, 1992). Important changes in behavior were observed and documented in various homosexually-active populations including those living in San Francisco, New York and Boston. These observed changes included a reduction in average sexual activity, a decline in risky behavior, and a decrease in the rate of unprotected sexual contacts (see Baldwin and Baldwin 1988, Curran et al. 1988, Fineberg 1988, Evans et al. 1989, Martin 1987, Saltzman et al. 1987, Shechter et al. 1988, van Griensven et al. 1989a,b, Wilkenstein 1988, and Wiktor et al. 1990). Recent data (see below) seem to support the view that at best these changes were not as prevalent as one would have liked to and at worst they may have been kept only for a short period of time.

The combined effects of short- and long-term reductions of risk behaviors on the transmission dynamics of sexually-transmitted diseases (STDs) has been explored in limited ways via the use of simple models (Blythe et al. 1993a, 1993b, 1997; Brauer et al. 1998; Hadeler and Castillo-Chavez, 1995; Hsieh and Cooke 2000; Velasco et al. 1996; and Velasco and Hsieh, 1994). The theoretical results and predictions obtained via mathematical models have helped define a nontrivial landscape (based on the qualitative dynamics of these models) in which appropriate evaluations of the effective-

---

\* Biometrics, Statistics and Theoretical and Applied Mechanics Departments, 432 Warren Hall, Cornell University, Ithaca, NY 14853-7801.

† Department of Mathematical Sciences, University of Alabama in Huntsville, Huntsville, AL 23599.

ness of disease management strategies may be tested. An effective program of evaluation must be capable of tracing the origin of behavioral changes. If changes are long term and the result of a systematic implementation of educational efforts then the program would be rated as highly effective. However, whenever such changes are temporary then the effectiveness of such program must necessarily be questioned. Educational programs that "work" only when disease prevalence is high are not effective since its "effectiveness" may decrease when individuals' risk and fear of infection goes down. Hadeler and Castillo-Chavez (1995) showed that educational programs "may not only fail to have a positive impact in core populations with low disease prevalence but may even increase their effective group size by implying a reduction in risk within the core."

Erin McClam from the *Associated Press* in her article "Syphilis outbreak alarms officials" (*Ithaca Journal*, February 23, 2001, page 10A) highlights the criticality of long-term behavioral changes. She cites several authorities: "HIV 'is no longer perceived to be the threat that it once was' said Dr. Ronald Valdeserri".... In fact, "Syphilis outbreaks in major cities 'show that the disease can make a comeback,' said Ken August'..." Ms. McClam includes some local statistics that show that out of the last 66 (from a total of 130) syphilis cases reported among gay men during the first semester of 2000 in four California Counties, "...33 reported they had anonymous sex, and 17 said they had met sex partners in bathhouses. Only one in five reported using a condom during his most recent sexual encounter, and two in five reported using illegal drugs." The connection to temporal changes in behavior and current policy is clear from Valdiserri's statement "...When we see reports of increasing risk behaviors, that's when we take action. We don't wait till we see the spike nationally' ..." Finally, it was noted in Ms. McClam's article that 34 out of 66 men who had syphilis were also HIV-positive while 9 of them had no idea about their HIV status. The memory of the fatality rates associated with HIV on homosexually-active men in San Francisco and similar cities seemed to have lost its impact on current generations of homosexually-active gay men. One of the possible consequences associated with such memory loss is the remergence of HIV (as predicted by Hadeler and Castillo-Chavez, 1995).

This paper is organized as follows: Section 1 introduces an age-structure model with prevalence dependent recruitment rates; Section 2 looks at the local stability of the infection-free distribution; Section 3 states the conditions for the existence of nonuniform endemic age distributions and gives an example where endemic age distributions are possible even though  $R_0 < 1$ . The occurrence of a backward (subcritical) bifurcation, several infected stationary states, and hysteresis phenomena (including abrupt changes in disease prevalence levels) may be possible in this model. Section 4 state our conclusions and discuss briefly possible extensions.

**1. The model.** Here we focus on disease dynamics within a core group population that recruits individuals from a non-core population. It is assumed that the rate of recruitment  $r$  from the non-core into the core depends on the level of disease prevalence within the core group. Specifically, it is assumed that recruitment slows down when prevalence is high and increases when it is low. The epidemic process is built on a simple demographic setting. It is assumed that the non-core group is completely inactive (no sexual activity), that is, it just serves as the source of new members for the core. It is further assumed that the average contact rate in the core group is age-dependent  $B(a)$ , that infected individuals are symptomatic, and that infected individuals may return (at a constant per capita rate  $\gamma$ ) to the susceptible class after treatment. Individuals are assumed to mix at random (proportionate mixing; Hethcote and Van Ark, 1987; Busenberg and Castillo-Chavez, 1991 or Castillo-Chavez et al. 1994) and the per capita mortality rate ( $\mu$ ) is assumed to be constant (independent of age). The model is formulated as a general homogeneous system (Hadeler et al. 1988, Hadeler 1992, Busenberg and Hadeler 1990) but only the case where the total population is either constant or asymptotically constant is discussed.

We let  $q(a, t)$  denote the non-core and  $c(a, t)$  the core group age-dependent densities. We sub-divide the core group into two epidemiological classes: susceptibles  $s(a, t)$  and infectives  $u(a, t)$ . Let  $r = r(\frac{U(t)}{C(t)})$  denote the per capita recruitment rate of susceptibles from the non-core (outside group with density  $p(a, t)$ ) where  $C(t) = \int_0^\infty c(t, a)da = \int_0^\infty [s(t, a) + u(t, a)]dt$  and  $U(t) = \int_0^\infty u(t, a)da$ . That is,  $C(t)$  and  $U(t)$  denote the total core group population and the population of infectives (within the core group) at time  $t$ , respectively. The disease dynamics in this setting can be described by the following set of equations:

$$(1.1) \quad \begin{aligned} (\frac{\partial}{\partial a} + \frac{\partial}{\partial t})q(t, a) &= -\mu q(t, a) - r(\frac{U(t)}{C(t)})q(t, a), \\ (\frac{\partial}{\partial t} + \frac{\partial}{\partial a})s(t, a) &= -\mu s(t, a) + \gamma u(t, a) + r(\frac{U(t)}{C(t)})q(t, a) \\ &\quad - \beta B(a)s(t, a)\Lambda(t, a), \\ (\frac{\partial}{\partial t} + \frac{\partial}{\partial a})u(t, a) &= -(\mu + \gamma)u(t, a) + \beta B(a)s(t, a)\Lambda(t, a), \end{aligned}$$

with boundary conditions

$$(1.2) \quad q(t, 0) = \mu K, \quad s(t, 0) = u(t, 0) = 0,$$

and initial conditions

$$q(0, a) = q_0, \quad s(0, a) = s_0, \quad u(0, a) = u_0.$$

$B \in L^\infty(\mathbb{R}^+)$  is nonnegative,  $\lim_{a \rightarrow \infty} B(a) = 0$ ,

$$\Lambda(t, a) = \frac{\int_0^\infty B(s)u(t, s)ds}{\int_0^\infty B(s)c(t, s)ds},$$

and  $q_0, s_0$  and  $u_0 \in L^1(\mathbb{R}^+)$  are all nonnegative. Furthermore, it is assumed that  $r(v)$  is decreasing and positive for  $v \in [0, 1]$ . Since  $p(t, a) = q(t, a) + s(t, a) + u(t, a)$  is the density of the total population then from the addition of the three equations in (1.1) we see that  $p(t, a)$  obeys the classical age-structured model:

$$(1.3) \quad \left( \frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) p(t, a) = -\mu p(t, a), \quad p(t, 0) = \mu K.$$

The integration of Equation (1.3) along characteristic lines gives

$$\lim_{t \rightarrow \infty} p(t, a) = \mu K e^{-\mu a}, \quad a \geq 0,$$

that is,  $p(t, a)$  converges to its stable age-distribution.

In order to simplify the analysis it will be assumed throughout this paper that the total population  $p(t, a)$  has reached its stable distribution, that is,

$$p(t, a) = p(a) = \mu K e^{-\mu a}, \quad a \geq 0.$$

Hence, it is implicitly assumed that the disease is nonfatal.

The substitution of  $q$  by  $K e^{-\mu a} - c$ , leads, using the second and third equations in (1.1), to the qualitatively equivalent simplified model

$$(1.4) \quad \begin{aligned} \left( \frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) c(t, a) &= -\mu c(t, a) + r \left( \frac{U(t)}{C(t)} \right) (\mu K e^{-\mu a} - c(t, a)), \\ \left( \frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) u(t, a) &= -\sigma u(t, a) + \beta B(a) (c(t, a) - u(t, a)) \Lambda(t, a), \end{aligned}$$

where  $\sigma = \mu + \gamma$ . The rest of this paper focuses on the study of the role of the basic reproductive number,  $R_0$ , on the local stability of the disease free distribution and on the possible existence of multiple (endemic) equilibria for System (1.4) (see Castillo-Chavez et al. 2001; this volume).

**2. Infection-free distribution and the threshold condition.** System(1.4) always admits the disease-free distribution  $(c^0, 0)$  as a solution, where  $c^0$  is the positive solution of the differential equation

$$(2.1) \quad \dot{c}(a) = -\mu c(a) + r(0)(\mu K e^{-\mu a} - c(a)), \quad 0 \leq a \leq \infty,$$

with initial condition  $c(0) = 0$ . Thus  $c^0$  is uniquely given by the explicit expression

$$c^0(a) = r_0 \int_0^a e^{-(\mu+r_0)(a-\theta)} \mu K e^{-\mu \theta} d\theta = \mu K e^{-\mu a} (1 - e^{-r_0 a}),$$

where  $r_0 = r(0)$ .

We now turn to study the stability of the disease free distribution. A straightforward computation shows that the linearized system at the disease free distribution is given by the system:

$$(2.2) \quad \begin{aligned} \left( \frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) x(t, a) &= -(\mu + r_0)x(t, a) + \nu(a) \int_0^\infty y(t, s)ds, \\ \left( \frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) y(t, a) &= -\sigma y(t, a) + \alpha(a) \int_0^\infty B(s)y(t, s)ds, \\ x(t, 0) &= y(t, 0) = 0, \end{aligned}$$

where

$$\nu(a) = \frac{r'(0)\mu K e^{-\mu a}}{\int_0^\infty c^0(\tau)d\tau}, \quad \alpha(a) = \frac{\beta B(a)c^0(a)}{\int_0^\infty B(\tau)c^0(\tau)d\tau}.$$

It is simple to show that the reproductive number (see Castillo-Chavez et al. 2001; this volume) is given by

$$R_0 = \int_0^\infty B(a) \int_0^a e^{-\sigma(a-\tau)} \alpha(\tau) d\tau da.$$

The basic reproductive number  $R_0$  (see Diekmann et al. 1990; Diekemann and Heesterbeek, 2000; and Castillo-Chavez et al. 2001) is used to settle the question of local asymptotic stability.

**THEOREM 2.1.** *If  $R_0 < 1$  and if  $B$  is uniformly Lipschitz continuous on  $\mathbb{R}^+$  then the disease free equilibrium is locally asymptotically stable.*

*Proof.* It is enough to show that the linear system (2.2) is asymptotically stable. First consider the second equation of (2.2). Let  $S(t), t \geq 0$  be the semigroup corresponding to this equation and let  $T(t), t \geq 0$  be a semigroup defined by the system

$$\left( \frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) y(t, a) = -\sigma y(t, a),$$

$$y(t, 0) = 0, \quad y(0, \cdot) = y_0 \in L^1(\mathbb{R}^+).$$

Since  $B$  is uniformly Lipschitz continuous and  $\lim_{a \rightarrow \infty} B(a) = 0$ , by the definition of  $\alpha(a)$  it can be verified that the operator  $L : L^1(\mathbb{R}^+) \rightarrow L^1(\mathbb{R}^+)$ :

$$[L\phi](a) = \alpha(a) \int_0^\infty B(s)\phi(s)ds$$

is compact. Therefore, it follows from the variation-of-constants formula that  $S(t)$  is a compact perturbation of  $T(t)$ . Hence (see G.F. Webb, Proposition 4.1, p.178-179; 1985)  $\gamma_{E\sigma}(T(t)) = \gamma_{E\sigma}(S(t))$ , where  $\gamma_{E\sigma}(L)$  denotes the radius of the essential spectrum of an operator  $L$ . Moreover, it is known (see, Webb, Example 4.10, p.175-176; 1985) that

$$(2.3) \quad \gamma_{E\sigma}(S(t)) = \gamma_{E\sigma}(T(t)) = e^{-\sigma t}.$$

Now we consider the eigenvalues of the infinitesimal generator  $A$  of the semigroup  $S(t)$  given by

$$[A\phi](a) = -\dot{\phi}(a) - \sigma\phi(a) + \alpha(a) \int_0^\infty B(s)\phi(s)ds,$$

with

$$\mathcal{D}(A) = \{\phi \in L^1(\mathbb{R}^+) : \dot{\phi} \in L^1(\mathbb{R}^+), \phi(0) = 0\}.$$

If  $\lambda$  is an eigenvalue of  $A$  and  $\phi$  its corresponding eigenvector then

$$\dot{\phi}(a) = -(\lambda + \sigma)\phi(a) + \alpha(a) \int_0^\infty B(s)\phi(s)ds.$$

Hence

$$\phi(a) = \int_0^a e^{-(\lambda + \sigma)(a-\theta)} \alpha(\theta) d\theta \int_0^\infty B(s)\phi(s)ds,$$

that is,

$$|\phi(a)| \leq \int_0^a e^{-(\Re\lambda + \sigma)(a-\theta)} \alpha(\theta) d\theta \int_0^\infty B(s)|\phi(s)|ds.$$

Multiplying both sides of the above inequality by  $B(a)$  and integrating it over the interval  $[0, \infty)$  gives

$$(2.4) \quad 1 \leq \int_0^\infty B(a) \int_0^a e^{-(\lambda + \sigma)(a-\theta)} \alpha(\theta) d\theta da.$$

Since

$$R_0 = \int_0^\infty B(a) \int_0^a e^{-\sigma(a-\theta)} \alpha(\theta) d\theta da < 1,$$

then Inequality (2.4) implies that there is a number  $w_0 > 0$  such that  $\Re\lambda \leq -w_0$ . It therefore follows that

$$(2.5) \quad \sup\{\Re\lambda : \lambda \in \sigma_p(A)\} \leq -w_0.$$

Hence  $\xi = \min\{\sigma, w_0\} > 0$  and (2.3), (2.5) imply the existence of an  $M > 0$  such that  $\|S(t)\| \leq M e^{-\xi t}$ , for all  $t \geq 0$ . Therefore,

$$(2.6) \quad y(t, \cdot) \leq M \|y_0\|_{L^1} e^{-\xi t}, \quad t \geq 0.$$

The first equation of (1.5) and the variation-of-constants formula implies that there exists a semigroup  $T_1(t)$  such that  $\|T_1(t)\| \leq e^{-(\mu+r_0)t}$ ,  $t \geq 0$  and such that

$$(2.7) \quad x(t, \cdot) = T_1(t)x_0 + \int_0^t T_1(t-s)\nu(\cdot)y(s, \cdot)ds.$$

From (2.6) and (2.7) we conclude that there is a positive number  $M_1$  such that

$$\|x(t, \cdot)\| \leq \|x_0\|e^{-(\mu+r_0)t} + M_1 \|y_0\|(e^{-\xi t} + e^{-(\mu+r_0)t}).$$

Hence System (1.5) is asymptotically stable.  $\square$

**3. Existence of endemic equilibria.** This section outlines the steps required in the proof of the existence of an endemic equilibrium when  $R_0 > 1$ . That is, we study the existence of positive solutions of the system of differential-integral equations

$$(3.1) \quad \begin{aligned} \dot{c}(a) &= -\mu c(a) + r\left(\frac{U}{C}\right)(\mu K e^{-\mu a} - c(a)) \\ \dot{u}(a) &= -\sigma u(a) + \frac{\beta B(a)(c(a) - u(a)) \int_0^\infty B(s)u(s)ds}{\int_0^\infty B(s)c(s)ds} \\ c(0) &= u(0) = 0, \end{aligned}$$

where  $U = \int_0^\infty u(a)da$ ,  $\int_0^\infty C = \int_0^\infty c(a)da$ .

Since  $0 \leq c(a) \leq p(a) = \mu K e^{-\mu a} \rightarrow 0$  as  $a \rightarrow \infty$  and  $c(0) = 0$  then the integration of both sides of (3.1) gives

$$\mu C = r\left(\frac{U}{C}\right)(K - C),$$

or equivalently

$$(3.2) \quad r\left(\frac{U}{C}\right) = \frac{\mu C}{K - C}.$$

In our search for a positive solution of (3.1)  $C$  is treated as a free parameter with  $C \in (0, K)$  and the function

$$f(C) = \frac{\mu C}{K - C}, \quad 0 < C < K$$

is considered. Whenever  $(c, u)$  is a solution of (3.1) then the use of the first equation of (3.1) gives

$$c(a) = h(C, a) := \mu K e^{-\mu a} (1 - e^{-f(C)a}).$$

Hence, a positive solution of (3.1) must be a solution of the following differential-integral equation

$$(3.3) \quad \begin{aligned} \dot{u}(a) &= -\sigma u(a) + \frac{\beta B(a)}{\int_0^\infty B(s)h(C, s)ds} [h(C, a) - u(a)] \int_0^\infty B(s)u(s)ds, \\ u(0) &= 0, \end{aligned}$$

subject to the constraint

$$(3.4) \quad r\left(\frac{\int_0^\infty u(a)da}{\int_0^\infty h(C, a)da}\right) = f(C).$$

The proof of the existence of positive solutions is technical and requires the proof of a series of lemmas. Here we list these lemmas without proof (see Huang and Castillo-Chavez 2001).

LEMMA 3.1. Let  $G(C, a) = \frac{\beta B(a)}{\int_0^\infty B(s)h(C, s)ds}$ . If we assume that  $B(a) > 0$  for all  $a \geq 0$  then the differential-integral equation (3.3) has a positive solution  $u_C \leq h(C, \cdot)$  if and only if

$$\int_0^\infty B(a) \int_0^a e^{-\sigma(a-s)} G(C, s) ds > 1.$$

Furthermore, the solution  $u_C$  depends continuously on  $C$ . Next we let  $C_0 = \frac{r_0 K}{\mu + r_0}$ .

LEMMA 3.2. There is a  $0 < \rho < 1$  such that for each  $0 < C \leq C_0$ , if  $u_C > 0$  is a solution of (3.2), then

$$\frac{U_C}{C} = \frac{\int_0^\infty u_C(a) da}{C} \leq \rho.$$

THEOREM 3.3. If  $\int_0^\infty B(a) \int_0^a e^{-\sigma(a-s)} G(C_0, s) ds > 1$  and  $B(a)$  is strictly positive then (3.1) has at least one positive solution.

*Proof.* Let

$$R(C) = \int_0^\infty B(a) \int_0^a e^{-\sigma(a-s)} G(C, s) ds$$

then  $R(C)$  is continuous on  $C$ . If we let

$$\bar{C} = \inf\{\hat{C} \in (0, C_0) : R(C) > 1 \text{ for all } C \in (\hat{C}, C_0]\},$$

then  $0 \leq \bar{C} < C_0$ , since  $R(C_0) > 1$ .

Case 1. If  $\bar{C} = 0$  then

$$f(C_0) = \frac{\mu C_0}{K - C_0} = r_0 = r(0) > r\left(\frac{U_{C_0}}{C_0}\right).$$

Lemma 3.2 yields

$$\lim_{C \rightarrow 0} r\left(\frac{U_C}{C}\right) \geq r(\rho) > r(1) \geq 0 = \lim_{C \rightarrow 0} \frac{\mu C}{K_C} = f(0).$$

Therefore, there is at least a  $C_* \in (0, C_0)$  such that

$$r\left(\frac{U_{C_*}}{C_*}\right) = f(C_*).$$

Case 2. If  $\bar{C} > 0$  then continuity implies that  $R(\bar{C}) = 1$  and that

$$\lim_{C \rightarrow \bar{C}} u_C = 0.$$

Hence  $f(C_0) > r\left(\frac{U_{C_0}}{C_0}\right)$  and

$$f(\bar{C}) = \lim_{C \rightarrow \bar{C}} \frac{\mu C}{K - C} < \frac{\mu C_0}{K - C_0} = r(0) = \lim_{C \rightarrow \bar{C}^+} r\left(\frac{U_C}{C}\right).$$

Consequently, there is a  $C_* \in (\bar{C}, C_0]$  such that

$$\frac{\mu C_*}{K - C_*} = r\left(\frac{U_{C_*}}{C_*}\right).$$

□

**Remark.** If  $r(1) > 0$  then Lemma 3.2 is not needed in the proof of Theorem 3.3.

**COROLLARY 3.4.** Suppose  $R_0 > 1$  then System (1.4) has at least an endemic equilibrium.

The proof can be found in Huang and Castillo-Chavez (2001).

**4. Endemic equilibria when  $R_0 < 1$ .** The ODE (non-age structured) model of Castillo-Chavez and Hadeler (1995) supports multiple endemic equilibria when  $R_0 < 1$  but its epidemiological structure is more complex than in our age structured model. The assumption of age-independent rates and integration over all age classes leads to an ODE model that is not capable of supporting multiple endemic equilibria. It is therefore of some interest to determine whether or not our age structured model can indeed support endemic equilibria when  $R_0 < 1$ . To show that Model (1.4) can support positive equilibria when  $R_0 < 1$  we need the following theorem which is stated without proof:

**THEOREM 4.1.** Given  $B > 0$ . For  $\alpha > 0$  let  $\hat{R}(\alpha)$  be defined by

$$(4.1) \quad \hat{R}(\alpha) = \frac{\int_0^\infty B(a) \int_0^a e^{-\sigma(a-s)} B(s) e^{-\mu s} (1 - e^{-\alpha s}) ds da}{\int_0^\infty B(a) e^{-\mu a} (1 - e^{-\alpha a}) da}.$$

If there is an  $\alpha_0 > 0$  such that  $\frac{d\hat{R}(\alpha_0)}{d\alpha} < 0$  then there exists a function  $r(v)$  of the rate of recruitment (from the non-core into the core) defined on  $[0, 1]$  such that  $r$  is decreasing and the corresponding reproductive number  $R_0$  is less than 1 but yet, there exists an endemic equilibrium.

**Example.** Let  $B(a) = ae^{-\eta a}$  where  $\eta > 0$ . Hence,  $B$  is bounded, positive and  $\lim_{a \rightarrow \infty} B(a) = 0$ . By the definition (4.1) we have

$$\frac{d\hat{R}(\alpha)}{d\alpha} = F(\alpha)[V(\alpha) - W(\alpha)],$$

where

$$\begin{aligned}
F(\alpha) &= \frac{1}{\left[\int_0^\infty B(s)e^{-\mu s}(1 - e^{-\alpha s})ds\right]^2}, \\
V(\alpha) &= \left( \int_0^\infty \left[ \int_s^\infty B(a)e^{-\sigma(a-s)}da \right] sB(s)e^{-(\mu+\alpha)s}ds \right) \\
&\quad \times \int_0^\infty B(a)e^{-\mu a}(1 - e^{-\alpha a})da, \\
W(\alpha) &= \left( \int_0^\infty \left[ \int_s^\infty B(a)e^{-\sigma(a-s)}da \right] B(s)e^{-\mu s}(1 - e^{-\alpha s})ds \right) \\
&\quad \times \int_0^\infty B(a)ae^{-(\mu+\alpha)a}da.
\end{aligned}$$

Let  $\sigma = \alpha = 2$ ,  $\mu = 1$ . A straightforward computation gives

$$\begin{aligned}
V(2) &= \frac{2(15 + 8\eta)(4 + 2\eta)}{(2 + \eta)^2(2\eta + 3)^3(3 + \eta)^2(2\eta + 3)(1 + \eta)^2}, \\
W(2) &= \frac{2(5 + 4\eta)(2\eta + 3)^3 - 2(7 + 4\eta)(2\eta + 1)^3}{(2 + \eta)^2(2\eta + 3)^3(3 + \eta)^2(2\eta + 1)^3(3 + \eta)}.
\end{aligned}$$

If we now let  $\eta = 0$  then

$$V(2) - W(2) = \frac{2[15 \cdot 4 - 128]}{2^3 \cdot 3^3 \cdot 3^3} < 0.$$

Then the continuity implies that

$$\frac{d\hat{R}(2)}{d\alpha} = F(2)[V(2) - W(2)] < 0$$

for all sufficiently small positive number  $\eta$ . Therefore it follows from Theorem 4.1 that we can find a decreasing recruitment function  $r$  such that its corresponding system has a stable disease free distribution, as well as an endemic equilibrium.

**5. Conclusion.** For natural selection to “advance” opportunities must be present. The growth of cities, the effects of mass transportation and international travel, the development of antibiotics, vaccines and other forms of treatment, the changes on values and costumes driven in part by globalization and modes of mass information are but some of the factors that make the study of disease dynamics challenging, complex and interesting. Every pressing question immediately brings to the forefront a multitude of choices: What should it be our epidemiological unit? What is the appropriate temporal or geographical scale? Does behavior matter? Is the population under study really isolated? Clearly, one cannot assume the existence of a fixed landscape in the study of disease evolution.

In the study of disease emergence or re-emergence, it seems that one must be aware of the possibilities. Such possibilities are often encoded in

the dynamical landscapes associated with models for the spread of a particular disease (landscapes are also closely connected to mechanisms). It seems to us that the occurrence of backward (subcritical) bifurcations, the existence of multiple infected stationary states, and hysteresis phenomena (including abrupt changes in disease prevalence levels) are but a few of the components that are supportive of disease reemergence. Models that generate this type of landscapes must be understood since they provide useful insights in the study of disease re-emrgence and evolution (see Castillo-Chavez et al. 1989, Huang et al. 1992, Hadeler et al. 1995 and Feng et al. 2000).

**Acknowledgments.** This research was partially supported by NSF and NSA awards to Carlos Castillo-Chavez. This work was completed while CCC and WH visited the IMA in Minnesota.

## REFERENCES

- BALDWIN J.D. AND BALDWIN, J.I. (1988). *Factors affecting AIDS-related sexual risk-taking behavior among college students*. J. Sex Research **25**, 181–196.
- BLYTHE, S.P., F. BRAUER, AND C. CASTILLO-CHAVEZ (1997). *Demographic Recruitment in Sexually Transmitted Disease Model*. Proc. First World Congress on Computational Medicine, Public Health and Biotechnology Part II Austin, TX, 1994. Ed. Matthew Witten. Series in Mathematical Biology and Medicine **5**, 1438–1457.
- BLYTHE, S.P., F. BRAUER, C. CASTILLO-CHAVEZ, AND J.X. VELASCO-HERNANDEZ (1993a). *Models for sexually transmitted diseases with recruitment*. Biometrics Unit Technical Report BU-1193-M, Cornell University.
- BLYTHE, S.P., K. COOKE, AND C. CASTILLO-CHAVEZ (1993b). *Autonomous risk-behavior change, and non-linear incidence rate, in models of sexually transmitted diseases*. Biometrics Unit Technical Report BU-1048-M, Cornell University, 1993b.
- BRAUER, F., C. CASTILLO-CHAVEZ, AND J.X. VELASCO-HERNANDEZ (1998). *Recruitment into Core Group and its effect on the spread of a sexually transmitted disease*. Advances in Mathematical Population Dynamics - Molecules, Cells, and Man (O. Arino, D. Axelrod, and M. Kimmel (eds), World Scientific Press, pp. 477–486, 1998.
- BUSENBERG, S. AND C. CASTILLO-CHAVEZ (1991). *A general solution of the problem of mixing subpopulations, and its application to risk- and age-structure epidemic models for the spread of AIDS*, IMA J. of Math. Appl. Med., **8**, 1–29.
- BUSENBERG, S. AND K.P. HADELER (1990). *Demography and epidemics*. Math. Biosc. **101**, 41–62.
- CASTILLO-CHAVEZ, C., ed. *Mathematical and Statistical Approaches to AIDS Epidemiology*, Vol. 83 of Lecture Notes in Biomathematics, Springer-Verlag, Berlin, 1989.
- CASTILLO-CHAVEZ, C., K. COOK, W. HUANG, AND S.A. LEVIN (1989). *Results on the dynamics for models for the sexual transmission of human immunodeficieny virus*. Appl. Math. Lett. **2**(4), 327–331.
- CASTILLO-CHAVEZ, C., Z. FENG, AND W. HUANG (2001). *On the computation of  $R_0$  and its role on global stability*, This volume.
- CASTILLO-CHAVEZ, C., J.X. VELASCO-HERNANDEZ, AND S. FRIDMAN (1994). *Modeling contact structures in Biology*. In: *Frontiers of Theoretical Biology*, S.A. Levin (ed.). Lecture notes in Biomathematics, **100**. Spring-Verlag, New York.
- CURRAN, J.W., JAFFE, H.W., HARDY, A.M., MORGAN, M.W., SELIK, R.M., AND DONDERO, T.J. (1988). *Epidemiology of infection and AIDS in the United States*. Science **293**, 610–616.

- DIEKMANN, O., J.A.P. HEESTERBEEK, AND J.A.J. METZ (1990). *On the definition and the computation of the basic reproduction ratio  $R_0$  in models for infectious diseases in heterogeneous population.* J. Math. Biol. **28**, 365–382.
- DIEKMANN, O. AND J.A.P. HEESTERBEEK (2000). *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation.* Wiley, New York.
- EVANS, B.A., MCLEAN, K.A., AND DAWSON, S.G. et al. (1989). Trends in sexual behavior and risk factors for the HIV infection among homosexual men, 1984–1987. Brit. Med. J. **298**, 215–218.
- FENG, Z., C. CASTILLO-CHAVEZ, AND A. CAPURRO (2000). *A model for TB with exogenous reinfection.* Theo. Pop. Biol. **578**, 237–247.
- FINEBERG, H.V. (1988). *Education to prevent AIDS: prospects and obstacles.* Science **239**, 592–596.
- HADELER, K.P., WALDSTÄTTER, R., WÖRZ-BUSEKROS, A. (1988). *Models for pair formation in bisexual populations.* J. Math. Biol. **26**, 635–649.
- HADELER, K.P. (1992). Periodic solutions of homogeneous equations. J. Diff. Equ. **95**, 183–202.
- HADELER, K.P. AND C. CASTILLO-CHAVEZ (1995). *A core group model for disease transmission,* Math. Biosci. **128**, 41–55.
- HETHCOTE H.W. AND J.W. VAN ARK (1987). *Epidemiological models with heterogeneous populations: Proportionate mixing, parameter estimation and immunization programs,* Math. Biosci. **84**, 85–118.
- HETHCOTE H.W. AND J.W. VAN ARK (1992). *Modeling HIV Transmission and AIDS in the United States,* Vol. **95** of Lecture Notes in Biomathematics, Springer-Verlag, Berlin.
- HETHCOTE, H.W. AND YORKE, J.A. (1984). *Gonorrhea: Transmission Dynamics and Control.* Lecture Notes in Biomathematics **56**, Springer-Verlag.
- HSIEH, Y. AND K. COOKE (2000). *Behaviour change and treatment of core groups: its effect on the spread of HIV/AIDS.* IMA J. of Math. Applied to Medicine and Biology **17**, 213–241, 2000.
- HUANG, W., K.L. COOKE, S., AND C. CASTILLO-CHAVEZ (1992). *Stability and Bifurcation for a multiple-group model for the dynamics of HIV-AIDS transmission,* SIAM J. Appl. Math. **52**, 835–854.
- HUANG, W. AND C. CASTILLO-CHAVEZ (2001), *On the role of disease driven recruitment on the dynamics of STDs in age-structured populations,* in preparation.
- MARTIN, J.L. (1987). *The impact of AIDS in gay male sexual behavior patterns in New York City.* Am. J. of Pub. Health **77** 578–581.
- MCKUSICK, L., WILEY, J.A., COATES, T.J., STALL, R., SAIKA, B., MORIN, S., HORTSMAN, C.K., AND CONANT, M.A. (1985). *Reported changes in the sexual behavior of men at risk for AIDS, San Francisco, 1983–1984: the AIDS behavioral research project.* Public Health Reports **100**, 622–629.
- SALTZMAN, S.P., STODDARD, A.M., MCCUSKER, J., MOON, M.W., AND MAYER, K.H. (1987). *Reliability of self-reported sexual behavior risk factors for HIV in homosexual men.* Public Health Report **102**, 692–697.
- MCCCLAM, E. (2001). “*Syphilis Outbreak alarms officials*”, Ithaca Journal, Section A., p. 10A, February 23, 2001.
- SCALIA-TOMBA, G. (1991). *The effect of structural behavior change on the spread of HIV.* Math. Bios. **107**, 547–556.
- SHECHTER, M.T., CRAIB, K.J.P., WILLOUGHBY, B., DOUGLAS B., ALASTAIR MCLEOD, W., MAYNARD, M., CONSTANCE, P., AND O'SHAUGHNESSY, M. (1988). *Patterns of sexual behavior and condom use in a cohort of homosexual men.* Amer. J. Publ. Health **78**, 1535–1538.
- SHILTS, R. (1987). *And the band played on.* San Martin's press, New York.
- WIKTOR, S.Z., BIGGAR, R.J., MELBY, M., EBBESEN, P., COLCLOUGH, G., DI GIOIA, R., SANCHEZ, W.C., GROSSMAN, R.J., AND GOEDERT, J. (1990). *Effects of knowledge of human immunodeficiency virus infection status on sexual activity among homosexual men.* J. of AIDS **3**, 62–68.

- VAN GRIENSVEN, G.J.P., DE VROOME, E.M.M., GOUDSMIT, J., AND COUTINHO, R.A. (1989a). *Changes in sexual behavior and the fall in incidence of HIV infection among homosexual men.* Br. Med. J. **298**, 218–221.
- VAN GRIENSVEN, G.J.P., DE VROOME, E.M.M., TIELMAN, R.A.P., GOUDSMIT, J.A.A.P., DE WOLF, F., VAN DER NOORDAA, J., AND COUTINHO, R.A. (1989b). *Effect of human immuno-deficiency virus (HIV) antibody knowledge on high-risk sexual behavior with steady and non-steady partners among homosexual men.* Am. J. Epidemiol. **129**, 596–603.
- VELASCO-HERNANDEZ, J.X., F. BRAUER, AND C. CASTILLO-CHAVEZ (1996). *Effects of Treatment and Prevalence-dependent Recruitment on the Dynamics of a fatal disease.* IMA Journal of Math. Medicine and Biology, **13**(3), 175–192.
- VELASCO-HERNANDEZ, J.X. AND Y. HSIEH (1994). *Modelling the effect of treatment and behavioral change in HIV transmission dynamics.* J. Math Biol. **32**, 233–249.
- WEBB, G.F. (1985). *Theory of Nonlinear Age-dependent Population Dynamics*, Marcel Dekker, New York.
- WILKENSTEIN, W. JR., WILEY, J.A., PADIAN, N.S., SAMUEL, M., SHIBOSKI, S., ASCHER, M. S., AND LEVY J. A. (1988). *The San Francisco Men's Health Study, continue decline in HIV seroconversion rates among homosexual/bisexual men.* Am. J. Public Health **78**, 1472–1474.

# GLOBAL DYNAMICS OF TUBERCULOSIS MODELS WITH DENSITY DEPENDENT DEMOGRAPHY

BAOJUN SONG\*, CARLOS CASTILLO-CHAVEZ\*, AND JUAN P. APARICIO†

**Abstract.** Mathematical models for Tuberculosis with linear and logistic growth rates are considered. The global dynamic structure for the logistic recruitment model is analyzed with the help of a strong version of the Poincaré-Bendixson Theorem. The nature of the global dynamics of the same model with a linear recruitment rate is established with the use of explicit threshold quantities controlling the absolute and relative spread of the disease and the likelihood of extinction or persistence of the total population. The classification of planar quadratic systems helps rule out the existence of closed orbits (limit cycles).

**Key words.** Tuberculosis, Global Stability, Monotone Systems, Density-dependent Recruitment Rates.

**1. Introduction.** Tuberculosis (TB) was the main cause of death in many places around the world until the recent past. Although the situation has changed dramatically in the past century, TB still remains the main cause of death by an infectious (communicable) disease. Two million deaths per year are still attributed to TB.

Tuberculosis is an infectious disease with singular features, that is, its epidemiology is quite different from the epidemiology of most communicable diseases. TB's progression is quite slow and treatment (costly and relative difficult to implement) is available for the latent and active phases of the disease. TB, caused by *Mycobacterium tuberculosis*, responds to a complex treatment schedule and recovery or treatment do not give immunity. Lack of treatment can lead to death and resistance to antibiotics is a serious problem (Blower and Gerberding, 1998; Castillo-Chavez and Feng, 1997). The case fatality of untreated individuals is about 50% for pulmonary tuberculosis; a percentage that rises to about 75% when cases are also sputum positive (Styblo, 1991). Since the average rate of progression from infected (non-infectious) to active (infectious) TB is very slow, most (particularly in developing nations) infected individuals never develop active-TB. That is, the dynamics of TB at the population level are slow with characteristic time-scales of decades. Consequently, demography plays an important role on the transmission dynamics of TB and its partial assessment on TB is the main focus of this paper. We look at two distinct demographic scenarios: exponential growth on a long time scale and exponential growth on a short time scale (quasi-exponential growth). The effect of TB-induced mortality is considered on both demographic settings. Mathematical studies of the impact of fatal diseases on populations with demography have

---

\*Department of Biometrics, Cornell University, Ithaca, NY 14853, USA.

†Universidad de Belgrano-CONICET, Zabala 1851, piso 12, 1428 Buenos Aires, Argentina.

been carried out by many researchers (see Brauer, 1989; Busenbarg and van den Driessche, 1990; Thieme, 1992; Lin and Hethcote, 1993; Iannelli, Miller, and Pugliese, 1992; Brauer and Castillo-Chavez, 2001) but not in the context of tuberculosis (but see Aparicio *et al.*, 2001a and 2001b).

Quasi-exponential growth, a process that can be modeled and fitted to data using a linear demographic model with time-dependent per-capital growth rates, has been studied in the past (Cohen, 1995; Aparicio *et al.*, 2001a). For example, the USA population exhibited a quasi-exponential phase until the middle of the 18th century, a phase that has been followed by an almost linear growth phase afterwards. The pattern of USA population growth from Colonial Times to our days has been fitted to a logistic model (see for example, Aparicio *et al.*, 2001a). As Cohen (1995) points out most models used to fit demographic data can only give reasonable predictions over short periods of time at best. Many of the reasons behind the failure of demographic models in predicting patterns of population growth over long time scales are outlined in Cohen's recent book. Many major cities in developed nations around the world that exhibited logistic growth had already reached (almost) stable values a few decades ago. The USA population growth pattern is different than those of developing nations (long-term quasi-exponential) or developed nations (no growth). The USA population is still growing albeit linearly. Hence, its growth is sort of intermediate between logistic and exponential. In this manuscript, we formulate a simple TB transmission model in a homogeneous population with demography. We show that demography does not impact the *qualitative* features of TB epidemics. That is, our results are qualitatively equivalent to those resulting from models for TB dynamics without demography (Blower *et al.*, 1995 and 1996; Castillo-Chavez and Feng, 1997 and 1998; Feng, Castillo-Chavez, and Huang, 2001). We establish the existence of a sharp "tipping point" with the help of natural non-dimensional thresholds that govern the transmission dynamics of TB and the nature of demographic growth.

The demographic setting is quite simple and well known (Brauer and Castillo-Chavez, 2001). We assume that the total population  $N(t)$  is either governed by

$$(1) \quad \frac{dN}{dt} = (b - \mu)N,$$

where  $b$  is the per-capita birth rate and  $\mu$  is per-capita mortality, both assumed constant (the total population  $N$  grows exponentially  $N(t) = N_0 e^{rt}$  where  $r = b - \mu$  is the net population growth rate and  $N_0$  initial population size); or that  $N(t)$  is modeled by

$$(2) \quad \frac{dN}{dt} = rN \left(1 - \frac{N}{K}\right),$$

where  $K$  is the carrying capacity.

It is shown that the *qualitative* dynamics of TB are “essentially” the same when  $N(t)$  is modeled by (1) or (2). In fact, the *qualitative* dynamics are identical to those without demography (Castillo-Chavez and Feng, 1997; Feng, Castillo-Chavez, and Huang, 2001). Our analysis is nevertheless useful as it identifies key thresholds in either case, that is, our analysis clarifies the role of demography. The global dynamic structure for the logistic recruitment model is studied with the help of a strong version of Poincaré-Bendixson Theorem while the nature of the global dynamics of the model with a linear recruitment rate is established with the use of explicit threshold quantities controlling the absolute and relative spread of the disease and the likelihood of extinction or persistence of the total population. The classification of planar quadratic systems is used to rule out the existence of closed orbits.

The rest of this paper is organized as follows: Section 2 introduces the epidemiological setting; Section 3 and 4 analyze the role of linear and logistic growth, respectively; Section 5 discusses the relevance of our results. Detailed analysis of the models, including the set up for the use of a strong version of the Poincaré-Bendixson Theorem, are included in the Appendix.

**2. Epidemiological model.** It is assumed that all immigrants and newborns are uninfected, that is, they are members of the susceptible class  $S$ . Infected individuals are divided into two classes: asymptomatic and non-infectious (latent-TB or inactive-TB), members of the class  $E$ ; and symptomatic infectious (active-TB), members of the class  $I$ . Treated individuals are moved into the class  $T$ . Individuals in either the  $E$ -class or the  $I$ -class may enter the  $T$ -class by treatment or natural recovery.

Typically, latent individuals remain latent (in  $E$ -class) for a long period of time before progressing into the infectious class  $I$ , but progression is not uniform in general. Risk of progression to active-TB is higher soon after infection. Those who progress to active-TB within the first five years after infection are classified as *primary tuberculosis cases* while those who progress later are classified as *secondary tuberculosis cases*. Late progression (secondary cases) may be due to endogenous reactivation of the initial infection or *exogenous re-infection* (Styblo, 1991; Feng *et al.*, 2000). Infected individuals who do not progress to active-TB within the first years following primary infection are at a low risk of progression. There are many different ways of modeling this differential risk of progression (Blower *et al.*, 1995; Vynnycky and Fine, 1997; Aparicio *et al.*, 2000; Feng, Castillo-Chavez, and Capurro, 2000; Thieme *et al.*, 1993; Feng, Castillo-Chavez, and Huang, 2001). The incorporation of primary tuberculosis and endogenous reactivation requires the introduction of age of infection (Vynnycky and Fine, 1997) but its incorporation adds complexity to the model. It may be reasonable to assume that, in the absence of re-infection, the distribution of new cases decays exponentially after the first infection (Styblo, 1991). In this manuscript, we ignore age-of-infection and assume a constant

per-capita progression rate. The nature of this assumption limits the generality of our results for TB. The mathematical analysis of a general model with long and variable period of time in the  $E$ -class suggest that this assumption may not be as limited as it appears to be (Feng, Castillo-Chavez, and Huang, 2001).

Our simple transmission model, which preserves some of the main features of tuberculosis epidemiology, is given by

$$(3) \quad \frac{dS}{dt} = B(N) - \beta cS \frac{I}{N} - \mu S,$$

$$(4) \quad \frac{dE}{dt} = \beta cS \frac{I}{N} - (\mu + k + \alpha)E + \beta' cT \frac{I}{N},$$

$$(5) \quad \frac{dI}{dt} = kE - (\mu + d + \rho)I,$$

$$(6) \quad \frac{dT}{dt} = \alpha E + \rho I - \beta' cT \frac{I}{N} - \mu T,$$

$$N = S + E + I + T,$$

where the recruitment rate  $B(N)$  is either  $bN$  or  $b_0 N(1 - \frac{N}{K})$ . [The form  $B(N) = \Lambda - \mu N$  was also used by Castillo-Chavez and Feng (1997 and 1998).] We let  $\beta$  and  $\beta'$  denote the average infected proportions of susceptible and treated individual contacted by one infectious individual per unit of time, respectively;  $c$  is the per-capita contact rate;  $\beta cS \frac{I}{N}$  and  $\beta' cT \frac{I}{N}$  denote the infection and reinfection rates, respectively;  $\mu$  denotes the per-capita mortality rate;  $d$  the TB-induced mortality rate;  $k$  the per-capita rate of progression to active-TB from latent-TB (class  $E$ );  $\alpha$  and  $\rho$  denote the treatment rates for the latent and infectious class, respectively.

Because TB increases mortality, both demography and epidemiology are incorporated into the equation that governs the dynamics of the total population, that is, we have that

$$(7) \quad \frac{dN}{dt} = B(N) - \mu N - dI$$

Currently, most deaths caused by TB represent but a small proportion of the deaths in most populations. In other words,  $d$  is often insignificant. Therefore, a linear recruitment rate  $B(N) = b_0 N$  with reasonable  $b_0$  values is likely to support exponential growth on a TB-infected population. The use of a logistic recruitment rate  $B(N) = b_0 N(1 - \frac{N}{K})$  to model the demography in general is also likely to result in logistic growth for the total population  $N$  in the presence of TB.

To simplify our analysis, we further assume that infected and reinfected proportions are equal,  $\beta' = \beta$ . Hence, the use of the variables,  $N$ ,  $E$  and  $I$ , is now enough, that is, Model (3–6) reduces to:

$$(8) \quad \frac{dN}{dt} = B(N) - \mu N - dI,$$

$$(9) \quad \frac{dE}{dt} = \beta c(N - E - I) \frac{I}{N} - (\mu + k + \alpha)E,$$

$$(10) \quad \frac{dI}{dt} = kE - (\mu + d + \rho)I.$$

Throughout this paper, we shall consistently use the following compressed notations  $m_r = b_0 + \rho + d$ ,  $n_r = b_0 + \alpha + k$ ,  $m_\mu = \mu + \rho + d$ ,  $n_\mu = \mu + \alpha + k$ , and  $\sigma = \beta c$  to simplify the discussions.

**3. Linear recruitment rate .** In this section, we study the dynamics of Model (8–10) with  $B(N) = b_0 N$ . That is, it is assumed that the total population exhibits exponential growth in the absence of TB (the net growth rate of the population, in the absence of the disease, is  $r \equiv b_0 - \mu$ ). Total population size increases (decreases) exponentially if  $b_0 > \mu$  ( $b_0 < \mu$ ), and remains constant if  $b_0 = \mu$ . The case where  $b_0 < \mu$  is trivial. Hence we assume that  $b_0 \geq \mu$ . In the presence of TB the total population may (theoretically) decrease exponentially even when  $b_0 > \mu$  provided that  $d$  is large enough. That is, technically, a fatal disease like TB can control population growth (see also May and Anderson, 1985; Busenberg and Hadeler, 1990). Realistic examples of situations where a disease has impacted or is likely to impact demographic growth can be found in the work on myxomatosis by Levin and Pimentel (1981) or in the work on HIV by Anderson, May, and Mclean (1988 and 1989).

Three non-dimensional threshold parameters provide a full characterization of the possible dynamical regimes of System (8–10):  $\mathcal{R}_0$ ,  $\mathcal{R}_1$  and  $\mathcal{R}_2$ .

The basic reproductive number given by

$$(11) \quad \mathcal{R}_0 = \left( \frac{\sigma}{\mu + \rho + d} \right) \left( \frac{k}{\mu + \alpha + k} \right),$$

gives the average number of secondary cases produced by a typical infectious individual during his/her entire life in a population of mostly susceptibles.  $\mathcal{R}_0 < 1$  implies that the infected populations goes to zero while  $\mathcal{R}_0 > 1$  implies that the infected populations grows (initially) exponentially (together with the total population  $N$ ). In this last case there are two possibilities:  $N$  grows faster than  $I$  or  $N$  does not grow faster than  $I$ . In the first case, the fraction  $u = \frac{I}{N}$  approaches zero as time increases and the additional threshold parameter

$$(12) \quad \mathcal{R}_1 = \left( \frac{\sigma}{b_0 + \rho + d} \right) \left( \frac{k}{b_0 + \alpha + k} \right)$$

plays a role.  $\mathcal{R}_1$  discriminates between the last two possibilities.  $\mathcal{R}_1 < 1$  implies that  $\lim_{t \rightarrow \infty} u = 0$  while  $\mathcal{R}_1 > 1$  implies that  $\lim_{t \rightarrow \infty} u = u^*$  where  $u^*$  is

a positive constant. Because by assumption  $b_0 > \mu$ , we always have that  $\mathcal{R}_0 > \mathcal{R}_1$ .

\ If the infectious ( $I$ ) population changes faster than the total population ( $N$ ) then a fatal disease can drive the population to extinction (even when  $\mathcal{R}_1 > 1$ ). The threshold parameter that decides this last situation is given by

$$(13) \quad \mathcal{R}_2 = \frac{b_0 - \mu}{du^*},$$

where  $u^*$  is a positive constant (independent of  $\mu$  (see (19)), that is,  $\mathcal{R}_2$  determines whether or not the total population size grows exponentially. It will be shown later that the population size decreases exponentially (because of TB) only if  $\mathcal{R}_2 < 1$ .

A detailed characterization of the dynamics of System (8–10) is provided in the rest of this section with the mathematical details included in the appendix. System (8–10) is homogeneous of degree one and, hence, it can support exponential solutions. Hadeler's theory for the study of the linear (local) stability of homogeneous systems (Hadeler, 1990 and 1992) applies albeit it does not address the issue of the global stability of solutions, the main focus of our analysis. Global analysis requires the rewriting of System (8–10) using the projections  $u = \frac{I}{N}$ , and  $v = \frac{E}{N}$ . The equations for  $u, v$  are given by the following quadratic system:

$$(14) \quad \frac{du}{dt} = -m_r u + kv + du^2,$$

$$(15) \quad \frac{dv}{dt} = \sigma u - n_r v + (d - \sigma)uv - \sigma u^2.$$

Note that both  $u$  and  $v$  are independent of  $N$  and  $\mu$ . It is easy to check that the subset

$$\Omega = \{(u, v) \in \mathbf{R}_2^+ | u + v \leq 1\}$$

is positively invariant. To further simplify the quadratic System (14–15), we introduce the new variables  $x$  and  $y$  and rescale time  $t$ . Specifically, we let

$$(16) \quad x = \frac{d}{m_r + n_r} u, \quad y = \frac{kd}{(m_r + n_r)^2} \left( \frac{n_r}{k} u + v \right), \quad \text{and} \quad \tau = (m_r + n_r)t.$$

The re-scaled system becomes

$$(17) \quad \frac{dx}{d\tau} = -x + y + x^2,$$

$$(18) \quad \frac{dy}{d\tau} = x(a_1 + a_2y + a_3x),$$

where

$$a_1 = \frac{m_r n_r (\mathcal{R}_1 - 1)}{(m_r + n_r)^2}, \quad a_2 = \frac{d - \sigma}{d}, \quad \text{and} \quad a_3 = \sigma \frac{n_r - k}{d(m_r + n_r)}.$$

In the new system,  $\Omega$  becomes

$$\Omega_1 = \left\{ (x, y) \in \mathbf{R}_2^+ \mid \frac{n_r}{m_r + n_r} x \leq y \leq \frac{dk}{(m_r + n_r)^2} + \frac{n_r - k}{m_r + n_r} x \right\}$$

which is positively invariant under the flow of System (17–18). This last transformation not only reduces the number of parameters but, more importantly, it fixes the horizontal isocline and decomposes the vertical isocline into a degenerate quadratic curve. Under the standard classification of Ye *et al.* (1986), System (17–18) is a *quadratic system of the second type*.

The following two theorems characterize the dynamics of System (17–18) and hence of (14–15). Proofs are in the Appendix.

**THEOREM 1.** *For System (17–18) with  $b_0 > \mu$ , the trivial equilibrium  $(0, 0)$  is globally asymptotically stable if  $\mathcal{R}_1 \leq 1$ . Furthermore there exists a unique positive equilibrium which is globally asymptotically stable if  $\mathcal{R}_1 > 1$ . The standard classification of planar quadratic differential systems rules out the existence of closed orbits or limit cycles. (Other approaches can be used to draw the same conclusion, for example, see Busenberg and van den Driessche, 1990; Lin and Hethcote, 1993). The full structure of the System (8–10) is characterized in Theorem 2 below:*

**THEOREM 2.** *Consider System (8–10) and assume that  $b_0 > \mu$ .*

1. *If  $\mathcal{R}_0 < 1$  then  $(\infty, 0, 0)$  is globally asymptotically stable.*
2. *If  $\mathcal{R}_1 < 1 < \mathcal{R}_0$  then  $(\infty, \infty, \infty)$  is globally asymptotically stable and*  $\lim_{t \rightarrow \infty} \frac{I}{N} = 0$ ,  $\lim_{t \rightarrow \infty} \frac{E}{N} = 0$ .
3. *If  $1 < \mathcal{R}_1 < \mathcal{R}_0$  then*

*(a)  $(0, 0, 0)$  is globally asymptotically stable and  $\lim_{t \rightarrow \infty} \frac{I}{N} = u^*$ ,*

$$\lim_{t \rightarrow \infty} \frac{E}{N} = v^* \text{ when } \mathcal{R}_2 < 1,$$

*(b)  $(\infty, \infty, \infty)$  is globally asymptotically stable and  $\lim_{t \rightarrow \infty} \frac{I}{N} = u^*$ ,*

$$\lim_{t \rightarrow \infty} \frac{E}{N} = v^* \text{ when } \mathcal{R}_2 > 1, \text{ where}$$

$$u^* = \frac{-[d(m_r + n_r) - \sigma(m_r + k)] + \sqrt{\delta}}{2d(\sigma - d)(k\sigma - m_r n_r)},$$

$$v^* = \frac{m_r(a_2 + a_3 + \Delta^{1/2}) - 2a_2 du^*{}^2}{2a_2 k},$$

$$\delta = [d(m_r + n_r) - \sigma(m_r + k)]^2 + 4d(\sigma - d)(k\sigma - m_r n_r),$$

$$\Delta = (a_2 + a_3)^2 + 4a_1 a_2 > 0.$$

(19)

Hence, whenever  $\mathcal{R}_0 < 1$  the disease dies out while the total population increases exponentially. Although the disease spreads when  $\mathcal{R}_1 < 1 < \mathcal{R}_0$ , the proportions  $\frac{I}{N}$  and  $\frac{E}{N}$  approach zero. From (c) one sees that disease-induced mortality can lead to the extinction of a population which would otherwise increase exponentially (a fatal disease can regulate a population). Note that  $\mathcal{R}_2$  is a positive number since  $u^*$  is positive and independent of  $\mu$ . We have also established that when  $b_0 < \mu$ ,  $(0, 0, 0)$  is globally asymptotically stable even though  $\lim_{t \rightarrow \infty} \frac{I}{N} = u^*$  and  $\lim_{t \rightarrow \infty} \frac{E}{N} = v^*$  when  $\mathcal{R}_1 > 1$ .  $\mathcal{R}_1 < 1$  implies that  $\lim_{t \rightarrow \infty} \frac{I}{N} = 0$ ,  $\lim_{t \rightarrow \infty} \frac{E}{N} = 0$ . Note that  $\mathcal{R}_1 < \mathcal{R}_0$  whenever  $b_0 > \mu$ . Theorem 2 provides a complete characterization of the dynamic structure of Model (8–10).

**4. Logistic recruitment rate.** In this section, we study the case where  $B(N) = b_0 N(1 - \frac{N}{K})$ . Since the total population  $N$  is now bounded, a threshold parameter like  $\mathcal{R}_1$ , which determines the asymptotic behavior of the proportions, is meaningless in this setting.

Re-scaling  $N$  by  $\frac{N}{K}$ ,  $I$  by  $\frac{I}{K}$  and  $E$  by  $\frac{E}{K}$  reduces Model (8–10) to

$$(20) \quad \frac{dN}{dt} = b_0 N(1 - N) - \mu N - dI,$$

$$(21) \quad \frac{dE}{dt} = \beta c(N - E - I) \frac{I}{N} - (\mu + k + \alpha)E,$$

$$(22) \quad \frac{dI}{dt} = kE - (\mu + d + \rho)I.$$

The dynamics of this model are characterized by the following theorem:

**THEOREM 3.** *For System (20–22), if  $\mathcal{R}_0 \leq 1$ , the disease-free equilibrium is globally asymptotically stable; while if  $\mathcal{R}_0 > 1$  and  $\mathcal{R}_2^* > 1$ , there exists a unique endemic equilibrium point where*

$$(23) \quad \mathcal{R}_2^* = \frac{b_0}{\mu + d \frac{k}{\mu + d + \rho + k} \frac{\mathcal{R}_0 - 1}{\mathcal{R}_0}}.$$

**Remark 1.** *In the proof (see the Appendix), we show that the disease not only dies out when the basic reproductive is less than or equal to one, but also that it dies out exponentially fast (see Equation (28) with an exponential rate of decay of the order of  $1 - \mathcal{R}_0$ ). The approach followed in Thieme (1993) can be used to show that the disease-free equilibrium is globally asymptotically stable; however, no result about the rate of convergence can be derived from this approach. The global stability of the trivial equilibrium is also established when  $\mathcal{R}_0 = 1$ .*

In order to show that the endemic equilibrium is globally asymptotically stable, we need to assume that  $b_0 > \mu + 2d\sigma(\sigma + \alpha - \rho - d)$ . This assumption does not conflict with the assumption that  $\mathcal{R}_0 > 1$  since  $\mathcal{R}_0$  does not depend on  $b_0$  at all. We collect the results in our last theorem:

**THEOREM 4.** *For System (20–22), if  $\mathcal{R}_0 > 1$  and  $\mathcal{R}_2^* > 1$ , then the endemic equilibrium is globally asymptotically stable, provided that  $\alpha > d$  and  $b_0 > \mu + 2d\sigma(\sigma + \alpha - \rho - d)$ .*

**5. Discussion and conclusions.** Slow progression rates from the latently-infected to the infectious stage are characteristic of tuberculosis, a disease with slow dynamics. A growing infected population may go hand in hand with a decreasing infected fraction (prevalence) whenever the population growth rate is greater than that of its infected subpopulation. In this last case, TB is not being eradicated but “diluted” by the populations fast demographic growth. When the total population is bounded, our model predicts stable levels for the infected populations (given that  $\mathcal{R}_0 > 1$ ). These constant levels are reached in a (quasi) monotonous way. This qualitative prediction is corroborated by epidemiological records. Damped oscillations or limit cycles are not found on TB data. However, there is still a need for more detailed models as shifts in epidemiological parameter values or the emergence of new diseases (like AIDS) can change, at least temporarily, the transmission dynamics of TB. Landscape changes may produce (sometimes dramatic) changes on the quantitative features of TB dynamics (Aparicio *et al.*, 2001a).

Reliable records associated with tuberculosis mortality go back two hundred years in many developed countries. From these records one can see that TB was not able to generate negative population growth rates in spite of the fact that tuberculosis was, in many places, the main cause of death. Its “limited” demographic impact may have been, in part, the result of (relatively) slow progression rates from the latent to the active (and often fatal) state. This is not surprising. The world population has experienced continuous steady (in most places) growth despite the impact of fatal diseases like tuberculosis and wars (Cohen, 1995). Most population growth patterns in the past have been quasi-exponential despite disease, famine, and wars. Hence, our analysis of the impact of TB on populations exhibiting exponential or quasi-exponential (logistic) growth covers most observed population growth patterns. Our results show that TB generates long-term and often short-term “boring” disease patterns. Furthermore, population growth combined with strong declines on TB progression rates (Aparicio *et al.*, 2001a) can explain the (often dramatic) quantitative changes observed on TB dynamics. Changes have had no impact on TB’s long-term qualitative features but strong impact on its quantitative dynamics. The study of the evolutionary dynamics of slow progressing diseases like TB must therefore include demography and more. Host heterogeneity, geography and social structure are some of the critical factors needed in the study of the evolution of slow diseases like TB. We hope to incorporate some form of host heterogeneity in order to take on some of these challenges.

**Acknowledgments.** *This work was partially supported by NSF and NSA grants to the Mathematical and Theoretical Biology Institute at Cor-*

nell University and the office of the Provost of Cornell University. JPA acknowledge support from CONICET Argentina.

## APPENDIX

### A. Appendix.

#### A.1. Proof of Theorem 1.

*Proof.* The proof is divided into three parts. First, we prove that the trivial equilibrium  $A_0(0, 0)$  of System (17–18) is globally asymptotically stable if  $\mathcal{R}_1 < 1$ . Then, it is proved that if  $\mathcal{R}_1 > 1$   $A_0(0, 0)$  is unstable and a unique positive equilibrium is born. Finally, we show that this positive equilibrium is globally asymptotically stable whenever it exists.

Part 1. If  $\mathcal{R}_1 < 1$ ,  $A_0(0, 0)$  is the trivial equilibrium of System (17–18) and it is locally asymptotically stable. To show that  $A_0$  is the unique positive equilibrium on  $\Omega_1$  we proceed as follows:  $\Omega_1$  is a triangle surrounded by  $x = 0$ ,  $y = \frac{n_r}{m_r+n_r}x$  and  $y = \frac{kd}{(m_r+n_r)^2} + \frac{n_r-k}{m_r+n_r}x$ . The equilibria of System (17–18) live at the intersections of the straight line  $a_1 + a_2y + a_3x = 0$  and the parabola  $y = x - x^2$ . After some tedious algebraic calculations, we find out that this straight line is outside  $\Omega_1$ , whenever  $\mathcal{R}_1 < 1$ ; that is, the trivial equilibrium is unique, whenever  $\mathcal{R}_1 < 1$ . Because  $A_0(0, 0)$  is located on the boundary of the positive invariant subset  $\Omega_1$ , there is no closed orbit around it. Thus,  $A_0(0, 0)$  is globally asymptotically stable.

Part 2.  $\mathcal{R}_1 > 1$  implies  $|J_{A_0}| < 0$  and thus  $A_0$  is a saddle. Let  $A_1(x^*, y^*)$  be an equilibrium of (17–18) in  $\Omega_1$ .  $x^*$  of (17–18) must be a positive solution of the quadratic equation  $f(x) = x - x^2 + \frac{a_3}{a_2}x + \frac{a_1}{a_2} = 0$ . If we let  $x_{2,4} = \frac{m_r n_r (1 - \mathcal{R}_1) d}{(m_r + n_r)(d n_r - k \sigma)}$  and  $x_{3,2} = \frac{d}{m_r + n_r}$  then  $f(x) = 0$  will have a unique root in the interval  $[0, \frac{d}{m_r + n_r}]$ . In fact, if  $\mathcal{R}_1 > 1$ , then  $0 < x_{2,4} < x_{3,2}$ , and  $\sigma > r + \rho + d > d$ . Hence,

$$f(0) = \frac{a_1}{a_2} = \frac{m_r n_r (\mathcal{R}_1 - 1) d}{(m_r + n_r)^2 (d - \sigma)} < 0,$$

$$f(x_{2,4}) = x_{2,4} \left( \frac{m_r}{m_r + n_r} - x_{2,4} \right) > x_{2,4} \left( \frac{m_r}{m_r + n_r} - x_{3,2} \right) = x_{2,4} \frac{r + \rho}{m_r + n_r} > 0,$$

and  $f(+\infty) = -\infty$ . Therefore, one solution of  $f(x) = 0$  is in  $[0, \frac{d}{m_r + n_r}]$ , and the other is in  $[\frac{d}{m_r + n_r}, +\infty)$ , located outside of  $\Omega_1$ . Explicitly,  $A_1(x^*, y^*) = \left( \frac{a_2 + a_3 + \Delta^{\frac{1}{2}}}{2a_2}, x^*(1 - x^*) \right)$  where  $\Delta = (a_2 + a_3)^2 + 4a_1a_2 > 0$ .

Part 3. We prove  $A_1(x^*, y^*)$  is globally asymptotically stable.

$$J_{A_1} = \begin{pmatrix} 2x^* - 1 & 1 \\ a_3 x^* & a_2 x^* \end{pmatrix}$$

is the Jacobian matrix of System (17–18) evaluated at  $A_1(x^*, y^*)$ .  $|J_{A_1}| = x^*(2a_2x^* - a_2 - a_3) = x^*\Delta^{\frac{1}{2}} > 0$  and  $-tr(J_{A_1}) = 1 - 2x^* - a_2x^* =$

$\frac{a_3 + \Delta^{\frac{1}{2}}}{-a_2} - a_2 x^* > 0$ , where the last inequality was derived from that fact that  $a_2 < 0$  (which is implied by  $\mathcal{R}_1 > 1$ ).  $A_1$  is thus a locally asymptotically stable equilibrium. Moreover it is a node, since

$$[-\text{tr}(J_{A_1})]^2 - 4|J_{A_1}| = [1 - (2 + a_2)x^*]^2 - 4x^*\Delta^{\frac{1}{2}} = \left(\frac{\Delta^{\frac{1}{2}}}{a_2} - a_2 x^*\right)^2 \geq 0.$$

Consequently, System (17–18) is a quadratic differential system of type two. It follows from Ye (1986) that there are no closed orbits around  $A_1$ . Hence,  $A_1$  is globally asymptotically stable.  $\square$

### A.2. Proof of Theorem 2.

*Proof.*  $\Omega_0 = \{(N, E, I) \in R_3^+ | E + I \leq N\}$  is a positive invariant subset of System (8–10). Part (a) of Theorem 2 is a particular case of Theorem 1 since  $\mathcal{R}_1 < \mathcal{R}_0$ . Thus we only need to prove part (b) and part (c). If  $\mathcal{R}_1 < 1$  (from Theorem 1) then  $\lim_{t \rightarrow \infty} \frac{I}{N} = 0$  and  $\lim_{t \rightarrow \infty} \frac{E}{N} = 0$ . Equation (8) can be rewritten as

$$\frac{1}{N} \frac{dN}{dt} = (b_0 - \mu) - d \frac{I}{N},$$

from which it follows that  $\lim_{t \rightarrow \infty} N(t) = +\infty$ . Hence, we only need to show that

$$\lim_{t \rightarrow \infty} I(t) = \lim_{t \rightarrow \infty} E(t) = +\infty.$$

Solving Equation (9) formally one sees that

$$E(t) = \frac{E_0 + \int_0^t (\sigma I(1 - \frac{I}{N}) \exp(n_\mu \zeta + \int_0^\zeta \sigma \frac{I}{N} ds)) d\zeta}{\exp(n_\mu t + \int_0^t \sigma \frac{I}{N} ds)}.$$

L'Hospital's rule (limit inferior) gives

$$(24) \quad E_\infty = \liminf_{t \rightarrow \infty} E(t) \geq \liminf_{t \rightarrow \infty} \left( \frac{\sigma}{n_\mu + \sigma \frac{I}{N}} I(1 + \frac{I}{N}) \right) = \sigma \frac{I_\infty}{n_\mu}.$$

Similarly, one establishes that

$$(25) \quad I_\infty \geq \frac{k E_\infty}{m_\mu}.$$

Combining (24) and (25) yields

$$(26) \quad \frac{\sigma I_\infty}{n_\mu} \leq E_\infty \leq \frac{m_\mu I_\infty}{k}.$$

From (26) it follows that  $I_\infty = 0$  if and only if  $E_\infty = 0$ . It also follows that  $I_\infty = +\infty$  if and only if  $E_\infty = +\infty$ . To show  $I_\infty > 0$  we analyze the trajectories of System (8–10). Evaluating  $\frac{dE}{dt}$  along  $kE - m_\mu I = 0$  gives

$$\frac{dE}{dt} \Big|_{kE - m_\mu I = 0} = I \left( \sigma - \frac{n_\mu m_\mu}{k} - \sigma \left( \frac{I}{N} + \frac{E}{N} \right) \right).$$

If a trajectory  $(N(t), E(t), I(t))$  intersects the plane  $kE - m_\mu I = 0$  (when time  $t$  is sufficiently large) then  $\frac{dE}{dt} \Big|_{kE - m_\mu I = 0} > 0$ . This last remark is true because  $\lim_{t \rightarrow \infty} \frac{I}{N} = 0$ ,  $\lim_{t \rightarrow \infty} \frac{E}{N} = 0$  and  $\mathcal{R}_0 > 1$ . That is, whenever  $t \gg 1$ , the trajectories cannot leave the set  $\Omega_0 - \bar{\Omega}_0$ , where  $\bar{\Omega}_0 = \{(N, E, I) \in R_3^+ | kE > m_\mu I\} \cap \Omega_0$ , that is, they must remain either in  $\bar{\Omega}_0$  or in  $\Omega_0 - \bar{\Omega}_0$ . If the former is true then  $\frac{dI}{dt} > 0$  which gives  $I_\infty > 0$ . If the latter is true then

$$\frac{dE}{dt} \geq I \left[ \left( \sigma - \frac{m_\mu n_\mu}{k} \right) - \sigma \left( \frac{I}{N} + \frac{E}{N} \right) \right] \geq 0$$

which yields  $E_\infty > 0$  and  $I_\infty > 0$ . However, according to (26), if  $I_\infty < +\infty$  then

$$0 < I_\infty \frac{(k\sigma - m_\mu n_\mu)}{kn_\mu} \leq 0,$$

which contradicts the fact that  $I_\infty = E_\infty = +\infty$ .

The proof of (c) is shorter. According to Theorem 1,  $\mathcal{R}_1 > 1$  leads to  $\lim_{t \rightarrow \infty} \frac{I}{N} = u^*$ . The limiting equation of Equation (8) is

$$(27) \quad \frac{dN}{dt} = Ndu^*(\mathcal{R}_2 - 1).$$

It follows from the theory of limiting equations (Thieme, 1992 and 1994; Thieme and Castillo-Chavez, 1995) that  $N(t)$  is asymptotically equal to  $e^{du^*(\mathcal{R}_2 - 1)t}$  by which (c) is established and the proof is complete.  $\square$

### A.3. Proof of Theorem 3.

*Proof.* The disease-free equilibrium is  $(\frac{b_0 - \mu}{b_0}, 0, 0)$ . It is straightforward to show that the endemic equilibrium is unique whenever  $\mathcal{R}_0 > 1$  and  $\mathcal{R}_2^* > 1$  and the disease-free equilibrium is locally stable whenever  $\mathcal{R}_0 \leq 1$ . Here, we only need to establish the global stability of the disease-free equilibrium under the assumption  $\mathcal{R}_0 \leq 1$ .

Let

$$f(t) = \gamma E(t) + 2\sigma I(t), \quad \text{where } \gamma = \sqrt{(m_\mu - n_\mu)^2 + 4k\sigma} + m_\mu - n_\mu.$$

It suffices to show  $\lim_{t \rightarrow \infty} f(t) = 0$ .

$$\begin{aligned}
\frac{df(t)}{dt} &= \gamma \frac{dE(t)}{dt} + 2\sigma \frac{dI(t)}{dt} \\
&\leq \gamma(\sigma I(t) - n_\mu E(t)) + 2\sigma(kE(t) - m_\mu I(t)) \\
&= (2\sigma k - \gamma n_\mu)E(t) + (\gamma\sigma - 2\sigma m_\mu)I(t) \\
&= (-n_\mu + \frac{2\sigma k}{\gamma})\gamma E(t) + (\frac{\gamma}{2} - m_\mu)2\sigma I(t) \\
&= (\gamma E(t) + 2\sigma I(t)) \left( \frac{\sqrt{(m_\mu - n_\mu)^2 + 4\sigma k} - (m_\mu + n_\mu)}{2} \right) \\
&= -\frac{(1 - \mathcal{R}_0)}{\sqrt{(m_\mu - n_\mu)^2 + 4\sigma k} + m_\mu + n_\mu} f(t).
\end{aligned}$$

This actually produces a differential inequality on the function  $f(t)$ , that is,

$$(28) \quad \frac{df(t)}{dt} < -\frac{(1 - \mathcal{R}_0)}{\sqrt{(m_\mu - n_\mu)^2 + 4\sigma k} + m_\mu + n_\mu} f(t).$$

It follows that  $\lim_{t \rightarrow \infty} f(t) = 0$  from the fact that  $\frac{(1 - \mathcal{R}_0)}{\sqrt{(m_\mu - n_\mu)^2 + 4\sigma k} + m_\mu + n_\mu} > 0$  when  $\mathcal{R}_0 < 1$ .

If  $\mathcal{R}_0 = 1$ ,  $f(t)$  no longer decays exponentially, but it still vanishes as time goes to infinite. We estimate the derivative of  $f(t)$  again.

$$\begin{aligned}
\frac{df(t)}{dt} &= -\sigma\gamma(E + I)\frac{I}{N} \\
&\leq -\sigma\gamma(E + I)I \\
&\leq -\gamma_1 I f(t), \quad \text{where } \gamma_1 = \frac{\min\{2\sigma, \gamma\}}{2} > 0.
\end{aligned}$$

This gives that  $f(t)$  is decreasing and that

$$(29) \quad f(t) \leq f(0)e^{-\gamma_1 \int_0^t I(s)ds}.$$

If  $\liminf_{t \rightarrow \infty} I(t) > 0$  then  $\lim_{t \rightarrow \infty} f(t) = 0$  by (29) which yields  $\lim_{t \rightarrow \infty} I(t) = 0$  from the definition of  $f(t)$ . Hence,  $\liminf_{t \rightarrow \infty} I(t) = 0$ . It follows that  $\liminf_{t \rightarrow \infty} E(t) = 0$  from the fluctuation lemma of Hirsch, Hanisch, and Gabriel (1985) and Proposition 2.2 by Thieme (1993). Consequently,  $\liminf_{t \rightarrow \infty} f(t) = 0$  and  $\lim_{t \rightarrow \infty} f(t) = 0$  because  $f(t)$  is decreasing.  $\square$

**A.4. Proof of Theorem 4.** The proof of the Theorem 4 is a consequence of four lemmas. Introducing a new variable  $Y = E + I$ , we arrive at an equivalent system to (20–22)

$$(30) \quad \frac{dN}{dt} = b_0 N(1 - N) - \mu N - dI,$$

$$(31) \quad \frac{dY}{dt} = \sigma(N - Y) \frac{I}{N} - (\mu + \alpha)Y + (\alpha - \rho - d)I,$$

$$(32) \quad \frac{dI}{dt} = kY - (\mu + d + \rho + k)I.$$

LEMMA 1. Let  $N(t)$ ,  $Y(t)$  and  $I(t)$  be the solution of System (30–32). If  $\alpha \geq d$ , then

$$\liminf_{t \rightarrow \infty} \left( \alpha - \rho - d + \sigma \frac{N(t) - Y(t)}{N(t)} \right) > 0.$$

*Proof.* Directly examine the long-term behavior of the function

$$f(t) = \alpha - \rho - d + \sigma \frac{N(t) - Y(t)}{N(t)} = \alpha - \rho - d + \sigma - \sigma \frac{Y(t)}{N(t)}.$$

By differentiation,

$$\begin{aligned} \frac{df}{dt} &= -\sigma \frac{\frac{dY}{dt}N - \frac{dN}{dt}Y}{N^2} \\ &= -\sigma \left( \frac{\sigma(N - Y)}{N} \frac{I}{N} - (\mu + \alpha) \frac{Y}{N} \right. \\ &\quad \left. + (\alpha - \rho - d) \frac{I}{N} - \frac{Y}{N} \left( b_0(1 - N) - \mu - d \frac{I}{N} \right) \right) \\ &= -\sigma \frac{I}{N} \left( \sigma \frac{(N - Y)}{N} + \alpha - \rho - d \right) + \sigma \frac{Y}{N} \left( \alpha + b_0(1 - N) - d \frac{I}{N} \right) \\ &\geq -\sigma \frac{I}{N} f(t) + \sigma \frac{Y}{N} \left( \alpha - d + b_0(1 - N) \right) \quad \left( \frac{I}{N} \leq 1 \right) \\ &\geq -\sigma \frac{I}{N} f(t), \quad (\alpha > d \text{ and } 1 > N) \end{aligned}$$

from which we obtain

$$f(t) \geq f(t_0) e^{-\sigma \int_{t_0}^t \frac{I(s)}{N(s)} ds} \geq f(t_0) e^{-\sigma(t - t_0)}.$$

Hence,

$$\liminf_{t \rightarrow \infty} f(t) \geq \liminf_{t \rightarrow \infty} \left( f(t_0) e^{-\sigma \int_{t_0}^t \frac{I(s)}{N(s)} ds} \right) = 0.$$

□

**Remark 2.** Using Lemma 1, it can be verified that any trajectory of System (30–32) will eventually enter the region

$$V = \left\{ (N, Y, I) \in R_+^3, 1 \geq N \geq Y \geq I, \alpha - \rho - d + \sigma \geq \sigma \frac{Y}{N} \right\}.$$

In fact,  $V$  is positive invariant on orbits in System (30–32).

LEMMA 2. Under the assumptions of Theorem 4, if  $N(t)$  is a periodic solution, then  $b_0 - \mu - 2b_0N(t) < 0$

*Proof.* From Lemma 1

$$\sigma \frac{Y}{N} \leq \sigma + \alpha - \rho - d\sigma$$

gives

$$I \leq Y \leq \frac{N}{\sigma}(\sigma + \alpha - \rho - d).$$

Consequently,

$$\begin{aligned} \frac{dN}{dt} &\geq b_0N(1 - N) - \mu N - \frac{d}{\sigma}(\sigma + \alpha - \rho - d)N \\ &= (b_0 - \mu - \frac{d}{\sigma}(\sigma + \alpha - \rho - d))N - b_0N^2. \end{aligned}$$

It follows from the comparison principle that

$$N(t) \geq \frac{1}{b_0}(b_0 - \mu - \frac{d}{\sigma}(\sigma + \alpha - \rho - d)) \text{ whenever } t \text{ is large enough,}$$

which holds for all  $t$  because  $N(t)$  is periodic (by assumption). Finally,

$$\begin{aligned} b_0 - \mu - 2b_0N &\leq b_0 - \mu - 2((b_0 - \mu - \frac{d}{\sigma}(\sigma + \alpha - \rho - d))) \\ &= -b_0 + \mu + \frac{2d}{\sigma}(\sigma + \alpha - \rho - d) < 0. \end{aligned}$$

Hence Lemma 2 is true.  $\square$

LEMMA 3. System (30–32) is equivalent to a monotone system in the region  $V$ . The proof of Lemma 3 can be done by choosing

$$\mathcal{E} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and verifying that  $\mathcal{E}^{-1}J\mathcal{E}$  is a non-positive matrix, where  $J$  is the Jacobian matrix of the System (30–32), that is,

$$J = \begin{pmatrix} b_0 - \mu - 2b_0N & 0 & -d \\ \frac{\sigma IY}{N^2} & -(\alpha + \mu + \frac{\sigma I}{N}) & \alpha + \sigma \frac{N-Y}{N} - \rho - d \\ 0 & k & -(\rho + d + \mu + k) \end{pmatrix}.$$

LEMMA 4. Under the assumptions of Theorem 4, any periodic solution of (30–32) is stable, if it exists.

*Proof.*  $J^{[2]}$  is the compound matrix of  $J$ . By the Theorem 4.2 in Muldowney (1990), it is enough to show the linear system

$$(33) \quad \frac{dX}{dt} = J^{[2]}X$$

with

$$J^{[2]} = \begin{pmatrix} b_0 - \mu - 2b_0N - \alpha - \sigma \frac{I}{N} & \alpha + \sigma \frac{I}{N} & d \\ k & b_0 - 2\mu - 2N - d - \rho - k & 0 \\ 0 & \sigma \frac{IY}{N^2} & -(m_\mu + n_\mu + \sigma \frac{I}{N}) \end{pmatrix}$$

is asymptotically stable. To establish this result, we write (33) explicitly, that is,

$$\begin{aligned} \frac{dx_1}{dt} &= \left( b_0 - \mu - 2b_0N(t) - \alpha - \sigma \frac{I(t)}{N(t)} \right) x_1 + \left( \alpha + \sigma \frac{N(t) - Y(t)}{N(t)} \right) x_2 + dx_3, \\ \frac{dx_2}{dt} &= kx_1 + \left( b_0 - \mu - 2b_0N(t) - d - \rho - k \right) x_2, \\ \frac{dx_3}{dt} &= \sigma \frac{I(t)Y(t)}{N(t)^2} x_2 - \left( m_\mu + n_\mu + \sigma \frac{I(t)}{N(t)} \right) x_3. \end{aligned}$$

Define  $D_+x(t)$  to be the right upper derivative of  $x(t)$  with respect to  $t$ , i.e.,

$$D_+x(t) = \limsup_{\Delta t \rightarrow 0^+} \frac{x(t + \Delta t) - x(t)}{\Delta t}.$$

Applying the operator  $D_+$  to each  $|x_i(t)|$ , we derive the following inequalities:

$$\begin{aligned} D_+|x_1(t)| &\leq \left( b_0 - \mu - 2b_0N(t) - \alpha - \sigma \frac{I(t)}{N(t)} \right) |x_1(t)| \\ &\quad + \left( \alpha + \sigma \frac{N(t) - Y(t)}{N(t)} \right) |x_2(t)| + d|x_3(t)|, \\ D_+|x_2(t)| &\leq k|x_1(t)| + \left( b_0 - \mu - 2b_0N(t) - d - \rho - k \right) |x_2(t)|, \\ D_+|x_3(t)| &\leq \sigma \frac{I(t)Y(t)}{N(t)^2} |x_2(t)| - \left( m_\mu + n_\mu + \sigma \frac{I(t)}{N(t)} \right) |x_3(t)|. \end{aligned}$$

To show that  $\lim_{t \rightarrow \infty} x_i(t) = 0$  holds simultaneously for  $i = 1, 2, 3$ , define

$$Q(t) = \max \left\{ |x_1(t)|, \frac{Y(t)}{I(t)} |x_2(t)|, \frac{N(t)}{I(t)} |x_3(t)| \right\}.$$

At any  $t$ ,  $Q(t)$  takes three possible values,  $|x_1(t)|$ ,  $\frac{Y(t)}{I(t)}|x_2(t)|$ , and  $\frac{N(t)}{I(t)}|x_3(t)|$ . If  $Q(t) = |x_1(t)|$ ,

$$\begin{aligned} D_+Q(t) &= D_+|x_1(t)| \\ &\leq \left( b_0 - \mu - 2b_0N - \alpha - \sigma \frac{I}{N} + (\alpha - \rho - d + \sigma \frac{N-Y}{N}) \frac{I}{Y} + d \frac{I}{N} \right) |x_1(t)| \\ &\leq Q \left( b_0 - \mu - 2b_0N - d \frac{I}{N} + (\alpha - \rho - d + \sigma \frac{N-Y}{N}) \frac{I}{Y} - (\mu + \alpha + d \frac{I}{N}) \right) \\ &= Q \left( \frac{dY}{dt} \frac{1}{Y} + b_0 - \mu - 2b_0N \right) \quad (\mathcal{R}_0 > 1 \text{ implies } \sigma > d) \end{aligned}$$

If  $Q(t) = \frac{Y(t)}{I(t)}|x_2(t)|$ ,

$$\begin{aligned} D_+Q(t) &= |x_2(t)| \frac{\frac{dY}{dt}I - \frac{dI}{dt}Y}{I^2} + \frac{Y(t)}{I(t)} D_+|x_2(t)| \\ &= |x_2(t)| \left( \frac{dY}{dt} \frac{1}{Y} - \frac{dI}{dt} \frac{1}{I} \right) \frac{Y}{I} \\ &\quad + \frac{Y}{I} \left( k|x_1(t)| \frac{Y}{I} + (b_0 - \mu - 2b_0N - (\mu + d + \rho + k)) |x_2(t)| \right) \\ &= \frac{Y}{I} |x_2(t)| \left( \frac{dY}{dt} \frac{1}{Y} - \frac{I}{dt} \frac{1}{I} + (k \frac{Y}{I} + (b_0 - \mu - 2b_0N - (\mu + d + \rho + k)) |x_2(t)|) \right) \\ &= Q \left( \frac{dY}{dt} \frac{1}{Y} - k \frac{Y}{I} + k \frac{Y}{I} + (\mu + d + \rho + k) + b_0 - \mu - 2rN - (\mu + d + \rho + k) \right) \\ &= Q \left( \frac{dY}{dt} \frac{1}{Y} + b_0 - \mu - 2rN \right). \end{aligned}$$

If  $Q(t) = \frac{N(t)}{I(t)}|x_3(t)|$ ,

$$\begin{aligned} D_+Q(t) &= |x_3(t)| \frac{\frac{dN}{dt}I - \frac{dI}{dt}N}{I^2} + \frac{N}{I} D_+|x_3(t)| \\ &= |x_3(t)| \frac{N}{I} \left( \frac{dN}{dt} - \frac{dI}{dt} \frac{1}{I} \right) + \frac{N}{I} \left( \sigma \frac{YI}{N^2} |x_2(t)| - (m_\mu + n_\mu + \sigma \frac{I}{N}) |x_3(t)| \right) \\ &= Q(t) \left( \frac{dN}{dt} - \frac{dI}{dt} \frac{1}{I} - (m_\mu + n_\mu) \right). \end{aligned}$$

Hence, in every case

$$D_+Q(t) \leq H(t)Q(t),$$

where

$$H(t) = \max \left\{ \frac{dY}{dt} \frac{1}{Y} + b_0 - \mu - 2b_0N, \frac{dN}{dt} - \frac{dI}{dt} \frac{1}{I} - (m_\mu + n_\mu) \right\}.$$

Hence

$$(34) \quad Q(t) \leq Q(t_0) e^{\int_{t_0}^t H(s) ds}.$$

Let  $\tau$  be the period of  $N(t)$ ,  $Y(t)$  and  $I(t)$ . Then for any integer  $n$

$$\int_{t_0}^{t_0+n\tau} \frac{dN}{ds} \frac{1}{N} ds = \int_{t_0}^{t_0+n\tau} \frac{dY}{ds} \frac{1}{Y} ds = \int_{t_0}^{t_0+n\tau} \frac{dI}{ds} \frac{1}{I} ds = 0.$$

This follows that

$$\begin{aligned} \int_{t_0}^t H(s) ds &= \int_{t_0}^{t_0+n\tau} H(s) ds + \int_{t_0+n\tau}^t H(s) ds \quad \left( n = \left[ \frac{t - t_0}{\tau} \right] \right) \\ &\leq \max \left\{ \int_{t_0}^{t_0+n\tau} \left( \frac{dY}{ds} \frac{1}{Y} + b_0 - \mu - 2b_0 N \right) ds, \right. \\ &\quad \left. \int_{t_0}^{t_0+n\tau} \left( \frac{dN}{ds} - \frac{dI}{ds} \frac{1}{I} - (m_\mu + n_\mu) \right) ds \right\} + \epsilon \\ &\quad \left( \int_{t_0+n\tau}^t H(s) ds < \epsilon \text{ for some } \epsilon > 0, \text{ because } H(t) \text{ is bounded} \right) \\ &= \max \left\{ \int_{t_0}^{t_0+n\tau} (b_0 - \mu - 2b_0 N) ds, -(m_\mu + n_\mu) n\tau \right\} + \epsilon \\ &\leq -\gamma_0 n\tau. \quad (\text{for some } \gamma_0 > 0) \end{aligned}$$

The last inequality is derived from Lemma 2. It follows from (34) that  $\lim_{t \rightarrow \infty} Q(t) = 0$ . This establishes Lemma 4.  $\square$

**Remark 3.** From Lemma 3 and Lemma 4, Theorem 4 follows using the strong Poincaré-Bendixson Theorem (see, for example, Smith, 1995).

## REFERENCES

- [1] ANDERSON, R.M., MAY, R.M., AND MCLEAN, A.R. (1988). Possible demographic impact of AIDS in developing countries, *Nature*, **332**:228–234.
- [2] APARICIO, J.P., CAPURRO, A.F., AND CASTILLO-CHAVEZ, C. (2000). Transmission and dynamics of tuberculosis on generalized households, *J. Theor. Biol.*, **206**: 327–341.
- [3] APARICIO, J.P., CAPURRO, A.F., AND CASTILLO-CHAVEZ, C. (2001a). Long-term dynamics and re-emergence of tuberculosis. (This volume.)
- [4] APARICIO, J.P., CAPURRO, A.F., AND CASTILLO-CHAVEZ, C. (2001b). Frequency dependent risk of infection and the spread of infectious diseases. (This volume.)
- [5] BLOWER, S.M., MCLEAN, A.R., PORCO, T., SANCHEZ, M., SMALL, P.M., HOPEWELL, P., AND MOSS, A. (1995). The intrinsic transmission dynamics of tuberculosis epidemics. *Nat. Med.*, **1**:815–821.
- [6] BLOWER, S.M., SMALL, P.M., AND HOPEWELL, P. (1996). Control Strategies for tuberculosis Epidemics: New models for old problems, *Science*, **272**:497–500.
- [7] BLOWER, S.M. AND GERBERDING, J.L. (1998). Understanding, predicting and controlling the emergence of drug-resistant tuberculosis: a theoretical framework, *J. Mol. Med.*, **76**:624–636.

- [8] BRAUER, F. (1989). Epidemic models in populations of varying size. In *Mathematical Approaches to Problems in Resource Management and Epidemiology*, Castillo-Chavez, C., Levin, S.A., and Shoemaker, C.A. (Eds.) Lecture Notes in Biomathematics, **81**, Springer-Verlag, Berlin-Heidelberg, New York, London, Paris, Tokyo, Hong Kong. pp. 109–123.
- [9] BRAUER, F. AND CASTILLO-CHAVEZ, C. (2001). *Mathematical Models in Population Biology and Epidemiology*, Springer-Verlag, New York, Berlin, Heidelberg.
- [10] BUSENBERG, S. AND HADELER, K.P. (1990). Demography and epidemics, *Math. Biosci.*, **101**:41–62.
- [11] BUSENBERG, S. AND VAN DEN DRIESSCHE, P. (1990). Nonexistence of Periodic Solutions for a Class of Epidemiological Models. In *Differential Equations Models in Biology, Epidemiology and Ecology*, Busenberg S. and Martelli M.(eds) , Lecture Notes in Biology, **92**, Springer-Verlag, Berlin-Heidelberg-New York, pp. 71–79.
- [12] CASTILLO-CHAVEZ, C. AND FENG, Z. (1997). To treat or not to treat: the case of tuberculosis, *J. Math. Biol.*, **35**: 629–656.
- [13] CASTILLO-CHAVEZ, C. AND FENG, Z. (1998). Global stability of an age-structure model for TB and its application to optimal vaccination strategies, *Math. Biosci.*, **151**:135–154
- [14] COHEN, J.E. (1995). How Many People Can the Earth Support? W.W. Norton and Company, New York, London.
- [15] FENG, Z., CASTILLO-CHAVEZ, C., AND CAPURRO, A.F. (2000). A model for tuberculosis with exogenous reinfection, *Theoretical Population Biology*, **57**:235–247.
- [16] FENG, Z., CASTILLO-CHAVEZ, C., AND HUANG, W. (2001). On the role of variable latent period in mathematical models for tuberculosis, *Journal of Dynamics and Differential Equations* (in process), Vol. **13**.
- [17] HADELER, K.P. AND NGOMA, K. (1990). Homogeneous models for sexually transmitted diseases, *Rocky Mountain J. Math.*, **20**:967–986.
- [18] HADELER, K.P. (1992). Periodic solutions of homogeneous equations, *J. Differential Equations*, **95**:183–202.
- [19] HIRSCH, W.M., HANISCH, H., AND GABRIEL, J.P. (1985). Differential equation modes for some parasitic infections; methods for the study of asymptotic behavior, *Comm. Pure Appl. Math.*, **38**:733–753.
- [20] IANNELLI, M., MILLER, F., AND PUGLIESE, A. (1992). Analytical and numerical results for the age-structured S-I-S epidemic model with mixed inter-intracohort transmission. *SIAM J. Appl. Math.*, **23**:662–688.
- [21] LEVIN, S.A. AND PIMENTEL, D. (1981). Selection of intermediate rates of increase in parasite-host systems. *Am. Nat.*, **117**:308–315.
- [22] LIN, X., HETHCOTE, W., AND VAN DEN DRIESSCHE P. (1993). An Epidemiological Models for HIV/AIDS with Proportional Recruitment, *Math. Biosci.*, **118**:181–195.
- [23] MAY, R.M. AND ANDERSON, R.M. (1985). Endemic infections in growing populations, *Math. Biosci.*, **77**:141–156.
- [24] MAY, R.M., ANDERSON, R.M., AND MCLEAN, A.R. (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*, Castillo-Chavez, C., Levin, S.A., and Shoemaker, C.A. (Eds.) Lecture Notes in Biomathematics, **81**, Springer-Verlag, Berlin-Heidelberg, New York, London, Paris, Tokyo, Hong Kong. 220–248.
- [25] MULDOWNEY, J.S. (1990). Compound matrix and ordinary differential equations. *Rocky Mountain J. Math.*, **20**:857–872.
- [26] SMITH, H.L. (1995). *Monotone Dynamical System*: an introduction to theory of competitive and cooperative systems. AMS Mathematical survey and monographs, **41**.
- [27] STYBLO, K. (1991). Selected Papers: *Epidemiology of tuberculosis*. Royal Netherlands Tuberculosis Association, **24**. The Hague, The Netherlands.

- [28] THIEME, R.H. (1992). Convergence results and a Poincaré-Bendixson trichotomy for asymptotically autonomous differential equation, *J. Math. Biol.*, **30**: 755–763.
- [29] THIEME, R.H. (1992). Epidemic and demographic interaction in the spread of potentially fatal disease in growing populations, *Math. Biosci.*, **111**:99–130.
- [30] THIEME, H.R. AND CASTILLO-CHAVEZ, C. (1993). How may infection-age-dependent infectivity affect the dynamics of HIV/AIDS? *SIAM J. Appl. Math.*, **53**(5):1447–1479.
- [31] THIEME, R.H. (1993). Persistence under relaxed point-dissipativity (with applications to an endemic models). *SIAM J. Math. Anal.*, **24**:407–435.
- [32] THIEME, R.H. (1994). Asymptotically autonomous differential equations in the plane, *Rocky Mt. J. Math.*, **24**:351–380.
- [33] THIEME, R.H. AND CASTILLO-CHAVEZ, C. (1995). Asymptotically Autonomous Epidemic Models, In *Mathematical Populations Dynamics: Analysis of Heterogeneity* (O. Arino, D.E. Axelrod, and M. Kimmel, eds.), pp. 33–50.
- [34] VYNNYCKY, E. AND FINE, P.E.M. (1997). The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. *Epidemiol. Infect.*, **119**:183–201.
- [35] YE, YAN-QIAN (eds) (1986). *Theory of limit cycles*, Translations of mathematical monographs By American Mathematical Society, **66**:245–260.

# GLOBAL STABILITY IN SOME SEIR EPIDEMIC MODELS

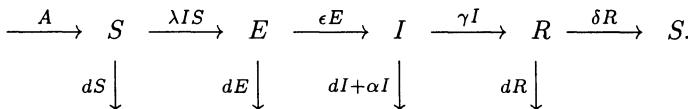
MICHAEL Y. LI\* AND LIANCHENG WANG†

**Abstract.** The dynamics of many epidemic models for infectious diseases that spread in a single host population demonstrate a threshold phenomenon. If the basic reproduction number  $R_0$  is below unity, the disease-free equilibrium  $P_0$  is globally stable in the feasible region and the disease always dies out. If  $R_0 > 1$ , a unique endemic equilibrium  $P^*$  is globally asymptotically stable in the interior of the feasible region and the disease will persist at the endemic equilibrium if it is initially present. In this paper, this threshold phenomenon is established for two epidemic models of SEIR type using two recent approaches to the global-stability problem.

**Key words.** Epidemic models, endemic equilibrium, latent period, global stability, compound matrices.

**AMS(MOS) subject classifications.** Primary 92D30.

**1. Introduction.** Epidemic models study the transmission dynamics of infectious diseases in host populations. In this paper, we deal with diseases that spread in a single host population through direct contact among hosts. Typically after the initial infection, a host stays in a latent period before becoming infectious. At the infectious stage a host may die from the disease or may recover with acquired immunity. The population can be partitioned into four compartments: susceptible, latent or exposed, infectious, and recovered, with sizes denoted by  $S$ ,  $E$ ,  $I$ , and  $R$ , respectively. The total population  $N = S + E + I + R$ . The dynamical transfer of hosts among compartments can be demonstrated in a diagram



The resulting model is of SEIRS type. The term  $A$  denotes the influx or recruitment of susceptibles. The constants  $d$  and  $\alpha$  denote the rates of natural and disease-caused death, respectively. The parameters  $\epsilon$ ,  $\gamma$ , and  $\delta$  denote the transfer rates between the corresponding compartments. Heuristically,  $1/\epsilon$  can be regarded as the mean latent period,  $1/\gamma$  the mean infectious period, and  $1/\delta$  the mean immune period. In the special case when  $\delta = 0$ , the immunity is permanent and there is no return from the  $R$  class to the  $S$  class, the resulting model is an SEIR model. Other special cases include SIRS ( $\epsilon \rightarrow \infty$ ) and SEIS ( $\delta \rightarrow \infty$ ) models, see [2, 3, 8, 12, 29, 32, 39]

\*Department of Mathematics and Statistics, Mississippi State University, Mississippi State, MS 39762.

†Current address: Department of Mathematics and Computer Science, Georgia Southern University, Statesboro, GA 30460-8093.

for detailed discussions. The bilinear incidence form  $\lambda IS$  assumes that the disease incidence is in proportion to the sizes of  $S$  and  $I$  classes ([1, 2]). Other incidence forms include the proportionate mixing incidence  $\beta IS/N$  ([2, 6]), nonlinear incidence  $\lambda I^p S^q$  ([16, 29]), and saturation incidences  $\lambda IS/(1 + aI)$  ([4, 16]) or  $\lambda IS/(1 + aS)$  ([2, 31]).

Using the transfer diagram, the following system of differential equations can be derived

$$(1.1) \quad \begin{aligned} S' &= A - dS - \lambda IS + \delta R \\ E' &= \lambda IS - (\epsilon + d)E \\ I' &= \epsilon E - (\gamma + \alpha + d)I \\ R' &= \gamma I - (\delta + d)R. \end{aligned}$$

From (1.1) and  $N(t) = S(t) + E(t) + I(t) + R(t)$  we have

$$(1.2) \quad N' = A - dN - \alpha I.$$

If  $A$  is a constant, the feasible region for (1.1) is

$$(1.3) \quad \Gamma = \{(S, E, I, R) \in \mathbf{R}_+^4 : S + E + I + R \leq A/d\},$$

since by (1.2),  $\limsup_{t \rightarrow \infty} N(t) \leq A/d$ , and thus the global attractor of (1.1) is contained in  $\Gamma$ . The dynamical behavior of (1.1) in  $\Gamma$  and the fate of the disease is determined by the basic reproduction number

$$(1.4) \quad R_0 = \frac{A\lambda\epsilon}{d(d + \epsilon)(\gamma + d + \alpha)}.$$

If  $R_0 \leq 1$ , (1.1) has only the disease-free equilibrium  $P_0 = (A/d, 0, 0, 0)$  and  $P_0$  is globally asymptotically stable in  $\Gamma$ . If  $R_0 > 1$ ,  $P_0$  becomes unstable and endemic equilibria  $P^* = (S^*, E^*, I^*, R^*)$  exist in  $\overset{\circ}{\Gamma}$ , the interior of  $\Gamma$ . Typically, if the endemic equilibrium is unique, it is globally asymptotically stable in  $\overset{\circ}{\Gamma}$ . Because of the high dimensionality of (1.1), it is highly non-trivial to prove the global stability of the unique endemic equilibrium  $P^*$ , and establish rigorously the threshold phenomena. Some earlier work deal with models that can be reduced to a 2-dimensional system. The global stability of  $P^*$  is proved using the classical Poincaré-Bendixson Theorem and periodic solutions are ruled out using Bendixson-Dulac conditions or a condition of Busenberg and van den Driessche [5], see [12, 14, 15, 32] for surveys of these results. In some recent work ([22, 25]), for SEIR models that can be reduced to a 3-dimensional monotone system, the global stability of  $P^*$  is proved using a Poincaré-Bendixson property due to Hirsch [17] and Smith [36], and periodic solutions are ruled out using a stability criterion of Muldowney [33] for periodic solutions in higher dimensions. This method is also used in a Dengue fever model by Esteva and Vargas [10]. In [26], the global stability of  $P^*$  is resolved for  $\delta$  small or  $\delta$  large in an SEIRS model with a constant population, in which the monotonicity is not

present. The proof uses a geometric approach to global-stability problems developed in Li and Muldowney [23] and Li [21], also see Smith [38] and Leonov *et al* [20]. This approach is also used in [28] to resolve the global stability of  $P^*$  for an SEIR model with vertical transmission.

In the present paper, we apply these two aforementioned approaches to prove the global stability of  $P^*$  in two SEIR models. The first model is (1.1) with  $\delta = 0$  and a constant recruitment  $A$ . The global stability of a unique  $P^*$  is proved using monotonicity and Muldowney's stability criterion. The second model is an SEIR model with a constant total population and  $A = bN$ . It also incorporates vertical transmission and vaccination. The proof of global stability of  $P^*$  in the second model uses the geometric approach of Li and Muldowney. These global-stability results have not been obtained before. Our main purpose is to demonstrate these new methods. We do not strive for the most generality in modeling considerations. The geometric approach of Li and Muldowney may be applied as in [26] to resolve the global stability of  $P^*$  when the immune period is sufficiently long ( $\delta$  small) or sufficiently short ( $\delta$  large).

Using a Lyapunov function, Mena-Lorca and Hethcote [32] prove the global stability of a unique endemic equilibrium for SIRS models with a varying size and constant recruitment when the disease does not cause fatality. Thieme and van den Driessche [40] consider SIRS models where the recovered class  $R$  has distributed stages. A unique endemic equilibrium is shown to be globally stable whenever it exists using techniques from differential-integral equations and a Lyapunov function.

Our paper is organized as follows, in the next section, we outline two general mathematical frameworks for resolving global-stability problems. The two SEIR models are analyzed in Sections 3 and 4. The paper ends with a brief discussion in Section 5.

**2. Mathematical frameworks.** In this section we outline two general mathematical frameworks for proving global-stability, and they will be applied to two SEIR models in Sections 3 and 4.

Let  $A$  be a linear operator on  $\mathbf{R}^n$  and also denote its matrix representation with respect to the standard basis of  $\mathbf{R}^n$ . Let  $\wedge^2 \mathbf{R}^n$  denote the exterior product of  $\mathbf{R}^n$ .  $A$  induces canonically a linear operator  $A^{[2]}$  on  $\wedge^2 \mathbf{R}^n$ : for  $u_1, u_2 \in \mathbf{R}^n$ , define

$$A^{[2]}(u_1 \wedge u_2) := A(u_1) \wedge u_2 + u_1 \wedge A(u_2)$$

and extend the definition over  $\wedge^2 \mathbf{R}^n$  by linearity. The matrix representation of  $A^{[2]}$  with respect to the canonical basis in  $\wedge^2 \mathbf{R}^n$  is called the *second additive compound matrix* of  $A$ . This is an  $\binom{n}{2} \times \binom{n}{2}$  matrix and satisfies the property  $(A + B)^{[2]} = A^{[2]} + B^{[2]}$ . In the special case when  $n = 2$ , we have  $A_{2 \times 2}^{[2]} = \text{tr}A$ . In general, each entry of  $A^{[2]}$  is a linear expression of those of  $A$ . For instance, when  $n = 3$ , the second additive compound matrix of  $A = (a_{ij})$  is

$$(2.1) \quad A^{[2]} = \begin{bmatrix} a_{11} + a_{22} & a_{23} & -a_{13} \\ a_{32} & a_{11} + a_{33} & a_{12} \\ -a_{31} & a_{21} & a_{22} + a_{33} \end{bmatrix}.$$

Let  $\sigma(A) = \{\lambda_1, \dots, \lambda_n\}$  be the spectrum of  $A$ . Then,  $\sigma(A^{[2]}) = \{\lambda_i + \lambda_j : 1 \leq i < j \leq n\}$  is the spectrum of  $A^{[2]}$ . For detailed discussions of general compound matrices and their properties we refer the reader to [11, 33]. A comprehensive survey on compound matrices and their relations to differential equations is given in [33].

Let  $x \mapsto f(x) \in \mathbf{R}^n$  be a  $C^1$  function for  $x$  in an open set  $D \subset \mathbf{R}^n$ . Consider the differential equation

$$(2.2) \quad x' = f(x).$$

Denote by  $x(t, x_0)$  the solution to (2.2) such that  $x(0, x_0) = x_0$ . A set  $K$  is said to be *absorbing* in  $D$  for (2.2) if  $x(t, K_1) \subset K$  for each compact  $K_1 \subset D$  and  $t$  sufficiently large. We make the following two basic assumptions:

- ( $H_1$ ) There exists a compact absorbing set  $K \subset D$ .
- ( $H_2$ ) Equation (2.2) has a unique equilibrium  $\bar{x}$  in  $D$ .

The equilibrium  $\bar{x}$  is said to be *globally stable* in  $D$  if it is locally stable and all trajectories in  $D$  converge to  $\bar{x}$ . The assumptions ( $H_1$ ) and ( $H_2$ ) are satisfied if  $\bar{x}$  is globally stable in  $D$ . For epidemic models and many other biological models where the feasible region is a bounded cone, ( $H_1$ ) is equivalent to the uniform persistence of (2.2) (see [7, 41]). The following global-stability problem is formulated in [23].

*Global-Stability Problem.* Under assumptions ( $H_1$ ) and ( $H_2$ ), find conditions on the vector field of (2.2) such that the local stability of  $\bar{x}$  implies its global stability in  $D$ .

When  $n = 2$ , the classical Poincaré-Bendixson theory allows the following two approaches to solve the global-stability problem. Approach I: if all periodic orbits  $\Omega$  of (2.2) in  $D$  can be shown to be orbitally asymptotically stable, using Poincaré's stability condition  $\int_{\Omega} \operatorname{div} f dt < 0$  ([13]) for instance, then the local asymptotic stability of  $\bar{x}$  also implies its global stability in  $D$ . Approach II: if  $D \subset \mathbf{R}^2$  is simply connected and that the Bendixson's criterion  $\operatorname{div} f < 0$  holds in  $D$ , then (2.2) has no nontrivial periodic orbits and  $\bar{x}$  is globally stable in  $D$ . Recent developments in the qualitative theory make it possible to use similar approaches in higher dimensions. We give a brief outline in the following two subsections.

**2.1. Proving global stability using the Poincaré-Bendixson property.** System (2.2) is said to satisfy the *Poincaré-Bendixson Property* if any nonempty compact omega limit set of (2.2) that contains no equilibria is a closed orbit.

Any autonomous system (2.2) in the plane satisfies the Poincaré-Bendixson Property by the classical Poincaré-Bendixson theory ([13]). It

is also known that a three-dimensional competitive system satisfies the Poincaré-Bendixson Property in a convex region, as shown by Hirsch [17] and Smith [36]. See [37] for a general definition of competitive systems.

**THEOREM 2.1.** *Assume that  $n = 3$  and  $D$  is convex. Suppose that (2.2) is competitive in  $D$ . Then it satisfies the Poincaré-Bendixson Property. (cf. [37, Chapter 3, Theorem 4.1].)*

For higher dimensional systems that satisfy the Poincaré-Bendixson Property, we prove the following global stability result.

**THEOREM 2.2.** *Assume that*

- (1) *assumptions  $(H_1)$  and  $(H_2)$  hold;*
- (2)  *$\bar{x}$  is locally asymptotically stable;*
- (3) *system (2.2) satisfies the Poincaré-Bendixson Property;*
- (4) *each periodic orbit of (2.2) in  $D$  is orbitally asymptotically stable.*

*Then the unique equilibrium  $\bar{x}$  is globally asymptotically stable in  $D$ .*

*Proof.* It suffices to show that  $\bar{x}$  attracts all points in  $D$ . Let  $U$  be the basin of attraction of  $\bar{x}$ , the set of all  $x_0$  such that  $x(t, x_0)$  converges to  $\bar{x}$ . Then  $U$  is nonempty and open by the asymptotic stability of  $\bar{x}$ . The theorem is proved if we establish that  $D \subset U$ . Assume the contrary; then the boundary  $\partial U$  of  $U$  has a nonempty intersection  $\mathcal{I}$  with  $D$ . Since both  $U$  and its closure  $\bar{U}$  are invariant and  $U$  is open,  $\partial U = \bar{U} - U$  is also invariant, and thus  $\mathcal{I}$  is positively invariant. Therefore  $\mathcal{I}$  contains a nonempty compact omega limit set  $\Omega$ . By the assumption  $(H_1)$ , we must have  $\Omega \cap \partial D = \emptyset$ . Since it contains no equilibria,  $\Omega$  is a closed orbit by the Poincaré-Bendixson Property, and is asymptotically orbitally stable by the assumption (4) of Theorem 2.2. We thus obtain a contradiction since  $\Omega$  belongs to the alpha limit set of a trajectory in  $U$ . This completes the proof.  $\square$

*Remark.* A similar proof was used in [25] in the context of epidemic models. We reproduce it here in a more general setting for the convenience of the reader.

The assumption (3) is satisfied if  $D$  is a convex region in  $\mathbf{R}^3$  and (2.2) is a competitive system in  $D$ . The orbital stability of periodic solutions in  $\mathbf{R}^n$  ( $n \geq 2$ ) can be verified using the following result of Muldowney [33], which generalizes a 2d condition of Poincaré.

**THEOREM 2.3.** *A periodic orbit  $\Omega = \{p(t) : 0 \leq t < \omega\}$  of (2.2) is orbitally asymptotically stable with asymptotic phase if the linear system*

$$(2.3) \quad z'(t) = \frac{\partial f^{[2]}}{\partial x}(p(t)) z(t)$$

*is asymptotically stable, where  $\frac{\partial f^{[2]}}{\partial x}$  is the second additive compound matrix of the Jacobian matrix  $\frac{\partial f}{\partial x}$  of  $f$ . (cf. [33, Theorem 4.2].)*

A matrix is *stable* if all its eigenvalues have negative real parts. The next result gives a criterion for the stability of matrices. A proof can be found in [27].

**THEOREM 2.4.** *An  $n \times n$  real matrix  $A$  is stable if and only if  $A^{[2]}$  is stable and  $(-1)^n \det(A) > 0$ .*

Using Theorems 2.2-2.4, we prove the following result.

**THEOREM 2.5.** *Assume that*

- (1) *assumptions  $(H_1)$  and  $(H_2)$  hold;*
- (2) *system (2.2) satisfies the Poincaré-Bendixson Property;*
- (3) *For each periodic solution  $x = p(t)$  to (2.2) with  $p(0) \in D$ , system (2.3) is asymptotically stable,*
- (4)  $(-1)^n \det \left( \frac{\partial f}{\partial x}(\bar{x}) \right) > 0$ .

*Then the unique equilibrium  $\bar{x}$  is globally asymptotically stable in  $D$ .*

*Proof.* It suffices to show that  $\bar{x}$  is locally asymptotically stable. Regard the equilibrium solution  $x = \bar{x}$  as a constant periodic solution. Then assumption (3) of the theorem implies that the matrix  $\frac{\partial f}{\partial x}^{[2]}(\bar{x})$  is stable. This and the assumption (4) of the theorem imply that the Jacobian matrix  $\frac{\partial f}{\partial x}(\bar{x})$  is stable, by Theorem 2.4. Therefore,  $\bar{x}$  is locally asymptotically stable, and thus Theorem 2.5 follows from Theorem 2.2.  $\square$

**2.2. Proving global stability using autonomous convergence theorems.** When a Poincaré-Bendixson Property is not known to exist for a system, a type of results known as autonomous convergence theorems (see [24, 38]) can be used to prove global stability.

For  $n \geq 2$ , by a *Bendixson criterion* we mean a condition satisfied by  $f$  which precludes the existence of nonconstant periodic solutions of (2.2). A Bendixson criterion is said to be *robust under  $C^1$  local perturbations of  $f$  at  $x_1 \in D$*  if, for sufficiently small  $\epsilon > 0$  and neighborhood  $U$  of  $x_1$ , it is also satisfied by  $g \in C^1(D \rightarrow \mathbf{R}^n)$  such that the support  $\text{supp}(f - g) \subset U$  and  $|f - g|_{C^1} < \epsilon$ , where

$$|f - g|_{C^1} = \sup \left\{ |f(x) - g(x)| + \left| \frac{\partial f}{\partial x}(x) - \frac{\partial g}{\partial x}(x) \right| : x \in D \right\}.$$

Such  $g$  will be called *local  $\epsilon$ -perturbations of  $f$  at  $x_1$* . It is easy to see that the classical Bendixson's condition  $\text{div } f(x) < 0$  for  $n = 2$  is robust under  $C^1$  local perturbations of  $f$  at each  $x_1 \in \mathbf{R}^2$ . Bendixson criterion for higher dimensional systems that are  $C^1$  robust are discussed in [21, 23, 24, 38].

A point  $x_0 \in D$  is *wandering* for (2.2) if there exists a neighborhood  $U$  of  $x_0$  and  $T > 0$  such that  $U \cap x(t, U)$  is empty for all  $t > T$ . Thus, for example, all equilibria and limit points are nonwandering. The following is a version of the local  $C^1$  Closing Lemma of Pugh (see [34, 35]) as stated in [18].

LEMMA 2.1. Let  $f \in C^1(D \rightarrow \mathbf{R}^n)$ . Suppose that  $x_0$  is a nonwandering point of (2.2) and that  $f(x_0) \neq 0$ . Then, for each neighborhood  $U$  of  $x_0$  and  $\epsilon > 0$ , there exists a  $C^1$  local  $\epsilon$ -perturbation  $g$  of  $f$  at  $x_0$  such that

- (1)  $\text{supp}(f - g) \subset U$ , and
- (2) the perturbed system  $x' = g(x)$  has a nonconstant periodic solution whose trajectory passes through  $x_0$ .

The following global-stability principle is established in Li and Muldowney [23] for autonomous systems in any finite dimension.

THEOREM 2.6. Suppose that assumptions  $(H_1)$  and  $(H_2)$  hold. Assume that (2.2) satisfies a Bendixson criterion that is robust under  $C^1$  local perturbations of  $f$  at all nonequilibrium nonwandering points for (2.2). Then  $\bar{x}$  is globally stable in  $D$  provided it is stable. (cf. [23, Theorem 2.3].)

The main idea of the proof in [23] for Theorem 2.6 is as follows. Suppose that system (2.2) satisfies a Bendixson criterion. Then it does not have any nonconstant periodic solutions. Moreover, the robustness of the Bendixson criterion implies that all nearby differential equations have no nonconstant periodic solutions. Thus by Lemma 2.1, all nonwandering points of (2.2) in  $D$  must be equilibria. In particular, each omega limit point in  $D$  must be an equilibrium. Therefore  $\omega(x_0) = \{\bar{x}\}$  for all  $x_0 \in D$  since  $\bar{x}$  is the only equilibrium in  $D$ .

The following Bendixson criterion is given in [23] and shown to have the robustness required by Theorem 2.6. Let  $x \mapsto P(x)$  be an  $(\binom{n}{2}) \times (\binom{n}{2})$  matrix-valued function that is  $C^1$  for  $x \in D$ . Assume that  $P^{-1}(x)$  exists and is continuous for  $x \in K$ , the compact absorbing set. A quantity  $\bar{q}_2$  is defined as

$$(2.4) \quad \bar{q}_2 = \limsup_{t \rightarrow \infty} \sup_{x_0 \in K} \frac{1}{t} \int_0^t \mu(B(x(s, x_0))) ds$$

where

$$(2.5) \quad B = P_f P^{-1} + P \frac{\partial f^{[2]}}{\partial x} P^{-1},$$

the matrix  $P_f$  is obtained by replacing each entry  $p$  of  $P$  by its derivative in the direction of  $f$ ,  $p_{ij,f}$ , and  $\mu(B)$  is the *Lozinskii measure* of  $B$  with respect to a vector norm  $|\cdot|$  in  $\mathbf{R}^N$ ,  $N = \binom{n}{2}$ , defined by (see [9], p.41)

$$\mu(B) = \lim_{h \rightarrow 0^+} \frac{|I + hB| - 1}{h}.$$

It is easy to see that  $\bar{q}_2$  is well defined. It is shown in [23] that, if  $D$  is simply connected, the condition  $\bar{q}_2 < 0$  rules out the presence of any orbit that gives rise to a simple closed rectifiable curve that is invariant for (2.2), such as periodic orbits, homoclinic orbits, and heteroclinic cycles. Moreover, it is robust under  $C^1$  local perturbations of  $f$  near any nonequilibrium point

that is nonwandering. In particular, the following global-stability result is proved in Li and Muldowney [23].

**THEOREM 2.7.** *Assume that  $D$  is simply connected and that the assumptions  $(H_1)$ ,  $(H_2)$  hold. Then the unique equilibrium  $\bar{x}$  of (2.2) is globally stable in  $D$  if  $\bar{q}_2 < 0$ . (cf. [23, Theorem 3.5].)*

**3. Global dynamics of an SEIR model with constant recruitment.** Let  $\delta = 0$  in (1.1). Then the first three equations in (1.1) contain no  $R$  terms. This allows the reduction of (1.1) to

$$(3.1) \quad \begin{aligned} S' &= A - dS - \lambda IS \\ E' &= \lambda IS - (\epsilon + d)E \\ I' &= \epsilon E - (\gamma + \alpha + d)I, \end{aligned}$$

and  $N$ ,  $R$  can be obtained from  $N' = A - dN - \alpha I$  and  $R = N - S - E - I$ . The feasible region of (3.1) is  $\Delta = \{(S, E, I) \in \mathbf{R}_+^3 : S + E + I \leq A/d\}$ , which is positively invariant for (3.1). Let  $R_0$  be defined as in (1.4). Straightforward calculations show that the disease-free equilibrium  $P_0 = (A/d, 0, 0)$  exists for all values of parameters. It is the only equilibrium in  $\Delta$  if  $R_0 \leq 1$ . If  $R_0 > 1$ , a unique endemic equilibrium  $P^* = (S^*, E^*, I^*) \in \overset{\circ}{\Delta}$  exists with

$$S^* = \frac{A}{dR_0}, \quad E^* = \frac{d}{\lambda} (R_0 - 1), \quad I^* = \frac{\epsilon}{(\gamma + d + \alpha)} E^*.$$

**THEOREM 3.1.** *The disease-free equilibrium  $P_0 = (A/d, 0, 0)$  of (3.1) is globally asymptotically stable in  $\Delta$  if  $R_0 \leq 1$ ; it is unstable if  $R_0 > 1$ , and the solutions of (3.1) starting sufficiently close to  $P_0$  in  $\Delta$  move away from  $P_0$  except that those starting on the invariant  $S$ -axis approach  $P_0$  along this axis.*

*Proof.* Set  $L = \epsilon E + (\epsilon + d)I$ . Then, if  $R_0 \leq 1$ ,

$$L' = I [\lambda \epsilon S - (\epsilon + d)(\gamma + d + \alpha)] = \frac{\lambda \epsilon A I}{d} \left( \frac{d}{A} S - \frac{1}{R_0} \right) \leq 0,$$

and  $L' = 0$  if and only if  $I = 0$ . The largest compact invariant set in  $\{(S, E, I) \in \Delta : L' = 0\}$  is the singleton  $\{P_0\}$ . The global stability of  $P_0$  then follows from LaSalle's Invariance Principle ([19, Chapter 2, Theorem 6.4]). If  $R_0 > 1$ , then  $L' > 0$  for  $S$  sufficiently close to  $A/d$  except when  $E = I = 0$ . Solutions starting sufficiently close to  $P_0$  leave a neighborhood of  $P_0$  except those on the invariant  $S$ -axis, on which (3.1) reduces to  $S' = A - ds$  and thus  $S(t) \rightarrow A/d$ , as  $t \rightarrow \infty$ . This establishes the theorem.  $\square$

Theorem 3.1 completely determines the global dynamics of (3.1) in  $\Delta$  for the case  $R_0 \leq 1$ . Its epidemiological implication is that the infected

population (the sum of the latent and the infectious population) vanish in time so the disease dies out. When  $R_0 > 1$ , the local behavior of (3.1) near  $P_0$  as described in Theorem 3.1 allows us to use a similar argument as in the proof of Proposition 3.3 in [22] and show that system (3.1) is uniformly persistent, namely, there exists constant  $0 < c < 1$  such that any solution  $(S(t), E(t), I(t))$  with  $(S(0), E(0), I(0)) \in \overset{\circ}{\Delta}$  satisfies

$$(3.2) \quad \min \{ \liminf_{t \rightarrow \infty} S(t), \liminf_{t \rightarrow \infty} E(t), \liminf_{t \rightarrow \infty} I(t) \} > c.$$

The boundedness of  $\Delta$  and condition (3.2) imply that (3.1) has a compact absorbing set  $K \subset \overset{\circ}{\Delta}$  (see [7, 41]).

**THEOREM 3.2.** *If  $R_0 > 1$ , then the unique endemic equilibrium  $P^*$  is globally asymptotically stable in  $\overset{\circ}{\Delta}$ .*

*Proof.* By examining the Jacobian matrix of (3.1), it can be verified that (3.1) is competitive in the convex region  $\overset{\circ}{\Delta}$ , with respect to the partial ordering defined by the orthant  $\{(S, E, I) \in \mathbf{R}^3 : S \leq 0, E \geq 0, I \leq 0\}$  (see [37]). By Theorem 2.1, (3.1) satisfies the Poincaré-Bendixson Property, and thus conditions (1) and (2) of Theorem 2.5 hold. The second compound system of (3.1) along a periodic solution  $(S(t), E(t), I(t))$  is

$$(3.3) \quad \begin{aligned} X' &= -(2d + \lambda I + \epsilon) X + \lambda S Y + \lambda S Z \\ Y' &= \epsilon X - (2d + \lambda I + \gamma + \alpha) Y \\ Z' &= \lambda I Y - (2d + \epsilon + \gamma + \alpha) Z. \end{aligned}$$

To show that (3.3) is asymptotically stable, consider a Lyapunov function

$$(3.4) \quad V(X, Y, Z; S, E, I) = \sup \left\{ |X|, \frac{E}{I} (|Y| + |Z|) \right\}.$$

The orbit  $\mathcal{O}$  of the periodic solution  $(S(t), E(t), I(t))$  is at a positive distance from the boundary  $\partial\Delta$  by the uniform persistence. Thus there exists a constant  $c_1 > 0$  such that

$$(3.5) \quad V(X, Y, Z; S, E, I) \geq c_1 \sup\{|X|, |Y|, |Z|\}$$

for all  $(X, Y, Z) \in \mathbf{R}^3$  and  $(S, E, I) \in \mathcal{O}$ . The right derivative of  $V$  along a solution  $(X(t), Y(t), Z(t))$  to (3.3) and  $(S(t), E(t), I(t))$  can be estimated as follows.

$$(3.6) \quad \begin{aligned} D_+|X(t)| &\leq -(2d + \lambda I + \epsilon) |X(t)| + \lambda S (|Y(t)| + |Z(t)|) \\ &= -(2d + \lambda I + \epsilon) |X(t)| + \frac{\lambda I S}{E} E (|Y(t)| + |Z(t)|), \end{aligned}$$

and

$$(3.7) \quad \begin{aligned} D_+|Y(t)| &\leq \epsilon |X(t)| - (2d + \lambda I + \gamma + \alpha) |Y(t)| \\ D_+|Z(t)| &\leq \lambda I |Y(t)| - (2d + \epsilon + \gamma + \alpha) |Z(t)|. \end{aligned}$$

Therefore

$$\begin{aligned} D_+ \frac{E}{I} (|Y(t)| + |Z(t)|) \\ = \left( \frac{E'}{E} - \frac{I'}{I} \right) \frac{E}{I} (|Y(t)| + |Z(t)|) + \frac{E}{I} D_+ (|Y(t)| + |Z(t)|) \\ \leq \frac{\epsilon E}{I} |X(t)| + \left( \frac{E'}{E} - \frac{I'}{I} - 2d - \gamma - \alpha \right) \frac{E}{I} (|Y(t)| + |Z(t)|). \end{aligned}$$

Relations (3.6) and (3.7) lead to

$$(3.8) \quad D_+ V(t) \leq \max\{g_1(t), g_2(t)\} V(t),$$

where

$$(3.9) \quad g_1(t) = -2d - \lambda I - \epsilon + \frac{\lambda I S}{E},$$

$$(3.10) \quad g_2(t) = \frac{E'}{E} - \frac{I'}{I} - 2d - \gamma - \alpha + \frac{\epsilon E}{I}.$$

Rewriting (3.1), we find that

$$(3.11) \quad \frac{\lambda I S}{E} = \frac{E'}{E} + \epsilon + d, \quad \frac{\epsilon E}{I} = \frac{I'}{I} + \gamma + d + \alpha.$$

From (3.9)–(3.11),  $\max\{g_1(t), g_2(t)\} \leq \frac{E'(t)}{E(t)} - d$ , and thus

$$\int_0^\omega \max\{g_1(t), g_2(t)\} dt \leq \log E(t) \Big|_0^\omega - d = -d,$$

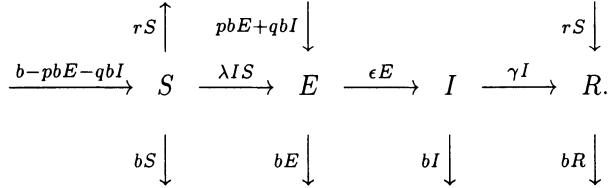
since  $E(t)$  is periodic of minimal period  $\omega$ . This relation and (3.8) imply that  $V(t) \rightarrow 0$  as  $t \rightarrow \infty$ , and in turn that  $(X(t), Y(t), Z(t)) \rightarrow 0$  as  $t \rightarrow \infty$  by (3.5). As a result, the second compound system (3.3) is asymptotically stable if the minimal period  $\omega > 0$ . The same estimates also hold when  $\omega \rightarrow 0$ . This verifies the condition (3) of Theorem 2.5. Let  $J(P^*)$  be the Jacobian matrix of (3.1) at  $P^*$ . Then

$$\begin{aligned} \det(J(P^*)) &= \begin{vmatrix} -\lambda I^* - d & 0 & -\lambda S^* \\ \lambda I^* & -\epsilon - d & \lambda S^* \\ 0 & \epsilon & -\gamma - \alpha - d \end{vmatrix} \\ &= -(\lambda I^* + d)(\epsilon + d)(\gamma + \alpha + d) + \lambda \epsilon S^* d. \end{aligned}$$

Using the equilibrial equations for  $P^* = (S^*, E^*, I^*)$  we can derive that  $\lambda \epsilon S^* = (\epsilon + d)(\gamma + \alpha + d)$ . Therefore,  $\det(J(P^*)) = -\lambda I^*(\epsilon + d)(\gamma + \alpha + d) < 0$ . This verifies the condition (4) of Theorem 2.5. Hence  $P^*$  is globally stable in  $\overset{\circ}{\Delta}$  by Theorem 2.5.  $\square$

#### 4. An SEIR model with vertical transmission and vaccination.

In this section, we consider an SEIR model that has an exponential birth. We assume that the disease spreads through both horizontal and vertical transmission. The transfer diagram for the model is



We assume that the disease is not fatal ( $\alpha = 0$ ) and that the birth and death rates are equal and denoted by  $b$ . Hence the total population is constant;  $N = S + E + I + R = 1$ , and thus the natural birth recruitment  $A = bN = b$ . For the vertical transmission, we assume that a fraction  $p$  and a fraction  $q$  of the offspring from the exposed and the infectious classes, respectively, are born into the exposed class  $E$ . Consequently, the birth flux into the exposed class is given by  $pbE + qbI$  and the birth flux into the susceptible class is given by  $b - pbE - qbI$ . Naturally,  $0 \leq p \leq 1$  and  $0 \leq q \leq 1$ . For the vaccination, we assume that all susceptible individuals are vaccinated at a constant per capita rate  $r$  and that the vaccination has no effect on infected individuals. When  $r = 0$ , no vaccination is considered, and the model reduces to that considered in [28]. The transfer diagram leads to the following system of differential equations

$$\begin{aligned}
 S' &= b - \lambda IS - pbE - qbI - bS - rS \\
 E' &= \lambda IS + pbE + qbI - (\epsilon + b)E \\
 I' &= \epsilon E - (\gamma + b)I \\
 R' &= \gamma I - bR + rS.
 \end{aligned} \tag{4.1}$$

As in Section 3, we study the reduced 3d system

$$\begin{aligned}
 S' &= b - \lambda IS - pbE - qbI - bS - rS \\
 E' &= \lambda IS + pbE + qbI - (\epsilon + b)E \\
 I' &= \epsilon E - (\gamma + b)I
 \end{aligned} \tag{4.2}$$

in its feasible region  $\Sigma = \{(S, E, I) \in \mathbf{R}_+^3 : S + E + I \leq 1\}$ . The dynamics of (4.2) is determined by the following vaccination modified basic reproduction number

$$R_0 = \frac{\lambda\epsilon}{(b + \epsilon)(b + \gamma) - bp(b + \gamma) - bq\epsilon} \frac{b}{b + r}. \tag{4.3}$$

If  $R_0 \leq 1$ , the disease-free equilibrium  $P_0 = (b/(b+r), 0, 0)$  is the only equilibrium. If  $R_0 > 1$ , there is a unique endemic equilibrium  $P^* = (S^*, E^*, I^*)$  with  $S^* = 1/R_0$ . The following result can be proved as Proposition 3.1 by using a global Lyapunov function  $L = \epsilon E + (\epsilon + b - pb)I$ .

**THEOREM 4.1.** (a) If  $R_0 \leq 1$ , then  $P_0$  is the only equilibrium in  $\Sigma$  and it is globally stable in  $\Sigma$ . (b) If  $R_0 > 1$ , then  $P_0$  is unstable and there exists a unique endemic equilibrium  $P^*$ . Furthermore, the system (4.2) is uniformly persistent in  $\Sigma$  if  $R_0 > 1$ .

From Theorem 4.1, we know that system (4.2) satisfies the assumptions  $(H_1)$  and  $(H_2)$ . Using Theorem 2.7, we can prove that  $P^*$  is globally stable in  $\overset{\circ}{\Sigma}$  if  $R_0 > 1$ .

**THEOREM 4.2.** Assume that  $R_0 > 1$ . Then the unique endemic equilibrium  $P^*$  is globally stable in  $\overset{\circ}{\Sigma}$ .

*Proof.* The Jacobian matrix  $J = \frac{\partial f}{\partial x}$  associated with a general solution  $(S(t), E(t), I(t))$  to (4.2) is

$$J = \begin{bmatrix} -\lambda I - b - r & -pb & -\lambda S - qb \\ \lambda I & pb - b - \epsilon & \lambda S + qb \\ 0 & \epsilon & -b - \gamma \end{bmatrix}$$

and its second additive compound matrix  $J^{[2]}$  is, by (2.1),

$$(4.4) \quad \begin{bmatrix} -\lambda I - \epsilon - 2b - r + pb & \lambda S + qb & \lambda S + qb \\ \epsilon & -\lambda I - \gamma - 2b - r & -pb \\ 0 & \lambda I & -\epsilon - \gamma - 2b + pb \end{bmatrix}.$$

Set the function  $P(x) = P(S, E, I)$  in (2.5) as

$$P(S, E, I) = \begin{bmatrix} a_1 & 0 & 0 \\ 0 & (1 - a_2)\frac{E}{I} & 0 \\ 0 & a_2\frac{E}{I} & \frac{E}{I} \end{bmatrix}$$

where  $1 < a_1 < 1 + \lambda c^2/(\lambda + b)$ ,  $c$  is a uniform persistence constant and

$$(4.5) \quad a_2 = \begin{cases} 0, & \text{if } \epsilon \geq pb, \\ 1 - \frac{\epsilon}{pb}, & \text{if } \epsilon < pb. \end{cases}$$

Then  $P_f P^{-1} = \text{diag}(0, E'/E - I'/I, E'/E - I'/I)$ , and the matrix  $B = P_f P^{-1} + P J^{[2]} P^{-1}$  in (2.5) can be written in block form  $B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$  with  $B_{11} = -\lambda I - \epsilon - 2b - r + pb$ ,

$$B_{12} = a_1 \left[ \left( \lambda S + qb \right) \frac{I}{E}, \quad \left( \lambda S + qb \right) \frac{I}{E} \right], \quad B_{21} = \frac{1}{a_1} \left[ \begin{array}{c} (1 - a_2) \frac{\epsilon E}{I} \\ a_2 \frac{\epsilon E}{I} \end{array} \right],$$

and  $B_{22}$  being the following

$$\begin{bmatrix} \frac{E'}{E} - \frac{I'}{I} - \lambda I - \gamma - 2b - r + a_2 pb & -(1 - a_2)pb \\ \lambda I + \frac{a_2 [\epsilon - (1 - a_2)pb]}{1 - a_2} & \frac{E'}{E} - \frac{I'}{I} - \epsilon - \gamma - 2b + (1 - a_2)pb \end{bmatrix},$$

which simplifies to

$$\begin{bmatrix} \frac{E'}{E} - \frac{I'}{I} - \lambda I - \gamma - 2b - r + a_2 pb & -(1 - a_2)pb \\ \lambda I & \frac{E'}{E} - \frac{I'}{I} - \epsilon - \gamma - 2b + (1 - a_2)pb \end{bmatrix},$$

since  $a_2[\epsilon - (1 - a_2)pb] = 0$  by (4.5). Let  $(u, v, w)$  denote the vectors in  $\mathbf{R}^3 \cong \mathbf{R}^{(3)}$ , we select a norm in  $\mathbf{R}^3$  as  $|(u, v, w)| = \max\{|u|, |v| + |w|\}$  and let  $\mu$  denote the Lozinskii measure with respect to this norm. Following the method in [30], we have the estimate  $\mu(B) \leq \sup\{g_1, g_2\}$ , where

$$g_1 = \mu_1(B_{11}) + |B_{12}|, \quad \text{and} \quad g_2 = |B_{21}| + \mu_1(B_{22}),$$

$|B_{12}|, |B_{21}|$  are matrix norms with respect to the  $l_1$  vector norm, and  $\mu_1$  denotes the Lozinskii measure with respect to the  $l_1$  norm, see [9, p.41]. More specifically,  $\mu_1(B_{11}) = -\lambda I - \epsilon - 2b - r + pb$ ,  $|B_{12}| = a_1(\lambda S + qb)I/E$ ,  $|B_{21}| = \epsilon E/(a_1 I)$ . To calculate  $\mu_1(B_{22})$ , add the absolute value of the off-diagonal elements to the diagonal one in each column of  $B_{22}$ , and then take the maximum of two sums, see [9, p.41]. We thus obtain

$$\begin{aligned} \mu_1(B_{22}) &= \frac{E'}{E} - \frac{I'}{I} - \gamma - 2b \\ &\quad + \max\{a_2 pb - r, (1 - a_2)pb - \epsilon + (1 - a_2)pb\} \\ &\leq \frac{E'}{E} - \frac{I'}{I} - \gamma - 2b + pb, \end{aligned}$$

since  $0 \leq a_2 < 1$  and  $(1 - a_2)pb - \epsilon \leq 0$  from (4.5). Therefore

$$(4.6) \quad g_1 = -\lambda I - \epsilon - 2b - r + pb + a_1(\lambda S + qb)\frac{I}{E},$$

$$(4.7) \quad g_2 \leq \frac{E'}{E} - \frac{I'}{I} - \gamma - 2b + pb + \frac{1}{a_1}\frac{\epsilon E}{I}.$$

Rewriting (4.2), we have

$$(4.8) \quad \frac{E'}{E} + b + \epsilon - pb = (\lambda S + qb)\frac{I}{E}, \quad \frac{I'}{I} + b + \gamma = \frac{\epsilon E}{I}.$$

The uniform persistence constant  $c$  can be adjusted so that there exists  $T > 0$  independent of  $(S(0), E(0), I(0)) \in K$ , the compact absorbing set, such that

$$(4.9) \quad I(t) > c \quad \text{and} \quad E(t) > c \quad \text{for } t > T.$$

Substituting (4.8) into (4.6) and (4.7) and using (4.9), we obtain, for  $t > T$ ,

$$(4.10) \quad \begin{aligned} g_1 &\leq -\lambda I - b + \frac{E'}{E} + (a_1 - 1)(\lambda S + qb)\frac{I}{E} \\ &\leq \frac{E'}{E} - \lambda c - b + (a_1 - 1)\frac{\lambda + b}{c} \leq \frac{E'}{E} - b \end{aligned}$$

and

$$(4.11) \quad \begin{aligned} g_2 &\leq \frac{E'}{E} - b + pb + \left(\frac{1}{a_1} - 1\right)\frac{\epsilon E}{I} \\ &\leq \frac{E'}{E} - b + pb + \frac{(1 - a_1)\epsilon c}{a_1} \leq \frac{E'}{E} - \frac{(a_1 - 1)\epsilon c}{a_1}, \end{aligned}$$

since  $0 \leq p \leq 1$ . Therefore  $\mu(B) \leq E'/E - \bar{b}$  for  $t > T$  by (4.10) and (4.11), where  $\bar{b} = \min\{b, (a_1 - 1)\epsilon c/a_1\} > 0$ . Along each solution  $(S(t), E(t), I(t))$  to (4.2) such that  $(S(0), E(0), I(0)) \in K$  and for  $t > T$ , we have

$$\frac{1}{t} \int_0^t \mu(B) ds \leq \frac{1}{t} \int_0^T \mu(B) ds + \frac{1}{t} \log \frac{E(t)}{E(T)} - \bar{b} \frac{t - T}{t},$$

which implies  $\bar{q}_2 \leq -\bar{b}/2 < 0$  from (2.4), proving Theorem 4.2.  $\square$

If  $r = 0$ , then  $R_0$  reduces to the threshold parameter  $R_0(p, q)$  in [28] for an SEIR model with no vaccination. In fact,  $R_0 = \frac{b}{b+r}R_0(p, q)$ . This relation clearly shows that vaccination lowers the basic reproduction number. When  $r = p = q = 0$ , then  $R_0 = \frac{\lambda\epsilon}{(\epsilon + b)(\gamma + b)}$ , which is the basic reproduction number (see [2, 12]) or the contact number (see [15, 29]) for the SEIR or SEIRS models with only horizontal transmission. In the limiting case when  $\epsilon \rightarrow \infty$ ,  $R_0$  gives the basic reproduction number in [4] for a SIR model.

Theorem 4.2 contains a global stability result in [28] for an SEIR model with vertical transmission but with no vaccination ( $r = 0$ ). It also contains a global result in [25], in which no vertical transmission and vaccination are assumed ( $r = p = q = 0$ ). Since our model contains SIR models as special cases ( $\epsilon \rightarrow \infty$ ), Theorem 4.1 also generalizes some earlier results on SIR models with vertical transmission and a constant population, see [4] and the references therein.

**5. Discussion.** In this paper, two recent mathematical approaches are used to establish the global stability of the unique endemic equilibrium in two epidemic models of SEIR type. One model has constant recruitment and exponential natural and disease-caused death. The total population size varies in time. The incidence is of bilinear form. In the second model, a balanced exponential birth and natural death is assumed and the disease causes no fatality so that the total population is constant. Both horizontal

and vertical transmission modes are considered. The horizontal transmission has a bilinear incidence. For the vertical transmission, we assume that a fraction  $p$  and a fraction  $q$  of the offspring from the exposed and the infectious classes, respectively, are born into the exposed class  $E$ . We have also considered the effects of vaccination in the second model.

In both models, the basic reproduction number  $R_0$  is identified and is established as a sharp threshold parameter. If  $R_0 \leq 1$ , the disease-free equilibrium  $P_0$  is globally stable in the feasible region and the disease always dies out. If  $R_0 > 1$ , a unique endemic equilibrium  $P^*$  exists and is globally stable in the interior of the feasible region, and once the disease appears, it eventually persists at the unique endemic equilibrium level. The global stability of  $P_0$  when  $R_0 \leq 1$  is proved using a Lyapunov function that has been widely used in the literature of epidemic models (see [29]). The mathematical difficulty lies in the proof of the global stability of  $P^*$  when  $R_0 > 1$ , since both models reduce to nonlinear ordinary differential equations in  $\mathbf{R}^3$ .

The global stability of  $P^*$  in the first model is proved using the approach of Li and Muldowney in [25], which takes the advantage of the monotonicity property that is present in the model and a Poincaré-Bendixson Property for 3-dimensional monotone systems due to Hirsch [17] and Smith [36]. Periodic solutions are ruled out using a stability criterion of Muldowney [33] for periodic orbits in higher dimensions. In the second model, the monotonicity is no longer present. The global stability of  $P^*$  is proved using a geometrical approach of Li and Muldowney in [23]. We expect that these approaches can be applied to solve global-stability problems in many other models.

**Acknowledgments.** Research of ML was supported in part by NSF grant DMS-9626128 and by a Ralph E. Powe Junior Faculty Enhancement Award from Oak Ridge Associated Universities (ORAU). Research of LW was supported in part by a Graduate Research Fellowship from Department of Mathematics and Statistics, Mississippi State University. Both authors wish to thank the IMA for financial support during the workshop.

## REFERENCES

- [1] R.M. ANDERSON AND R.M. MAY, *Population biology of infectious diseases I*, *Nature*, **180** (1979), pp. 361–367.
- [2] R.M. ANDERSON AND R.M. MAY, *Infectious Diseases of Humans, Dynamics and Control*, Oxford University Press, Oxford, 1992.
- [3] F. BRAUER, *Models for the spread of universally fatal diseases*, *J. Math. Biol.* **28** (1990), pp. 451–462.
- [4] S. BUSENBERG AND K. COOKE, *Vertically Transmitted Diseases*, Biomathematics, vol. 23, Springer-Verlag, Berlin, 1993.
- [5] S.N. BUSENBERG AND P. VAN DEN DRIESSCHE, *A method for proving the non-existence of limit cycles*, *J. Math. Anal. Appl.* **172** (1993), pp. 463–479.

- [6] S.N. BUSENBERG AND P. VAN DEN DRIESSCHE, *Analysis of a disease transmission model in a population with varying size*, J. Math. Biol. **28** (1990), pp. 257–270.
- [7] G.J. BUTLER AND P. WALTMAN, *Persistence in dynamical systems*, Proc. Amer. Math. Soc. **96** (1986), pp. 425–430.
- [8] K.L. COOK AND P. VAN DEN DRIESSCHE, *Analysis of an SEIRS epidemic model with two delays*, J. Math. Biol. **35** (1996), pp. 240–260.
- [9] W.A. COPPEL, *Stability and Asymptotic Behavior of Differential Equations*, Health, Boston, 1965.
- [10] L. ESTEVA AND C. VARGAS, *A model for dengue disease with variable human population*, J. Math. Biol. **38** (1999), pp. 220–240.
- [11] M. FIEDLER, *Additive compound matrices and inequality for eigenvalues of stochastic matrices*, Czech. Math. J. **99** (1974), pp. 392–402.
- [12] D. GREENHALGH, *Hopf bifurcation in epidemic models with a latent period and nonpermanent immunity*, Math. Comput. Modelling **25** (1997), pp. 85–107.
- [13] J.K. HALE, *Ordinary Differential Equations*, John Wiley & Sons, New York, 1969.
- [14] H.W. HETHCOTE AND S.A. LEVIN, *Periodicity in epidemiological models*, in *Applied Mathematical Ecology*, L. Gross and S.A. Levin (eds.), Springer, New York, 1989, pp. 193–211.
- [15] H.W. HETHCOTE, H.W. STECH, AND P. VAN DEN DRIESSCHE, *Periodicity and stability in epidemic models: a survey*, in *Differential Equations and Applications in Ecology, Epidemics, and Population Problems*, K. L. Cook (ed.), Academic Press, New York, 1981, pp. 65–85.
- [16] H.W. HETHCOTE AND P. VAN DEN DRIESSCHE, *Some epidemiological models with nonlinear incidence*, J. Math. Biol. **29** (1991), pp. 271–287.
- [17] M.W. HIRSCH, *Systems of differential equations which are competitive or cooperative IV: Structural stability in three dimensional systems*, SIAM J. Math. Anal. **21** (1990), pp. 1225–1234.
- [18] M.W. HIRSCH, *Systems of differential equations that are competitive or cooperative. VI: A local  $C^r$  closing lemma for 3-dimensional systems*, Ergod. Th. & Dynam. Sys. **11** (1991), pp. 443–454.
- [19] J.P. LASALLE, *The Stability of Dynamical Systems*, Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, 1976.
- [20] G.A. LEONOV, D.V. PONOMARENKO, AND V.B. SMIRNOVA, *Frequency-Domain Methods for Nonlinear Analysis, Theory and Applications*, World Scientific, Singapore, 1996.
- [21] M.Y. LI, *Dulac criteria for autonomous systems having an invariant affine manifold*, J. Math. Anal. Appl. **199** (1996), pp. 374–390.
- [22] M.Y. LI, J.R. GRAEF, L.C. WANG, AND J. KARSAI, *Global dynamics of a SEIR model with a varying total population size*, Math. Biosci. **160** (1999), pp. 191–213.
- [23] M.Y. LI AND J.S. MULDOWNEY, *A geometric approach to the global-stability problems*, SIAM J. Math. Anal. **27** (1996), pp. 1070–1083.
- [24] M.Y. LI AND J.S. MULDOWNEY, *On R.A. Smith's autonomous convergence theorem*, Rocky Mount. J. Math. **25** (1995), pp. 365–379.
- [25] M.Y. LI AND J.S. MULDOWNEY, *Global stability for the SEIR model in epidemiology*, Math. Biosci. **125** (1995), pp. 155–164.
- [26] M.Y. LI, J.S. MULDOWNEY, AND P. VAN DEN DRIESSCHE, *Global stability for SEIRS models in epidemiology*, Canadian Appl. Math. Quart., to appear.
- [27] M.Y. LI AND L. WANG, *A criterion for stability of matrices*, J. Math. Anal. Appl. **225** (1998), pp. 249–264.
- [28] M.Y. LI, H.L. SMITH, AND L. WANG, *Global stability of an SEIR model with vertical transmission*, SIAM J. Appl. Math., to appear.
- [29] W-M. LIU, H.W. HETHCOTE AND S.A. LEVIN, *Dynamical behavior of epidemiological models with nonlinear incidence rate*, J. Math. Biol. **25** (1987), pp. 359–380.

- [30] R.H. MARTIN, JR., *Logarithmic norms and projections applied to linear differential systems*, J. Math. Anal. Appl. **45** (1974), pp. 432–454.
- [31] R.M. MAY AND R.M. ANDERSON, *Regulation and stability of host-parasite population interactions. II. Destabilizing process*, J. Anim. Ecol. **47** (1978), pp. 219–267.
- [32] J. MENA-LORCA AND H.W. HETHCOTE, *Dynamic models of infectious diseases as regulator of population sizes*, J. Math. Biol. **30** (1992), pp. 693–716.
- [33] J.S. MULDOWNEY, *Compound matrices and ordinary differential equations*, Rocky Mount. J. Math. **20** (1990), pp. 857–872.
- [34] C.C. PUGH, *An improved closing lemma and the general density theorem*, Amer. J. Math. **89** (1967), pp. 1010–1021.
- [35] C.C. PUGH AND C. ROBINSON, *The  $C^1$  closing lemma including Hamiltonians*, Ergod. Th. & Dynam. Sys. **3** (1983), pp. 261–313.
- [36] H.L. SMITH, *Periodic orbits of competitive and cooperative systems*, J. Differential Equations **65** (1986), pp. 361–373.
- [37] H.L. SMITH, *Monotone Dynamical Systems, An Introduction to the Theory of Competitive and Cooperative Systems*, Amer. Math. Soc., Providence, 1995.
- [38] R.A. SMITH, *Some applications of Hausdorff dimension inequalities for ordinary differential equations*, Proc. Roy. Soc. Edinburgh **104A** (1986), pp. 235–259.
- [39] H. THIEME, *Epidemic and demographic interaction in the spread of potentially fatal diseases in growing populations*, Math. Biosci. **111** (1992), pp. 99–130.
- [40] H.R. THIEME AND P. VAN DEN DRIESSCHE, *Global stability in cyclic epidemic models with disease fatalities*, in *Differential Equations with Applications to Biology* (Halifax, NS, 1997), pp. 459–472, Fields Inst. Commun., vol. 21, Amer. Math. Soc., Providence, 1999.
- [41] P. WALTMAN, *A brief survey of persistence*, in *Delay Differential Equations and Dynamical Systems*, S. Busenberg and M. Martelli (eds.), Springer-Verlag, New York, 1991, pp. 31–40.

# THE GLOBAL STABILITY ANALYSIS FOR AN SIS MODEL WITH AGE AND INFECTION AGE STRUCTURES

YICANG ZHOU<sup>†\*</sup>, BAOJUN SONG<sup>†</sup>, AND ZHIEN MA<sup>‡</sup>

**Abstract.** A general SIS model with chronological age and infection age structures is formulated. We analyze the global dynamics of the model with a constructive iteration procedure. The basic reproductive number  $R_0$  is calculated using the next generation operator approach.  $R_0$  plays a sharp threshold role in determining the global dynamics, i.e., the endemic steady-state is globally asymptotically stable if  $R_0 > 1$ , while the disease-free steady-state is globally asymptotically stable if  $R_0 \leq 1$ . The basic reproductive number is over estimated where the infection age is ignored.

**Key words.** Age-structure, infection age, proportionate mixing, basic reproductive number, global stability.

**1. Introduction.** Virtually all epidemic models for communicable diseases in homogeneous populations have been inspired by the model derived by Kermack and Mckendrick [1,2]. Epidemiologically, this model, in its general form only became well-known in the 1970s [3,4,5]. Chronological and infection age might be the most important factors in disease spread. Vynnycky and Fine [6], for instance, have shown that tuberculosis(TB) infection is at low rate for individuals less than 10 years old, but dramatically increases when the individual's age is between 10 and 20 years. The mixing structure of a population is often closely related to age structure of the population. Contact rates are highly dependent on age. Many diseases, for example, childhood diseases, may be mostly transmitted to individuals of the same age. Vaccination strategies for specific diseases are naturally applied to different age groups( see Castillo-Chavez and Feng [7] and references therein). Hence, during the last two decades of the 20th century, age-structured epidemic models have been extensively studied [8,9,10,11,12,13,14,15,16]. Dietz and Schenzle studied an epidemic model with special forms of age and infectious age dependent contact rate [17]. Thieme and Castillo-Chavez formulated a model and explored the role of variable infectivity in combination with a variable incubation period in the dynamics of HIV transmission in homogeneously mixing population [18]. Proportionate mixing has been extensively used in age-structure models. This mixing structure has been useful in the study of dynamics of childhood diseases. The use of proportionate mixing also makes it easier to get an explicit formula for the basic reproductive number [19].

Previous dynamical analysis for many age-structure models has been incomplete. The local stability for disease-free steady-state is easy to estab-

---

\*Author to whom correspondence should be addressed. E-mail: zhouyc@mail.xjtu.edu.cn.

<sup>†</sup>Department of Biometrics, Cornell University, Ithaca, NY 14853-7801, USA.

<sup>‡</sup>Department of Mathematics, Xi'an Jiaotong University, Xi'an, 710049, China.

lish for most age-structured models when the basic reproductive number is less than a unity. The globally asymptotic stability of a stable age distribution, however, is very difficult in general. In this paper, we focus on the study of the global dynamics of two-age structured model. The study is theoretical in nature, but the framework and the approach may be applicable to specific diseases.

The paper is organized as follows: Section 2 introduces the epidemic model. The basic reproductive number is computed and fundamental assumptions are spelled out. Section 3 establishes the global asymptotic stability of the disease-free steady-state and the endemic non-uniform age-distributions. An iteration procedure is used. The appendix shows that the solutions of the epidemic model are nonnegative for relevant initial distributions.

**2. Model and basic reproductive number.** The framework of our model is from Busenberg and Castillo-Chavez [20]. SIS models with both chronological age and infection age structures are formulated. Demographically the population is stratified by chronological age and, epidemiologically, it is partitioned into susceptible and infective. Here  $s(a, t)$  denotes the density of susceptibles at time  $t$  and  $i(a, c, t)$  the density of infectives at time  $t$ , where  $a$  is the chronological age and  $c$  is the infection age, that is, the time span since infection. The variable  $c$  does not distinguish individuals and can never be greater than the chronological age  $a$ . The total numbers of the susceptibles  $S(t)$  and infectives  $I(t)$  at time  $t$  are given by  $S(t) = \int_0^A s(a, t)da$ ,  $I(t) = \int_0^A \int_0^a i(a, c, t)dc da$ , respectively, where  $A$  is the maximum age. The total population size is given by  $P(t) = S(t) + I(t) = \int_0^A s(a, t)da + \int_0^A \int_0^a i(a, c, t)dc da$ . It is assumed that all newborns are susceptibles and the disease is not fatal. Thus, we ignore the disease-related mortality. The model describing the dynamics of an SIS infectious disease takes the form of a nonlinear hyperbolic system of PDEs:

$$(1a) \quad \frac{\partial s}{\partial a} + \frac{\partial s}{\partial t} = -\mu(a)s(a, t) - G(a, t) + \gamma(a) \int_0^a i(a, c, t)dc,$$

$$(1b) \quad s(0, t) = \int_{A_1}^{A_2} b(a, P(t))p(a, t)da,$$

$$(1c) \quad s(a, 0) = s_0(a), \quad s(A, t) = 0,$$

$$(2a) \quad \frac{\partial i}{\partial a} + \frac{\partial i}{\partial c} + \frac{\partial i}{\partial t} = -(\mu(a) + \gamma(a))i(a, c, t),$$

$$(2b) \quad i(a, 0, t) = G(a, t) = C(a)s(a, t) \int_0^A \int_0^{a'} \beta(a', c) \frac{i(a', c, t)}{p(a', t)} \rho(a, a', t) dc da',$$

$$(2c) \quad i(A, c, t) = 0,$$

$$(2d) \quad i(a, c, 0) = i_0(a, c),$$

where  $p(a, t) = s(a, t) + \int_0^a i(a, c, t)dc$  is the entire population density at time  $t$ ;  $P(t) = \int_0^A p(a, t)da$  the total population size at time  $t$ ;  $b(a, P(t))$  the density-dependent age-specific birth rate(birth modulus) of the population;  $s_0(a)$  and  $i_0(a, c)$  are the initial distributions;  $[A_1, A_2]$  is the fecundity period,  $0 < A_1 < A_2 < A$ ;  $\mu(a)$  the age-specific mortality rate;  $\gamma(a)$  the age-specific recovery rate of the infective individuals;  $C(a)$  the age-specific contact rate;  $\beta(a, c)$  the age-specific probability that a susceptible becomes infected given that it had a contact with an infectious of  $(a, c)$  type;  $\rho(a, a', t)$  the probability that an individual of age  $a$  has contact with an individual of  $a'$  given that it has a contact. We implicitly assume that the population mixes proportionately [20], that is,  $\rho(a, a', t) = \rho(a', t) = \frac{C(a')p(a', t)}{\int_0^A C(a)p(a, t)da}$ , The force of infection(incidence), that is, the rate at which susceptibles individuals of age  $a$  move over into the infective class per capita and per unit of time, is given by  $G(a, t)$ ;

In the most cases (except for the case where backward bifurcation takes place) the basic reproductive number regulates the local dynamics of disease transmission., i.e., if it is less than a unity, the disease dies out, while if it is greater than one the disease establishes itself. The main result of this paper is the establishment of the global nature of this transcritical bifurcation. The next generation operator is used to find out the basic reproductive number. This approach formulated by Diekmann et al. [19] has been widely used recently. Feng, Castillo-Chavez, and Huang have a detailed version of this method and specific applications in this volume. First, we compute the demographic steady-state and expected infectivity. The demographic steady-state is calculated from Equation (1a) by letting  $i(a, a', t) = 0$  ( details are provided later in this section). It turns out that the demographic steady-state

$$p_\infty(a) = P_\infty \frac{\exp(-\int_0^a \mu(\tau)d\tau)}{\int_0^A \exp(-\int_0^a \mu(\tau)d\tau)da}, \quad \text{where } P_\infty \text{ is a constant.}$$

Let  $E(a, a', c)$  denote the expected infectivity of an infectious individual of  $(a', c)$  type towards a susceptible of age  $a$  at the demographic steady-state.  $E(a, a', c)$  is determined by a routine infection process thought that follows the pattern of an infective who first survives from  $a' - c$  to  $a'$ ; has contacts with susceptibles; and transmits the disease by chance. It is assumed that these events happen independently. Hence,  $E(a, a', c)$  is the product of three terms: the survival probability of an infected individual surviving from  $a' - c$  to  $a'$ ; the contact rate of individuals of age  $a$  with individuals of age  $a'$ ; and the transmission probability. Therefore

$$E(a, a' - c, c) = C(a)\rho(a')\beta(a', c)\frac{\pi(a', c)}{p_\infty(a')},$$

where  $\rho(a') = \frac{C(a')p_\infty(a')}{\int_0^A C(a)p_\infty(a)da}$  and  $\pi(a', c) = e^{-\int_{a'-c}^{a'} (\mu(a)+\gamma(a))da}$  (the survival probability of an infected individual survives from  $a' - c$  to  $a'$ ). Hence

$$(3) \quad \begin{aligned} R_0 &= \int_0^A \int_0^{a'} C(a' - c)p_\infty(a' - c)\rho(a')\beta(a', c)\frac{\pi(a', c)}{p_\infty(a')}dcda' \\ &= \int_0^A \int_0^{a'} C(a' - c)p_\infty(a' - c)\frac{C(a')\beta(a', c)}{\int_0^A C(a')p_\infty(a')da'}\pi(a', c)dcda'. \end{aligned}$$

### 3. Dynamics analysis.

**3.1. Assumptions.** From the epidemiological and mathematical point of view, we make the following assumptions:

- A1.  $\mu(a) > \mu_0 > 0$  is a positive continuous function on  $[0, A]$ , and  $\int_0^A \mu(a)da = +\infty$ . Set  $M(a) = \exp(-\int_0^a \mu(\tau)d\tau)$  for  $0 \leq a < A$ , and  $M(A) = 0$ . It is obvious that  $M(a)$  is a continuous, decreasing function on  $[0, A]$ , and  $0 \leq M(a) \leq 1$  for  $0 \leq a \leq A$ .
- A2.  $b(a, P)$  is a nonnegative continuous function, and  $b(a, P) > 0$  for  $A_1 < a < A_2$ ,  $b(a, P) = 0$  for  $a \leq A_1$  or  $a \geq A_2$ .
- A3.  $p_0(a) \geq 0$  is a continuous function,  $p_0(a) > 0$  if  $a \in [0, A]$ , and  $p_0(A) = 0$ . The initial data  $p_0(a)$  satisfies the compatible condition

$$p_0(0) = \int_0^A b\left(a, \int_0^A p_0(\tau)d\tau\right)p_0(a)da,$$

which is simply the requirement that the initial data be consistent with the birth process.

- A4.  $\gamma(a)$  is a nonnegative and continuous function on  $[0, A]$ .
- A5. The effects of infection age follows an exponential distribution with an expected infection age  $\frac{1}{\delta}$  and  $\beta(a', c)$  takes a form of variable-separated,  $\beta(a', c) = \beta_1(a')e^{-\delta c}$ , where  $\beta_1(a')$  is assumed to be bounded, positive and continuous. To keep notation simpler, we set  $\lambda_1(a) = \frac{C(a)}{\int_0^A C(a')p_\infty(a')da'}$ ,  $\lambda_2(a') = \beta_1(a')C(a')$ ,  $\lambda(a, a', c) = \lambda_1(a)\lambda_2(a')e^{-\delta c}$ , and  $\lambda^* = \max_{0 \leq a \leq A, 0 \leq a' \leq A} \lambda(a, a', c)$ .
- A6.  $i_0(a, c)$  is bounded, nonnegative and continuous function on  $0 \leq c \leq a \leq A$ . And  $i_0(a, c)$  satisfies the continuous compatible condition

$$\left(p_0(a) - \int_0^a i_0(a, c)dc\right) \int_0^A \int_0^{a'} \lambda(a, a', c)i_0(a', c)dcda' = i_0(a, 0).$$

The lack of disease-induced mortality in our model implies that the total population density  $p(a, t)$  is governed by the following demographic evolution equation [21,22]:

$$(4a) \quad \frac{\partial p}{\partial a} + \frac{\partial p}{\partial t} = -\mu(a)p(a, t),$$

$$(4b) \quad p(0, t) = \int_{A_1}^{A_2} b(a, P(t))p(a, t)da,$$

$$(4c) \quad p(a, 0) = p_0(a), \quad p(A, t) = 0,$$

where  $p_0(a) = s_0(a) + \int_0^a i_0(a, c)dc$ . Under assumptions (A1)~(A3), the age-structured population model (4) is well-posed [23,24]. From the net reproductive number  $n(P) = \int_0^A b(a, P)M(a)da$ , we see that, if the equation  $n(P) = 1$  has a positive root  $P_\infty$ , then the total population density  $p(a, t)$  has a steady-state  $p_\infty(a) = P_\infty M(a)/\int_0^A M(a)da$ [23]. Sufficient conditions for the local stability for the positive steady-state  $p_\infty(a)$  can be found in [25,26,27]. We focus on the global stability of non-uniform epidemic steady-state throughout the rest of this paper.

We assume that the total population is at its demographic steady-state  $p_\infty(a)$  i.e.,  $p(a, t) = p_\infty(a) = P_\infty M(a)/\int_0^A M(a)da$ , thus replacing  $s(a, t)$  by  $p_\infty(a) - \int_0^a i(a, c, t)dc$  in the force of infection,  $G(a, t)$ , we arrive at a single equation in terms of  $i(a, c, t)$ .

$$(5a) \quad \frac{\partial i}{\partial a} + \frac{\partial i}{\partial c} + \frac{\partial i}{\partial t} = -(\mu(a) + \gamma(a))i(a, c, t),$$

$$(5b) \quad \begin{aligned} i(a, 0, t) &= G(a, t) \\ &= \left( p_\infty(a) - \int_0^a i(a, c, t)dc \right) \int_0^A \int_0^{a'} \lambda(a, a', c)i(a', c, t)dc da'. \end{aligned}$$

$$(5c) \quad i(A, c, t) = 0,$$

$$(5d) \quad i(a, c, 0) = i_0(a, c).$$

Using a similar idea by Tucker et al., it can be shown that the Equation (5) has a unique continuous solution for all  $t \geq 0$ , provided that the assumptions (A1)~(A6) hold [28]. Are the solutions to the Equation (5) nonnegative? Theorem 1 below gives a positive answer.

**Theorem 1.** *Assume that (A1)~(A6) hold. Let  $i(a, c, t)$  be the solution of (5) with the initial distribution  $i_0(a, c)$  satisfying  $\int_0^a i_0(a, c)dc \leq p_\infty(a)$ . Then  $i(a, c, t) \geq 0$  for all  $t \geq 0$ .*

The proof of Theorem 1 can be found in Appendix.

**3.2. Global stability of disease-free steady-state.** Equation (5) can be explicitly solved along the characteristic lines,

$$(6) \quad \begin{aligned} i(a, c, t) &= G(a - c, t - c)\pi(a, c), \\ \text{where } G(a - c, t - c) &= i_0(a - t, c - t) \frac{\pi(a, t)}{\pi(a, c)}, \quad \text{if } t \leq c. \end{aligned}$$

Imposing the boundary condition (5b) upon (6) gives the following nonlinear equation for  $G(a, t)$

$$(7) \quad \begin{aligned} G(a, t) = & \left( p_\infty(a) - \int_0^a G(a - c, t - c) \pi(a, c) dc \right) \\ & \times \int_0^A \int_0^{a'} \lambda(a, a', c) G(a' - c, t - c) \pi(a', c) dc da'. \end{aligned}$$

Since  $\pi(a', c) = \exp\left(-\int_{a'-c}^{a'} (\mu(\tau) + \gamma(\tau)) d\tau\right)$  is bounded, positive, and continuous, it follows from (6) that the asymptotic behavior of  $i(a, c, t)$  is completely determined by the asymptotic behavior of  $G(a, t)$ , which satisfies Equation (7). Therefore, Equation (7) is the main target in our analysis.

The specific expression for the basic reproductive number for Model (5) under the assumptions (A1)~(A6) is

$$(8) \quad R_0 = \int_0^A \int_0^{a'} \lambda_1(a' - c) \lambda_2(a') e^{-\delta c} p_\infty(a' - c) \pi(a', c) dc da'.$$

To see the effect of infection age, we look at  $R_0$  as a function of the expected infection age,  $\frac{1}{\delta}$ .  $R_0(\frac{1}{\delta})$  increases as does  $\frac{1}{\delta}$ . An extreme case  $\frac{1}{\delta} = \infty$  corresponds to the situation where the infection age has no effect.  $R_0(\frac{1}{\delta})$  approaches its maximum value  $R_0 = \int_0^A \int_0^{a'} \lambda_1(a' - c) \lambda_2(a') p_\infty(a' - c) \pi(a', c) dc da'$  that is the corresponding basic reproductive number to the case where only chronological age is considered. Hence the basic reproductive number in single age-structure model over estimates the severity of the epidemic.

We now examine the stability of the disease-free steady-state and the existence of an endemic steady-state.

**Theorem 2.** *Assume that (A1)~(A6) hold. Then the disease-free steady-state is globally asymptotically stable if  $R_0 \leq 1$ , whereas, it is unstable and there exists a unique endemic steady-state if  $R_0 > 1$ .*

**Proof.** The local stability of the disease-free steady-state is directly derived from the definition of  $R_0$ . What remain now is to show that it is a global attractor. Defining

$$\begin{aligned} \lambda_3(a', c) &= \lambda_2(a') e^{-\delta c}, \\ w(t) &= \int_0^A \int_0^{a'} \lambda_3(a', c) i(a', c, t) dc da', \end{aligned}$$

one can see that

$$w(t) = \int_0^A \int_0^{a'} \lambda_3(a', c) G(a' - c, t - c) \pi(a', c) dc da', \quad t \geq A.$$

According to Equation (7), we arrive at

$$(9) \quad \begin{aligned} w(t) &= \int_0^A \int_0^{a'} \lambda_1(a' - c) \lambda_3(a', c) p_\infty(a' - c) \pi(a', c) w(t - c) dc da' \\ &\quad - \int_0^A \int_0^{a'} \lambda_1(a' - c) \lambda_3(a', c) \\ &\quad \times \int_0^{a'-c} G(a' - c - \tau, t - c - \tau) \pi(a', c - \tau) w(t - c) d\tau dc da', \end{aligned}$$

and

$$(10) \quad w(t) \leq \int_0^A \int_0^{a'} \lambda_1(a' - c) \lambda_3(a', c) p_\infty(a' - c) \pi(a', c) w(t - c) dc da'.$$

Let  $\|p_\infty(a)\|$  and  $\|i_0(a, c)\|$  be the maximum norm of the continuous functions, that is,  $\|p_\infty(a)\| = \max_{0 \leq a \leq A} p_\infty(a)$ ,  $\|i_0(a, c)\| = \max_{0 \leq c \leq a \leq A} i_0(a, c)$ . The assumption (A6) and the expression of  $p_\infty(a) = P_\infty M(a) / \int_0^A M(a) da$  imply that  $\|p_\infty(a)\|$  and  $\|i_0(a, c)\|$  are well-defined positive numbers. The fact that the disease-free steady-state  $i(a, c, t) = 0$  is a trivial solution to (5) and the regularity result of solutions implies that  $i(a, c, t)$ , the solutions of (5) with the initial distribution  $i_0(a, c)$ , are continuously dependent on  $i_0(a, c)$ . Hence, for any given positive number  $\varepsilon$ , there exists a positive number  $\delta$ , such that  $i(a, c, t) < \varepsilon$  if  $\|i_0(a, c)\| < \delta$  and  $t \leq A$ . From (5b) it follows that  $0 \leq G(a, t) \leq c_1 \varepsilon$  for  $0 \leq t \leq A$ , here  $c_1$  is a positive constant. From the definition of  $w(t)$ , there also exists a positive constant  $c_0$ , such that  $w(t) < c_0 \varepsilon$  for  $0 \leq t \leq A$ .

If  $R_0 < 1$ , Inequality (10) implies that  $w(A) \leq R_0 c_0 \varepsilon < \frac{R_0+1}{2} c_0 \varepsilon < c_0 \varepsilon$ . We claim that  $w(t) < \frac{R_0+1}{2} c_0 \varepsilon$  for all  $A \leq t \leq 2A$ . Otherwise, there exists at least one  $t_0 \in (A, 2A]$ , such that  $w(t_0) = \frac{R_0+1}{2} c_0 \varepsilon$ , and  $w(t) < \frac{R_0+1}{2} c_0 \varepsilon$  for all  $A \leq t < t_0$ . From Inequality (10) it follows that

$$\begin{aligned} w(t_0) &\leq \int_0^A \int_0^{a'} \lambda_1(a' - c) \lambda_3(a', c) p_\infty(a' - c) \pi(a', c) w(t_0 - c) dc da' \\ &\leq R_0 c_0 \varepsilon < \frac{R_0 + 1}{2} c_0 \varepsilon. \end{aligned}$$

This contradiction implies that  $w(t) < \frac{R_0+1}{2} c_0 \varepsilon$  for all  $A \leq t \leq 2A$ . Mathematical induction gives

$$(11) \quad w(t) \leq \left( \frac{R_0 + 1}{2} \right)^n c_0 \varepsilon \quad \text{for } nA \leq t \leq (n+1)A.$$

From (5b) and (11) we obtain

$$(12) \quad G(a, t) \leq \left( \frac{R_0 + 1}{2} \right)^n c_2 \varepsilon, \quad \text{for } nA \leq t \leq (n+1)A.$$

Finally the expression (6) gives rise to

$$(13) \quad i(a, c, t) \leq \left( \frac{R_0 + 1}{2} \right)^n c_3 \varepsilon, \quad \text{for } nA \leq t \leq (n+1)A.$$

$c_2$  and  $c_3$  in (12) and (13) are positive constants. Inequality (13) says that the disease-free steady-state  $i(a, c, t) = 0$  is a globally asymptotically attractive. Thus it is globally asymptotically stable since we have known it is local stable.

To prove the existence of unique stable age distribution, we need to look for the time-independent solution  $G_\infty(a)$  of the equation (7) when  $R_0 > 1$ .  $G_\infty(a)$  satisfies the equation

$$(14a) \quad G_\infty(a) = \lambda_1(a) \left( p_\infty(a) - \int_0^a G_\infty(\tau) \frac{N(a)}{N(\tau)} d\tau \right) w^*,$$

$$(14b) \quad \text{where } w^* = \int_0^A \int_0^{a'} \lambda_3(a', c) G_\infty(a' - c) \pi(a', c) dc da',$$

$$(14c) \quad N(a) = \exp \left( - \int_0^a (\mu(\tau) + \gamma(\tau)) d\tau \right).$$

Defining  $g(a) = \frac{G_\infty(a)}{\lambda_1(a)}$ , it can be verified that  $g(a)$  is the solution of the equation

$$(15a) \quad g(a) = \left( p_\infty(a) - \int_0^a g(\tau) \lambda_1(\tau) \frac{N(a)}{N(\tau)} d\tau \right) w^*,$$

$$(15b) \quad \text{where } w^* = \int_0^A \int_0^{a'} \lambda_3(a', c) g(a' - c) \lambda_1(a' - c) \pi(a', c) dc da'.$$

Changing the integral equation for  $g(a)$  into a differential equation, solving the resulting equation, a closed form for  $g(a)$  is obtained,

$$(16) \quad \begin{aligned} g(a) = & w^* N(a) \exp \left( - w^* \int_0^a \lambda_1(\tau) d\tau \right) \\ & \times \left( p_\infty(0) + \int_0^a \frac{\gamma(\tau) p_\infty(\tau) \exp(w^* \int_0^\tau \lambda_1(\theta) d\theta)}{N(\tau)} d\tau \right). \end{aligned}$$

The substitution of (16) into (15b) leads to that (16) is the solution of (15) if and only if  $w^*$  is the solution of the equation  $f(w^*) = 1$ , where

$$\begin{aligned} f(w^*) = & \int_0^A \int_0^{a'} \lambda_1(a' - c) \lambda_3(a', c) N(a') \exp(-w^* \int_0^{a'-c} \lambda_1(\tau) d\tau) \\ & \times \left( p_\infty(0) + \int_0^{a'-c} \frac{\gamma(\tau) p_\infty(\tau) \exp(w^* \int_0^\tau \lambda_1(\theta) d\theta)}{N(\tau)} d\tau \right) dc da' \\ = & p_\infty(0) \int_0^A \int_0^{a'} \lambda_1(a' - c) \lambda_3(a', c) N(a') \exp(-w^* \int_0^{a'-c} \lambda_1(\tau) d\tau) dc da' \\ & + p_\infty(0) \int_0^A \int_0^{a'} \lambda_1(a' - c) \lambda_3(a', c) N(a') \\ & \times \int_0^{a'-c} \gamma(\tau) \exp \left( -w^* \int_\tau^{a'-c} \lambda_1(\theta) d\theta + \int_0^\tau \gamma(\theta) d\theta \right) d\tau dc da'. \end{aligned}$$

$f(w^*)$  is a monotonic decreasing function of  $w^*$ , with  $\lim_{w^* \rightarrow +\infty} f(w^*) = 0$ , and

$$\begin{aligned} f(0) &= \int_0^A \int_0^{a'} \lambda_1(a' - c) \lambda_3(a', c) p_\infty(0) \\ &\quad \times \left( 1 + \int_0^{a'-c} \gamma(\tau) \exp \left( \int_0^\tau \gamma(\theta) d\theta \right) d\tau \right) N(a') dc da' \\ &= \int_0^A \int_0^{a'} \lambda_1(a' - c) \lambda_3(a', c) p_\infty(0) \exp \left( \int_0^{a'-c} \gamma(\tau) d\tau \right) N(a') dc da' \\ &= \int_0^A \int_0^{a'} \lambda_1(a' - c) \lambda_3(a', c) p_\infty(a' - c) \pi(a', c) dc da' = R_0 > 1. \end{aligned}$$

This follows that there exists one and only one solution to equation  $f(w^*) = 1$  by using the intermediate theorem for continuous function. Hence the existence of a unique steady age distribution is established.

Now turn to prove that the disease-free steady-state is unstable when  $R_0 > 1$ . If it is stable then for the positive number  $\varepsilon = \frac{R_0 - 1}{2A^3\lambda^*}$ , there exists a  $\delta > 0$ , such that,  $0 \leq i_0(a, c) < \delta$  implies that

$$(17) \quad 0 \leq i(a, c, t) < \varepsilon, \quad 0 \leq c \leq a \leq A, \quad t \geq 0.$$

By the definition of  $w(t)$  and (17), we know

$$(18) \quad 0 < w(t) = \int_0^A \int_0^{a'} \lambda_3(a', c) i(a', c, t) dc da' \leq A^2 \lambda^* \varepsilon, \quad t \geq 0.$$

On the other hand, if  $t \geq A$ , (6) and (9) yield

$$\begin{aligned} w(t) &= \int_0^A \int_0^{a'} \lambda_3(a', c) \left( p_\infty(a' - c) \right. \\ &\quad \left. - \int_0^{a'-c} i(a' - c, \tau, t - c) d\tau \right) w(t - c) \pi(a', c) dc da' \\ (19) \quad &\geq \int_0^A \int_0^{a'} \lambda_3(a', c) \left( p_\infty(a' - c) \right. \\ &\quad \left. - \int_0^{a'-c} i(a' - c, \tau, t - c) d\tau \right) \pi(a', c) dc da' \min_{t-A \leq \tau \leq t} w(\tau) \\ &\geq (R_0 - A^3 \lambda^* \varepsilon) \min_{t-A \leq \tau \leq t} w(\tau) = \frac{R_0 + 1}{2} \min_{t-A \leq \tau \leq t} w(\tau), \end{aligned}$$

from which we can have

$$(20) \quad w(t) \geq \left( \frac{R_0 + 1}{2} \right)^n \min_{0 \leq \tau \leq A} w(\tau), \quad nA \leq t \leq (n+1)A.$$

The initial distribution  $i_0(a, c)$  can be chosen such that  $\min_{0 \leq \tau \leq A} w(\tau) = w^0 > 0$ . Consequently, Inequality (18) and (20) can not be true simultaneously. This contradiction implies the instability of the disease free equilibrium. The proof of Theorem 2 is complete.

The threshold result (Theorem 2) presents a satisfactory answer to the stability of the disease-free steady-state and the existence and uniqueness of nontrivial steady-state. It is natural to investigate the stability of the unique endemic steady-state when  $R_0 > 1$ .

**3.3. The global stability of the endemic steady-state.** The two-age-structured system is replaced by two coupled single-age-structured system. A comparison theorem then is established for the coupled models. Finally, we prove the global stability results for the SIS model (2).

We first define the age and weighted age density functions

$$j(a, t) = \int_0^a i(a, c, t) dc, \quad k(a, t) = \int_0^a e^{-\delta c} i(a, c, t) dc,$$

which are governed by

$$(21a) \quad \frac{\partial j}{\partial a} + \frac{\partial j}{\partial t} = -(\mu(a) + \gamma(a))j + (p_\infty(a) - j(a, t))\lambda_1(a) \int_0^A \lambda_2(a')k(a', t)da',$$

$$(21b) \quad \frac{\partial k}{\partial a} + \frac{\partial k}{\partial t} = -(\mu(a) + \gamma(a) + \delta)k + (p_\infty(a) - j(a, t))\lambda_1(a) \int_0^A \lambda_2(a')k(a', t)da'$$

$$(21c) \quad j(a, 0) = j_0(a) = \int_0^a i_0(a, c) dc,$$

$$(21d) \quad k(a, 0) = k_0(a) = \int_0^a e^{-\delta c} i_0(a, c) dc,$$

$$(21e) \quad j(0, t) = 0, \quad j(A, t) = 0,$$

$$(21f) \quad k(0, t) = 0, \quad k(A, t) = 0.$$

Going one step further, we perform the normalization,

$$u(a, t) = e^{-\delta a} j(a, t) / p_\infty(a), \quad v(a, t) = k(a, t) / p_\infty(a),$$

thus

$$0 \leq u(a, t) \leq v(a, t) \leq 1, \quad 0 \leq e^{\delta a} u(a, t) \leq 1.$$

A new system in terms  $u(a, t)$  and  $v(a, t)$  is

$$(22a) \quad \frac{\partial u}{\partial a} + \frac{\partial u}{\partial t} = F(u, v)(a, t),$$

$$(22b) \quad \frac{\partial v}{\partial a} + \frac{\partial v}{\partial t} = G(u, v)(a, t),$$

$$(22c) \quad u(a, 0) = u_0(a) = e^{-\delta a} j_0(a) / p_\infty(a),$$

$$(22d) \quad v(a, 0) = v_0(a) = k_0(a) / p_\infty(a),$$

$$(22e) \quad u(0, t) = 0, \quad v(0, t) = 0,$$

where

$$\begin{aligned} F(u, v)(a, t) &= -(\gamma(a) + \delta)u(a, t) \\ &\quad + (e^{-\delta a} - u(a, t))\lambda_1(a) \int_0^A \lambda_2(a')p_\infty(a')v(a', t)da', \\ G(u, v)(a, t) &= -(\gamma(a) + \delta)v(a, t) \\ &\quad + (1 - e^{\delta a}u(a, t))\lambda_1(a) \int_0^A \lambda_2(a')p_\infty(a')v(a', t)da'. \end{aligned}$$

When  $R_0 > 1$ , it is straightforward to show that the unique endemic steady-state of (2) corresponds to the unique endemic steady-state of (22):

$$\begin{aligned} u_\infty(a) &= \int_0^a V e^{-\delta \tau} \lambda_1(\tau) \exp\left(-\int_\tau^a (\gamma(\theta) + \delta + V \lambda_1(\theta))d\theta\right) d\tau \\ v_\infty(a) &= \int_0^a V (1 - e^{\delta \tau} u_\infty(\tau)) \lambda_1(\tau) \exp\left(-\int_\tau^a (\gamma(\theta) + \delta)d\theta\right) d\tau, \end{aligned}$$

where

$$V = \int_0^A \lambda_2(a)p_\infty(a)v_\infty(a)da.$$

A comparison theorem for the constructed System of PDEs (22) is established below:

**Theorem 3 (Comparison Theorem).** *Assume that (A1)~(A6) hold. Let  $u_1(a, t)$ ,  $u_2(a, t)$ ,  $v_1(a, t)$ ,  $v_2(a, t)$  be the solutions of (22) with the initial conditions  $u_1(a, 0) = u_{10}(a)$ ,  $u_2(a, 0) = u_{20}(a)$ ,  $v_1(a, t) = v_{10}(a)$ ,  $v_2(a, 0) = v_{20}(a)$ , respectively. Then (1)  $u_1(a, t) \leq u_2(a, t)$ ,  $v_1(a, t) \leq v_2(a, t)$  if  $0 \leq u_{20}(a) - u_{10}(a) \leq v_{20}(a) - v_{10}(a)$ . (2)  $\xi u_1(a, t) \leq u_\xi(a, t)$  and  $\xi v_1(a, t) \leq v_\xi(a, t)$ , where  $u_\xi(a, t)$  and  $v_\xi(a, t)$  are the solutions of (22) with the initial conditions  $u_\xi(a, 0) = \xi u_{10}(a)$ ,  $v_\xi(a, 0) = \xi v_{10}(a)$ , and  $\xi$  is a constant  $0 \leq \xi < 1$ .*

**Proof.** Let  $\eta$  be a positive constant such that

$$\eta \max_{0 \leq a \leq A} \left( \gamma(a) + 2\delta + \lambda_1(a) \exp(2\delta A) \int_0^A \lambda_2(a)p_\infty(a)da \right) < 1.$$

The characteristics method implies that (22) is equivalent to the following system of integral equations

$$\begin{aligned} (23a) \quad u(a, t) &= \exp\left(\frac{-t}{\eta}\right)u_0(a-t) + \frac{1}{\eta} \int_0^t \exp\left(\frac{-(t-\tau)}{\eta}\right) \left( u(a-t+\tau, \tau) \right. \\ &\quad \left. + \eta F(u, v)(a-t+\tau, \tau) \right) d\tau, \quad a \geq t, \end{aligned}$$

$$(23b) \quad u(a, t) = \frac{1}{\eta} \int_0^a \exp\left(\frac{-(a-\tau)}{\eta}\right) \left( u(\tau, t-a+\tau) + \eta F(u, v)(\tau, t-a+\tau) \right) d\tau, \quad a < t,$$

$$(23c) \quad v(a, t) = \exp\left(\frac{-t}{\eta}\right) v_0(a-t) + \frac{1}{\eta} \int_0^t \exp\left(\frac{-(t-\tau)}{\eta}\right) \left( v(\tau, t-a+\tau) + \eta G(u, v)(a-t+\tau, \tau) \right) d\tau, \quad a \geq t,$$

$$(23d) \quad v(a, t) = \frac{1}{\eta} \int_0^a \exp\left(\frac{-(a-\tau)}{\eta}\right) \left( v(\tau, t-a+\tau) + \eta G(u, v)(\tau, t-a+\tau) \right) d\tau, \quad a < t.$$

Construct two sequences by iteration

$$(24a) \quad u^{(0)}(a, t) = u_0(a),$$

$$u^{(n+1)}(a, t) = \exp\left(\frac{-t}{\eta}\right) u_0(a-t)$$

$$(24b) \quad + \frac{1}{\eta} \int_0^t \exp\left(\frac{-(t-\tau)}{\eta}\right) \left( u^{(n)}(a-t+\tau, \tau) + \eta F(u^{(n)}, v^{(n)})(a-t+\tau, \tau) \right) d\tau, \quad a \geq t,$$

$$(24c) \quad u^{(n+1)}(a, t) = \frac{1}{\eta} \int_0^a \exp\left(\frac{-(a-\tau)}{\eta}\right) \left( u^{(n)}(\tau, t-a+\tau) + \eta F(u^{(n)}, v^{(n)})(\tau, t-a+\tau) \right) d\tau, \quad a < t.$$

$$(24d) \quad v^{(0)}(a, t) = v_0(a),$$

$$v^{(n+1)}(a, t) = \exp\left(\frac{-t}{\eta}\right) v_0(a-t)$$

$$(24e) \quad + \frac{1}{\eta} \int_0^t \exp\left(\frac{-(t-\tau)}{\eta}\right) \left( v^{(n)}(a-t+\tau, \tau) + \eta G(u^{(n)}, v^{(n)})(a-t+\tau, \tau) \right) d\tau, \quad a \geq t,$$

$$(24f) \quad v^{(n+1)}(a, t) = \frac{1}{\eta} \int_0^a \exp\left(\frac{-(a-\tau)}{\eta}\right) \left( v^{(n)}(\tau, t-a+\tau) + \eta G(u^{(n)}, v^{(n)})(\tau, t-a+\tau) \right) d\tau, \quad a < t,$$

It is not difficult to show that  $u^{(n)}(a, t)$  and  $v^{(n)}(a, t)$  converge uniformly to  $u(a, t)$  and  $v(a, t)$ , the solutions of the integral equations (22). If  $0 \leq u^{(0)}(a, t) \leq v^{(0)}(a, t) \leq 1$ ,  $0 \leq e^{\delta a} u^0(a, t) \leq 1$ ,  $0 \leq u^{(n)}(a, t) \leq v^{(n)}(a, t) \leq 1$ , and  $0 \leq e^{\delta a} u^{(n)}(a, t) \leq 1$ , then from (24b) and the selection for  $\eta$ , we have that

$$\begin{aligned}
0 &\leq u^{(n+1)}(a, t) \\
&= \exp\left(\frac{-t}{\eta}\right)u_0(a-t) + \frac{1}{\eta} \int_0^t \exp\left(\frac{-(t-\tau)}{\eta}\right) \left( u^{(n)}(a-t+\tau, \tau) \right. \\
&\quad \left. - \eta e^{-\delta(a-t+\tau)} (\gamma(a-t+\tau) + \delta) \right. \\
&\quad \left. + \eta(e^{-\delta(a-t+\tau)} - u^{(n)}(a-t+\tau, \tau)) (\gamma(a-t+\tau) + \delta) \right. \\
&\quad \left. + \lambda_1(a-t+\tau) \int_0^A \lambda_2(a') p_\infty(a') v^{(n)}(a', \tau) da' \right) d\tau \\
&\leq \exp\left(\frac{-t}{\eta} - \delta a\right) + \frac{1}{\eta} \int_0^t \exp\left(\frac{-(t-\tau)}{\eta}\right) \left( u^{(n)}(a-t+\tau, \tau) \right. \\
&\quad \left. - \eta e^{-\delta(a-t+\tau)} (\gamma(a-t+\tau) + \delta) + e^{-\delta(a-t+\tau)} - u^{(n)}(a-t+\tau, \tau) \right) d\tau \\
&\leq \exp\left(\frac{-t}{\eta} - \delta a\right) + \frac{1}{\eta} \int_0^t \exp\left(\frac{-(t-\tau)}{\eta}\right) (1 - \eta\delta) e^{-\delta(a-t+\tau)} d\tau \\
&\leq \exp\left(\frac{-t}{\eta} - \delta a\right) + (1 - \exp\left(\frac{-t}{\eta} + \delta t\right)) e^{-\delta a} \\
&\leq \exp\left(\frac{-t}{\eta} - \delta a\right) + (1 - \exp\left(\frac{-t}{\eta}\right)) e^{-\delta a} \leq e^{-\delta a} \leq 1, \quad a \geq t.
\end{aligned}$$

A similar procedure implies that

$$(25) \quad 0 \leq e^{\delta a} u^{(n+1)}(a, t) \leq 1, \quad 0 \leq v^{(n+1)}(a, t) \leq 1.$$

Based on (24), a relationship between both sequences is deduced

$$\begin{aligned}
&u^{(n+1)}(a, t) - v^{(n+1)}(a, t) \\
&= \exp\left(\frac{-t}{\eta}\right) (u_0(a-t) - v_0(a-t)) \\
(26a) \quad &+ \frac{1}{\eta} \int_0^t \left( \exp\left(\frac{-(t-\tau)}{\eta}\right) \left( (1 - \eta(\gamma(a-t+\tau) + \delta)) (u^{(n)}(a-t+\tau, \tau) \right. \right. \\
&\quad \left. \left. - v^{(n)}(a-t+\tau, \tau)) + \eta(e^{-\delta(a-t+\tau)} - 1) (1 - e^{\delta(a-t+\tau)} u^{(n)}(a-t+\tau, \tau)) \right) \right. \\
&\quad \times \lambda_1(a-t+\tau) \int_0^A \lambda_2(a') p_\infty(a') v^{(n)}(a', \tau) da' \Big) d\tau \leq 0, \quad a \geq t,
\end{aligned}$$

$$\begin{aligned}
&u^{(n+1)}(a, t) - v^{(n+1)}(a, t) \\
&= \frac{1}{\eta} \int_0^a \left( \exp\left(\frac{-(a-\tau)}{\eta}\right) \left( (1 - \eta(\gamma(\tau) + \delta)) (u^{(n)}(\tau, a-t+\tau) \right. \right. \\
(26b) \quad &\quad \left. \left. - v^{(n)}(\tau, a-t+\tau)) + \eta(e^{-\delta\tau} - 1) (1 - e^{\delta\tau} u^{(n)}(\tau, a-t+\tau)) \right) \right. \\
&\quad \times \lambda_1(\tau) \int_0^A \lambda_2(a') p_\infty(a') v^{(n)}(a', t-a+\tau) da' \Big) d\tau \leq 0, \quad a \leq t,
\end{aligned}$$

which implies  $0 \leq u^{(n)}(a, t) \leq v^{(n)}(a, t) \leq 1$ . It can be seen that  $0 \leq e^{\delta a} u^{(n)}(a, t) \leq 1$  for all integer  $n$  by induction. Hence,

$$(27a) \quad \lim_{n \rightarrow \infty} u^{(n)}(a, t) = u(a, t) \leq e^{-\delta a},$$

$$(27b) \quad \lim_{n \rightarrow \infty} v^{(n)}(a, t) = v(a, t) \leq 1,$$

$$(27c) \quad 0 \leq u(a, t) \leq v(a, t) \leq 1,$$

$$(27d) \quad 0 \leq k(a, t) \leq j(a, t) \leq p_\infty(a), \quad 0 \leq a \leq A, \quad 0 \leq t \leq T.$$

We claim that

$$u_1^{(n)}(a, t) \leq u_2^{(n)}(a, t) \quad \text{and} \quad v_1^{(n)}(a, t) \leq v_2^{(n)}(a, t)$$

hold for any integer  $n$ . Now we apply induction again to ensure the claim. Assume they are true for  $n$ , i.e.,

$$0 \leq u_1^{(n)}(a, t) \leq v_1^{(n)}(a, t) \leq 1, \quad 0 \leq u_2^{(n)}(a, t) \leq v_2^{(n)}(a, t) \leq 1,$$

and  $0 \leq u_2^{(n)}(a, t) - u_1^{(n)}(a, t) \leq e^{\delta a}(v_2^{(n)}(a, t) - v_1^{(n)}(a, t))$ .

Then a direct computation gives

$$\begin{aligned} & u_2^{(n+1)}(a, t) - u_1^{(n+1)}(a, t) \\ &= \exp\left(\frac{-t}{\eta}\right)(u_2^{(0)}(a, t) - u_1^{(0)}(a, t)) \\ & \quad + \frac{1}{\eta} \int_0^t \exp\left(\frac{-(t-\tau)}{\eta}\right) \left( (u_2^{(n)}(a-t+\tau, \tau) - u_1^{(n)}(a-t+\tau, \tau)) \right. \\ (28a) \quad & \quad \times \left( 1 - \eta \left( \gamma(a-t+\tau) + \delta + \lambda_1(a-t+\tau) \int_0^A \lambda_2(a') p_\infty(a') v_2^{(n)}(a', \tau) da' \right) \right) \\ & \quad + \eta e^{-\delta(a-t+\tau)} \lambda_1(a-t+\tau) (1 - e^{\delta(a-t+\tau)} u_1^{(n)}(a-t+\tau, \tau)) \\ & \quad \times \left. \int_0^A \lambda_2(a') p_\infty(a') (v_2^{(n)}(a', \tau) - v_1^{(n)}(a', \tau)) da' \right) d\tau \geq 0, \quad a \geq t, \end{aligned}$$

and

$$\begin{aligned} & u_2^{(n+1)}(a, t) - u_1^{(n+1)}(a, t) \\ &= \frac{1}{\eta} \int_0^a \exp\left(\frac{-(a-\tau)}{\eta}\right) \left( (u_2^{(n)}(\tau, t-a+\tau) - u_1^{(n)}(\tau, t-a+\tau)) \right. \\ (28b) \quad & \quad \times \left( 1 - \eta \left( \gamma(\tau) + \delta + \lambda_1(\tau) \int_0^A \lambda_2(a') p_\infty(a') v_2^{(n)}(a', t-a+\tau) da' \right) \right) \\ & \quad + \eta e^{-\delta\tau} \lambda_1(\tau) (1 - e^{\delta\tau} u_1^{(n)}(\tau, t-a+\tau)) \int_0^A \lambda_2(a') p_\infty(a') \\ & \quad \times \left. (v_2^{(n)}(a', t-a+\tau) - v_1^{(n)}(a', t-a+\tau)) da' \right) d\tau \geq 0, \quad t \geq a. \end{aligned}$$

The selection for  $\eta$  and  $-(u_2^{(n)}(a, t) - u_1^{(n)}(a, t)) \geq -e^\delta(v_2^{(n)}(a, t) - v_1^{(n)}(a, t))$  result in

$$\begin{aligned}
& v_2^{(n+1)}(a, t) - v_1^{(n+1)}(a, t) \\
&= \exp\left(\frac{-t}{\eta}\right)(v_2^{(0)}(a, t) - v_1^{(0)}(a, t)) + \frac{1}{\eta} \int_0^t \exp\left(\frac{-(t-\tau)}{\eta}\right) \\
&\quad \times \left( (v_2^{(n)}(a-t+\tau, \tau) - v_1^{(n)}(a-t+\tau, \tau))(1 - \eta(\gamma(a-t+\tau) + \delta)) \right. \\
(29a) \quad &\quad \left. - \eta e^{\delta(a-t+\tau)} \lambda_1(a-t+\tau) (u_2^{(n)}(a-t+\tau, \tau) \right. \\
&\quad \left. - u_1^{(n)}(a-t+\tau, \tau)) \int_0^A \lambda_2(a') p_\infty(a') v_2^{(n)}(a', \tau) da' \right. \\
&\quad \left. + \eta \lambda_1(a-t+\tau) (1 - e^{\delta(a-t+\tau)} u_1^{(n)}(a-t+\tau, \tau)) \right. \\
&\quad \left. \times \int_0^A \lambda_2(a') p_\infty(a') (v_2^{(n)}(a', \tau) - v_1^{(n)}(a', \tau)) da' \right) d\tau \geq 0, \quad a \geq t,
\end{aligned}$$

and

$$\begin{aligned}
& v_2^{(n+1)}(a, t) - v_1^{(n+1)}(a, t) \\
&= \frac{1}{\eta} \int_0^a \exp\left(\frac{-(a-\tau)}{\eta}\right) \left( (v_2^{(n)}(\tau, t-a+\tau) \right. \\
&\quad \left. - v_1^{(n)}(\tau, t-a+\tau))(1 - \eta(\gamma(\tau) + \delta)) \right. \\
(29b) \quad &\quad \left. - \eta e^{\delta\tau} \lambda_1(\tau) (u_2^{(n)}(\tau, t-a+\tau) \right. \\
&\quad \left. - u_1^{(n)}(\tau, t-a+\tau)) \int_0^A \lambda_2(a') p_\infty(a') v_2^{(n)}(a', t-a+\tau) da' \right. \\
&\quad \left. + \eta \lambda_1(\tau) (1 - e^{\delta\tau} u_1^{(n)}(\tau, t-a+\tau)) \int_0^A \lambda_2(a') p_\infty(a') \right. \\
&\quad \left. \times (v_2^{(n)}(a', t-a+\tau) - v_1^{(n)}(a', t-a+\tau)) da' \right) d\tau \geq 0, \quad t \geq a.
\end{aligned}$$

From the above expressions, we deduce that

$$\begin{aligned}
& u_2^{(n+1)}(a, t) - u_1^{(n+1)}(a, t) \\
&\leq \exp\left(\delta a + \frac{-t}{\eta}\right)(v_2^{(0)}(a, t) - v_1^{(0)}(a, t)) \\
&\quad + \frac{1}{\eta} \int_0^t \exp\left(\frac{-(t-\tau)}{\eta}\right) \left( \exp(\delta(a-t+\tau)) (v_2^{(n)}(a-t+\tau, \tau) \right. \\
(30a) \quad &\quad \left. - v_1^{(n)}(a-t+\tau, \tau)) (1 - \eta(\gamma(a-t+\tau) + \delta) \right. \\
&\quad \left. + \lambda_1(a-t+\tau) \int_0^A \lambda_2(a') p_\infty(a') v_2^{(n)}(a', \tau) da' \right) \right) \\
&\quad + \eta e^{-\delta(a-t+\tau)} \lambda_1(a-t+\tau) (1 - e^{\delta(a-t+\tau)} u_1^{(n)}(a-t+\tau, \tau)) \\
&\quad \times \int_0^A \lambda_2(a') p_\infty(a') (v_2^{(n)}(a', \tau) - v_1^{(n)}(a', \tau)) da' d\tau \\
&\leq e^{\delta a} (v_2^{(n+1)}(a, t) - v_1^{(n+1)}(a, t)), \quad a \geq t,
\end{aligned}$$

thus

$$(30b) \quad u_2^{(n+1)}(a, t) - u_1^{(n+1)}(a, t) \leq e^{\delta a} (v_2^{(n+1)}(a, t) - v_1^{(n+1)}(a, t)), \quad t \geq a$$

Therefore,

$$u_1(a, t) \leq u_2(a, t) \quad \text{and} \quad v_1(a, t) \leq v_2(a, t).$$

Hence given any constant  $0 \leq \xi < 1$ , it is readily found that

$$\xi F(u, v)(a, t) \leq F(\xi u, \xi v)(a, t), \quad \xi G(u, v)(a, t) \leq G(\xi u, \xi v)(a, t).$$

From the construction of the iterative sequences in (24), we have that

$$\xi u_1(a, t) \leq u_\xi(a, t), \quad \xi v_1(a, t) \leq v_\xi(a, t),$$

and Theorem 3 is proved.

**Theorem 4** *Assume that (A1)~(A6) hold. The positive equilibrium solution  $u_\infty(a), v_\infty(a)$  of (22) is globally stable if  $R_0 > 1$ .*

**Proof.** Using the method of characteristics, System (22) can be solved explicitly:

$$(31a) \quad \begin{aligned} u(a, t) = & u_0(a-t) \exp\left(-\int_0^t \sigma_1(a-t+\tau, \tau) d\tau\right) \\ & + \int_0^t e^{-\delta(a-t+\tau)} \lambda_1(a-t+\tau) \omega(\tau) \\ & \times \exp\left(-\int_\tau^t \sigma_1(a-t+\theta, \theta) d\theta\right) d\tau, \quad a \geq t, \end{aligned}$$

$$(31b) \quad \begin{aligned} u(a, t) = & \int_0^a e^{-\delta\tau} \lambda_1(\tau) \omega(t-a+\tau) \\ & \times \exp\left(-\int_\tau^a \sigma_1(\theta, t-a+\theta) d\theta\right) d\tau, \quad a \leq t, \end{aligned}$$

$$(31c) \quad \begin{aligned} v(a, t) = & v_0(a-t) \exp\left(-\int_0^t \sigma_2(a-t+\tau) d\tau\right) \\ & + \int_0^t (1 - e^{\delta(a-t+\tau)} u(a-t+\tau, \tau)) \lambda_1(a-t+\tau) \omega(\tau) \\ & \times \exp\left(-\int_\tau^t \sigma_2(a-t+\theta) d\theta\right) d\tau, \quad a \geq t, \end{aligned}$$

$$(31d) \quad \begin{aligned} v(a, t) = & \int_0^a (1 - e^{\delta\tau} u(\tau, t-a+\tau)) \lambda_1(\tau) \omega(t-a+\tau) \\ & \times \exp\left(-\int_\tau^a \sigma_2(\theta) d\theta\right) d\tau, \quad a \leq t, \end{aligned}$$

where,

$$\omega(t) = \int_0^A \lambda_2(a) p_\infty(a) v(a, t) da,$$

$$\begin{aligned}\sigma_1(a, t) &= \gamma(a) + \delta + \lambda_1(a)\omega(t), \\ \sigma_2(a) &= \gamma(a) + \delta.\end{aligned}$$

$\omega(t) > 0$  for  $0 \leq t \leq A$  holds from its definition. Combining (31b), (31d) and  $e^{\delta a}u(a, t) < 1(0 < a < A)$  yields  $u(a, A) > 0$  and  $v(a, A) > 0$  for  $0 < a < A$ . Continuing this process leads to  $\omega(t) > 0$  for  $0 \leq t \leq 3A$ . Again from (31b) and (31d), we obtain

$$\begin{aligned}(32a) \quad u(a, 2A) &\geq \int_0^a e^{-\delta\tau} \lambda_1(\tau)\omega_0 \\ &\times \exp\left(-\int_\tau^a \sigma_1^*(\theta, 2A - a + \theta)d\theta\right)d\tau, \quad 0 \leq a \leq A,\end{aligned}$$

$$\begin{aligned}(32b) \quad v(a, 2A) &\geq \int_0^a (1 - e^{\delta\tau}u(\tau, 2A - a + \tau))\lambda_1(\tau)\omega_0 \\ &\times \exp\left(-\int_\tau^a \sigma_2(\theta)d\theta\right)d\tau, \quad 0 < a \leq A,\end{aligned}$$

where

$$\omega_0 = \min_{A \leq t \leq 2A} \omega(t),$$

$$\sigma_1^*(a, 2A) = \gamma(a) + \delta + \lambda_1(a) \max_{A \leq t \leq 2A} \int_0^A \lambda_2(\tau)p_\infty(\tau)v(\tau, t)d\tau.$$

Moreover, from the expressions of  $u_\infty(a)$  and  $v_\infty(a)$  together (31) and (32),  $\xi$  can be chosen such that

$$\xi u_\infty(a) \leq u(a, 2A) \leq e^{-\delta a}, \quad \xi v_\infty(a) \leq v(a, 2A) \leq 1.$$

Let  $u_\xi(a, t)$ ,  $v_\xi(a, t)$ ,  $u_\delta(a, t)$ ,  $v_1(a, t)$  be the solutions of (22) satisfying  $u_\xi(a, 0) = \xi u_\infty(a)$ ,  $v_\xi(a, 0) = \xi v_\infty(a)$ ,  $u_\delta(a, 0) = e^{-\delta a}$ ,  $v_1(a, 0) = 1$ , respectively. By virtue of Theorem 3, it is clear that

$$\begin{aligned}u_\xi(a, t) &\leq u(a, t + 2A) \leq u_\delta(a, t), \\ v_\xi(a, t) &\leq v(a, t + 2A) \leq v_1(a, t).\end{aligned}$$

The fact that  $u_\infty(a)$  and  $v_\infty(a)$  are equilibrium solution implies that

$$\begin{aligned}\xi u_\infty(a) &\leq u_\xi(a, t), \quad u_\delta(a, t) \leq e^{-\delta a}, \\ \xi v_\infty(a) &\leq v_\xi(a, t), \quad v_1(a, t) \leq 1,\end{aligned}$$

which further imply that  $u_\xi(a, t)$ ,  $v_\xi(a, t)$  are increasing, and conversely  $u_\delta(a, t)$ ,  $v_1(a, t)$  are decreasing. Therefore,  $\omega_\xi(t) = \int_0^A \lambda_2(a)p_\infty(a)v_\xi(a, t)da$  is increasing and  $\omega_1(t) = \int_0^A \lambda_2(a)p_\infty(a)v_1(a, t)da$  is decreasing. Hence  $\omega_\xi(t)$  and  $\omega_1(t)$  must approach the same limit  $\omega_\infty^*$ , otherwise, the positive steady-state for (6) would not be unique. Hence the positive steady-state of (22) is globally asymptotically stable. This ends the proof of Theorem 4.

From the assumptions and (2b),  $G(a, t)$  can be expressed in terms of  $u(a, t)$  and  $v(a, t)$  as

$$\begin{aligned} G(a, t) &= (p_\infty(a) - j(a, t)) \int_0^A \lambda_1(a) \lambda_2(a') k(a', t) da' \\ &= \lambda_1(a) p_\infty(a) (1 - e^{\delta a} u(a, t)) \int_0^A \lambda_2(a') p_\infty(a') v(a', t) da'. \end{aligned}$$

Consequently, the solution of (2) is of the form

$$i(a, c, t) = G(a - c, t - c) \pi(a', c), \quad t \geq A.$$

where  $G(a, t)$  is given above. Hence, the positive steady-state of (2) is globally asymptotically stable whenever  $R_0 > 1$ .

**4. Concluding remarks.** A general age-structured epidemic model is modified via the introduction of the infection age. The stability of steady-states and the uniqueness and existence of endemic steady-state for an SIS epidemic model are established. The basic reproductive number increases our understanding of the effect of infection age. Traditional single-age-structured model attempts to over estimate the disease severity. A two-age-structured SI epidemic model recently studied by Brauer [29] has also indicated this over estimation when the ratio of mean age at infection to the mean life span is very large.

It has been found that endemic equilibria are globally stable for epidemic models given by monotone ODE systems by Feng, Castillo-Chavez and Huang [30], Song and Castillo-Chavez [31], and Li [32]. Our work is an extension of this global dynamics to PDE system. We also expand the work of Thieme, Busenberg, Iannelli [9] to PDE systems with two-age structures. The approach of constructing iterate sequences to establish the global dynamics here may be useful in similar analysis of general epidemic models.

In the future, we intend to extend this approach to more realistic situations. For instance, we would like to apply it to fatal diseases and to situations that can handle general age distribution infection.

**Acknowledgments.** *The valuable comments from Carlos Castillo-Chavez on our two manuscripts have improved this paper to a large extent. This work was partially supported by NSF and NSA grants to the Mathematical and Theoretical Biology Institute at Cornell University and the office of the Provost of Cornell University and by CNSF Grant 19971066 to Zhien Ma.*

## 5. Appendix.

A rigorous proof of Theorem 1 is provided here.

**Proof.** Expression (6) helps us to equivalently show  $G(a, t) \geq 0$ . For the solution  $i(a, c, t)$  of (5) with the initial distribution satisfying the condition in Theorem 1, we define the age-structured distribution of the infectives by

$$j(a, t) = \int_0^a i(a, c, t) dc = \int_0^a G(a - c, t - c) \frac{N(a)}{N(a - c)} dc$$

$$= \int_0^a G(\tau, t - a + \tau) \frac{N(a)}{N(\tau)} d\tau,$$

which satisfies

$$(33a) \quad \frac{\partial j}{\partial a} + \frac{\partial j}{\partial t} = -(\mu(a) + \gamma(a))j(a, t) + G(a, t),$$

$$(33b) \quad j(0, t) = 0,$$

$$(33c) \quad j(a, 0) = j_0(a) = \int_0^a i_0(a, c) dc.$$

$u(a, t) = j(a, t)/p_\infty(a)$  is a normalization of  $j(a, t)$ , then

$$(34a) \quad \frac{\partial u}{\partial a} + \frac{\partial u}{\partial t} = -\gamma(a)u(a, t) + (1 - u(a, t)) \int_0^A \int_0^{a'} \lambda(a, a', c)i(a', c, t) dc da',$$

$$(34b) \quad u(0, t) = 0,$$

$$(34c) \quad u(a, 0) = u_0(a) = j_0(a)/p_\infty(a).$$

For any given  $T (0 < T < A)$  it follows from the given conditions of Theorem 1 that  $i(a, c, t)$  is bounded for  $0 \leq c \leq a \leq A$ ,  $0 \leq t \leq T$ , and  $0 \leq u_0(a) \leq 1$ . Choose any positive constant  $\eta$  and rewrite Equation (34a) as

$$\begin{aligned} \frac{\partial u}{\partial a} + \frac{\partial u}{\partial t} &= -\frac{u(a, t)}{\eta} + \frac{u(a, t)}{\eta} - \gamma(a)u(a, t) \\ &\quad + (1 - u(a, t)) \int_0^A \int_0^{a'} \lambda(a, a', c)i(a', c, t) dc da', \end{aligned}$$

from which an equivalent integral equations is deduced

$$\begin{aligned} (35a) \quad u(a, t) &= \exp\left(-\frac{t}{\eta}\right) u_0(a - t) + \frac{1}{\eta} \int_0^t \exp\left(-\frac{(t - \tau)}{\eta}\right) \left( u(a - t + \tau, \tau) \right. \\ &\quad \left. + \eta \left( -\gamma(a - t + \tau)u(a - t + \tau, \tau) + (1 - u(a - t + \tau, \tau)) \right. \right. \\ &\quad \left. \times \int_0^A \int_0^{a'} \lambda(a - t + \tau, a', c)i(a', c, \tau) dc da' \right) d\tau, \quad a \geq t, \end{aligned}$$

$$\begin{aligned} (35b) \quad u(a, t) &= \frac{1}{\eta} \int_0^a \exp\left(-\frac{(a - \tau)}{\eta}\right) \left( u(\tau, t - a + \tau) \right. \\ &\quad \left. + \eta \left( -\gamma(\tau)u(\tau, t - a + \tau) + (1 - u(\tau, t - a + \tau)) \right. \right. \\ &\quad \left. \times \int_0^A \int_0^{a'} \lambda(\tau, a', c)i(a', c, t - a' + \tau) dc da' \right) d\tau, \quad a < t, \end{aligned}$$

where  $\eta$  is a positive constant, such that  $\eta \left( \gamma(a) + \int_0^A \int_0^{a'} \lambda(a, a', c) |i(a', c, t)| dc da' \right) < 1$ . We apply the routine iterative procedure

$$(36a) \quad u^{(0)}(a, t) = u_0(a),$$

$$\begin{aligned}
(36b) \quad & u^{(n+1)}(a, t) = \\
& \exp\left(\frac{-t}{\eta}\right) u_0(a-t) + \frac{1}{\eta} \int_0^t \exp\left(\frac{-(t-\tau)}{\eta}\right) \left( u^{(n)}(a-t+\tau, \tau) \right. \\
& \left. + \eta \left( -\gamma(a-t+\tau) u^{(n)}(a-t+\tau, \tau) + (1-u^{(n)}(a-t+\tau, \tau)) \right. \right. \\
& \times \int_0^A \int_0^{a'} \lambda(a-t+\tau, a', c) i(a', c, \tau) dc da' \Big) d\tau, \quad a \geq t,
\end{aligned}$$

$$\begin{aligned}
(36c) \quad & u^{(n+1)}(a, t) = \frac{1}{\eta} \int_0^a \exp\left(\frac{-(a-\tau)}{\eta}\right) \left( u^{(n)}(\tau, t-a+\tau) \right. \\
& \left. + \eta \left( -\gamma(\tau) u^{(n)}(\tau, t-a+\tau) + (1-u^{(n)}(\tau, t-a+\tau)) \right. \right. \\
& \times \int_0^A \int_0^{a'} \lambda(\tau, a', c) i(a', c, t-a+\tau) dc da' \Big) d\tau, \quad a < t.
\end{aligned}$$

One can easily show that the iterative sequence  $u^{(n)}(a, t)$  converges uniformly to the solution of the integral equation (35).

If  $0 \leq u^{(n)}(a, t) \leq 1$ , then we can obtain from (36b) and (36c) that

$$\begin{aligned}
(37a) \quad & u^{(n+1)}(a, t) \leq \exp\left(\frac{-t}{\eta}\right) u_0(a-t) \\
& + \frac{1}{\eta} \int_0^t \exp\left(\frac{-(t-\tau)}{\eta}\right) \left( u^{(n)}(a-t+\tau, \tau) \right. \\
& \left. + \eta \left( \gamma(a-t+\tau) (1-u^{(n)}(a-t+\tau, \tau)) \right. \right. \\
& \left. \left. + (1-u^{(n)}(a-t+\tau, \tau)) \right. \right. \\
& \times \int_0^A \int_0^{a'} \lambda(a-t+\tau, a', c) |i(a', c, \tau)| dc da' \Big) d\tau \\
& \leq \exp\left(\frac{-t}{\eta}\right) u_0(a-t) + \frac{1}{\eta} \int_0^t \exp\left(\frac{-(t-\tau)}{\eta}\right) \left( u^{(n)}(a-t+\tau, \tau) \right. \\
& \left. + (1-u^{(n)}(a-t+\tau, \tau)) \right. \eta \left( \gamma(a-t+\tau) \right. \\
& \left. + \int_0^A \int_0^{a'} \lambda(a-t+\tau, a', c) |i(a', c, \tau)| dc da' \right) d\tau \\
& \leq \exp\left(\frac{-t}{\eta}\right) + \frac{1}{\eta} \int_0^t \exp\left(\frac{-(t-\tau)}{\eta}\right) \left( u^{(n)}(a-t+\tau, \tau) \right. \\
& \left. + (1-u^{(n)}(a-t+\tau, \tau)) \right) d\tau \\
& \leq \exp\left(\frac{-t}{\eta}\right) + \frac{1}{\eta} \int_0^t \exp\left(\frac{-(t-\tau)}{\eta}\right) d\tau \leq 1, \quad a \geq t,
\end{aligned}$$

$$(37b) \quad u^{(n+1)}(a, t) \leq \frac{1}{\eta} \int_0^a \exp\left(\frac{-(a-\tau)}{\eta}\right) d\tau \leq 1, \quad a > t.$$

where Inequality (37b) is deduced by the same process as does for (37a). Hence, induction implies that  $u^{(n)}(a, t) \leq 1$  for all integer  $n$ , by which we arrive at

$$(38a) \quad \lim_{n \rightarrow \infty} u^{(n)}(a, t) = u(a, t) \leq 1,$$

$$(38b) \quad j(a, t) \leq p_\infty(a), \quad 0 \leq a \leq A, \quad 0 \leq t \leq T.$$

Recalling Equation (6) and (7), we find

$$(39a) \quad G(a, t) = f(a, t) + h(a, t) \int_0^t \int_c^A \lambda(a, a', c) G(a' - c, t - c) \pi(a', c) da' dc,$$

where

$$(39b) \quad f(a, t) = h(a, t) \int_t^A \int_c^A \lambda(a, a', c) i_0(a' - t, c - t) \pi(a', t) da' dc,$$

$$(39c) \quad h(a, t) = p_\infty(a) - j(a, t), \quad 0 \leq a \leq A, \quad 0 \leq t \leq A.$$

$f(a, t)$  and  $h(a, t)$  are nonnegative continuous and bounded on  $D = [0, A] \times [0, A]$ . In order to prove the nonnegativity of  $G(a, t)$  we built a new sequence by iteration as follows

$$(40a) \quad G^{(0)}(a, t) = i_0(a, t), \quad 0 \leq t \leq a \leq A,$$

$$(40b) \quad G^{(0)}(a, t) = i_0(a, a), \quad 0 \leq a \leq t \leq A,$$

$$(40c) \quad \begin{aligned} G^{(n)}(a, t) &= f(a, t) + h(a, t) \\ &\quad \times \int_0^t \int_c^A \lambda(a, a', c) G^{(n-1)}(a' - c, t - c) \pi(a', c) da' dc, \\ &\quad (a, t) \in D, \quad n = 1, 2, \dots \end{aligned}$$

Owing to the fact that  $G^{(0)}(a, t)$  is a nonnegative and continuous on  $D$ ,  $G^{(n)}(a, t)$  are also nonnegative and continuous on  $D$  for all  $n$ . Denoting  $h^* = \max_{(a,t) \in D} h(a, t)$ , we can estimate the norm between two consecutive terms in the sequence, resulting in

$$\|G^{(n+1)}(a, t) - G^{(n)}(a, t)\| \leq h^* A \lambda^* \|G^{(n)}(a, t) - G^{(n-1)}(a, t)\| t.$$

Inequality (41) together with the induction imply that

$$\|G^{(n+1)}(a, t) - G^{(n)}(a, t)\| \leq (h^* A \lambda^*)^n \|G^{(1)}(a, t) - G^{(0)}(a, t)\| t^n / n!,$$

from which it follows that  $G^{(n)}(a, t)$  converges uniformly on  $D$ . And thus  $G(a, t) = \lim_{n \rightarrow \infty} G^{(n)}(a, t) \geq 0$  holds for  $0 \leq t \leq A$ . This is also true for all  $t$  by repeating the process. We finish the proof of Theorem 1.

## REFERENCES

- [1] W.O. KERMACK AND A.G. MCKENDRICK, Contributions to the theory of epidemics, *Proc. Roy. Soc., A*, 115, 700–721 (1927).
- [2] A.G. MCKENDRICK, Applications of mathematics to medical problems. *Proc. Edinburgh Math. Soc.*, 44, 98–130 (1926).
- [3] F. HOPPENSTEADT, An age dependent epidemic model, *J. Franklin Inst.*, 297, 325–333 (1974).
- [4] F. HOPPENSTEADT, Mathematical theories of populations: demographics, genetics and epidemics, SIAM, Philadelphia (1975).
- [5] P. WALTMAN, Deterministic threshold models in the theory of epidemics, *Lecture Notes in Biomathematics 1*, Springer-Verlag, Berlin, New York (1974).
- [6] E. VYNNYCKY AND P.E. FINE, The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. *Epidemiol. Infect.*, 19, 183–201 (1997).
- [7] C. CASTILLO-CHAVEZ AND Z. FENG, Global stability of an age-structure model for TB and its application to optimal vaccination strategies, *Math. Bioscie.*, 151, 135–154 (1998).
- [8] S. BUSENBERG, K. COOKE, AND M. IANNELLI, Endemic thresholds and stability in a class of age-structured epidemics, *SIAM J. Appl. Math.*, 48(6), 1379–1395 (1988).
- [9] S. BUSENBERG, M. IANNELLI, AND H.R. THIEME, Global behavior of an age-structured epidemic model, *SIAM J. Math. Anal.*, 22(4), 1065–1080 (1991).
- [10] M. IANNELLI, F.A. MILNER, AND A. PUGLIESE, Analytical and numerical results for the age-structured S-I-S epidemic model with mixed inter-intracohort transmission, *SIAM J. Math. Anal.*, 23(3), 662–688 (1992).
- [11] Y. CHA, M. IANNELLI, AND F.A. MILNER, Existence and uniqueness of endemic states for the age-structured S-I-R epidemic model, *Math. Bioscie.*, 150, 177–190 (1998).
- [12] H.W. HETHCOTE, A thousand and one epidemic models, in: *Frontiers in Mathematical Biology*, Simon A. Levin eds, *Lecture Notes in Biomathematics 100*, Springer-Verlag, Berlin, Heidelberg, 504–515 (1994).
- [13] V. CAPASSO, Mathematical structures of epidemic systems, *Lecture Notes in Biomathematics 97*, Springer-Verlag, Berlin, Heidelberg (1993).
- [14] M. IANNELLI, Mathematical theory of age-structured population dynamics, Giardini Editori, E Stampatori, in Pisa, 1995.
- [15] C. CASTILLO-CHAVEZ, H.W. HETHCOTE, V. ANDREASEN, S.A. LEVIN, AND W.M. LIU, Epidemiological models with age-structure, proportionate mixing, and cross-immunity, *J. Math. Biol.*, 27, 233–258 (1989).
- [16] H.W. HETHCOTE AND J.A. YORKE, Gonorrhea, transmission, dynamics, and control, *Lecture Notes in Biomathematics, 56*, Springer-Verlag, Berlin, New York (1984).
- [17] K. DIETZ AND D. SCHENZLE, Proportionate mixing models for age-dependent infection transmission, *J. of Math. Biol.*, 22, 117–120 (1985).
- [18] H.R. THIEME AND C. CASTILLO-CHAVEZ, How may infection-age-dependent infectivity affect the dynamics of HIV/AIDS? *SIAM J. Appl. Math.*, 53(5), 1447–1479 (1993).
- [19] O. DIEKMANN, J.A.P. HEESTERBEEK, AND J.A.J. METS, On the definition and the computation of basic reproductive ratio  $R_0$  in models for infectious diseases in heterogeneous population, *J. Math. Biol.* 28, 365–382 (1990).
- [20] S. BUSENBERG AND C. CASTILLO-CHAVEZ, 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 (1990).
- [21] SONG JIAN AND YU JINGYUAN, Population system control, Springer-Verlag, Berlin (1987).

- [22] J.M. CUSHING, An introduction to structured population dynamics, Society for Industrial and Applied Mathematics, Philadelphia (1998).
- [23] M.E. GURTIN AND R.C. MACCAMY, Nonlinear age-dependent population dynamics, *Archive for Rational Mechanics and Analysis*, 54(3), 281–300 (1974).
- [24] G.F. WEBB, Theory of nonlinear age-dependent population dynamics, Marcel Dekker, New York (1985).
- [25] J. PRUSS, Stability analysis for equilibria in an age-specific population dynamics, *Nonlinear Analysis TMA*, 7(12), 1291–1313 (1983).
- [26] Y. ZHOU, The continuous solutions and average net reproductive number of an age structured population model, in: Advanced topics in biomathematics, Lansun Chen, Shigui Ruan, Jun Zhu, eds., World Scientific, Singapore, 315–320 (1998).
- [27] Y. ZHOU AND Z. MA, The global stability of the positive equilibrium of a nonlinear age-structured population model, *J. of Xi'an Jiaotong University*, 33(10), 87–90, 1999.
- [28] S.L. TUCKER AND S.O. ZIMMERMAN, A nonlinear model of population dynamics containing an arbitrary number of continuous structure variables, *SIAM J. Appl. Math.*, 48(3), 549–591 (1988).
- [29] F. BRAUER, Infectious diseases models with chronological age structure and epidemiological age structure (this volume).
- [30] Z. FENG, C. CASTILLO-CHAVEZ, AND W. HUANG, On the role of variable latent period in mathematical models for tuberculosis, *Journal of Dynamics and Differential Equations* (in process), Vol. 13 (2001).
- [31] B. SONG, C. CASTILLO-CHAVEZ, AND J.P. APARICIO, Global Dynamics of TB Models with Density Dependent Demography (this volume).
- [32] M.Y. LI, J.R. GRAEF, L.C. WANG, AND J. KARSAI, Global dynamics of an SEIR model with varying total population size, *Math. Biosci.*, 160, 191–213 (1999).

# ENDEMIC THRESHOLD AND STABILITY IN AN EVOLUTIONARY EPIDEMIC MODEL

HISASHI INABA\*

**Abstract.** In this paper, our main purpose is to investigate mathematical aspects of the Pease's evolutionary epidemic model for type A influenza. First we formulate the Pease model as an abstract semilinear Cauchy problem and construct the semigroup solution. Next we prove existence and uniqueness of the endemic steady state and show endemic threshold phenomena. Subsequently by using semigroup approach, we investigate the local stability of endemic steady state. We prove that the endemic steady state is locally asymptotically stable if its prevalence is greater than fifty percent. Finally we discuss some possible extensions of the Pease's model and open problems.

**1. Introduction.** In most traditional epidemic models, genetic or evolutionary changes in the virus are not taken into account. In fact for common childhood diseases such as measles, mumps and rubella, it is a reasonable assumption that the genetic characters of the virus which could affect to its epidemics does not change.

However, if we consider epidemics like type A influenza, genetic changes in the virus are thought to play an important role in causing recurrent epidemic. In the type A influenza epidemic, the virus changes genetically, and hence immunologically from one epidemic to the next. Therefore a descendant virus strain can infect hosts who are immune to the progenitor strain diseases, and hence reinvoke communities that recently suffered an epidemic of the progenitor strain. It is also observed that the more a virus has changed genetically from its progenitor, the more easily it will be able to reinfect a host that is immune to its progenitor. The evolutionary mechanism would be one of most important factors which reemerge infectious diseases.

In the influenza virus the terms *drift* and *shift* are used to distinguish two alternative mechanism of evolution. Drift occurs by point mutation and possibly by short deletions and insertions, and it causes continual, gradual changes in the influenza antigens. Shift occurs irregularly and causes abrupt and large changes in the influenza antigens. Pease (1987) has first proposed a mathematical model which can take into account drift effect, and suggested that the evolutionary mechanism could lead a dampening epidemic oscillation. After that, several authors have tried to develop evolutionary epidemic models (Andreasen, *et al.* 1996, 1997; Inaba 1998), but many open problems still remain for the evolutionary epidemic models.

In this paper, our main purpose is to investigate mathematical aspects of the Pease's evolutionary epidemic model. In the next section, we formulate the Pease model as an abstract semilinear Cauchy problem to construct

---

\*Department of Mathematical Sciences, University of Tokyo, 3-8-1 Komaba Meguro-ku, Tokyo 153-8914, Japan; E-mail: inaba@ms.u-tokyo.ac.jp.

the semigroup solution, by which we establish mathematical well-posedness of the model. In section 3, we prove existence and uniqueness of the endemic steady state and show endemic threshold phenomena, that is, we can define the basic reproduction number such that if it is less than unity, there is no endemic steady state and the disease is eradicated, otherwise there exists a unique endemic steady state. Subsequently again by using semigroup approach, we study the local stability of endemic steady state. We prove that the endemic steady state is locally asymptotically stable if its prevalence is greater than fifty percent, while we conjecture that for realistic parameter values, there is the possibility that the steady state could be destabilized and lead to a periodic solution. Finally we discuss some possible extensions of the Pease's model and open problems. We will show that the Pease model can be seen as a kind of the variable susceptibility model which was first proposed by Kermack and McKendrick in the 1930s. Some technical proofs of propositions are given in Appendix.

**2. The evolutionary epidemic model.** According to Pease (1987) we make three major biological assumptions to formulate the evolutionary epidemic model for type A influenza: First the probability of reinfection is a monotone increasing function of the number of amino acid substitutions between the immunizing and challenge virus strains. Though Pease's original assumption is that the probability is proportional to the number of amino acid substitutions, we assume that the infection rate is upper bounded, since the arbitrarily large susceptibilities seem unrealistic as Pease pointed out. Second, only one virus strain circulates in a human community at any one time. Third, random drift, and not frequency-dependent selection by the host, causes amino acid substitutions to occur in the influenza virus. Random drift occurs continually and causes gradual changes in the virus antigens, thereby genetic changes in the pathogen from epidemic to epidemic cause previously immune hosts to become susceptible. Let  $I(t)$  be the number of infected hosts at time  $t$  and let  $S(t, a)$  be the density of uninfected hosts, so that

$$\int_{a_0}^{a_1} S(t, a) da,$$

is the number of uninfected hosts that were last infected by a virus which differed by more than  $a_0$  and less than  $a_1$  amino acid substitution from the virus strain prevailing at time  $t$ . We assume that the number of amino acid substitution is a continuous variable, and it is causing the antigenic drift in the virus strain. The host population size  $N(t)$  at time  $t$  is given by

$$(2.1) \quad N(t) = \int_0^\infty S(t, a) da + I(t).$$

Then the evolutionary epidemic model is formulated by the following integrodifferential equations:

$$(2.2) \quad \frac{\partial S(t, a)}{\partial t} + k \frac{\partial S(t, a)}{\partial a} = -\gamma(a)S(t, a)I(t), \quad t > 0, \quad a > 0,$$

$$(2.3) \quad \frac{dI(t)}{dt} = -vI(t) + I(t) \int_0^\infty \gamma(a)S(t, a)da, \quad t > 0,$$

$$(2.4) \quad kS(t, 0) = vI(t), \quad t > 0,$$

$$(2.5) \quad S(0, a) = S_0(a), \quad a > 0,$$

where  $v$  is the rate at which infected hosts recover,  $k$  is the (constant) rate at which amino acid substitutions occur in the virus population,  $\gamma(a)$  specifies how amino acid substitutions affect the probability of reinfection and  $S_0(a)$  is a given initial data. Note that demography of the host population is neglected in the above modeling. In order to reflect the basic biological assumptions for the infection rate, we adopt the following mathematical assumption:

**ASSUMPTION 2.1.** *The transmission rate  $\gamma(a)$  is not identically zero, and it is a nonnegative, non-decreasing continuous function with a finite upper bound, that is, there exists a number  $\gamma(\infty) := \lim_{a \rightarrow \infty} \gamma(a) > 0$ .*

If we are concerned with the classical solution of (2.2)–(2.5), it is natural to assume that  $t \rightarrow S(t, *)$  takes a value in  $W_+^{1,1}(\mathbf{R}_+)$ , where  $W_+^{1,1}(\mathbf{R}_+)$  denotes the positive cone of the Sobolev space  $W^{1,1}(\mathbf{R}_+)$ , which is the set of integrable functions on  $\mathbf{R}_+ = [0, \infty)$  that have first order generalized derivative belonging to  $L^1$ . In this case, we have  $S(t, \infty) = 0$  and the total size of the host population is constant,  $N(t) = N$ . The reader may refer to Appendix for its proof. Thus we can write

$$(2.6) \quad I(t) = N - \int_0^\infty S(t, a)da,$$

then (2.2)–(2.5) can be reduced to the following initial-boundary value problem of an equation for the susceptible hosts:

$$(2.7) \quad \frac{\partial S(t, a)}{\partial t} + k \frac{\partial S(t, a)}{\partial a} = -\gamma(a)S(t, a) \left[ N - \int_0^\infty S(t, a)da \right],$$

$$(2.8) \quad kS(t, 0) = v \left[ N - \int_0^\infty S(t, a)da \right],$$

$$(2.9) \quad S(0, a) = S_0(a).$$

Conversely, if we define  $I(t)$  by (2.6), we can recover the original model (2.2)–(2.5) from (2.7)–(2.9). Therefore, instead of (2.2)–(2.5), we can consider the above system (2.7)–(2.9). From the biological meanings, the state space of the above initial-boundary value problem is given as

$$(2.10) \quad X := \{u \in L_+^1(\mathbf{R}_+) : \|u\|_{L^1} \leq N\},$$

where  $L_+^1(\mathbf{R}_+)$  is the positive cone of the space of integrable functions on  $\mathbf{R}_+$ .

In order to establish the well-posedness of the initial-boundary value problem (2.7)–(2.9), we consider the system (2.7)–(2.9) as a semilinear Cauchy problem on a Banach space  $L^1(\mathbf{R}_+)$ . First observe a trivial case. If an initial data  $S(0, a) = S_0(a) \in X$  satisfies the condition  $\int_0^\infty S_0(a)da = N$ , that is, there is no infected population at initial time, (2.7)–(2.9) has a trivial solution as

$$(2.11) \quad S(t, a) = \begin{cases} S_0(a - kt), & a > kt, \\ 0, & a < kt. \end{cases}$$

Next suppose that the initial data satisfies the condition  $\int_0^\infty S_0(a)da \neq N$ . Let us choose one trivial solution  $V(t, a)$  of (2.7)–(2.9) with initial data  $V_0(a) \in X$  such that  $\int_0^\infty V_0(a)da = N$ ,  $V_0(a) \neq S_0(a)$  and introduce a new variable  $x(t, a)$  as

$$(2.12) \quad x(t, a) := V(t, a) - S(t, a).$$

Then it is easy to see that the system (2.7)–(2.9) can be rewritten into the following new system with homogeneous boundary condition:

$$(2.13) \quad \frac{\partial x(t, a)}{\partial t} + k \frac{\partial x(t, a)}{\partial a} = \gamma(a)(V(t, a) - x(t, a)) \int_0^\infty x(t, a)da,$$

$$(2.14) \quad x(t, 0) = -\frac{v}{k} \int_0^\infty x(t, a)da,$$

$$(2.15) \quad x(0, a) = V_0(a) - S_0(a).$$

To construct the semigroup solutions for (2.13)–(2.15) in  $L^1(\mathbf{R}_+)$ , let us define a linear operator  $A$  with domain  $\mathcal{D}(A)$  as follows:

$$(2.16) \quad (Af)(a) := -kf'(a),$$

$$\mathcal{D}(A) := \left\{ f \in W^{1,1}(\mathbf{R}_+) : f(0) = -\frac{v}{k} \int_0^\infty f(a)da \right\}.$$

Moreover we define a time-dependent bounded operator  $B(t) : L^1(\mathbf{R}_+) \rightarrow L^1(\mathbf{R}_+)$  and a nonlinear operator  $F : L^1(\mathbf{R}_+) \rightarrow L^1(\mathbf{R}_+)$  by

$$(2.17) \quad (B(t)f)(a) := \gamma(a)V(t, a) \int_0^\infty f(a)da,$$

$$(2.18) \quad F(f)(a) := -\gamma(a)f(a) \int_0^\infty f(a)da.$$

Then under the Assumption 2.1 it is easy to see that the operator  $F$  is locally Lipschitz continuous, that is, there exists an increasing function  $L(r)$  such that  $\|F(f) - F(g)\|_{L^1} \leq L(r)\|f - g\|_{L^1}$  for all  $f, g \in \{f \in$

$L^1(\mathbf{R}_+) : \|f\|_{L^1} \leq r\}$ . And for any  $f \in L^1$ ,  $B(t)f$  is strongly continuous for  $t \geq 0$ .

In order that  $S(t, *) = V(t, *) - x(t, *)$  takes a value in  $X$ , the value  $x(t, *)$  should be included in the following subset for all  $t \geq 0$ :

$$(2.19) \quad \Omega_t := \left\{ f \in L^1(\mathbf{R}_+) : V(t, a) \geq f(a), \int_0^\infty f(a)da \geq 0 \right\}.$$

Under the above setting, the initial-boundary value problem (2.13)–(2.15) can be formulated as an abstract semilinear time-inhomogeneous Cauchy problem in  $L^1(\mathbf{R}_+)$ :

$$(2.20) \quad x'(t) = Ax(t) + B(t)x(t) + F(x(t)), \quad x(0) = x_0 \in \Omega_0.$$

For the linear operator  $A$ , we can prove the following technical lemma, its proof is given in Appendix:

**LEMMA 2.2.** *The operator  $A$  is a densely defined closed linear operator in  $L^1(\mathbf{R}_+)$  and it generates a strongly continuous, bounded semigroup.*

Since  $B(t)$  is a bounded linear operator and the nonlinear term  $F$  is continuously Fréchet differentiable on  $L^1$ , it is well known that for each  $x_0 \in L^1$  there exists a maximal interval of existence  $[0, t_0)$  and a unique continuous mild solution  $x(t; x_0) \in L^1, t \in [0, t_0)$  for (2.20) such that

$$(2.21) \quad x(t; x_0) = e^{tA}x_0 + \int_0^t e^{(t-s)A} \{B(s)x(s; x_0) + F(x(s; x_0))\} ds,$$

for all  $t \in [0, t_0)$  and either  $t_0 = \infty$  or  $\lim_{t \uparrow t_0} \|x(t; x_0)\|_{L^1} = \infty$ . Moreover if we restrict the initial data as  $x_0 \in \mathcal{D}(A)$ , then  $x(t; x_0) \in \mathcal{D}(A)$  and  $x(t; x_0)$  is continuously differentiable and satisfies (2.20) on  $[0, t_0)$ .

Let  $x(t; x_0)$  be a classical solution of (2.20) with  $x_0 \in \Omega_0 \cap \mathcal{D}(A)$  and let us introduce a function  $S(t, a)$  by  $S(t, a) := V(t, a) - x(t; x_0)(a)$ . Then it is easily seen that for almost all  $(t, a) \in \mathbf{R}_+ \times \mathbf{R}_+$ ,  $S(t, a)$  satisfies the following system:

$$(2.22) \quad \frac{\partial S(t, a)}{\partial t} + k \frac{\partial S(t, a)}{\partial a} = -\gamma(a)S(t, a)I(t),$$

$$(2.23) \quad S(t, 0) = \frac{v}{k}I(t),$$

where  $I(t)$  is defined by

$$(2.24) \quad I(t) := \int_0^\infty x(t; x_0)(a)da.$$

It follows from (2.22)–(2.23) that  $S(t, a)$  is expressed as

$$(2.25) \quad S(t, a) = \begin{cases} \frac{v}{k}I(t - \frac{a}{k})e^{-\int_0^{\frac{a}{k}} \gamma(k\sigma)I(t - \frac{a}{k} + \sigma)d\sigma}, & a < kt, \\ S_0(a - kt)e^{-\int_0^t \gamma(a - kt + k\sigma)I(\sigma)d\sigma}, & a > kt. \end{cases}$$

Then it follows from (2.24)–(2.25) that we obtain an integral equation for  $I(t)$  as follows:

$$(2.26) \quad \begin{aligned} I(t) = & N - \frac{v}{k} \int_0^{kt} I\left(t - \frac{a}{k}\right) e^{-\int_0^{\frac{a}{k}} \gamma(k\sigma) I(t - \frac{a}{k} + \sigma) d\sigma} da \\ & - \int_{kt}^{\infty} S_0(a - kt) e^{-\int_0^a \gamma(a - kt + k\sigma) I(\sigma) d\sigma} da. \end{aligned}$$

As is shown in Inaba (1998), integral equation (2.26) has a unique positive continuous solution  $I(t) \geq 0$  for all  $t \geq 0$  if  $x_0 \in \Omega_0$ . Hence we know that  $S(t, a) = V(t, a) - x(t; x_0) \geq 0$  if  $x_0 \in \Omega_0$ . Then we conclude that  $x(t; x_0) \in \Omega_t$  for all  $t \geq 0$  if  $x_0 \in \Omega_0$ . From this fact, we know that the solution  $x(t; x_0)$  with  $x_0 \in \Omega_0$  is in fact a global solution. In summary, we know that the following well-posedness result holds:

**PROPOSITION 2.3.** *For any initial data  $x_0 \in \Omega_0 \cap \mathcal{D}(A)$ , the semilinear time-inhomogeneous problem (2.20) has a unique global classical solution  $x(t; x_0) \in \Omega_t$  and  $S(t, a) := V(t, a) - x(t; x_0)$  gives a global solution for (2.7)–(2.9).*

**3. The endemic threshold.** In this section first we consider the existence and uniqueness of endemic steady state of the evolutionary epidemic model. Next we introduce the idea of prevalence of the epidemic and show the relation between the prevalence and the time scale of the epidemic. Thirdly we consider the invasion condition by using the effective reproduction number.

Let  $(S^*(a), I^*)$  be a steady state solution for the system (2.7)–(2.9). That is, we have

$$(3.1) \quad \frac{dS^*(a)}{da} = -\frac{\gamma(a) I^*}{k} S^*(a),$$

$$(3.2) \quad kS^*(0) = vI^*,$$

$$(3.3) \quad I^* = N - \int_0^{\infty} S^*(a) da.$$

From (3.1)–(3.2), we easily obtain that

$$(3.4) \quad S^*(a) = \frac{vI^*}{k} e^{-\frac{I^*}{k} \int_0^a \gamma(\sigma) d\sigma}.$$

Then we know that there is no disease-free steady state, which is an important character different from traditional epidemic models. From (3.3)–(3.4) we arrive at the following equation for  $I^*$ :

$$(3.5) \quad N = I^* + \int_0^{\infty} S^*(a) da = I^* \left( 1 + \frac{v}{k} \int_0^{\infty} e^{-\frac{I^*}{k} \int_0^a \gamma(\sigma) d\sigma} da \right).$$

Therefore if the equation (3.5) has a positive root  $I^* > 0$ , the corresponding endemic steady state is given by (3.4).

Since there is no disease-free steady state, it seems that the standard threshold argument in epidemic modeling could not be applied to our model. However it is reasonable to assume that the upper bound  $\gamma(\infty)$  is the transmission rate from infecteds to the susceptibles who have never been infected, hence the basic reproduction number of the epidemic could be defined by

$$(3.6) \quad R_0 = \frac{\gamma(\infty)N}{v}.$$

By using the basic reproduction number (3.6), in fact we can state the threshold criteria for the existence of endemic steady states of (2.7)–(2.9) as follows, though the proof of Proposition 3.1 is long and technical, it is given in Appendix:

**PROPOSITION 3.1.** *If  $R_0 \leq 1$ , the disease is eradicated naturally as time evolves, on the other hand if  $R_0 > 1$ , there exists a unique endemic steady state.*

If  $R_0 \geq 1$ , we can define the prevalence  $\pi$  at the endemic steady state as follows:

$$(3.7) \quad \pi := \frac{I^*}{N} = \frac{I^*}{I^* + \int_0^\infty S^*(a)da}.$$

The prevalence is related to the basic reproduction number as follows:

$$(3.8) \quad \pi \leq 1 - \frac{1}{R_0}.$$

In fact, we can observe that

$$(3.9) \quad \begin{aligned} \int_0^\infty S^*(a)da &= \frac{vI^*}{k} \int_0^\infty e^{-\frac{I^*}{k} \int_0^a \gamma(\sigma)d\sigma} da \\ &\geq \frac{vI^*}{k} \int_0^\infty e^{-\frac{I^*}{k} \gamma(\infty)a} da = \frac{v}{\gamma(\infty)} = \frac{N}{R_0}. \end{aligned}$$

Then we obtain

$$\pi = \frac{I^*}{I^* + \int_0^\infty S^*(a)da} \leq \frac{I^*}{I^* + \frac{N}{R_0}},$$

and (3.8) follows immediately.

More important information about the prevalence in the influenza epidemic can be derived from the time scale argument. In the endemic steady state, the ratio

$$(3.10) \quad \frac{S^*(a)}{S^*(0)} = e^{-\frac{I^*}{k} \int_0^a \gamma(\sigma)d\sigma},$$

can be interpreted as a survival rate in a susceptible cohort, that is, the rate that a newly recovered host will be uninfected at duration  $a/k$ . Therefore the average number of amino acid substitutions, denoted by  $A_0$ , which occurs before a host individual can become reinfected is calculated as

$$(3.11) \quad A_0 = \int_0^\infty a \frac{I^* \gamma(a)}{k} e^{-\frac{I^*}{k} \int_0^a \gamma(\sigma) d\sigma} da = \int_0^\infty e^{-\frac{I^*}{k} \int_0^a \gamma(\sigma) d\sigma} da.$$

Hence the length of time that takes for  $A_0$  amino acid substitutions to occur is  $A_0/k$ . On the other hand  $1/v$  is the mean length of time that a host is infected and infectious. Since  $A_0/k$  is typically measured in years while  $1/v$  is measured in days, so it would be reasonable to assume that  $A_0/k \gg 1/v$  in the real influenza epidemics. Then the ratio of the long to short time scales  $q$  is given by

$$q := \frac{A_0 v}{k} = \frac{v}{k} \int_0^\infty e^{-\frac{I^*}{k} \int_0^a \gamma(\sigma) d\sigma} da = \frac{1}{I^*} \int_0^\infty S^*(a) da = \frac{1 - \pi}{\pi}.$$

Then it follows that

$$(3.12) \quad \pi = \frac{1}{1 + q}.$$

For example, if we assume that  $q$  may be greater than about fifty in reality, the prevalence would be at most several percent under the evolutionary mechanism.

Though we have so far discussed the basic reproduction number as the endemic threshold, of course it is also a condition for disease invasion. For our system (2.2)–(2.5), the effective reproduction number can be defined as

$$(3.13) \quad R(t) = \frac{1}{v} \int_0^\infty \gamma(a) S(t, a) da.$$

By using the effective reproduction number, the growth rate of the infected population is

$$\frac{1}{I(t)} \frac{dI(t)}{dt} = v(R(t) - 1).$$

Therefore, the invasion condition is given by  $R(t) > 1$ . If the initial population is composed of entirely susceptible population, it follows from (2.11) that

$$R(t) = \frac{1}{v} \int_{kt}^\infty \gamma(a) S_0(a - kt) da = \frac{1}{v} \int_0^\infty \gamma(a + kt) S_0(a) da.$$

Under the Assumption 2.1, we obtain

$$\lim_{t \rightarrow \infty} R(t) = R_0.$$

Thus if  $R_0 > 1$ , the susceptible host population will be ultimately invaded by the disease as time passes, even though  $R(0) < 1$ .

**4. Local stability of the endemic steady state.** For type A influenza epidemic, we observe that between pandemics caused by shift (introduction of genetically very distinct influenza strains), antigenic drift occurs continuously in dominant strains and it may lead recurrent local outbreaks. Empirical studies have found undamped oscillation in the incidence of influenza over the long term, while Pease shows that small perturbations in the steady state of the evolutionary epidemic model leads to damped oscillations.

It is well known in the traditional epidemic model (SIR model) that the damped oscillation may be transformed into sustained oscillation by seasonal variation in the transmission rate, by stochastic effects, by time lags, etc. (Anderson and May 1991, 6.5). While the sustained oscillations in influenza could be explained in a similar manner, it would be a most interesting question whether there is a possibility that the evolutionary mechanism itself produces sustained oscillations without any additional assumption. Though we cannot answer this question completely, we here examine the stability problem of the endemic steady state and consider a possible way to sustained oscillations.

In the following we assume that  $R_0 > 1$ , so there exists an endemic steady state. Let us introduce a new variable  $\zeta(t, a)$  as

$$(4.1) \quad S(t, a) = S^*(a) + \zeta(t, a),$$

where  $S^*(a)$  is a stationary solution of (2.7)–(2.9). Then the system (2.7)–(2.9) can be written as the following new system of  $\zeta(t, a)$ :

$$(4.2) \quad \frac{\partial \zeta(t, a)}{\partial t} + k \frac{\partial \zeta(t, a)}{\partial a} = -\gamma(a) I^* \zeta(t, a) + \gamma(a) [S^*(a) + \zeta(t, a)] \int_0^\infty \zeta(t, a) da,$$

$$(4.3) \quad k \zeta(t, 0) = -v \int_0^\infty \zeta(t, a) da.$$

Again to make use of the semigroup approach, let us formulate (4.2)–(4.3) as a semilinear Cauchy problem in  $L^1$  as follows:

$$(4.4) \quad \zeta'(t) = A_1 \zeta(t) + C \zeta(t) - F(\zeta(t)),$$

where the differential operator  $A_1$  and the linear operator  $C$  are defined by

$$(4.5) \quad \begin{aligned} (A_1 f)(a) &:= -k f'(a) - \gamma(a) I^* f(a), \\ \mathcal{D}(A_1) &:= \left\{ f \in W^{1,1}(\mathbf{R}_+) : f(0) = -\frac{v}{k} \int_0^\infty f(a) da \right\}, \end{aligned}$$

$$(4.6) \quad (C f)(a) := \gamma(a) S^*(a) \int_0^\infty f(a) da.$$

Since the Fréchet derivative of  $F$  at the equilibrium  $\zeta = 0$  is zero, the linearized equation of (4.4) is given by

$$(4.7) \quad \zeta'(t) = (A_1 + C)\zeta(t).$$

In order to investigate the stability of the equilibrium  $\zeta = 0$ , let us consider the resolvent equation for  $A_1 + C$ :

$$(4.8) \quad (\lambda - (A_1 + C))f = g, \quad f \in \mathcal{D}(A_1), \quad g \in L^1, \quad \lambda \in \mathbf{C}.$$

Then we obtain

$$(4.9) \quad kf'(a) + (\lambda + \gamma(a)I^*)f(a) = g(a) + \gamma(a)S^*(a) \int_0^\infty f(z)dz,$$

$$(4.10) \quad f(0) = -\frac{v}{k} \int_0^\infty f(z)dz.$$

By formal integration, we have the following expression:

$$(4.11) \quad \begin{aligned} f(a) &= \frac{1}{k} \int_0^a e^{-\frac{\lambda}{k}(a-s)-\frac{I^*}{k} \int_s^a \gamma(\sigma)d\sigma} g(s)ds \\ &\quad + \frac{1}{1-\Delta(\lambda)} \frac{1}{k} \int_0^\infty \int_0^a e^{-\frac{\lambda}{k}(a-s)-\frac{I^*}{k} \int_s^a \gamma(\sigma)d\sigma} g(s)dsda \\ &\quad \times \left\{ -\frac{v}{k} e^{-\frac{\lambda}{k}a-\frac{I^*}{k} \int_0^a \gamma(\sigma)d\sigma} + \frac{1}{k} \int_0^a e^{-\frac{\lambda}{k}(a-s)-\frac{I^*}{k} \int_s^a \gamma(\sigma)d\sigma} \gamma(s)S^*(s)ds \right\}, \end{aligned}$$

where the mapping  $\Delta(\lambda) : \mathbf{C} \rightarrow \mathbf{C}$  is defined by

$$(4.12) \quad \begin{aligned} \Delta(\lambda) &= \frac{1}{k} \int_0^\infty \int_0^a e^{-\frac{\lambda}{k}(a-s)-\frac{I^*}{k} \int_s^a \gamma(\sigma)d\sigma} \gamma(s)S^*(s)dsda \\ &\quad - \frac{v}{k} \int_0^\infty e^{-\frac{\lambda}{k}a-\frac{I^*}{k} \int_0^a \gamma(\sigma)d\sigma} da. \end{aligned}$$

As is well known in linearized analysis for stability of equilibrium, we can expect that the location of roots of the characteristic equation  $\Delta(\lambda) = 1$  may determine the stability of the steady state. In order to state the principle of linearized stability rigorously, let us introduce some notations and results from functional analysis. For more information, the reader may refer to Webb (1985).

Let  $\sigma(A)$  be the spectrum of an operator  $A$ , let  $P_\sigma(A)$  be the point spectrum of  $A$ , let  $E\sigma(A)$  be the essential spectrum of  $A$  and let  $\rho(A)$  be the resolvent set of  $A$ . In case that the operator  $A$  generates a strongly continuous semigroup,  $\omega_0(A)$  denotes the growth bound of the semigroup  $e^{tA}$  and  $\omega_1(A)$  denotes the  $\alpha$ -growth bound of  $e^{tA}$ . For the linear operator  $A_1 + C$ , we obtain the following result. The proof is in the Appendix:

#### PROPOSITION 4.1.

$$(1) \quad \omega_1(A_1 + C) \leq -\gamma(\infty)I^*,$$

- (2) If  $\Re \lambda > -\gamma(\infty)I^*$  and  $\Delta(\lambda) = 1$ , then  $\lambda \in P_\sigma(A_1 + C)$ ,  
(3) If  $\Re \lambda > -\gamma(\infty)I^*$ , then  $\lambda \in \sigma(A + C)$  if and only if  $\lambda \in P_\sigma(A + C)$   
if and only if  $\lambda \in \Lambda := \{\lambda \in \mathbf{C} : \Re \lambda > -\gamma(\infty)I^*, \Delta(\lambda) = 1\}$ , and it follows  
that for  $\lambda \in \rho(A_1 + C)$ ,

$$(4.13) \quad \begin{aligned} ((\lambda - (A_1 + C))^{-1}f)(a) &= \frac{1}{k} \int_0^a e^{-\frac{\lambda}{k}(a-s) - \frac{I^*}{k} \int_s^a \gamma(\sigma)d\sigma} g(s)ds \\ &+ \frac{1}{1 - \Delta(\lambda)} \frac{1}{k} \int_0^\infty \int_0^a e^{-\frac{\lambda}{k}(a-s) - \frac{I^*}{k} \int_s^a \gamma(\sigma)d\sigma} g(s)dsda \\ &\times \left\{ -\frac{v}{k} e^{-\frac{\lambda}{k}a - \frac{I^*}{k} \int_0^a \gamma(\sigma)d\sigma} + \frac{1}{k} \int_0^a e^{-\frac{\lambda}{k}(a-s) - \frac{I^*}{k} \int_s^a \gamma(\sigma)d\sigma} \gamma(s)S^*(s)ds \right\}. \end{aligned}$$

Since  $1 - \Delta(\lambda)$  is holomorphic in the domain  $\Re \lambda > -\gamma(\infty)I^*$ , we know that the set of characteristic root  $\Lambda$  consists of distinct points. Moreover we can observe that the characteristic equation  $\Delta(\lambda) = 1$  can be rewritten as follows:

$$(4.14) \quad \Delta(\lambda) = - \int_0^\infty e^{-\frac{\lambda}{k}a} \Phi(a) da = 1,$$

where

$$(4.15) \quad \Phi(a) := \frac{v}{k} e^{-\frac{I^*}{k} \int_0^a \gamma(\sigma)d\sigma} \left( 1 - \frac{I^*}{k} \int_0^\infty \gamma(s) e^{-\frac{I^*}{k} \int_a^{a+s} \gamma(\sigma)d\sigma} ds \right).$$

In fact, by exchanging the order of integrals in (4.12), we obtain that

$$\begin{aligned} \frac{1}{k} \int_0^\infty \int_0^a e^{-\frac{\lambda}{k}(a-s) - \frac{I^*}{k} \int_s^a \gamma(\sigma)d\sigma} \gamma(s)S^*(s)dsda \\ &= \frac{1}{k} \int_0^\infty \gamma(s)S^*(s)ds \int_s^\infty e^{-\frac{\lambda}{k}(a-s) - \frac{I^*}{k} \int_s^a \gamma(\sigma)d\sigma} da \\ &= \frac{1}{k} \int_0^\infty \gamma(s)S^*(s)ds \int_0^\infty e^{-\frac{\lambda}{k}z - \frac{I^*}{k} \int_s^{s+z} \gamma(\sigma)d\sigma} dz \\ &= \frac{1}{k} \int_0^\infty \gamma(s)ds \int_0^\infty e^{-\frac{\lambda}{k}z} S^*(s+z)dz \\ &= \frac{1}{k} \int_0^\infty e^{-\frac{\lambda}{k}a} \int_0^\infty S^*(s+a)\gamma(s)dsda. \end{aligned}$$

Therefore, using the expression (3.4), we arrive at the following expression:

$$(4.16) \quad \Delta(\lambda) = - \int_0^\infty e^{-\frac{\lambda}{k}a} \frac{S^*(a)}{I^*} \left[ 1 - \frac{I^*}{k} \int_0^\infty \gamma(s) \frac{S^*(a+s)}{S^*(a)} ds \right] da.$$

Again using (3.4), now we arrive at the expression (4.14)–(4.15). Next let us define

$$\phi(a) := 1 - \frac{I^*}{k} \int_0^\infty \gamma(s) e^{-\frac{I^*}{k} \int_a^{a+s} \gamma(\sigma)d\sigma} ds.$$

Then it follows from the monotonicity of  $\gamma(a)$  that  $\phi(a)$  is non-decreasing, and we can observe that

$$\phi(0) = 1 + \int_0^\infty \frac{d}{ds} \left[ e^{-\frac{I^*}{k} \int_0^s \gamma(\sigma) d\sigma} \right] ds = 0.$$

Hence it follows that  $\phi(a) \geq 0$  for all  $a \geq 0$ . Then  $\Phi(a) \geq 0$  for all  $a \geq 0$  and  $\Delta(\lambda) \leq 0$  for real  $\lambda$ , we can conclude that the characteristic equation  $\Delta(\lambda) = 1$  has no real root. And it is clear that if  $\lambda \in \Lambda$ , then  $\bar{\lambda} \in \Lambda$ , hence  $\Lambda$  consists of complex conjugate pairs. In summary, we can state the following, though its proof of the second part is given in Appendix:

**PROPOSITION 4.2.**  *$\Lambda \cap \mathbf{R} = \emptyset$  and  $\Lambda$  consists of distinct, complex conjugate pairs. For any  $\alpha > -\gamma(\infty)I^*$ ,  $\Lambda$  has only finitely many elements in the strip  $\Re \lambda \geq \alpha$ .*

Under the above preparation, we can state the principle of linearized stability for the semilinear problem (4.4) as follows, its proof is given in Appendix:

**PROPOSITION 4.3.** *Suppose that  $\Re \lambda < 0$  for any  $\lambda \in P_\sigma(A_1 + C) \cap \{\lambda \in \mathbf{C} : \Re \lambda > -\gamma(\infty)I^*\}$ . Then the endemic steady state is locally asymptotically stable.*

From the above propositions, we can conclude that if there is no eigenvalue for  $A_1 + C$  or its dominant eigenvalue is located in the left half plane, the endemic steady state is locally asymptotically stable.

Let us rewrite the characteristic equation as follows:

$$(4.17) \quad \hat{K}(\lambda) = \int_0^\infty e^{-\lambda\tau} K(\tau) d\tau = \frac{1}{\Delta(0)},$$

where  $K(\tau)$  is defined by

$$K(\tau) := \frac{k\Phi(k\tau)}{\int_0^\infty \Phi(a) da} = -\frac{k\Phi(k\tau)}{\Delta(0)}.$$

Note that  $\hat{K}(0) = 1$ . Then we can prove the following stability result:

**PROPOSITION 4.4.** *If  $\Delta(0) \geq -1$ , all characteristic roots have negative real part, the endemic steady state is locally asymptotically stable.*

*Proof.* Suppose that there is a characteristic root  $\lambda = \alpha + i\beta \neq 0$  with  $\alpha \geq 0$ . Then it follows from the characteristic equation (4.17) that

$$\left| \frac{1}{\Delta(0)} \right| \leq \int_0^\infty e^{-\alpha\tau} |K(\tau)| |\cos \beta\tau| d\tau < \int_0^\infty |K(\tau)| d\tau = 1.$$

Therefore  $\Delta(0) < -1$  is a necessary condition for the existence of a characteristic root with nonnegative real part. From Proposition 4.3, we conclude that the endemic steady state is locally asymptotically stable if  $\Delta(0) \geq -1$ .  $\square$

It is interesting that the prevalence of disease is related to the stability of the endemic steady state, that is, we can show the following *fifty percent prevalence rule*:

**COROLLARY 4.5.** *If  $\pi \geq \frac{1}{2}$ , the endemic steady state is locally asymptotically stable.*

*Proof.* From Proposition 4.4, it is sufficient to show that  $\Delta(0) \geq -1$  if  $\pi \geq \frac{1}{2}$ . The condition  $\pi \geq \frac{1}{2}$  implies that

$$\int_0^\infty S^*(a)da \leq I^*.$$

Then it follows from  $0 \leq \phi(a) < 1$  that

$$-\Delta(0) = \frac{v}{k} \int_0^\infty e^{-\frac{I^*}{k} \int_0^a \gamma(\sigma)d\sigma} \phi(a)da \leq \frac{1}{I^*} \int_0^\infty S^*(a)da \leq 1.$$

This completes our proof.  $\square$

However as is seen in the previous section, the prevalence for the realistic parameter values may be very small (at most several percent), the fifty percent prevalence rule would not cover the domain of realistic parameter values for type A influenza epidemic. So an interesting question is whether the evolutionary epidemic model could have a periodic solution in the parameter region escaped from the fifty percent prevalence rule.

A possible mechanism to create a periodic solution is a Hopf bifurcation of an equilibrium, which would occur if a pair of complex conjugate characteristic root crosses the imaginary axis from the left half plain to the right half plain (Prüss 1983b, Metz and Diekmann 1986). The characteristic equation (4.17) is written as

$$(4.18) \quad \Delta(0) \int_0^\infty e^{-\lambda\tau} K(\tau)d\tau = 1, \quad K(\tau) \geq 0.$$

This type of characteristic equation has been investigated by several authors (Frauenthal 1975, Diekmann and van Gils 1984, Metz and Diekmann 1986), and it is shown that if  $\Delta(0) < -1$ , there exists the possibility that some complex conjugate pairs of characteristic roots lie in the right half plane. Therefore if the prevalence is less than fifty percent, we could conjecture that the destabilization of the endemic steady state may be possible and it would lead to a periodic solution.

**5. Discussion.** Finally let us consider possible extensions of the Pease model. In the Pease model, the  $i$ -state (individual state) is characterized by parameter  $a$  (the amount of amino acid substitutions), and it plays essentially the same rule as duration since the last infection, because the rate of amino acid substitutions is equal to the mutation rate and it is assumed to be constant. So the  $i$ -state of susceptibles is determined locally independent of the state of epidemic. However if amino acid substitutions are

caused by frequency-dependent selection and several types of virus strain can circulate at the same time in a community, we have to deal with much more complex evolutionary models in which the  $i$ -state will depend on global state of epidemic.

From practical point of view, it is also important to consider the effects of introducing some kind of vaccination policy. In Inaba (1998) I have already studied a modified Pease model with simple vaccination term, in which a constant vaccination rate is assumed and the vaccinated susceptibles are identified with the individuals recovered from real infection. However in reality, the vaccination in the evolutionary epidemic would be much more complex phenomena. For example, if we use the immunizing virus isolated in the past, according to the amount of the amino acid substitutions occurring between the immunizing virus strain and the strain circulating at given time  $t$ , the induced immunity will be incomplete, that is, the result of vaccination is to make the susceptible host "younger" with respect to the "age"  $a$ , but it does not necessarily reset the "age" to zero. Therefore it seems that the possible rate of vaccination  $\delta(a)$  is expressed as

$$\delta(a) = \delta_c(a) + \delta_p(a),$$

where  $\delta_c(a)$  denotes the rate of complete immunization and  $\delta_p(a)$  is the rate of partial immunization. Let  $\beta(a, \alpha)$  be the probability per unit amino acid substitution that a partial vaccination in a susceptible host of "age"  $\alpha$  leads to a susceptible host with "age"  $a$  ( $a \leq \alpha$ ). We assume that  $\beta(a, \alpha) = 0$  for  $a > \alpha$  and for all  $\alpha$

$$\int_0^\alpha \beta(a, \alpha) da = 1.$$

If we combine the vaccination term defined above with the basic model (2.2)–(2.5), we obtain the following new model:

$$(5.1) \quad \frac{\partial S(t, a)}{\partial t} + k \frac{\partial S(t, a)}{\partial a} = -\gamma(a)S(t, a) \left[ N - \int_0^\infty S(t, a) da \right] \\ -\delta(a)S(t, a) + \int_a^\infty \beta(a, \alpha) \delta_p(\alpha) S(t, \alpha) d\alpha,$$

$$(5.2) \quad kS(t, 0) = v \left[ N - \int_0^\infty S(t, a) da \right] + \int_0^\infty \delta_c(a) S(t, a) da.$$

An important point different from the original model (2.2)–(2.5) is that in this vaccination model, there is a disease-free steady state. It can be shown that the stationary state for (5.1)–(5.2) is determined by solving a Fredholm integral equation in the same manner as is shown in the metapopulation

model by Hastings (1995). The analysis of the threshold phenomena of this vaccination model will be a future project.

In the Pease model, the demography of host population is neglected. If we take into account the birth and death of the host population, we have to add a new class in which the host individuals are never infected. A possible demographic extension of the Pease model can be written as follows:

$$(5.3) \quad \frac{\partial S_0(t, \tau)}{\partial t} + \frac{\partial S_0(t, \tau)}{\partial \tau} = -\mu S_0(t, \tau) - S_0(t, \tau) \gamma(\infty) \int_0^\infty \theta(\tau) I(t, \tau) d\tau,$$

$$(5.4) \quad S_0(t, 0) = B$$

$$(5.5) \quad \frac{\partial S(t, \tau)}{\partial t} + \frac{\partial S(t, \tau)}{\partial \tau} = -\mu S(t, \tau) - S(t, \tau) \gamma(\tau) \int_0^\infty \theta(\tau) I(t, \tau) d\tau,$$

$$(5.6) \quad S(t, 0) = \int_0^\infty v(\tau) I(t, \tau) d\tau,$$

$$(5.7) \quad \frac{\partial I(t, \tau)}{\partial t} + \frac{\partial I(t, \tau)}{\partial \tau} = -(\mu + v(\tau)) I(t, \tau),$$

$$(5.8) \quad I(t, 0) = \int_0^\infty \{ \gamma(\infty) S_0(t, \tau) d\tau + \gamma(\tau) S(t, \tau) \} d\tau \int_0^\infty \theta(\tau) I(t, \tau) d\tau,$$

where  $S_0(t, \tau)$  is the density of never infected population with complete susceptibility ("virgin" population in terminology of Kermack and McKendrick) at time  $t$  and duration  $\tau$ ,  $I(t, \tau)$  the density of infected/infectious population at time  $t$  and duration  $\tau$ ,  $S(t, \tau)$  be the density of recovered population with partial susceptibility at time  $t$  and duration  $\tau$ ,  $\mu$  the natural death rate,  $B$  the birth rate,  $v(\tau)$  the recovery rate at duration  $\tau$ , and  $\gamma(\tau)\theta(\tau')$  the infection rate from the infected individual at duration  $\tau'$  to the recovered individual at duration  $\tau$ . The infection rate from the infected individual at duration  $\tau'$  to the virgin population is given by  $\gamma(\infty)\theta(\tau')$ . The function  $\gamma(\tau)$ , which reflects the susceptibility of the host, is assumed to be a bounded, nonnegative, monotone increasing function. If we assume that  $\tau = a/k$ ,  $B = \mu = 0$ ,  $\theta(\tau)$  and  $v(\tau)$  are duration independent and  $S_0 \equiv 0$ , the above extended model is reduced to the Pease model.

Again it is easy to see that if we seek for the classical solution in  $W_+^{1,1}(\mathbf{R}_+)$ , the total size of host population is constant  $N = B/\mu$ . Then (5.3)–(5.8) has a disease-free steady state  $(S_0^*, S^*, I^*) = (Be^{-\mu\tau}, 0, 0)$ , and the basic reproduction number is given by

$$(5.9) \quad R_0 = N \gamma(\infty) \int_0^\infty \theta(\tau) e^{-\mu\tau - \int_0^\tau v(\sigma) d\sigma} d\tau.$$

Though the system (5.2)–(5.8) looks very complex, we can prove the existence and uniqueness of endemic steady state if  $R_0 > 1$  in the same manner

as the proof of Proposition 3.1, and it is locally asymptotically stable if the prevalence is small enough (Inaba 2000, 2001).

Historically speaking, it is remarked that this extended model (the variable susceptibility model) was already proposed by Kermack and McKendrick (1932, 1933). In contrast to Kermack's and McKendrick's models developed in the most famous paper in 1927, the more general model proposed in the papers of 30s has been so far almost neglected in the field of mathematical epidemiology. As is shown in the above, the Pease model is a kind of the variable susceptibility model, and it suggests that the Kermack-McKendrick's general formulation in the 1930s could be a useful tool to take into account the heterogeneity (immunological status) of the host population in epidemic modeling.

At present-day world, the evolutionary mechanism in epidemic has been becoming more and more important factor responsible for emerging and reemerging infectious diseases, it makes the control of infectious diseases more difficult. Therefore to develop mathematical modeling for the evolutionary epidemic phenomena would be a most important challenge in mathematical epidemiology

**6. Appendix.** First we make up the argument in the second section. Let us prove that if  $f \in W_+^{1,1}(\mathbf{R}_+)$ , then we have  $f(\infty) := \lim_{x \rightarrow \infty} f(x) = 0$ . In fact, for any  $0 < x < y < \infty$ , it follows that

$$(6.1) \quad |f(y) - f(x)| = \left| \int_x^y \frac{\partial f(z)}{\partial z} dz \right| \leq \int_x^y \left| \frac{\partial f(z)}{\partial z} \right| dz.$$

It follows from  $f \in W_+^{1,1}(\mathbf{R}_+)$  that the right hand side of the above inequality goes to zero if  $x, y \rightarrow \infty$ . Thus  $f(x)$  forms a Cauchy sequence when  $x \rightarrow \infty$ , it has a nonnegative limit  $f(\infty) \geq 0$ . If  $f(\infty) > 0$ , we can choose a small number  $\epsilon$  such that  $f(\infty) > \epsilon > 0$ , and there exists a number  $x_0(\epsilon) > 0$  such that for all  $x > x_0(\epsilon)$ ,  $f(x) > f(\infty) - \epsilon > 0$ . Hence we have

$$\int_0^\infty f(x) dx > \int_{x_0(\epsilon)}^\infty [f(\infty) - \epsilon] dx = \infty.$$

This is a contradiction, so we conclude that  $f(\infty) = 0$ . Hence for the solution  $S(t, a)$  of (2.2)–(2.5), if  $S(t, *) \in W_+^{1,1}(\mathbf{R}_+)$ , then  $S(t, \infty) = 0$  and we can calculate as

$$\begin{aligned} \frac{\partial}{\partial t} \int_0^\infty S(t, a) da &= \int_0^\infty \frac{\partial S(t, a)}{\partial t} da \\ &= \int_0^\infty \left( -k \frac{\partial S(t, a)}{\partial a} - \gamma(a) S(t, a) I(t) \right) da \\ &= kS(t, 0) - I(t) \int_0^\infty \gamma(a) S(t, a) da = -\frac{dI(t)}{dt}. \end{aligned}$$

Therefore we know that  $N(t) = I(t) + \int_0^\infty S(t, a)da$  is constant.

**Proof of Lemma 2.1** It is not difficult to see that the operator  $A$  is a closed linear operator in  $L^1$ , though we omit the proof. Let us consider the resolvent equation

$$(\lambda - A)f = g, \quad f \in \mathcal{D}(A), \quad g \in L^1.$$

Then it is easily verified that for  $\Re\lambda > 0$ ,  $(\lambda - A)^{-1}$  exists and it is expressed as follows:

$$(6.2) \quad (\lambda - A)^{-1}g = J(\lambda)g + K(\lambda)g,$$

where  $J(\lambda)$  and  $K(\lambda)$  are defined by

$$\begin{aligned} (J(\lambda)g)(a) &:= \frac{1}{k} \int_0^a e^{-\frac{\lambda}{k}(a-s)} g(s)ds. \\ (K(\lambda)g)(a) &:= -\frac{v}{k(\lambda+v)} e^{-\frac{\lambda}{k}a} \int_0^\infty g(s)ds, \end{aligned}$$

Observe that

$$\|\lambda(\lambda - A)^{-1}g - g\|_{L^1} \leq \|\lambda J(\lambda)g - g\|_{L^1} + \|\lambda K(\lambda)g\|_{L^1},$$

where it is easy to see that  $\lim_{\lambda \rightarrow \infty} \|\lambda K(\lambda)g\|_{L^1} = 0$ . On the other hand, note that  $J(\lambda)$  is the resolvent of a closed linear operator  $A_0$  defined by

$$(6.3) \quad (A_0f)(a) = -kf'(a), \quad \mathcal{D}(A_0) = \{f \in W^{1,1}(\mathbf{R}_+) : f(0) = 0\}.$$

Since the operator  $A_0$  is the infinitesimal generator of a translation semigroup, it follows that  $\lim_{\lambda \rightarrow \infty} \|\lambda J(\lambda)g - g\|_{L^1} = 0$ . This shows that the domain  $\mathcal{D}(A)$  is dense in  $L^1$ , because  $\lambda(\lambda - A)^{-1}g \in \mathcal{D}(A)$  for any  $g \in L^1$  and  $\lim_{\lambda \rightarrow \infty} \lambda(\lambda - A)^{-1}g = g$ . Moreover it follows from the above estimate that if  $\lambda > v$ ,

$$(6.4) \quad \|(\lambda - A)^{-1}\| \leq \frac{\lambda + 2v}{\lambda(\lambda + v)} < \frac{1}{\lambda - v}.$$

Therefore from the Hille-Yosida Theorem, we know that the densely defined closed operator  $A$  is the infinitesimal generator of a strongly continuous semigroup  $T(t) = \exp(tA)$ ,  $t \geq 0$ . The semigroup  $e^{tA}$  can be expressed as follows:

$$(6.5) \quad (e^{At}\phi)(a) = \begin{cases} \phi(a - kt), & a > kt \\ \psi(t - \frac{a}{k}), & a < kt \end{cases},$$

where the boundary value  $\psi(t)$  satisfies the following equation:

$$(6.6) \quad \psi(t) = -\frac{v}{k} \int_0^{kt} \psi(t - \frac{a}{k})da - \frac{v}{k} \int_{kt}^\infty \phi(a - kt)da.$$

Then it is easy to see that  $\psi(t)$  is given by

$$(6.7) \quad \psi(t) = -e^{-vt} \frac{v}{k} \int_0^\infty \phi(a) da.$$

Hence we can observe that

$$\begin{aligned} \|e^{At}\phi\|_{L^1} &\leq \frac{v}{k} \int_0^{kt} e^{-v(t-\frac{a}{k})} da \int_0^\infty |\phi(a)| da + \int_{kt}^\infty |\phi(a-kt)| da \\ &\leq \left( \frac{v}{k} \int_0^{kt} e^{-v(t-\frac{a}{k})} da + 1 \right) \|\phi\|_{L^1} = (2 - e^{-vt}) \|\phi\|_{L^1}. \end{aligned}$$

Then we conclude that  $\|e^{At}\| \leq 2$ . This completes our proof.  $\square$

**Proof of Proposition 3.1** It follows from (2.3) that for all  $t > 0$

$$I'(t) \leq -vI(t) + I(t)\gamma(\infty)(N - I(t)) = v(R_0 - 1)I(t) - \gamma(\infty)I^2(t).$$

Then it is easily seen that if  $R_0 \leq 1$ ,  $\lim_{t \rightarrow \infty} I(t) = 0$ , so the infected population will be eradicated and there is no endemic steady state. Next assume that  $R_0 > 1$ . Let us define a function  $F(x)$ ,  $x \in (0, N]$  by

$$(6.8) \quad F(x) := x \left( 1 + \frac{v}{k} \int_0^\infty e^{-\frac{x}{k} \int_0^a \gamma(\sigma) d\sigma} da \right).$$

From (3.5), an endemic steady state corresponds to a positive solution of equation  $F(x) = N$ . Under the Assumption 2.1, for any small  $\epsilon > 0$  there exists a number  $a_0$  such that  $\gamma(a) \geq \gamma(\infty) - \epsilon$  for  $a \geq a_0$ . Then we can observe that

$$\begin{aligned} \int_0^\infty e^{-\frac{x}{k} \int_0^a \gamma(\sigma) d\sigma} da &= \int_0^{a_0} e^{-\frac{x}{k} \int_0^a \gamma(\sigma) d\sigma} da + \int_{a_0}^\infty e^{-\frac{x}{k} \left\{ \int_0^{a_0} + \int_{a_0}^a \right\} \gamma(\sigma) d\sigma} da \\ &\leq a_0 + \int_{a_0}^\infty e^{-\frac{x}{k} (\gamma(\infty) - \epsilon)(a - a_0)} da = a_0 + \frac{k}{x(\gamma(\infty) - \epsilon)}. \end{aligned}$$

Therefore we have

$$\limsup_{x \rightarrow +0} F(x) \leq \frac{v}{\gamma(\infty) - \epsilon}.$$

Since  $R_0 > 1$ , we can choose a small  $\epsilon$  as  $0 < \epsilon < \gamma(\infty) - \frac{v}{N}$  in advance, hence we can conclude that  $\limsup_{x \rightarrow +0} F(x) < N$ . Since  $F(N) > N$ , we know that  $F(x) = N$  has at least one root in  $(0, N]$ , which corresponds to the endemic steady state  $I^*$ .

Finally we show the uniqueness of the endemic steady state. First let us consider the case that  $\gamma(0) > 0$ . In this case,  $\gamma(a) > 0$  for all  $a \geq 0$ , so

we can rewrite  $F(x)$  as follows:

$$(6.9) \quad \begin{aligned} F(x) &= x - \int_0^\infty \frac{v}{\gamma(a)} \frac{d}{da} e^{-\frac{x}{k} \int_0^a \gamma(\sigma) d\sigma} da \\ &= x + \frac{v}{\gamma(0)} - v \int_0^\infty \frac{\gamma'(a)}{\gamma^2(a)} e^{-\frac{x}{k} \int_0^a \gamma(\sigma) d\sigma} da. \end{aligned}$$

Thanks to the nonnegativity of  $\gamma'(a)$ , we know that  $F(x)$  is monotone increasing, so the endemic steady state uniquely exists. Next we consider the case that  $\gamma(0) = 0$ . In this case, under the Assumption 2.1, without loss of generality we can assume that there exists a number  $a_0 \geq 0$  such that  $\gamma(a) = 0$  for  $a \in [0, a_0]$  and  $\gamma(a) > 0$  for  $a > a_0$ . That is,  $[0, a_0]$  is a time interval during which the recovered individuals have a complete immunity. Let  $h > 0$  be an arbitrary small positive number. Then  $\gamma(a) > 0$  for all  $a \geq a_0 + h$ . Observe that

$$\frac{vx}{k} \int_0^\infty e^{-\frac{x}{k} \int_0^a \gamma(\sigma) d\sigma} da = \frac{vx}{k} \left\{ \int_0^{a_0} + \int_{a_0}^{a_0+h} + \int_{a_0+h}^\infty \right\} e^{-\frac{x}{k} \int_0^a \gamma(\sigma) d\sigma} da.$$

In the right hand side of the above equation, we can calculate three integrals as

$$\begin{aligned} J_1(x) &:= \frac{vx}{k} \int_0^{a_0} e^{-\frac{x}{k} \int_0^a \gamma(\sigma) d\sigma} da = \frac{vx}{k} a_0, \\ J_2(x) &:= \frac{vx}{k} \int_{a_0}^{a_0+h} e^{-\frac{x}{k} \int_0^a \gamma(\sigma) d\sigma} da = \frac{vx}{k} \int_{a_0}^{a_0+h} e^{-\frac{x}{k} \int_{a_0}^a \gamma(\sigma) d\sigma} da, \\ J_3(x) &:= \frac{vx}{k} \int_{a_0+h}^\infty e^{-\frac{x}{k} \int_{a_0}^a \gamma(\sigma) d\sigma} da \\ &= \frac{v}{\gamma(a_0+h)} e^{-\frac{x}{k} \int_{a_0}^{a_0+h} \gamma(\sigma) d\sigma} - \int_{a_0+h}^\infty \frac{\gamma'(a)v}{\gamma^2(a)} e^{-\frac{x}{k} \int_{a_0}^a \gamma(\sigma) d\sigma} da. \end{aligned}$$

Therefore we have the expression as follows:

$$F(x) = x + g(x) - \int_{a_0+h}^\infty \frac{\gamma'(a)v}{\gamma^2(a)} e^{-\frac{x}{k} \int_{a_0}^a \gamma(\sigma) d\sigma} da,$$

where  $g(x)$  is given by

$$g(x) := J_1(x) + J_2(x) + \frac{v}{\gamma(a_0+h)} e^{-\frac{x}{k} \int_{a_0}^{a_0+h} \gamma(\sigma) d\sigma}$$

In order to see the monotonicity of  $F(x)$ , it is sufficient to show that  $1+g'(x)$  is positive. Observe that

$$g'(x) = \frac{va_0}{k} + \frac{v}{k} \int_{a_0}^{a_0+h} e^{-\frac{x}{k} \int_{a_0}^a \gamma(\sigma) d\sigma} da$$

$$(6.10) \quad -\frac{vx}{k} \int_{a_0}^{a_0+h} \frac{1}{k} \left( \int_{a_0}^a \gamma(\sigma) d\sigma \right) e^{-\frac{\pi}{k} \int_{a_0}^a \gamma(\sigma) d\sigma} da \\ -\frac{v}{\gamma(a_0+h)} \frac{1}{k} \int_{a_0}^{a_0+h} \gamma(\sigma) d\sigma e^{-\frac{\pi}{k} \int_{a_0}^{a_0+h} \gamma(\sigma) d\sigma}.$$

Hence it follows that

$$\left| g'(x) - \frac{va_0}{k} \right| \leq \frac{v}{k} \left( h + \frac{x\gamma(\infty)h^2}{2k} + \frac{1}{\gamma(a_0+h)} \int_{a_0}^{a_0+h} \gamma(\sigma) d\sigma \right).$$

Let  $0 < \eta < 1$ . Then we can choose a sufficiently small  $h > 0$  in advance such that

$$\left| g'(x) - \frac{va_0}{k} \right| \leq \frac{v}{k} \left( 2 + \frac{N\gamma(\infty)h}{2k} \right) h < \eta < 1.$$

Then it follows that for all  $x \in [0, N]$

$$1 + g'(x) > 1 - \eta + \frac{va_0}{k} \geq 1 - \eta > 0.$$

That is,  $1 + g'(x)$  is positive, again  $F(x)$  is monotone. Therefore we know that the endemic state is unique.  $\square$

**Proof of Proposition 4.1** Note that  $C$  is one-dimensional, so it is a compact perturbation. Then it follows that  $\omega_1(A_1 + C) = \omega_1(A_1)$ . Let us define the mappings  $T_j(t)$  ( $j = 1, 2$ ) as follows:

$$(6.11) \quad (T_1(t)\phi)(a) = \begin{cases} \phi(a - kt)e^{-I^* \int_0^t \gamma(a - kt + k\sigma) d\sigma}, & a > kt \\ 0, & a < kt \end{cases},$$

$$(6.12) \quad (T_2(t)B)(a) = \begin{cases} 0, & a > kt \\ B(t - \frac{a}{k})e^{-I^* \int_0^{\frac{a}{k}} \gamma(k\sigma) d\sigma}, & a < kt \end{cases},$$

where the boundary value  $B(t)$  is given by the solution of the renewal integral equation as

$$(6.13) \quad B(t) = -\frac{v}{k} \int_0^{kt} B(t - \frac{a}{k})e^{-I^* \int_0^{\frac{a}{k}} \gamma(k\sigma) d\sigma} da \\ -\frac{v}{k} \int_{kt}^{\infty} \phi(a - kt)e^{-I^* \int_0^t \gamma(a - kt + k\sigma) d\sigma} da,$$

For any  $t > 0$ , the solution  $B(t)$  of the above integral equation is given by

$$(6.14) \quad B(t) = ((I - V)^{-1}G\phi)(t),$$

where the operators  $G : L^1(\mathbf{R}_+) \rightarrow L^1(\mathbf{R}_+)$  and  $V : L^1(\mathbf{R}_+) \rightarrow L^1(\mathbf{R}_+)$  are defined by

$$(6.15) \quad \begin{aligned} (G\phi)(t) &:= -\frac{v}{k} \int_{kt}^{\infty} \phi(a - kt) e^{-I^* \int_0^t \gamma(a - kt + k\sigma) d\sigma} da \\ &= -\frac{v}{k} \int_0^{\infty} \phi(a) e^{-I^* \int_0^t \gamma(a + k\sigma) d\sigma} da, \end{aligned}$$

$$(6.16) \quad \begin{aligned} (V\phi)(t) &:= -\frac{v}{k} \int_0^{kt} \phi(t - \frac{a}{k}) e^{-I^* \int_0^{\frac{a}{k}} \gamma(k\sigma) d\sigma} da \\ &= -v \int_0^t \phi(a) e^{-I^* \int_0^{t-a} \gamma(k\sigma) d\sigma} da. \end{aligned}$$

Then the semigroup  $e^{A_1 t}$  can be expressed as follows:

$$(6.17) \quad e^{A_1 t} \phi = T_1(t)\phi + T_2(t)(I - V)^{-1}G\phi.$$

For any fixed  $t_0 > 0$ , it is easily seen that the operator  $G$  is a compact operator from  $L^1(\mathbf{R}_+)$  into  $L^1(0, t_0)$ , hence the mapping  $T_2(t)(I - V)^{-1}G$  is a compact operator, too (Prüss 1983). Then we have  $\alpha(T_2(t)(I - V)^{-1}G) = 0$  and

$$\alpha(e^{A_1 t}) \leq \alpha(T_1(t)) \leq \|T_1(t)\|.$$

It follows from the monotonicity of  $\gamma(a)$  that

$$\|T_1(t)\phi\|_{L^1} = \int_0^{\infty} \phi(a) e^{-I^* \int_0^t \gamma(a + k\sigma) d\sigma} da \leq \|\phi\|_{L^1} e^{-I^* \int_0^t \gamma(k\sigma) d\sigma}.$$

From the Assumption 2.1, for any small  $\epsilon > 0$  there exists a number  $a_0 > 0$  such that  $\gamma(a) \geq \gamma(\infty) - \epsilon$  for all  $a \geq a_0$ . Hence we can observe that if  $t > \frac{a_0}{k}$ , then

$$\int_0^t \gamma(k\sigma) d\sigma \geq (\gamma(\infty) - \epsilon)(t - \frac{a_0}{k}).$$

Therefore we obtain the estimate as follows:

$$(6.18) \quad \|T_1(t)\| \leq e^{-I^*(\gamma(\infty) - \epsilon)(t - \frac{a_0}{k})}, \quad \text{for } t > \frac{a_0}{k}.$$

Then it follows that

$$\begin{aligned} \omega_1(A_1) &= \lim_{t \rightarrow \infty} \frac{\log \alpha(e^{t A_1})}{t} \leq \lim_{t \rightarrow \infty} \frac{\log \alpha(T_1(t))}{t} \\ &\leq \lim_{t \rightarrow \infty} \frac{\log \|T_1(t)\|}{t} \leq -I^*(\gamma(\infty) - \epsilon). \end{aligned}$$

Since  $\epsilon > 0$  is any small positive number, we can conclude that  $\omega_1(A_1) \leq -I^*\gamma(\infty)$ . This completes the proof of (1)

Next it is easily seen that if  $\Re\lambda > -\gamma(\infty)I^*$ , the mapping  $\lambda \rightarrow \Delta(\lambda)$  defines a holomorphic function, hence  $1 - \Delta(\lambda)$  has only isolated zeros of finite order in  $\{\lambda \in \mathbf{C} : \Re\lambda > -\gamma(\infty)I^*\}$  and  $(1 - \Delta(\lambda))^{-1}$  is meromorphic in this domain. For  $\lambda \in \Lambda$ , it follows that

$$f_\lambda(a) = \frac{1}{k} \int_0^a e^{-\frac{\lambda}{k}(a-s) - \frac{I^*}{k} \int_s^a \gamma(\sigma)d\sigma} \gamma(s) S^*(s) ds - \frac{v}{k} e^{-\frac{\lambda}{k}a - \frac{I^*}{k} \int_0^a \gamma(\sigma)d\sigma},$$

is an eigenfunction associated with eigenvalue  $\lambda$ . Then the result (2) follows.

Thirdly suppose that  $\lambda \in \sigma(A_1 + C)$  and  $\Re\lambda > -\gamma(\infty)I^*$ . Since  $\omega_1(A_1 + C) \leq -\gamma(\infty)I^*$ , then we have  $\lambda \in \sigma(A_1 + C) \setminus E\sigma(A_1 + C)$  because  $\sup_{\lambda \in E\sigma(B)} \Re\lambda \leq \omega_1(B)$  for an infinitesimal generator  $B$  (see Webb 1985, Prop. 4.15). Moreover if  $\lambda \in \sigma(A_1 + C) \setminus E\sigma(A_1 + C)$ , then it follows that  $\lambda \in P_\sigma(A_1 + C)$  and  $\lambda$  is a pole of the resolvent  $(\lambda - (A_1 + C))^{-1}$  (Webb 1985, Prop. 4.11). Then we obtain the result (3), since a pole of the resolvent is no other than a pole of  $1/(1 - \Delta(\lambda))$ .  $\square$

**Proof of Proposition 4.2** It is sufficient to show the latter half of statement. First if  $\lambda \in \Lambda$  has the positive real part, it follows that

$$1 = |\Delta(\lambda)| \leq \int_0^\infty e^{-\frac{\Re\lambda}{k}a} \Phi(a) da \leq \frac{v}{\Re\lambda}.$$

Then we know that  $\Lambda \subset \{\lambda \in \mathbf{C} : -\gamma(\infty)I^* < \Re\lambda \leq v\}$ . Hence if we assume that there exists an infinite number of roots  $\lambda_n = \alpha_n + i\beta_n \in \Lambda$  ( $n = 1, 2, \dots$ ) with  $\alpha \leq \alpha_n \leq v$ , we can choose a subsequence,  $\lambda_{n(j)} = \alpha_{n(j)} + i\beta_{n(j)}$  ( $k = 1, 2, \dots$ ) such that  $\lim_{j \rightarrow \infty} \alpha_{n(j)} = \alpha^*$ , where  $\alpha \leq \alpha_{n(j)} \leq v$  and  $\lim_{j \rightarrow \infty} \beta_{n(j)} = \infty$ . Then it follows that

$$\begin{aligned} |\Delta(\lambda_{n(j)}) - \Delta(\alpha^* + i\beta_{n(j)})| &= |1 - \Delta(\alpha^* + i\beta_{n(j)})| \\ &\leq \int_0^\infty |e^{-\frac{\alpha_{n(j)}}{k}a} - e^{-\frac{\alpha^*}{k}a}| \Phi(a) da \rightarrow 0 \quad (j \rightarrow \infty). \end{aligned}$$

On the other hand, from Riemann-Lebesgue lemma, we have  $\lim_{j \rightarrow \infty} \Delta(\alpha^* + i\beta_{n(j)}) = 0$ , which contradicts the above result. This completes our proof.  $\square$

**Proof of Proposition 4.3** First note that for any small number  $\epsilon > 0$  there can be only finitely many  $\lambda$  such that  $\lambda \in P_\sigma(A_1 + C) \cap \{\lambda \in \mathbf{C} : \Re\lambda \geq -\gamma(\infty)I^* + \epsilon\}$ . Hence under our assumptions it follows that

$$\lambda_0 := \sup_{\lambda \in \sigma(A_1 + C) \setminus E\sigma(A_1 + C)} \Re\lambda < 0.$$

Therefore it follows from Proposition 4.13 of Webb (1985) that

$$\omega_0(A_1 + C) = \max\{\omega_1(A_1 + C), \lambda_0\} < 0.$$

Then it follows immediately that  $\lim_{t \rightarrow \infty} \|e^{t(A_1 + C)}\| = 0$ . By the principle of linearized stability, we can conclude that the endemic steady state is locally asymptotically stable.  $\square$

## REFERENCES

- [1] ANDERSON, R.M. AND R.M. MAY (1991), *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, Oxford.
- [2] ANDREASEN, V., S. LEVIN, AND J. LIN (1996), A model of influenza A drift evolution, *Z. Angew. Math. Mech.* **76**(S2): 421–424.
- [3] 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.
- [4] DIEKmann, O. AND S.A. VAN GILS (1984), Invariant manifolds for Volterra integral equations of convolution type, *J. Diff. Equ.* **54**: 139–180.
- [5] FRAUENTHAL, J.C. (1975), A dynamical model for human population growth, *Theor. Popul. Biol.* **8**: 64–73.
- [6] HASTINGS, A. (1995), A metapopulation model with population jumps of varying sizes, *Math. Biosci.* **128**: 285–298.
- [7] INABA, H. (1998), Mathematical analysis for an evolutionary epidemic model, in *Mathematical Models in Medical and Health Sciences*, Vanderbilt University Press, Nashville, pp. 213–236.
- [8] INABA, H. (2000), Revisiting Kermack and McKendrick, in *Mathematical Models and Functional Equations*, S. Sakata (ed.), Kokyuroku 1128, Research Institute for Mathematical Sciences, Kyoto University: 112–121.
- [9] INABA, H. (2001), Kermack and McKendrick revisited: The variable susceptibility model for infectious diseases (to appear in *Japan J. Indust. Appl. Math.* **18**(2)).
- [10] KERMACK, W.O. AND A.G. MCKENDRICK (1927), Contributions to the mathematical theory of epidemics-I, *Proceedings of the Royal Society* **115A**: 700–721 (reprinted in *Bulletin of Mathematical Biology* **53**(1/2): 33–55, 1991).
- [11] KERMACK, W.O. AND A.G. MCKENDRICK (1932), Contributions to the mathematical theory of epidemics-II. The problem of endemicity, *Proceedings of the Royal Society* **138A**: 55–83 (reprinted in *Bulletin of Mathematical Biology* **53**(1/2): 57–87, 1991).
- [12] KERMACK, W.O. AND A.G. MCKENDRICK (1933), Contributions to the mathematical theory of epidemics-III. Further studies of the problem of endemicity, *Proceedings of the Royal Society* **141A**: 94–122 (reprinted in *Bulletin of Mathematical Biology* **53**(1/2): 89–118, 1991).
- [13] METZ, J.A.J. AND O. DIEKmann (eds.) (1986), *The Dynamics of Physiologically Structured Populations*, Lecture Notes in Biomathematics **68**, Springer, Berlin.
- [14] PEASE, C.M. (1987), An evolutionary epidemiological mechanism, with applications to type A influenza, *Theor. Popul. Biol.* **31**: 422–452.
- [15] PRÜSS, J. (1983a), Stability analysis for equilibria in age-specific population dynamics, *Nonlinear Analysis, Theory, Methods and Applications* **7**(12): 1291–1313.
- [16] PRÜSS, J. (1983b), On the qualitative behaviour of populations with age-specific interactions, *Comp. and Maths. with Appl.* **9**(3): 327–339.
- [17] WEBB, G.F. (1985), *Theory of Nonlinear Age-Dependent Population Dynamics*, Marcel Dekker, New York and Basel.

## EPILOGUE

The evolution of epidemiological phenomena such as cross-immunity, differential susceptibility, differential disease progression and resistance do not take place on the same time scale as disease dynamics or demographic processes. Disease evolution models deal with processes with inherently distinct temporal epidemiological, demographic and evolutionary scales. The situation is more complex once the need to include changes in cultural and social landscapes is recognized. The accelerated birth and growth of urban centers, fast changes in public health policies, unpredictable demographic growth patterns, increases in the rates of antibiotic use and misuse, impact of globalization on populations' movement (growth of mass transportation, international travel and global warming), advances in genomics and the information revolution are but some of the factors that continuously shift the social and cultural landscape in which disease evolves.

These two volumes grew out of the discussions and presentations that began during the Workshop on Emerging and Reemerging Diseases sponsored by the Institute for Mathematics and its Applications (IMA) at the University of Minnesota in May 17–21, 1999. Workshop activities focused on specific diseases and issues connected to disease emergence or reemergence. Researchers' presentations dealt with possible manifestations of such phenomena while group discussions tried to identify the factors that must be considered in their study. Workshop participants did not attempt to precisely define these concepts.

Challenges posed by particular diseases must often be studied at different levels of complexity. The selection of an appropriate epidemiological unit, for example, has to be tailored to specific questions. The dynamics of HIV are studied (at least) at two levels: the level of the individual (immunology) and the population level (epidemiology). The search for answers to HIV questions posed at both levels has brought together teams of immunologists, sociologists, demographers, economists, mathematicians, statisticians, public health officials and evolutionary biologists. The development of preventive and timely vaccines for communicable diseases like influenza, a highly variable and potentially deadly virus, must now include increases in international and regional travel in its development. Renewed threats are being posed by recently observed increases in the number of active cases of tuberculosis (TB) and the number of infected individuals with resistant strains of TB. Factoring the costs associated with the treatment and control of diseases like HIV or TB makes disease control one of the most pressing and challenging issues of our time.

The articles in these two volumes illustrate the role of mathematics

in the study of particular diseases and, in the process, highlight important mathematical methods and modeling approaches. It is unfortunate that, for the most part, these volumes have ignored the study of vector-transmitted (malaria, Chagas, etc.) and macroparasitic diseases. There are no articles addressing the study of one of the most scariest diseases, Ebola, the study of animal diseases or the role of humans on the spread of diseases that can devastate local, national and global economies (“foot and mouth disease”). The methods and modeling approaches found in these volumes, particularly those in the tutorial papers, provide a useful starting point for the development of models that may help us gain some understanding on the transmission dynamics and control of these diseases that are ignored in these volumes.

We hope that those interested in the study of communicable diseases will find the methods and tools provided in these volumes useful.

Carlos Castillo-Chavez, Sally Blower, Pauline van den Driessche,  
Denise Kirschner, and Abdul-Aziz Yakubu.

## **LIST OF TUTORIAL/WORKSHOP PARTICIPANTS**

- Linda Allen, Department of Mathematics, Texas Tech
- Kevin Anderson, Institute for Mathematics and its Applications
- Viggo Andreasen, Department of Matematik, Roskilde University
- Bruce Ayati, Institute for Mathematics and its Applications
- Frank Ball, School of Mathematical Sciences, University of Nottingham
- Joan Bechtold, Midwest Orthopedic Research Foundation
- Robert Berry, Department of Mathematics, University of Houston
- Sally Blower, Department of Biomathematics, UCLA School of Medicine
- Mary E. Bradley, Department of Mathematics, University of Louisville
- Fred Brauer, Department of Mathematics, University of British Columbia
- Angel Capurro, Departamento de Investigacion, Universidad de Belgrano
- Carlos Castillo-Chavez, Departments of Biometrics, Statistics and Theoretical and Applied Mechanics, Cornell University
- Lester F. Caudill, Department Mathematics and Computer Science, University of Richmond
- Kenneth Cooke, Pomona College
- Fred Dulles, Institute for Mathematics and its Applications
- Jonathan Dushoff, Institute of Physics, Academia Sinica
- Lourdes Esteva, Departamento de Matematicas, Fac. de Ciencias, University Nac Autonoma de Mexico
- Zhilan Feng, Department of Mathematics, Purdue University
- Neil Ferguson, Imperial College
- K. Renee Fister, Department of Mathematics and Statistics, Murray State University
- John Gardner, University of Houston,
- Don Gaver, Naval Postgraduate School
- M. Gabriela M. Gomes, Department of Biological Sciences, University of Warwick
- Kimber Gross, Department of Mathematics, University of Houston
- Mats Gyllenberg, Department of Applied Mathematics, University of Turku
- Carlos Hernandez, Department of Biomathematics, Cornell University

- Herb Hethcote, Department of Mathematics, University of Iowa
- Janda Holmes, Department of Mathematics, University of Tennessee
- Sarah Holte, Department of Biostatistics, Fred Hutchinson Cancer Research Center
- Mary Ann Horn, Department of Mathematics, Vanderbilt University
- Shu-Fang Hsu-Schmidt, Institute of Mathematical Statistics, University of Bern
- Wenzhang Huang, Department of Mathematical Sciences, University of Alabama, Huntsville
- Hisashi Inaba, Department of Mathematical Sciences, University of Tokyo
- Yoh Iwasa, Department of Biology, Kyushu University
- John Kemper, University of St. Thomas
- Denise Kirschner, Department of Microbiology and Immunology, University of Michigan Medical School
- Steve Krone, Department of Mathematics, University of Idaho
- Suzanne Lenhart, Department of Mathematics, University of Tennessee
- Jia Li, Department of Mathematical Science, University of Alabama, Huntsville
- Michael Li, Department of Mathematical Sciences, University of Alberta, Edmonton
- Tom Lietman, F.I. Proctor Foundation, University of California-San Francisco
- Frithjof Lutscher, Lehrstuhl fur Biomathematik, Universitat Tbingen
- Maia Martcheva, Department of Mathematics, Polytechnic University
- Gustavo Martinez-Meckler, Centro de Ciencias Fiscas, UNAM
- Douglas Meade, Department of Mathematics, University of South Carolina
- Willard Miller, Institute for Mathematics and its Applications
- Fabio Milner, Department of Mathematics, Purdue University
- Olivier Monzin, MORF
- Johannes Mueller, Biomathematics - C5P13, University of Tbingen
- Ingemar Nasell, Department of Mathematics, Royal Institute of Technology
- Patrick Nelson, Department of Mathematics, University of Michigan
- Douglas Norton, Mathematical Sciences, Villanova University
- Mouzin Olivier, Biomechanics Orthopaedic Laboratory

- Andrea Pugliese, Dipartimento di Matematica, Universita degli Studi di Trento
- Kathleen Rogers, Mathematics Department, University of Maryland Baltimore County
- Sonja Sandberg, Department of Mathematics, Framingham State College
- Lisa Sattenspiel, Department of Anthropology, University of Missouri
- Matthew Schuette, Applied Mathematics and Computational Sciences, University of Iowa
- Carl P. Simon, Department of Mathematics, University of Michigan
- Zachariah Sinkala, Middle Tennessee State University
- Baojun Song, Department of Biometrics, Cornell University
- Angela Stevens, Max-Planck-Institute for Mathematics in the Sciences
- Moxun Tang, School of Mathematics, University of Minnesota
- Horst Thieme, Department of Mathematics, Arizona State University
- Dean T. Tsukayama, Department of Medicine, Hennepin County Medical Center
- Pauline van den Driessche, Department of Mathematics and Statistics, University of Victoria
- Jorge X. Velasco-Hernandez, Department of Mathematics, UAM-Iztapalapa
- Liancheng Wang, Department of Mathematics and Statistics, Mississippi State University
- James Watmough, Department of Mathematics and Statistics, University of New Brunswick
- Glenn F. Webb, Department of Mathematics, Vanderbilt University
- Kathryn Weld, Department of Mathematics and Computer Science, Manhattan College
- Ralf Wittenberg, Department of Mathematics, University of Michigan
- Lih-Ing Wu, Department of Mathematics, Purdue University
- Abdul-Aziz Yakubu, Department of Mathematics, Howard University
- Zhou Yicang, Department of Applied Mathematics, Xi'an Jiaotong University
- Mary Lou Zeeman, Department of Mathematics and Statistics, University of Texas at San Antonio

IMA VOLUME 125 CONTENTS  
**MATHEMATICAL APPROACHES FOR EMERGING AND  
REEMERGING INFECTIOUS DISEASES: AN  
INTRODUCTION**

EDITORS: CARLOS CASTILLO-CHAVEZ WITH SALLY BLOWER,  
PAULINE VAN DEN DRIESSCHE, DENISE KIRSCHNER,  
AND ABDUL-AZIZ YAKUBU

Foreword .....	v
Preface .....	vii–viii
New directions in the mathematics of infectious disease.....	1–5
<i>Simon A. Levin</i>	
Fred Brauer: The man and his mathematics.....	7–20
<i>Christopher M. Kribs-Zaleta</i>	
Kenneth L. Cooke: Researcher, educator par excellence.....	21–30
<i>P. van den Driessche</i>	
Basic ideas of mathematical epidemiology .....	31–65
<i>Fred Brauer</i>	
Extensions of the basic models .....	67–95
<i>Fred Brauer</i>	
New vaccination strategies for pertussis .....	97–118
<i>Herbert W. Hethcote</i>	
Time delay in epidemic models .....	119–128
<i>P. van den Driessche</i>	
Nonlocal response in a simple epidemiological model .....	129–151
<i>K.R. Heiderich, W. Huang, and C. Castillo-Chavez</i>	
Discrete-time S-I-S models with simple and complex population dynamics.....	153–163
<i>Carlos Castillo-Chavez and Abdul-Aziz Yakubu</i>	
Intraspecific competition, dispersal and disease dynamics in discrete-time patchy environments.....	165–181
<i>Carlos Castillo-Chavez and Abdul-Aziz Yakubu</i>	
The impact of long-range dispersal on the rate of spread in population and epidemic models.....	183–197
<i>Linda J.S. Allen and Ralynn K. Ernest</i>	
Endemicity, persistence, and quasi-stationarity .....	199–227
<i>Ingemar Nåsell</i>	

On the computation of $\mathcal{R}_0$ and its role in global stability.....	229–250
<i>Carlos Castillo-Chavez, Zhilan Feng, and Wenzhang Huang</i>	
Nonlinear mating models for populations with discrete generations.....	251–268
<i>Carlos Castillo-Chavez, Abdul-Aziz Yakubu, Horst Thieme, and Maia Martcheva</i>	
Center manifolds and normal forms in epidemic models.....	269–286
<i>Christopher M. Kribs-Zaleta</i>	
Remarks on modeling host-viral dynamics and treatment .....	287–308
<i>Jorge X. Velasco-Hernández, José A. García and Denise E. Kirschner</i>	
A multiple compartment model for the evolution of HIV-1 after highly active antiretroviral therapy .....	309–323
<i>G. Cocho, L. Huerta, G. Martinez-Mekler, and C. Villarreal</i>	
Modeling cancer as an infectious disease: The epidemiology of <i>Helicobacter pylori</i> .....	325–339
<i>H.B. Gershengorn and S.M. Blower</i>	
Frequency dependent risk of infection and the spread of infectious diseases.....	341–350
<i>Juan P. Aparicio, Angel F. Capurro, and Carlos Castillo-Chavez</i>	
Long-term dynamics and re-emergence of tuberculosis.....	351–360
<i>Juan P. Aparicio, Angel F. Capurro, and Carlos Castillo-Chavez</i>	
Epilogue.....	361–362
List of tutorial/workshop participants.....	363–365
IMA volume 126 contents: Mathematical approaches for emerging and reemerging infectious diseases: models, methods and theory .....	367–368

## **IMA SUMMER PROGRAMS**

- 1987 Robotics  
1988 Signal Processing  
1989 Robust Statistics and Diagnostics  
1990 Radar and Sonar (June 18–29)  
New Directions in Time Series Analysis (July 2–27)  
1991 Semiconductors  
1992 Environmental Studies: Mathematical, Computational, and Statistical Analysis  
1993 Modeling, Mesh Generation, and Adaptive Numerical Methods for Partial Differential Equations  
1994 Molecular Biology  
1995 Large Scale Optimizations with Applications to Inverse Problems, Optimal Control and Design, and Molecular and Structural Optimization  
1996 Emerging Applications of Number Theory (July 15–26)  
Theory of Random Sets (August 22–24)  
1997 Statistics in the Health Sciences  
1998 Coding and Cryptography (July 6–18)  
Mathematical Modeling in Industry (July 22–31)  
1999 Codes, Systems, and Graphical Models (August 2–13, 1999)  
2000 Mathematical Modeling in Industry: A Workshop for Graduate Students (July 19–28)  
2001 Geometric Methods in Inverse Problems and PDE Control (July 16–27)  
2002 Special Functions in the Digital Age (July 22–August 2)

## **IMA “HOT TOPICS” WORKSHOPS**

- Challenges and Opportunities in Genomics: Production, Storage, Mining and Use, April 24–27, 1999
- Decision Making Under Uncertainty: Energy and Environmental Models, July 20–24, 1999
- Analysis and Modeling of Optical Devices, September 9–10, 1999
- Decision Making under Uncertainty: Assessment of the Reliability of Mathematical Models, September 16–17, 1999
- Scaling Phenomena in Communication Networks, October 22–24, 1999
- Text Mining, April 17–18, 2000
- Mathematical Challenges in Global Positioning Systems (GPS), August 16–18, 2000
- Modeling and Analysis of Noise in Integrated Circuits and Systems, August 29–30, 2000
- Mathematics of the Internet: E-Auction and Markets, December 3–5, 2000
- Analysis and Modeling of Industrial Jetting Processes, January 10–13, 2001

- Wireless Networks, August 6–10 2001

## **SPRINGER LECTURE NOTES FROM THE IMA:**

*The Mathematics and Physics of Disordered Media*

Editors: Barry Hughes and Barry Ninham  
(Lecture Notes in Math., Volume 1035, 1983)

*Orienting Polymers*

Editor: J.L. Ericksen  
(Lecture Notes in Math., Volume 1063, 1984)

*New Perspectives in Thermodynamics*

Editor: James Serrin  
(Springer-Verlag, 1986)

*Models of Economic Dynamics*

Editor: Hugo Sonnenschein  
(Lecture Notes in Econ., Volume 264, 1986)

## The IMA Volumes in Mathematics and its Applications

---

### *Current Volumes:*

- 1 **Homogenization and Effective Moduli of Materials and Media**  
J. Ericksen, D. Kinderlehrer, R. Kohn, and J.-L. Lions (eds.)
- 2 **Oscillation Theory, Computation, and Methods of Compensated Compactness** C. Dafermos, J. Ericksen, D. Kinderlehrer, and M. Slemrod (eds.)
- 3 **Metastability and Incompletely Posed Problems**  
S. Antman, J. Ericksen, D. Kinderlehrer, and I. Muller (eds.)
- 4 **Dynamical Problems in Continuum Physics**  
J. Bona, C. Dafermos, J. Ericksen, and D. Kinderlehrer (eds.)
- 5 **Theory and Applications of Liquid Crystals**  
J. Ericksen and D. Kinderlehrer (eds.)
- 6 **Amorphous Polymers and Non-Newtonian Fluids**  
C. Dafermos, J. Ericksen, and D. Kinderlehrer (eds.)
- 7 **Random Media** G. Papanicolaou (ed.)
- 8 **Percolation Theory and Ergodic Theory of Infinite Particle Systems** H. Kesten (ed.)
- 9 **Hydrodynamic Behavior and Interacting Particle Systems**  
G. Papanicolaou (ed.)
- 10 **Stochastic Differential Systems, Stochastic Control Theory, and Applications** W. Fleming and P.-L. Lions (eds.)
- 11 **Numerical Simulation in Oil Recovery** M.F. Wheeler (ed.)
- 12 **Computational Fluid Dynamics and Reacting Gas Flows**  
B. Engquist, M. Luskin, and A. Majda (eds.)
- 13 **Numerical Algorithms for Parallel Computer Architectures**  
M.H. Schultz (ed.)
- 14 **Mathematical Aspects of Scientific Software** J.R. Rice (ed.)
- 15 **Mathematical Frontiers in Computational Chemical Physics**  
D. Truhlar (ed.)
- 16 **Mathematics in Industrial Problems** A. Friedman
- 17 **Applications of Combinatorics and Graph Theory to the Biological and Social Sciences** F. Roberts (ed.)
- 18 ***q*-Series and Partitions** D. Stanton (ed.)
- 19 **Invariant Theory and Tableaux** D. Stanton (ed.)
- 20 **Coding Theory and Design Theory Part I: Coding Theory**  
D. Ray-Chaudhuri (ed.)
- 21 **Coding Theory and Design Theory Part II: Design Theory**  
D. Ray-Chaudhuri (ed.)
- 22 **Signal Processing Part I: Signal Processing Theory**  
L. Auslander, F.A. Grünbaum, J.W. Helton, T. Kailath, P. Khargonekar, and S. Mitter (eds.)

- 23 **Signal Processing Part II: Control Theory and Applications of Signal Processing** L. Auslander, F.A. Grünbaum, J.W. Helton, T. Kailath, P. Khargonekar, and S. Mitter (eds.)
- 24 **Mathematics in Industrial Problems, Part 2** A. Friedman
- 25 **Solitons in Physics, Mathematics, and Nonlinear Optics**  
P.J. Olver and D.H. Sattinger (eds.)
- 26 **Two Phase Flows and Waves**  
D.D. Joseph and D.G. Schaeffer (eds.)
- 27 **Nonlinear Evolution Equations that Change Type**  
B.L. Keyfitz and M. Shearer (eds.)
- 28 **Computer Aided Proofs in Analysis**  
K. Meyer and D. Schmidt (eds.)
- 29 **Multidimensional Hyperbolic Problems and Computations**  
A. Majda and J. Glimm (eds.)
- 30 **Microlocal Analysis and Nonlinear Waves**  
M. Beals, R. Melrose, and J. Rauch (eds.)
- 31 **Mathematics in Industrial Problems, Part 3** A. Friedman
- 32 **Radar and Sonar, Part I**  
R. Blahut, W. Miller, Jr., and C. Wilcox
- 33 **Directions in Robust Statistics and Diagnostics: Part I**  
W.A. Stahel and S. Weisberg (eds.)
- 34 **Directions in Robust Statistics and Diagnostics: Part II**  
W.A. Stahel and S. Weisberg (eds.)
- 35 **Dynamical Issues in Combustion Theory**  
P. Fife, A. Liñán, and F.A. Williams (eds.)
- 36 **Computing and Graphics in Statistics**  
A. Buja and P. Tukey (eds.)
- 37 **Patterns and Dynamics in Reactive Media**  
H. Swinney, G. Aris, and D. Aronson (eds.)
- 38 **Mathematics in Industrial Problems, Part 4** A. Friedman
- 39 **Radar and Sonar, Part II**  
F.A. Grünbaum, M. Bernfeld, and R.E. Blahut (eds.)
- 40 **Nonlinear Phenomena in Atmospheric and Oceanic Sciences**  
G.F. Carnevale and R.T. Pierrehumbert (eds.)
- 41 **Chaotic Processes in the Geological Sciences** D.A. Yuen (ed.)
- 42 **Partial Differential Equations with Minimal Smoothness and Applications** B. Dahlberg, E. Fabes, R. Fefferman, D. Jerison, C. Kenig, and J. Pipher (eds.)
- 43 **On the Evolution of Phase Boundaries**  
M.E. Gurtin and G.B. McFadden
- 44 **Twist Mappings and Their Applications**  
R. McGehee and K.R. Meyer (eds.)

- 45 **New Directions in Time Series Analysis, Part I**  
D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt,  
and M.S. Taqqu (eds.)
- 46 **New Directions in Time Series Analysis, Part II**  
D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt,  
and M.S. Taqqu (eds.)
- 47 **Degenerate Diffusions**  
W.-M. Ni, L.A. Peletier, and J.-L. Vazquez (eds.)
- 48 **Linear Algebra, Markov Chains, and Queueing Models**  
C.D. Meyer and R.J. Plemmons (eds.)
- 49 **Mathematics in Industrial Problems, Part 5** A. Friedman
- 50 **Combinatorial and Graph-Theoretic Problems in Linear Algebra**  
R.A. Brualdi, S. Friedland, and V. Klee (eds.)
- 51 **Statistical Thermodynamics and Differential Geometry  
of Microstructured Materials**  
H.T. Davis and J.C.C. Nitsche (eds.)
- 52 **Shock Induced Transitions and Phase Structures in General  
Media** J.E. Dunn, R. Fosdick, and M. Slemrod (eds.)
- 53 **Variational and Free Boundary Problems**  
A. Friedman and J. Spruck (eds.)
- 54 **Microstructure and Phase Transitions**  
D. Kinderlehrer, R. James, M. Luskin, and J.L. Ericksen (eds.)
- 55 **Turbulence in Fluid Flows: A Dynamical Systems Approach**  
G.R. Sell, C. Foias, and R. Temam (eds.)
- 56 **Graph Theory and Sparse Matrix Computation**  
A. George, J.R. Gilbert, and J.W.H. Liu (eds.)
- 57 **Mathematics in Industrial Problems, Part 6** A. Friedman
- 58 **Semiconductors, Part I**  
W.M. Coughran, Jr., J. Cole, P. Lloyd, and J. White (eds.)
- 59 **Semiconductors, Part II**  
W.M. Coughran, Jr., J. Cole, P. Lloyd, and J. White (eds.)
- 60 **Recent Advances in Iterative Methods**  
G. Golub, A. Greenbaum, and M. Luskin (eds.)
- 61 **Free Boundaries in Viscous Flows**  
R.A. Brown and S.H. Davis (eds.)
- 62 **Linear Algebra for Control Theory**  
P. Van Dooren and B. Wyman (eds.)
- 63 **Hamiltonian Dynamical Systems: History, Theory,  
and Applications**  
H.S. Dumas, K.R. Meyer, and D.S. Schmidt (eds.)
- 64 **Systems and Control Theory for Power Systems**  
J.H. Chow, P.V. Kokotovic, R.J. Thomas (eds.)
- 65 **Mathematical Finance**  
M.H.A. Davis, D. Duffie, W.H. Fleming, and S.E. Shreve (eds.)

- 66      **Robust Control Theory** B.A. Francis and P.P. Khargonekar (eds.)  
67      **Mathematics in Industrial Problems, Part 7** A. Friedman  
68      **Flow Control** M.D. Gunzburger (ed.)  
69      **Linear Algebra for Signal Processing**  
        A. Bojanczyk and G. Cybenko (eds.)  
70      **Control and Optimal Design of Distributed Parameter Systems**  
        J.E. Lagnese, D.L. Russell, and L.W. White (eds.)  
71      **Stochastic Networks** F.P. Kelly and R.J. Williams (eds.)  
72      **Discrete Probability and Algorithms**  
        D. Aldous, P. Diaconis, J. Spencer, and J.M. Steele (eds.)  
73      **Discrete Event Systems, Manufacturing Systems,  
        and Communication Networks**  
        P.R. Kumar and P.P. Varaiya (eds.)  
74      **Adaptive Control, Filtering, and Signal Processing**  
        K.J. Åström, G.C. Goodwin, and P.R. Kumar (eds.)  
75      **Modeling, Mesh Generation, and Adaptive Numerical Methods  
        for Partial Differential Equations** I. Babuska, J.E. Flaherty,  
        W.D. Henshaw, J.E. Hopcroft, J.E. Oliger, and T. Tezduyar (eds.)  
76      **Random Discrete Structures** D. Aldous and R. Pemantle (eds.)  
77      **Nonlinear Stochastic PDEs: Hydrodynamic Limit and Burgers'  
        Turbulence** T. Funaki and W.A. Woyczyński (eds.)  
78      **Nonsmooth Analysis and Geometric Methods in Deterministic  
        Optimal Control** B.S. Mordukhovich and H.J. Sussmann (eds.)  
79      **Environmental Studies: Mathematical, Computational,  
        and Statistical Analysis** M.F. Wheeler (ed.)  
80      **Image Models (and their Speech Model Cousins)**  
        S.E. Levinson and L. Shepp (eds.)  
81      **Genetic Mapping and DNA Sequencing**  
        T. Speed and M.S. Waterman (eds.)  
82      **Mathematical Approaches to Biomolecular Structure and Dynamics**  
        J.P. Mesirov, K. Schulten, and D. Sumners (eds.)  
83      **Mathematics in Industrial Problems, Part 8** A. Friedman  
84      **Classical and Modern Branching Processes**  
        K.B. Athreya and P. Jagers (eds.)  
85      **Stochastic Models in Geosystems**  
        S.A. Molchanov and W.A. Woyczyński (eds.)  
86      **Computational Wave Propagation**  
        B. Engquist and G.A. Kriegsmann (eds.)  
87      **Progress in Population Genetics and Human Evolution**  
        P. Donnelly and S. Tavaré (eds.)  
88      **Mathematics in Industrial Problems, Part 9** A. Friedman  
89      **Multiparticle Quantum Scattering With Applications to Nuclear,  
        Atomic and Molecular Physics** D.G. Truhlar and B. Simon (eds.)

- 90   **Inverse Problems in Wave Propagation** G. Chavent,  
G. Papanicolaou, P. Sacks, and W.W. Symes (eds.)
- 91   **Singularities and Oscillations** J. Rauch and M. Taylor (eds.)
- 92   **Large-Scale Optimization with Applications, Part I:**  
**Optimization in Inverse Problems and Design**  
L.T. Biegler, T.F. Coleman, A.R. Conn, and F. Santosa (eds.)
- 93   **Large-Scale Optimization with Applications, Part II:**  
**Optimal Design and Control**  
L.T. Biegler, T.F. Coleman, A.R. Conn, and F. Santosa (eds.)
- 94   **Large-Scale Optimization with Applications, Part III:**  
**Molecular Structure and Optimization**  
L.T. Biegler, T.F. Coleman, A.R. Conn, and F. Santosa (eds.)
- 95   **Quasiclassical Methods**  
J. Rauch and B. Simon (eds.)
- 96   **Wave Propagation in Complex Media**  
G. Papanicolaou (ed.)
- 97   **Random Sets: Theory and Applications**  
J. Goutsias, R.P.S. Mahler, and H.T. Nguyen (eds.)
- 98   **Particulate Flows: Processing and Rheology**  
D.A. Drew, D.D. Joseph, and S.L. Passman (eds.)
- 99   **Mathematics of Multiscale Materials** K.M. Golden, G.R. Grimmett,  
R.D. James, G.W. Milton, and P.N. Sen (eds.)
- 100   **Mathematics in Industrial Problems, Part 10** A. Friedman
- 101   **Nonlinear Optical Materials** J.V. Moloney (ed.)
- 102   **Numerical Methods for Polymeric Systems** S.G. Whittington (ed.)
- 103   **Topology and Geometry in Polymer Science** S.G. Whittington,  
D. Sumners, and T. Lodge (eds.)
- 104   **Essays on Mathematical Robotics** J. Baillieul, S.S. Sastry,  
and H.J. Sussmann (eds.)
- 105   **Algorithms For Parallel Processing** M.T. Heath, A. Ranade,  
and R.S. Schreiber (eds.)
- 106   **Parallel Processing of Discrete Problems** P.M. Pardalos (ed.)
- 107   **The Mathematics of Information Coding, Extraction, and  
Distribution** G. Cybenko, D.P. O'Leary, and J. Rissanen (eds.)
- 108   **Rational Drug Design** D.G. Truhlar, W. Howe, A.J. Hopfinger,  
J. Blaney, and R.A. Dammkohler (eds.)
- 109   **Emerging Applications of Number Theory** D.A. Hejhal,  
J. Friedman, M.C. Gutzwiller, and A.M. Odlyzko (eds.)
- 110   **Computational Radiology and Imaging: Therapy and Diagnostics**  
C. Börgers and F. Natterer (eds.)
- 111   **Evolutionary Algorithms** L.D. Davis, K. De Jong, M.D. Vose,  
and L.D. Whitley (eds.)
- 112   **Statistics in Genetics** M.E. Halloran and S. Geisser (eds.)

- 113 **Grid Generation and Adaptive Algorithms** M.W. Bern,  
J.E. Flaherty, and M. Luskin (eds.)
- 114 **Diagnosis and Prediction** S. Geisser (ed.)
- 115 **Pattern Formation in Continuous and Coupled Systems: A Survey Volume** M. Golubitsky, D. Luss, and S.H. Strogatz (eds.)
- 116 **Statistical Models in Epidemiology, the Environment, and Clinical Trials** M.E. Halloran and D. Berry (eds.)
- 117 **Structured Adaptive Mesh Refinement (SAMR) Grid Methods**  
S.B. Baden, N.P. Chrisochoides, D.B. Gannon, and M.L. Norman (eds.)
- 118 **Dynamics of Algorithms**  
R. de la Llave, L.R. Petzold, and J. Lorenz (eds.)
- 119 **Numerical Methods for Bifurcation Problems and Large-Scale Dynamical Systems**  
E. Doedel and L.S. Tuckerman (eds.)
- 120 **Parallel Solution of Partial Differential Equations**  
P. Bjørstad and M. Luskin (eds.)
- 121 **Mathematical Models for Biological Pattern Formation**  
P.K. Maini and H.G. Othmer (eds.)
- 122 **Multiple-Time-Scale Dynamical Systems**  
C.K.R.T. Jones and A. Khibnik (eds.)
- 123 **Codes, Systems, and Graphical Models**  
B. Marcus and J. Rosenthal (eds.)
- 124 **Computational Modeling in Biological Fluid Dynamics**  
L.J. Fauci and S. Gueron (eds.)
- 125 **Mathematical Approaches for Emerging and Reemerging Infectious Diseases: An Introduction**  
C. Castillo-Chavez with S. Blower, P. van den Driessche, D. Kirschner, and A.A. Yakubu (eds.)
- 126 **Mathematical Approaches for Emerging and Reemerging Infectious Diseases: Models, Methods, and Theory**  
C. Castillo-Chavez with S. Blower, P. van den Driessche, D. Kirschner, and A.A. Yakubu (eds.)
- 127 **Mathematics of the Internet: E-Auction and Markets**  
B. Dietrich and R.V. Vohra (eds.)

**Forthcoming Volumes:**

1998–1999: Mathematics in Biology

Membrane Transport and Renal Physiology

Decision Making Under Uncertainty: Energy and Environmental Models

1999–2000: Reactive Flow and Transport Phenomena

Confinement and Remediation of Environmental Hazards and Resource Recovery

Atmospheric Modeling

Dispersive Corrections to Transport Equations, Simulation of Transport in Transition Regimes/Multiscale Models for Surface Evolution and Reacting Flows

2000-2001: Mathematics in Multimedia

Mathematical Foundations of Speech Processing and Recognition/Mathematical Foundations of Natural Language Modeling

Mathematical Methods in Speech and Image Analysis/ Image Processing and Low Level Vision/Image Analysis and High Level Vision

Fractals in Multimedia

2001 Summer Program

Geometric Methods in Inverse Problems and PDE Control