

*Joseph Lemaître*

# Modeling infectious disease dynamics towards informed public health interventions: applications on COVID-19 and cholera

OCTOBER 24, 2021

REMOVE THIS TITLE PAGE AND REPLACE WITH OFFICIAL, PLEASE :)



*Vedrò con mio diletto  
L'alma dell'alma mia  
Il core del mio cor  
Pien di contento  
Vedrò con mio diletto  
L'alma dell'alma mia  
Il cor di questo cor  
Pien di contento  
E se dal caro oggetto  
Lungi convien che sia  
Sospirerò penando  
Ogni momento  
Vedrò con mio diletto  
L'alma dell'alma mia  
Il core del mio cor  
Pien di contento  
Vedrò con mio diletto  
L'alma dell'alma mia  
Il cor di questo cor  
Pien di contento*  
- Anastasio



# ACKNOWLEDGEMENTS

The present thesis is the result of the past 4.5 years spent studying the dynamics of cholera and COVID-19. While challenging, this journey has been incredibly rewarding – in no small part thanks to amazing colleagues. I am grateful to Prof. Andrea Rinaldo for warmly welcoming me to science and academia, and for allowing me to explore a subject that I grew to love. Thank you for encouraging me to pursue my research interests while benefiting from your unconditional support.

My most sincere thanks go to Damiano Pasetto for deeply caring about every aspect of my Ph.D. experience and for either pushing or carrying me through these years. It has been wonderful to evolve under your supervision. Scientifically, this thesis has benefited enormously from Javier Perez-Saez. Thank you for your insights on modeling, inference, and life in general. Thank you Mario Zanon for making learning complicated optimal control algorithms a pleasure. Many thanks to Jacques Fellay and Andrew Azman for your scientific insights and for introducing me to new lines of work.

Upon my arrival at the ECHO lab in 2017, I've been welcomed by wonderful friends and colleagues. Thanks Silvia for the best tea cookies ever; Giezi for caring so much about the lab and organizing our pasta Thursdays; and Luca because sports and meta-science are meant to go together. I wish to thank Mitra for sharing delicious Iranian delicacies and a unique perspective on life; Paolo for the support at all time and spontaneous lunch discussions; Filippo for teaching me experimental science and theoretical parenting. Finally Cristiano: I am truly grateful to have you as an office mate, colleague, and friend. It did change my last year at EPFL.

From my intense and abbreviated stay in Baltimore, I would like to thank Prof. Justin Lessler for welcoming me and for trusting me on the scenario pipeline project. Thank you Elizabeth, because wine and sports mix so well with science. Thanks, Hannah, Kyu, Shaun, Steve, and Quifang, for the inspirational science and the warm welcome in Baltimore. We loved our time there. Thanks to the COVID Scenario Pipeline team, working at night is much more pleasant with you all; especially to Josh for long hours pair debugging sessions.

Merci à mes ami-e-s et à ma famille pour votre soutien et tous les moments de joie; les cafés en EL, les longues discussions à toute heure et toutes ces belles réunions.

Merci à mes parents de m'avoir ouvert au monde et de m'avoir tant encouragé et soutenu au cours de cette thèse. À Céline, Eugène et Olivier, merci évidemment pour tout; sans votre folie et votre aide, cela n'aurait pas été pareil.

Finalement, merci Marion ! Merci d'avoir traversé avec moi toutes ces épreuves, merci d'avoir donné un sens à tout ça, merci d'avoir rendu ces quatre ans aussi doux. Merci pour tout le reste évidemment. Merci au merveilleux petit bébé qui bouge pour m'encourager et que l'on a si hâte de rencontrer !

## STANDING ON THE SHOULDER OF GIANTS

During this thesis, I have been fortunate to participate to the GTFCC annual meeting, IDDconf, and SISMID. These conferences were an incredible learning experience, and I would like to thank the infectious disease modeling community for being so welcoming, inclusive, and patient. I learned and still learn every day from many of you.

This thesis is built on many datasets, from rainfall to reported cholera cases. I am very grateful to the workers all along the chain from collection to curation who have made my job possible.

Thanks to the reviewers and co-authors from whom I have learned both scientific concepts and valuable writing tips.

Tools shape the way you think about problems. This thesis was not only made possible but also sculpted by the languages and libraries that were used. I want to acknowledge all the open-source maintainers and contributors who tirelessly make, document, and maintain powerful tools, enabling everybody to use cutting-edge methods and to fill bug-reports.



# SUMMARY

Emerging and existing infectious diseases pose a constant threat to individuals and communities across the world. In many cases, the burden of these diseases is preventable through public health interventions. However, taking the right decisions and designing effective policies is an intricate task: epidemics are complex phenomena resulting from the interaction between the environment, pathogens, individuals, and societies. Modeling offers a principled way to reason about infectious disease dynamics from scarce and biased information and to guide decision-makers towards effective policies.

This thesis tackles selected topics in cholera and COVID-19 modeling towards informed public-health decisions. These two contrasting diseases were associated by a twist of fate, but also through the lens of a common modeling approach: compartmental, SIR-based, models are conditioned on the available evidence using computer-age statistical inference frameworks. A set of five models is developed, each tackling a different facet of the spread and control of these two infectious diseases. Each model aims at answering questions related to either the understanding of the mechanisms behind disease transmission, the projection of the future dynamics under different scenarios, or the assessment of the effectiveness of past interventions. Moreover, a novel application of epidemiological models to the formal design of control policies is proposed. Optimal control provides a rigorous framework to identify the most effective control measures under a set of operational constraints, providing a benchmark on what it is possible to achieve with the available resources.

The results presented in this thesis range from scientific insight on the relationship between cholera and rainfall in Juba, South Sudan to the COVID Scenario Pipeline which produces reports used to inform the response to the COVID-19 pandemic of different governmental entities. Furthermore, the effectiveness of the non-pharmaceutical interventions against COVID-19 in Switzerland is evaluated; and so is the probability of eliminating cholera from Haiti under different scenarios of mass vaccination campaigns. Finally, the development of an optimal control framework towards the effective spatial allocation of vaccines against SARS-CoV-2 in Italy closes this conversation of models.

The present thesis demonstrates how infectious disease modeling enables informed decision-making by projecting the uncertainties under the light of the available evidence. It also highlights the effort needed to tailor the models and inference methods to the specificities of the transmission setting and the research question considered. From insights on transmission pathways to weekly reports aimed at decision-makers, it explores different applications of infectious disease modeling. Methods developed along the way enrich the toolbox available to modelers, to guide policy decisions further towards a reduction of the burden of infectious diseases on communities.

**KEYWORDS** cholera, COVID-19, epidemiology, public health, ecohydrology, infectious disease dynamics, SIR, mathematical modeling, statistical inference, optimal control.



# RÉSUMÉ

Existantes ou émergentes, les maladies infectieuses constituent une menace constante sur les individus et les communautés du monde entier. Dans de nombreux cas, le fardeau de ces maladies peut être évité grâce à des interventions de santé publique. Cependant, prendre les bonnes décisions et concevoir des politiques efficaces sont des tâches difficiles : les épidémies sont des phénomènes complexes résultant de l'interaction entre l'environnement, les agents pathogènes, les individus et les sociétés. La modélisation offre un moyen pour raisonner sur la dynamique de ces maladies à partir d'informations rares et biaisées, et ainsi guider les décideurs vers des politiques efficaces.

Cette thèse aborde la modélisation du choléra et de la COVID-19 pour la prise de décisions de santé publique éclairées. Ces deux maladies très différentes ont été associées ici par le hasard des événements, mais aussi par le biais d'une approche analytique commune: des modèles compartimentaux, basés sur le SIR, sont conditionnés par les données disponibles en utilisant des méthodes d'inférence statistique modernes. Un ensemble de cinq modèles y est développé, chacun abordant des facettes différentes de la propagation et du contrôle de ces deux maladies. Chaque modèle cherche à répondre à des questions liées soit à la compréhension des mécanismes de transmission, soit à la projection de la dynamique future selon différents scénarios, ou encore à l'évaluation de l'efficacité d'interventions passées. En outre, une nouvelle application des modèles épidémiologiques y est proposée: la conception formelle de politiques d'allocation de ressource de contrôle. La commande optimale offre un cadre rigoureux pour identifier les mesures de contrôle les plus efficaces tout en respectant les contraintes opérationnelles, ce qui permet de déterminer ce qu'il est possible d'atteindre avec les ressources disponibles.

Les résultats présentés dans cette thèse vont de la compréhension scientifique de la relation entre le choléra et les précipitations à Juba, au Sud-Soudan, à la COVID Scenario Pipeline qui produit des rapports utilisés par différentes entités gouvernementales pour répondre à la pandémie de COVID-19 . En outre, l'efficacité des interventions non-pharmaceutiques contre la COVID-19 en Suisse est évaluée, de même que la probabilité d'éliminer le choléra d'Haïti selon différents scénarios de campagnes de vaccination de masse. Enfin, le développement d'un outil de commande optimale qui produit une allocation spatiale efficace des vaccins contre le SRAS-CoV-2 en Italie clôt cette conversation de modèles.

La présente thèse montre comment la modélisation des maladies infectieuses permet une prise de décision éclairée en projetant les incertitudes à la lumière des données disponibles. Elle souligne également l'effort nécessaire pour adapter les modèles et les méthodes d'inférence aux spécificités du contexte de transmission et de la question de recherche considérée. De la compréhension des voies de transmission aux rapports hebdomadaires destinés aux décideurs, elle explore les différentes applications de la modélisation des maladies infectieuses. Les méthodes

MOT-CLÉS choléra, COVID-19, épidémiologie, santé publique, écohydrologie, dynamique des maladies infectieuses, SIR, modélisation mathématique, inférence statistique, commande optimale.

développées en cours de route enrichissent la boîte à outils mise à disposition de la communauté scientifique, afin de mieux orienter la prise de décisions avec pour but une réduction durable du fardeau des maladies infectieuses sur les communautés.

# CONTENTS

ACKNOWLEDGEMENTS	5
SUMMARY	7
RÉSUMÉ	9
LIST OF FIGURES	12
LIST OF TABLES	14
INTRODUCTION	17
CONTEXT	17
MODELING INFECTIOUS DISEASES	18
THESIS AIM AND OUTLINE	20
1 A MODELLING ORIENTED PRIMER ON CHOLERA	23
HISTORY AND EPIDEMIOLOGY	23
PATHOGEN	24
CHOLERA AND THE HUMAN	25
INTERVENTIONS AGAINST CHOLERA	26
ECHO's CHOLERA MODEL	27
2 RAINFALL AS A DRIVER OF EPIDEMIC CHOLERA: COMPARATIVE ASSESSMENTS OF THE EFFECT OF INTRA-SEASONAL PRECIPITATION EVENTS	31
RAINFALL AND THE TRANSMISSION OF CHOLERA	31
CASE STUDY: THE 2015 CHOLERA OUTBREAK IN JUBA, SOUTH SUDAN	33
CHOLERA MODELS	34
GENERALIZED CHOLERA MODEL	34
COMPETING TRANSMISSION MODELS	35
MODELS CALIBRATION	36
RESULTS	37
MODEL SELECTION OF RAINFALL EFFECTS ON CHOLERA TRANSMISSION	37
INTRA-SEASONAL RAINFALL EVENTS AND THE 2015 JUBA EPIDEMIC	38
DISCUSSION	40
3 ESTIMATING THE PROBABILITY OF ELIMINATING CHOLERA FROM HAITI THROUGH A MASS VACCINATION CAMPAIGN	43
DESCRIPTION OF THE MODEL	43
MODEL DYNAMICS	44
MODEL EQUATIONS	47
CALIBRATION OF THE MODEL	49

RESULTS AND DISCUSSION	51
4 A SCENARIO MODELING PIPELINE FOR COVID-19 EMERGENCY PLANNING	55
INTRODUCTION	55
METHODS	56
PIPELINE AT A GLANCE	56
MODULE 1: EPIDEMIC SEEDING AND INITIALIZATION.	57
MODULE 2: TRANSMISSION MODEL AND INTERVENTION SCENARIOS.	57
MODULE 3: CALCULATION OF HEALTH OUTCOMES	60
MODULE 4: SUMMARIZATION OF MODEL OUTPUTS	62
MODEL SPECIFICATION	63
RESULTS	63
EXAMPLE OF SCENARIO PLANNING REPORT FOR CANTON DE VAUD	63
DISCUSSION	66
5 ASSESSING THE IMPACT OF NON-PHARMACEUTICAL INTERVENTIONS ON SARS-CoV-2 TRANSMISSION IN SWITZERLAND	69
INTRODUCTION	69
METHODS	71
MODEL AND ASSUMPTIONS	71
DATA AND INFERENCE	72
RESULTS	74
DISCUSSION	77
6 OPTIMIZING THE SPATIO-TEMPORAL ALLOCATION OF COVID-19 VACCINES: ITALY AS A CASE STUDY	81
INTRODUCTION	81
MATERIALS AND METHODS	83
OBJECTIVE FUNCTION	83
CONSTRAINTS	83
COVID-19 TRANSMISSION MODEL	84
FORMULATION OF THE OPTIMAL CONTROL PROBLEM	87
RESULTS	90
COMPARISON OF VACCINATION STRATEGIES	90
ANALYSIS OF THE OPTIMAL VACCINE ALLOCATION	93
DISCUSSION	94
CONCLUSION AND PERSPECTIVES	99
APPENDICES	103
APPENDIX TO CHAPTER 5	105
CANTON DE VAUD HOSPITALIZATION DATA	105
APPENDIX TO CHAPTER 6	109
COMMENT ON THE SIMPLIFICATIONS	109
DETAILS OF THE ALTERNATIVE STRATEGIES	110
ADDITIONAL RESULTS	111
BIBLIOGRAPHY	114
CURRICULUM VITAE	133

# LIST OF FIGURES

1	Processes for infectious disease modeling	19
1.1	Map of the clusters of cholera cases in London, 1854	23
1.2	Image of the <i>Vibrio cholerae</i> bacteria	24
2.1	Similarities between daily cholera incidence and rainfall in Haiti	32
2.2	Cholera cases and precipitation during the 2015 cholera epidemic in Juba	33
2.3	Transition diagram for the competing cholera models	35
2.4	Non-linear rainfall effect on transmission	35
2.5	Likelihood ratio tests of model nesting	39
2.6	Fit of the different models	40
3.1	Incident cholera cases and rainfall in Haiti from 2010 to 2019	44
3.2	Schematic diagram of the cholera transmission processes in a single department	46
3.3	Visual assessment of model fit for the selected parameters set	50
3.4	Rate of vaccine doses deployment in each considered scenarios	51
3.5	Model results: probability of cholera elimination after mass vaccination campaigns	52
4.1	Overview of the pipeline	56
4.2	Tessellation of airports in China	57
4.3	Time series of the daily number of hospital beds needed across scenarios	59
4.4	Health outcome risks and logistical needs for fictional counties	66
4.5	County-level COVID-19 risk for three scenarios	67
5.1	COVID-19 epidemic curve in Switzerland and timing of interventions	70
5.2	Schematic diagram of COVID-19 transmission and hospitalization processes	72
5.3	Estimates of changes in the basic reproduction number $R_0$	74
5.4	Changes in mobility patterns and $R_0$	75
5.5	Timing between changes in $R_0$ and mobility	76
5.6	Modelled proportion of people infected with SARS-CoV-2 in Switzerland	77
5.7	Google trends for COVID-19 and changes in $R_0$ in Switzerland	78
6.1	Visualization of the local maximum vaccination rate constraint	83
6.2	Diagram of the COVID-19 model compartments and transitions	84
6.3	Mobility network connecting the 107 provinces of Italy	85
6.4	Model fit with data assimilation and projected scenarios for optimization	86
6.5	Comparison between different vaccine allocation strategies	90
6.6	Optimal vaccine allocation for the baseline scenario	92
6.7	Temporal vaccine allocation for the pessimistic scenario	93
6.8	Spatial vaccine distribution patterns	94
6.9	Analysis of the optimal solution	95

A1	Age distribution of patients hospitalized for COVID-19 in the canton of Vaud	105
A2	Key data of hospitalization events in the canton of Vaud	106
A3	Survival functions of death and discharge for hospitalized patients and patients in ICU	106
A4	Simplification of the mobility matrix to obtain a sparse and tractable problem	110
A5	Comparison of different allocation strategies	113

## LIST OF TABLES

1	Summary of the models described in this thesis	21
1.1	Characteristic of existing oral cholera vaccines	27
2.1	Parameters considered in the eight compared models	36
2.2	Model comparison statistics	38
3.1	Model parameters	49
4.1	Health outcome risk parameters	61
6.1	Averted infections per dose for different allocation strategies	91
A2	Observed hospitalization time distributions	107
A3	Estimated parameters of hospitalization time distributions	107
A4	Absolute number of averted infections for each scenario	112



# INTRODUCTION

## CONTEXT

Centuries after the first cholera pandemics and 160 years after the realization that safe drinking water and adequate sanitation prevent its transmission, cholera remains a threat to millions living in hotspots or at-risk areas. The recent emergence of the new coronavirus disease 2019, COVID-19, and the strain it put on the world's most advanced healthcare systems recalls the constant risks posed by emerging diseases.

Public-health policies have proven the effectiveness of interventions against infectious diseases, showing that many deaths are preventable, and in some cases elimination is possible. In the fight against infectious diseases, a series of successes – attributable to *e.g.* hygiene, vaccines, antibiotics, safe drinking water, ... – brought the hope of a global and durable reduction of the burden. Especially in privileged communities, long-term improvements have been achieved for many of the diseases that have shaped the history of humanity. While these progresses show that infectious diseases are not a necessary fate, setbacks on the control of existing and emerging pathogens remind us of the ongoing threat they pose to public health. Indeed, the current global health picture is marked by inequalities in the distribution of the burden, which disproportionately piles up on already impoverished communities, in conflict zones, or after natural disasters. Today, communicable diseases cause approx. 15% of global deaths every year<sup>1</sup>, and nearly 1/3 of all child deaths are caused by pneumonia and diarrhoea alone<sup>2</sup>.

As a consequence of the COVID-19 pandemic, the present thesis associate two antipodean diseases. Cholera, one of the most ancient recorded disease<sup>3</sup>, has caused 7 pandemics in the modern era. In contrast, COVID-19 earliest known onset of symptoms is on December 1, 2019. The pathogen for cholera is a bacteria, *Vibrio Cholerae*, responsible for heavy watery diarrhea, whereas SARS-CoV-2, COVID-19's pathogen, is a virus responsible for respiratory infections. Cholera belongs to the neglected tropical diseases, a class of understudied infections while the COVID-19 pandemic has sparked an unprecedented accumulation of scientific evidence<sup>4</sup>. Other differences between the two diseases include the posited transmission routes (fecal-oral vs. respiratory) and the affected communities ("poorest of the poor", higher severity on children vs. global, higher severity on elders). Despite these differences, cholera and COVID-19 shares a burden that echoes unequal access to care, disproportionately affecting stigmatized communities, and mechanisms that make their transmission sustainable in populations and cause pandemics with a regrettably high toll on human lives. Both cholera and COVID-19 are infectious diseases – with the potential of starting epidemics and pandemics.

Epidemics – the rapid spread of an infectious disease in a population – are

<sup>1</sup> Roth et al., "Global, Regional, and National Age-Sex-Specific Mortality for 282 Causes of Death in 195 Countries and Territories, 1980–2017: A Systematic Analysis for the Global Burden of Disease Study 2017" in: *The Lancet* (2018), tab. 1, excl. non-transmissible neonatal and maternal diseases and nutritional diseases, pre-COVID-19 estimates.

<sup>2</sup> WHO, *Ending Preventable Child Deaths from Pneumonia and Diarrhoea by 2035 - The Integrated Global Action Plan for Pneumonia and Diarrhoea (GAPPD)* (2013), *i.e.* 2M deaths among under 5, every year.

<sup>3</sup> History of pre-pandemic cholera is uncertain but numerous accounts of the disease are supposed from as early as 400 BCE. See Byrne, *Encyclopedia of Pestilence, Pandemics, and Plagues: A-M* (2008), p. 95.

<sup>4</sup> With more than 190'000 peer-reviewed papers and many more preprints, websites, and reports published at the time of writing. Estimation from: COVID-19 Open Access Project, *Living Evidence on COVID-19* (2020).

complex phenomena resulting from the interactions between pathogens, environment, societies, and individuals<sup>5</sup>. The mitigation of the spread of infectious disease epidemics presents challenges across every dimension of environmental and human health; to prevent spillover events, to block transmission routes, and to protect or treat every person appropriately. One of the challenges of designing effective public-health policies is dealing with the uncertainties that plague every facet of disease transmission. Only biased and sometimes scarce information is available to reason on complex systems with multi-factorial interactions.

Models – conceptual representations of systems – are tools for us to reason about the world. Historically, conceptual models of the propagation of diseases, from divine retribution to miasma theory, have motivated more (quarantine) or less (persecution) effective approaches to the control of these pests. Scientific breakthroughs in biology and medicine, with the identification of pathogens and their transmission routes, provided a new look at contamination and opened the path for improved preventive interventions and treatments. Novel statistical modeling approaches<sup>6</sup> developed in the 20th century – and continuously improved ever since<sup>7</sup> – provide a formal framework to reason about the propagation of a disease in a population. Models enable one to deal with biases on data collection, to account for epistemic uncertainties, to encompass uncertainties in the transmission dynamics. It becomes possible to simulate the disease spread in the affected populations and to study the impact of intervention policies in a principled way. The toolbox was further reinforced by advances in mechanistic modeling applied to disease transmission, starting from SIR compartmental models<sup>8</sup>, which divide the population depending on their status with respect to the disease. And recent advances in computing power proved a paradigm shift in dealing with the available evidence to accurately simulate complex systems.

## MODELING INFECTIOUS DISEASES

The present thesis explores compartmental models as an approach to infectious disease transmission, and as tools to guide intervention strategies. Each of the presented research works strives to answer a series of research questions, all being variations of the same theme: how do infectious diseases spread? how can we prevent them from spreading? These questions are answered using extensively computer-age modeling and inference methods, in an interactive process represented in fig. 1. Given a question and information in the form of observations (data) and knowledge, statistical inference is performed in four steps:

**MODEL DESIGN** One or more model(s) of disease transmission are designed. Choices on what to include and how to express the supposed dynamics depend on the underlying knowledge of the processes and the available data. An epidemiological model considers a subset of the known transmission processes that are relevant to answer the research or policy questions<sup>9</sup>. A set of 5 models, see tab. 1, is proposed in this thesis, with features that depend on the disease, the context, the possible control measures, and the uncertainties associated with the observations and the processes by themselves. Some models have stochastic transitions while others are mostly deterministic, some models consider human mobility as fluxes between regions whereas others assume a well-mixed population. Despite the foregoing differences, all the models presented in this thesis are compartmental

<sup>5</sup> Rinaldo, Gatto, et al., *River Networks as Ecological Corridors: Species, Populations, Pathogens* (2020); Buckee et al., “Thinking Clearly about Social Aspects of Infectious Disease Transmission” in: *Nature* (2021); Hesterbeek et al., “Modeling Infectious Disease Dynamics in the Complex Landscape of Global Health” in: *Science* (2015).

<sup>6</sup> Freedman, “From Association to Causation: Some Remarks on the History of Statistics” in: *Statistical Science* (1999).

<sup>7</sup> Gelman et al., *What Are the Most Important Statistical Ideas of the Past 50 Years?* (2021).

<sup>8</sup> Kermack et al., “A Contribution to the Mathematical Theory of Epidemics” in: *Proceedings of the Royal Society A* (1927); Anderson et al., “Population Biology of Infectious Diseases: Part I” in: *Nature* (1979).

<sup>9</sup> “Since all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers aboard.” from Box, “Science and Statistics” in: *Journal of the American Statistical Association* (1976).

with some mechanistic components<sup>10</sup>; all are based on the SIR model. It means that individuals are characterized by their status with respect to the disease. In a population, some individuals start susceptible  $S$  to the disease and might become infected and infectious  $I$ . After some time they turn recovered  $R$ : they are immune to infection and do not contribute to transmission anymore. Additional compartments may be defined depending on the objective of the exercise. In tab. 1, the principal compartments of the considered models are indicated: exposed  $E$  (incubating, infectious or not), asymptomatic  $A$  (infectious with no symptoms),  $H$  indicate compartments that represent the healthcare facilities (hospitalization, ICUs),  $V$  the compartments for vaccinated individuals, and finally  $B$  the modeling of an environmental bacteria reservoir for cholera models<sup>11</sup>. In these models, some states are observed through a reporting process (*e.g.* infected  $I$  may be reported as incident cases), but most are unobserved. Likewise, some parameters are fixed to values (or distribution of values) that are known more or less precisely, while many others are left unknown.

<sup>10</sup> The opposite would be empirical models, which aim at reproducing observations rather than the relationship between the parts of the system modeled.

<sup>11</sup> In this table, the exponents denote the number of sub-compartments of the same type used to model non-exponential distributions of the residence times in that compartment, using the linear-chain trick.

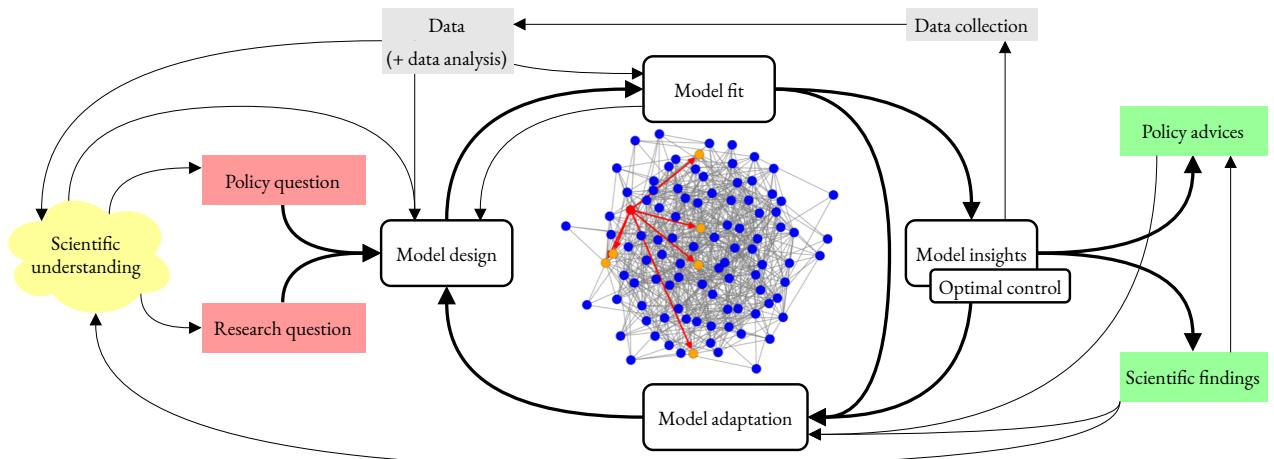


fig. 1: Processes for infectious disease modeling. The numerous feedback loops make this procedure iterative. All boxes save for data collection were explored in this thesis. The central figure represents an agent-based model of disease transmission in a random graph (by Thomas Fry, a master student supervised during this thesis, with permission).

**MODEL FIT** *Fitting* conditions the model on the observed data. It allows one to infer unobserved states and parameters from reported quantities, such as the reduction of transmission during lockdown from hospitalization and deaths. A panel of methods for fitting a model to data has been developed. Usually, computer-age statistical inference relies on thousands of model simulations to explore the high dimensional parameter and state spaces; each realization being evaluated against the data. The inner working of the fitting algorithms varies depending on the formal framework it is built on, but also on numerous computational methods it uses to make inference tractable for complex problems. In this thesis, two classes of methods are used: Bayesian methods based on Markov chain Monte Carlo (MCMC) derivatives and frequentist Iterated Filtering (IF) methods. Fitting the model allows one to uncover previously unknown transmission parameters and unobservable states. As such, it requires a high degree of care as inferred features may be caused by modeling artifacts instead of real epidemic features. Moreover, results may misrepresent the uncertainties associated with the process and these methods fail more or less gracefully, thus it is possible to obtain results that seem reliable but are invariably wrong.

**MODEL ADAPTATION** An evaluation of the *fit* of the model and the implications on the results is done by performing visuals and formal checks. More often than not, it is necessary to adapt either the fitting procedure or the model structure. It could be the inclusion or the removal of a process from the model. Or perhaps choosing to account for a particular dynamic in a more or less explicit way. Models are lossy compressions of reality (from an information theory point of view), and choices on the processes to be included, and how to include them are the cornerstone of epidemiological modeling. Indeed despite its recent formalism, statistical inference remains an art, uncomfortably dependent on the practitioners, their backgrounds, and their expectation<sup>12</sup>. The cycle *design*→*fit*→*adapt* is then repeated until the predictive accuracy is deemed satisfying for the goal of the exercise.

**MODEL INSIGHTS** Armed with a reliable model whose parameters and states are consistent with data and knowledge about an epidemic, it is now possible to answer the research questions. A model that reproduces observed dynamics may accurately represent epidemiological processes. It can be used as a substitute for experiments: it enables to simulate different intervention scenarios and to evaluate the impact of past or planned policies on a virtual system. This sets the range of expectations one can have with regard to the real-world consequences of modeled policies. In some cases, the model is designed to replicate the mechanistic relationships between the epidemiological processes, and several different models might be compared to identify which transmission routes are responsible for the observed dynamics. Finally, the communication of model results ought to come with proper awareness of the limitations, as models are always incomplete and biased in reproducing the real epidemiological dynamics. It provides an additional opportunity for feedback towards an eventual adaptation of the model<sup>13</sup>.

**OPTIMAL CONTROL** Another application of epidemiological models explored in this thesis is their integration with optimal control methods to search for the *best* intervention against an infectious disease. Optimal control methods are recent mathematical optimization procedures extending the calculus of variations to enable the derivation of control policies for dynamical systems. While technical adaptations are necessary to perform optimal control on epidemiological models, this rigorous framework identifies the most effective control measures under a set of operational constraints, discovering non-intuitive features and policies. These optimal interventions exploit every feature of the complex interactions modeled, allowing to *e.g.* best allocate a limited vaccine supply in space and time. Such tools, currently in their infancy and requiring accurate models, are promising as reasoning aid, uncovering another facet of disease transmission, and for the effective allocation of control resources in the fight against epidemics.

## THESIS AIM AND OUTLINE

The present thesis has been developed within the Swiss National Foundation project “Optimal control of intervention strategies for waterborne disease epidemics (SNF 200021-172578)”. Its initial goal was to develop a decision support system for the real-time design of optimal intervention strategies against cholera, which includes an operational forecasting tool coupled with an optimal control

<sup>12</sup> Multi-modeling studies – where several modelers are tasked to answer the same question – uncover the importance of the choices while performing statistical inference and mitigate the issues linked to opinionated model design.

<sup>13</sup> Heesterbeek et al., “Modeling Infectious Disease Dynamics in the Complex Landscape of Global Health” in: *Science* (2015).

CHAPTER	Disease	Compartments	Processes	$N_{\text{spatial}}$	Fit	Aim	Reference
2	Cholera	SEIR+B	Stochastic	–	IF-like	explain	Lemaitre, Pasetto, Perez-Saez, et al., “Rainfall as a Driver of Epidemic Cholera: Comparative Model Assessments of the Effect of Intra-Seasonal Precipitation Events” in: <i>Acta Tropica</i> (2019)
2	Cholera	SEIR+B	Deterministic	–	MCMC-like	explain	Lemaitre, Pasetto, Perez-Saez, et al., “Rainfall as a Driver of Epidemic Cholera: Comparative Model Assessments of the Effect of Intra-Seasonal Precipitation Events” in: <i>Acta Tropica</i> (2019)
3	Cholera	SEAIR <sup>3</sup> +V	Stochastic	10	IF-like	project (scenarios)	Lee et al., “Achieving Coordinated National Immunity and Cholera Elimination in Haiti through Vaccination: A Modelling Study” in: <i>The Lancet Global Health</i> (2020)
4	COVID-19	SEI <sup>3</sup> R+VH	Stochastic	3'000+	MCMC-like	project (scenarios)	Lemaitre, Gratzl, et al., “A Scenario Modelling Pipeline for COVID-19 Emergency Planning” in: <i>Scientific Reports</i> (2021)
5	COVID-19	SEI <sup>3</sup> R+H	Stochastic	–	IF-like	infer	Lemaitre, Perez-Saez, et al., “Assessing the Impact of Non-Pharmaceutical Interventions on SARS-CoV-2 Transmission in Switzerland” in: <i>Swiss Medical Weekly</i> (2020)
6	COVID-19	SEIAR+VH	Deterministic	107	MCMC-like	optimal control	Lemaitre, Pasetto, Zanon, et al., “Optimizing the Spatio-Temporal Allocation of COVID-19 Vaccines: Italy as a Case Study” in: <i>medRxiv</i> (2021)

solver. This framework has been developed, albeit for COVID-19 transmission in Italy, and is presented in CHAPTER 6.

The work presented extensively builds on the ECHO laboratory expertise on the spatially-explicit modeling of cholera transmission<sup>14</sup>. The group has developed over a decade a rainfall-mediated, spatially explicit cholera model that has inspired each of the other models presented in this thesis. The original model is presented in CHAPTER 1 with a short introduction on the ancient disease that is cholera and its sorrowful history.

Cholera is the focus of two additional chapters. The explanatory power of two existing models linking cholera transmission with rainfall is compared in CHAPTER 2, through the analysis of an outbreak in Juba, South Sudan. In CHAPTER 3, a scenario planning estimation of the probability of eliminating cholera from Haiti through a mass vaccination campaign is presented; this work has been carried within the framework of a multi-modeling study with other groups, and input from the ministry of health of Haiti.

The remaining chapters focus on COVID-19. In CHAPTER 4, some facets of dealing with uncertainties from an emerging disease pandemic are uncovered, with a focus on a pipeline for scenario modeling that has been used to inform governments, in the United States and other countries. Using data collected while participating in the COVID-19 response, it was possible to estimate the impact of early interventions against COVID-19 in Switzerland, this study is presented in CHAPTER 5. And finally, in CHAPTER 6 a spatially explicit epidemiological model for the Italian epidemic is integrated within an optimal control framework in order to discover the optimal strategy for vaccine allocation.

tab. 1: Summary of the compartmental models described in this thesis.

<sup>14</sup> and waterborne diseases in general; see: Rinaldo, Gatto, et al., *River Networks as Ecological Corridors: Species, Populations, Pathogens* (2020); Rinaldo, Bertuzzo, Mari, et al., “Reassessment of the 2010–2011 Haiti Cholera Outbreak and Rainfall-Driven Multiseason Projections” in: *Proceedings of the National Academy of Sciences* (2012)



# CHAPTER I

## A MODELLING ORIENTED PRIMER ON CHOLERA

Cholera is an acute intestinal infection causing severe diarrhea that may lead to dehydration, and sometimes death. The global burden of cholera is difficult to estimate as the majority of cases are not reported. It is estimated that 3 million cases and 95'000 deaths occur every year, with millions more at risk in endemic areas (around 50 countries)<sup>1</sup>. Despite its household name, cholera belongs to the neglected tropical diseases group: our understanding of many important aspects of cholera clinical course and transmission is still limited. A political will to eliminate this ancient disease has recently arisen. The World Health Organization (WHO) initiated the Global Task Force for Cholera Control (GTFCC), which provides a concrete path towards the elimination of cholera by 2030<sup>2</sup>. The consensus is that to reach this goal in endemic countries, there must be substantial long-term improvements in safe water distribution systems, adequate sanitation, and accessible hygiene education. Moreover, in the event of a cholera outbreak, timely interventions such as vaccination campaigns are crucial to limit the spread of the disease, and access to proper treatments reduces the toll of the disease on communities.

### HISTORY AND EPIDEMIOLOGY

Humanity and cholera share a long history, with supposed mentions as early as the 5th century BCE. The disease became more widely known in the modern era. From 1817 to 1923, six successive pandemics occurred, all originated in the delta of the Ganges river but taking different paths across the world. Cholera spread around the world owing to the nascent mobility, leaving 10s of millions dead across countries and continents. Scientific developments sped up during the third cholera pandemic (1846-60). In 1854, a cholera outbreak in Broad Street (Soho, London) was studied by physician John Snow who led an impressive early work in spatial epidemiology applied to public health. Analyzing the contamination pattern among residents (fig. 1.1), Snow postulated that cholera spreads through water contaminated by an infectious agent, instead of foul air<sup>3</sup>. Simultaneously, Italian microbiologist Filippo Pacini isolated the bacterium in Florence<sup>4</sup>. Thirty years after Pacini, German scientist Robert Koch independently rediscovered the cholera pathogen after investigations in Egypt and India. He first posited *Vibrios* causative relationship with the disease. Finally, a hundred years later, Indian researcher Sambhu Nath De discovered the cholera toxin in 1959<sup>5</sup>.

Throughout a seventh pandemic from 1961 onward, cholera spread in several

<sup>1</sup> Ali et al., “Updated Global Burden of Cholera in Endemic Countries” in: *PLoS Neglected Tropical Diseases* (2015).

<sup>2</sup> Elimination being defined as a 90% reduction of cholera deaths per year, see [gtfcc.org](http://gtfcc.org).

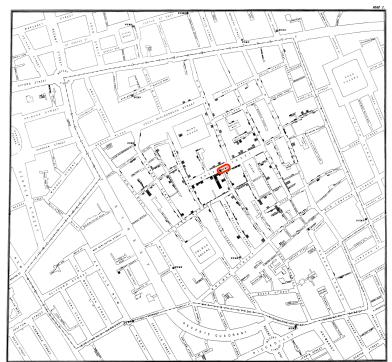


fig. 1.1: Original map by John Snow. Stacked rectangles represent cholera cases of the 1854 Broad Street outbreak (London). The work of John Snow convinced the authorities to close the contaminated water pump (circled in red), leading to a decrease in mortality. Lithography by Charles Cheffins, in Snow, *On the Mode of Communication of Cholera* (1855), p. 54.

<sup>3</sup> At the time, the accepted mode of contamination for cholera and many other diseases was through miasma, *bad air* contaminated by organic matter. Chasing odor justified urban planning along streets and river banks in Paris and London. But also in Lausanne, where rivers Flon and Louve were covered in 1832 in response to a cholera outbreak. Cholera is the name of a savory dish from Valais, a testimony of the strong impression cholera left on the Swiss.

<sup>4</sup> Pacini, *Osservazioni microscopiche e deduzioni patologiche sul Cholera asiatico* (1854).

<sup>5</sup> De et al., “An Experimental Study of the Action of Cholera Toxin” in: *The Journal of Pathology and Bacteriology* (1951).

waves, through Asia in the 1960s, reaching Africa and the Middle East in the 1970s and the Americas in 1991<sup>6</sup>. Improvement in sanitation and hygiene spared higher-income countries from the disease, and cholera became a burden of the poorest communities of the Global South. Series of outbreaks (*e.g.* Zimbabwe 2008, Haïti 2010, Yemen 2016), continuous transmission in endemic countries, and millions at risk keep this ancient disease an ongoing public health issue.

The span of this thesis was marked by a cholera outbreak in Yemen (2016–2021) – a humanitarian crisis with 2.5M suspected cases and nearly 4'000 deaths – an outbreak in Zimbabwe, a flare-up in Algeria, in addition to many seasonal outbreaks and sustained endemic cholera transmission in Asia and sub-Saharan Africa. But also the last confirmed cholera case of Haiti, in early 2019, bringing an end to a 9 years epidemic and the hope of cholera elimination from its last foothold in the Americas.

While the number of cholera cases doubled from 2018 to 2019 to nearly a million, reported cholera deaths decreased to less than 2'000 in 2019, with Africa reporting its lowest numbers since the 2000s. Efforts towards the improvement of sanitary conditions and reactive vaccinations campaigns (2.4M doses of cholera vaccines were distributed in 2019) aim at further reducing the toll of cholera<sup>7</sup>.

The following sections highlight some relevant aspects of cholera transmission, while further details about the biological and medical features of cholera are outside the scope of this thesis.

## PATHOGEN

**PATHOGEN** Cholera is an infection caused by a waterborne bacteria: the *Vibrio cholerae* (fig. 1.2). While many serogroups of *V. cholerae* can secrete the cholera toxin responsible for massive watery diarrhea, only serogroups O139 and O1 are responsible for disease epidemics. O1 is causing most of the recent epidemics and is divided into two biotypes: Classical O1 and El Tor, which are both divided into three serotypes: Ogawa, Inaba, and the rare Hikojima<sup>8</sup>. Cholera classification has its importance as it affects many epidemiological characteristics *e.g.*, El Tor survives longer in water and has a higher asymptomatic/symptomatic ratio<sup>9</sup>. The current cholera pandemic is mainly caused by El Tor, while the Classical derivative caused the previous fifth and sixth pandemics, and its transmission is now limited to the Ganges delta<sup>10</sup>.

**ENVIRONMENTAL RESERVOIR** *V. cholerae* has for natural habitat some aquatic ecosystems, in particular brackish waters and estuaries. There is no marine host, but complex ecological association processes take place in the aquatic medium, and natural genetic transformation is enabled by chitin, the polymer constituting the crustacean exoskeleton<sup>11</sup>. There is no consensus on how long *V. cholerae* remains infectious in water and under which conditions it can reproduce<sup>12</sup>. In cholera epidemics, it is difficult to isolate the importance of the natural bacteria reservoir compared to the freshly introduced *Vibrios* excreted by infected persons, though the latter is suspected to drive transmission during outbreaks.

**CLIMATIC DRIVERS** Cholera epidemics express marked seasonal patterns that might have different shapes in different countries. A complex and unclear association between precipitation and cholera infections has been put forward in many

<sup>6</sup> Mutreja et al., “Evidence for Several Waves of Global Transmission in the Seventh Cholera Pandemic” in: *Nature* (2011).

<sup>7</sup> see WHO, “Cholera, 2019” in: *Weekly Epidemiological Record* (2020) and previous *Weekly Epidemiological Records* about cholera.

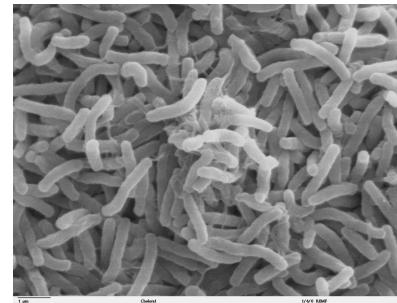


fig. 1.2: Scanning electron microscope image of *Vibrio cholerae*, a gram-negative rod-shaped bacteria (Public domain image by Ronald Taylor, Tom Kirn, Louisa Howard).

<sup>8</sup> Kaper et al., “Cholera” in: *Clinical Microbiology Reviews* (1995).

<sup>9</sup> WHO, “Cholera Vaccines: WHO Position Paper” (2017).

<sup>10</sup> Nair et al., “Cholera Due to Altered El Tor Strains of *Vibrio Cholerae* O1 in Bangladesh” in: *Journal of Clinical Microbiology* (2006); Domman et al., “Defining Endemic Cholera at Three Levels of Spatiotemporal Resolution within Bangladesh” in: *Nature Genetics* (2018).

<sup>11</sup> Reidl et al., “*Vibrio Cholerae* and Cholera: Out of the Water and into the Host” in: *FEMS Microbiology Reviews* (2002); Meibom et al., “Chitin Induces Natural Competence in *Vibrio Cholerae*” in: *Science* (2005).

<sup>12</sup> Mavian et al., “Toxigenic *Vibrio Cholerae* Evolution and Establishment of Reservoirs in Aquatic Ecosystems” in: *Proceedings of the National Academy of Sciences* (2020).

research works. After Hurricane Matthew and heavy rainfall in October 2016, cholera flared up after a low transmission episode in Haiti<sup>13</sup>. Similarly, cholera in many African countries follows a seasonal trend, with the higher transmission during the rainy season<sup>14</sup>. Rainfall might play a major role in water contamination, for instance through the washout of open-air defecation and raw sewage circulation in the environment. A literature review about studies investigating the relationship between cholera and rainfall is provided in CHAPTER 2. Temperature, water acidity, sunlight, and other environmental factors have also been shown to affect the survival and reproduction of *V. cholerae* in water bodies. Hence, macroclimate phenomena such as the El Niño Southern Oscillation have been associated with changes in transmission, even if no causal link could be established<sup>15,16</sup>.

## CHOLERA AND THE HUMAN

**DISEASE** A human host becomes infected through the ingestion of a critical dose of *V. cholerae*<sup>17,18</sup>. The susceptibility depends on many factors such as gastric acidity and age, with children under 5 much more likely to become infected<sup>19</sup>. *V. cholerae* colonizes the small intestine for an incubation period lasting 12 hours to 5 days<sup>20</sup> before symptoms. Then, a wide range of outcomes are possible. Most of the time the infection is asymptomatic. On the other end of the spectrum severe infection (or *cholera gravis*), characterized by vomiting and profuse rice water diarrhea, occurs in 1% to 15% of the cases. Many stages of mild infections lie in between these extremes. Unless tested in a laboratory, symptoms are indistinguishable from those of numerous other infections causing diarrhea<sup>17,18,21</sup>. The severity of the symptoms correlates with the quantity of *V. cholerae* ingested<sup>22</sup>, and depends on the cholera strain and personal characteristics, age, immunity, pregnancy, blood type<sup>23</sup>, ...Treatment is crucial: severely infected individuals might lose up to 20 liters of water through diarrhea in a day and may die within hours. If left untreated, severe cases mortality reaches up to 60% but proper therapy lowers it below 1%<sup>24</sup>.

**SHEDDING AND TRANSMISSION** The number of bacteria shed by an infectious individual varies with the intensity of the infection. It is estimated to range from  $10^3$  to  $10^9$  *Vibrios* per gram of stool for asymptomatic infected and severely infected individual respectively<sup>18</sup>. Similarly, the duration of the shedding period typically ranges from a day up to two weeks<sup>17,18</sup>. Transmission occurs along the fecal-oral route (consumption of contaminated water or food, contact with fomites), but contamination is also possible from the aquatic reservoirs, through ingestion of water or seafood. The principle mechanism of transmission is the intake of water contaminated by the untreated feces of other infectious individuals. In fact, freshly shed *Vibrios* may be in a hyper-infectious state, which could be of great importance in driving epidemic transmission<sup>25</sup>. Asymptomatic individuals remain mobile and are thought to transmit cholera, thus may be important vectors for cholera dissemination.

**HUMAN MOBILITY AND HYDROLOGICAL TRANSPORT** Spatial spread of cholera outbreaks may occur through two networks. *V. cholerae* may be transported through the river network. An example is the start of the 2010 Haiti epidemic along the Artibonite river<sup>26</sup>. Human mobility also plays a major role in the

<sup>13</sup> Pasetto, Finger, Camacho, et al., “Near Real-Time Forecasting for Cholera Decision Making in Haiti after Hurricane Matthew” in: *PLOS Computational Biology* (2018).

<sup>14</sup> Baracchini et al., “Seasonality in Cholera Dynamics: A Rainfall-Driven Model Explains the Wide Range of Patterns in Endemic Areas” in: *Advances in Water Resources* (2017).

<sup>15</sup> Pascual et al., “Cholera Dynamics and El Niño-Southern Oscillation” in: *Science* (2000).

<sup>16</sup> Marc Lipsitch and Cécile Viboud beautifully describe the difficulty evaluating the environmental factors in disease transmission, writing for influenza: “This potpourri of possible mechanisms places us in a kind of Popperian purgatory, in which data in support of every hypothesis exist, yet none of the hypotheses has been subjected to tests that are rigorous enough to reject it.” Quote from Lipsitch and Viboud, “Influenza Seasonality: Lifting the Fog” in: *Proceedings of the National Academy of Sciences* (2009)

<sup>17</sup> Kaper et al., “Cholera” in: *Clinical Microbiology Reviews* (1995).

<sup>18</sup> Nelson et al., “Cholera Transmission: The Host, Pathogen and Bacteriophage Dynamic” in: *Nature reviews. Microbiology* (2009).

<sup>19</sup> Sack, Sack, Nair, et al., “Cholera” in: *The Lancet* (2004).

<sup>20</sup> Azman, Rudolph, et al., “The Incubation Period of Cholera: A Systematic Review” in: *Journal of Infection* (2013).

<sup>21</sup> King, Ionides, et al., “Inapparent Infections and Cholera Dynamics” in: *Nature* (2008); van de Linde et al., “Observations on the Spread of Cholera in Hong Kong, 1961-63” in: *Bulletin of the World Health Organization* (1965); McCormack et al., “A Community Study of Inapparent Cholera Infections” in: *American Journal of Epidemiology* (1969)

<sup>22</sup> Brouwer et al., “Dose-Response Relationships for Environmentally Mediated Infectious Disease Transmission Models” in: *PLOS Computational Biology* (2017).

<sup>23</sup> WHO, “Cholera Vaccines: WHO Position Paper” (2017); Azman, Rudolph, et al., “The Incubation Period of Cholera: A Systematic Review” in: *Journal of Infection* (2013).

<sup>24</sup> Luquero, Rondy, et al., “Mortality Rates during Cholera Epidemic, Haiti, 2010–2011” in: *Emerging Infectious Diseases* (2016).

<sup>25</sup> Butler et al., “Cholera Stool Bacteria Repress Chemotaxis to Increase Infectivity” in: *Molecular Microbiology* (2006).

<sup>26</sup> Piarroux et al., “Understanding the Cholera Epidemic, Haiti” in: *Emerging Infectious Diseases* (2011).

spreading of the infections possibly due to the large proportion of asymptomatic that transport and transmit cholera across regions. Indeed, also in Haiti cholera was brought into the state by infected United Nations peacekeepers<sup>27</sup>.

**IMMUNITY** Infected individuals that recover from the infection are immunized against *V. cholerae* of the same serogroup. The duration of acquired immunity is difficult to estimate and depends on many factors. Acquired immunity has been reported to range from few months to several years, with most recent estimates ranging from 2 to 10 years, possibly depending on the virulence of the infection<sup>28</sup>.

### INTERVENTIONS AGAINST CHOLERA

Interventions against cholera may be preventive or concern the treatment of infected individuals. Case management and treatment plays an important role in reducing the transmission and the toll of the epidemic and mainly consists of:

**Oral (or Intravenous) Rehydration Therapy** The main treatment for cholera consists of replacing fluids as fast as they are lost. Despite its simplicity, it is very effective in reducing mortality. Fluids with the same electrolyte composition must be administered<sup>29</sup>. Re-hydration is usually done in treatment centers but may take place at the patient home. This differentiation might determine if stools contribute to the infection cycle or are properly disposed of.

**Antibiotics** reduce the severity and the duration of the infection. WHO recommends their use only for the most severe cases as antibiotic resistance of *V. cholerae* is raising worldwide<sup>30</sup>.

Prevention measures may be carried out before and during the outbreak, and are described below.

**SURVEILLANCE AND REPORTING** During an outbreak, surveillance consists of the timely detection and reporting of new cases. In many countries where outbreaks occur annually during the rainy season, the observation of past epidemics provides insight on the severity and timing of the new infections, that can be used for preparation<sup>31</sup>. While possible, environmental monitoring for *V. Cholerae* in the water has never succeeded to warn about an upcoming epidemic.

Since without laboratory equipment it is impossible to distinguish cholera from another pathogen in a patient with acute watery diarrhea, there has been an effort to standardize the clinical definition of a suspected cholera case. A suspected case combines acute watery diarrhea and severe dehydration, the latter condition being dropped in case of an outbreak. This diagnosis may be precised using rapid diagnosis tests (RTDs), with a pretty high sensibility but low sensitivity. Precise identification is obtained through culture, the current gold standard<sup>32</sup>. RTDs and culture are not always available, especially during an outbreak, and over-reporting due to the misclassification of other diarrheal diseases is possible. Conversely, under-reporting might also occur as transmission settings might be isolated or plagued with conflicts or natural disasters. WHO guidelines recommend that, when a patient enters a treatment center, his name, address, sex, age (over or below 5), and symptoms are recorded<sup>33</sup>.

**WASH** Water, Sanitation and Hygiene (WaSH) is a broad term that encompasses many intervention strategies which are key to the long-term control of

<sup>27</sup> Piarroux et al., "Understanding the Cholera Epidemic, Haiti" in: *Emerging Infectious Diseases* (2011).

<sup>28</sup> Levine et al., "Duration of Infection-Derived Immunity to Cholera" in: *The Journal of Infectious Diseases* (1981); Kaper et al., "Cholera" in: *Clinical Microbiology Reviews* (1995); Woodward, "Cholera Reinfection in Man" in: *The Journal of Infectious Diseases* (1971); Glass et al., "Seroepidemiological Studies of El Tor Cholera in Bangladesh: Association of Serum Antibody Levels with Protection" in: *The Journal of Infectious Diseases* (1985); Clemens et al., "Biotype as Determinant of Natural Immunising Effect of Cholera" in: *The Lancet* (1991); Leung et al., "Protection Afforded by Previous Vibrio Cholerae Infection against Subsequent Disease and Infection: A Review" in: *PLOS Neglected Tropical Diseases* (2021).

<sup>29</sup> Kühn et al., "Glucose- but Not Rice-Based Oral Rehydration Therapy Enhances the Production of Virulence Determinants in the Human Pathogen Vibrio Cholerae" in: *PLOS Neglected Tropical Diseases* (2014).

<sup>30</sup> Sack, Sack, and Chaignat, "Getting Serious about Cholera" in: *New England Journal of Medicine* (2006).

<sup>31</sup> Baracchini et al., "Seasonality in Cholera Dynamics: A Rainfall-Driven Model Explains the Wide Range of Patterns in Endemic Areas" in: *Advances in Water Resources* (2017).

<sup>32</sup> Camacho et al., "Cholera Epidemic in Yemen, 2016–18: An Analysis of Surveillance Data" in: *The Lancet Global Health* (2018) and CDC, *Diagnosis and Detection | Cholera | CDC* (2018)

<sup>33</sup> WHO, "First Steps for Managing an Outbreak of Acute Diarrhoea" (2010).

cholera. Improved sanitary conditions brought cholera elimination in first-world countries. WaSH is divided into short and long-term measures. Short term strategies involve sterilization, decontamination, hand washing, education sessions, and water purification and filtering (chlorination, ...)<sup>34</sup>. Long-term sanitation strategies involve the construction of infrastructures for fecal sludge management, sewage systems, toilets, and access to safe water sources. From a modeling point of view, WaSH reduces exposure (water purification, sari filtration<sup>35</sup>) and shedding (sewage and fecal sludge management). By its nature and the possible long-term effects, WaSH improvement is difficult to quantify in modeling frameworks.

**VACCINATION** is a safe and effective way to protect individuals from cholera and to reduce the propagation of the epidemic. It can be used in a preventive or reactive way. Several vaccines exist for cholera, with different characteristics. As of today, two main oral cholera vaccines (OCVs) are used in vaccination campaigns around the world: WC-rBS and BiWC<sup>36</sup>. The main characteristics of these two vaccines are shown in tab. 1.1.

Vaccines can either be administered in a targeted fashion or to the whole population in mass vaccination campaigns, either preventively or reactively during outbreaks<sup>37</sup>. Despite the efforts to build a worldwide vaccine stockpile, the demand for cholera vaccine vastly exceeds the supply<sup>38</sup>.

Generic Name	BiWC	WC-rBS
Commercial name	mORCVAX, Shanchol, Euvi-chol, Cholvax	Dukoral
Target strain O1	yes (classical, El Tor, Ogawa, Inaba)	yes (classical, El Tor, Ogawa, Inaba), also target a cholera toxin
Target strain O139	yes	no
Doses	2 doses, 2 weeks apart	2 doses (3 for children) 1–6 weeks apart
Field effectiveness	between 37% and 87% for two years	78% protection 1–6 months after vaccination
Age	> 1 year	> 2 year
Usage	Mass vaccination, Global OCV stockpile, 25M doses administered	Mainly for travelers (> 1M doses administered)
Protection length	3 years (1 dose: short term protection)	2 years
Constraints	–	needs buffer solution
Price per dose	1.85\$	5.25\$
Usage	Since 1998 in most recent outbreaks	Between 1997 and 2009 in Uganda, Tanzania, Indonesia, ...

## ECHO's CHOLERA MODEL

Cholera has been the subject of modeling studies since the first attempt by Capasso and Paveri-Fontana<sup>39,40</sup>, and cholera modeling has received renewed attention during the 2010 outbreak in Haiti. Most of recent cholera modeling literature focuses on phenomenological (or statistical) models with different degrees of mechanistic processes<sup>41</sup>, but numerous mechanistic cholera models are proposed, differing in the way they account for the epidemiological processes<sup>42</sup>.

<sup>34</sup> Rebaudet, Bulit, Gaudart, Michel, Gazin, Evers, Beaulieu, Abedi, Osei, Barrais, Pierre, Moore, Boncy, Adrien, Beigbeder, et al., “The National Alert-Response Strategy against Cholera in Haiti: A Four-Year Assessment of Its Implementation” in: *bioRxiv* (2018); Fewtrell et al., “Water, Sanitation, and Hygiene Interventions to Reduce Diarrhoea in Less Developed Countries: A Systematic Review and Meta-Analysis” in: *The Lancet Infectious Diseases* (2005).

<sup>35</sup> Colwell et al., “Reduction of Cholera in Bangladeshi Villages by Simple Filtration” in: *Proceedings of the National Academy of Sciences* (2003).

<sup>36</sup> Another vaccine, Vaxchora, was recently approved by the FDA, mostly for travelers.

<sup>37</sup> See this interesting thread on the surprisingly recent history of reactive vaccination.

<sup>38</sup> Parker et al., “Adapting to the Global Shortage of Cholera Vaccines: Targeted Single Dose Cholera Vaccine in Response to an Outbreak in South Sudan” in: *The Lancet Infectious Diseases* (2017); Seidlein et al., “Preventing Cholera Outbreaks through Early Targeted Interventions” in: *PLOS Medicine* (2018).

tab. 1.1: Characteristic of existing oral cholera vaccines. Field effectiveness and vaccine efficacy are difficult to evaluate and hence omitted from this table. See WHO, “Cholera Vaccines: WHO Position Paper” (2017), Bi, Ferreras, et al., “Protection against Cholera from Killed Whole-Cell Oral Cholera Vaccines: A Systematic Review and Meta-Analysis” in: *The Lancet Infectious Diseases* (2017)(and WHO 2017a; Azman, Rumunu, et al. 2016; Luquero, Grout, Ciglenecki, Sakoba, Traore, Heile, Dialo, et al. 2013; WHO 2009; Luquero, Grout, Ciglenecki, Sakoba, Traore, Heile, Diallo, et al. 2014; Qadri et al. 2018; Azman, Luquero, Ciglenecki, et al. 2015; Tohme et al. 2015).

<sup>39</sup> Capasso et al., “A Mathematical Model for the 1973 Cholera Epidemic in the European Mediterranean Region” in: *Revue D'épidémiologie Et De Santé Publique* (1979).

<sup>40</sup> For an overview of the history of cholera modeling, from SI to SIR to SIRB, the reader is referred to Rinaldo, Gatto, et al., *River Networks as Ecological Corridors: Species, Populations, Pathogens* (2020).

<sup>41</sup> Azman, Luquero, Rodrigues, et al. 2012; Finger et al. 2018; Camacho et al. 2018; Lessler, Moore, et al. 2018; Koelle et al. 2004.

<sup>42</sup> Kirpich, Weppelmann, Yang, Jr, et al. 2017; Tuite, Tien, et al. 2011; Chao et al. 2011; Kirpich, Weppelmann, Yang, Ali, et al. 2015.

A spatially-explicit model has been developed at the ECHO laboratory in the past 10 years<sup>43</sup>. It has been used to studies on the dynamics of several cholera epidemics, such as in South Africa in 2000<sup>44</sup>, Senegal in 2005, Haiti from 2010 to 2019<sup>45</sup>, Democratic Republic of the Congo from 2004 to 2011 and many others<sup>46</sup>. A complete formulation of the model is presented here, including vaccination and human and hydrologic mobility<sup>47</sup>. This model has inspired to a various degree all the other models described in this thesis.

The area potentially concerned by the epidemic is divided into  $n$  regions. The sub-regions, defined by political boundaries or geomorphological features (like watersheds<sup>48</sup>), are represented as connected nodes. Each of the  $n$  nodes represent a human community with population  $H_i, i = 1, \dots, n$ .

The model is a variation of the SIR model introduced first by Kermack and McKendrick<sup>49</sup>, with additional compartments for the vaccinated individuals and the bacteria concentration in the environment. At time  $t$  and for each node  $i$ , the individuals are divided into six compartments:

- $S_i(t)$ : susceptible individuals have no immunity and may enter in contact with the bacteria and become infected (symptomatic or not),
- $I_i(t)$ : infected and infectious individual, that shed bacteria into the community reservoir,
- $R_i(t)$ : recovered are temporally immune, and don't participate in disease transmission,
- $V_i^S(t)$ : vaccinated susceptible (as the oral cholera vaccines do no offer total protection against infection),
- $V_i^I(t)$ : vaccinated infected,
- $V_i^R(t)$ : vaccinated recovered.

In addition, the model considers the bacterial concentration of *V. cholerae* in the water reservoir of the community,  $B_i(t)$ .

The  $n$  nodes are connected by both human mobility and pathogen transport through water. Individuals commute from node  $i$  to node  $j$  with probability  $Q_{ij}$ . Bacteria are transported along with the river network from node  $i$  to node  $j$  with probability  $P_{ij}$ .

The cholera dynamics are described by the following set of coupled ordinary differential equations:

$$\frac{dI_i}{dt} = \sigma F_i(t)S_i - (\gamma + \mu + \alpha)I_i \quad (1.1)$$

$$\frac{dR_i}{dt} = (1 - \sigma)F_i(t)S_i + \gamma I_i - (\rho + \mu + \frac{\nu_i(t)}{S_i + R_i})R_i \quad (1.2)$$

$$\frac{dV_i^S}{dt} = \nu_i(t)\frac{S_i}{S_i + R_i} - \mu V_i^S \quad (1.3)$$

$$\frac{dV_i^I}{dt} = \sigma(1 - \eta)F_i(t)V_i^S - (\gamma + \mu + \alpha)V_i^I \quad (1.4)$$

$$\frac{dV_i^R}{dt} = \nu_i(t)\frac{R_i}{S_i + R_i} + (1 - \sigma)(1 - \eta)F_i(t)V_i^S + \gamma V_i^I - (\mu + \rho_v)V_i^R \quad (1.5)$$

$$\frac{dB_i}{dt} = -\mu_B B_i + \frac{p}{W_i} [1 + \lambda J_i(t)] \left( (1 - m)(I_i + V_i^I) + m \sum_{j=1}^n Q_{ij}(I_j + V_j^I) \right) - l \left( B_i - \sum_{j=1}^n P_{ji} \frac{W_j}{W_i} B_j \right). \quad (1.6)$$

<sup>43</sup> Bertuzzo, Azale, et al., “On the Space-Time Evolution of a Cholera Epidemic” in: *Water Resources Research* (2008).

<sup>44</sup> Mari, Bertuzzo, Righetto, et al., “Modelling Cholera Epidemics: The Role of Waterways, Human Mobility and Sanitation” in: *Journal of The Royal Society Interface* (2012).

<sup>45</sup> Bertuzzo, Mari L., et al., “Prediction of the Spatial Evolution and Effects of Control Measures for the Unfolding Haiti Cholera Outbreak” in: *Geophysical Research Letters* (2011); Bertuzzo, Finger, et al., “On the Probability of Extinction of the Haiti Cholera Epidemic” in: *Stochastic Environmental Research and Risk Assessment* (2016).

<sup>46</sup> Finger et al., “The Potential Impact of Case-Area Targeted Interventions in Response to Cholera Outbreaks: A Modeling Study” in: *PLOS Medicine* (2018).

<sup>47</sup> Bertuzzo, Finger, et al., “On the Probability of Extinction of the Haiti Cholera Epidemic” in: *Stochastic Environmental Research and Risk Assessment* (2016); Pasetto, Finger, Camacho, et al., “Near Real-Time Forecasting for Cholera Decision Making in Haiti after Hurricane Matthew” in: *PLOS Computational Biology* (2018).

<sup>48</sup> Bertuzzo, Finger, et al., “On the Probability of Extinction of the Haiti Cholera Epidemic” in: *Stochastic Environmental Research and Risk Assessment* (2016).

<sup>49</sup> Kermack et al., “A Contribution to the Mathematical Theory of Epidemics” in: *Proceedings of the Royal Society A* (1927).

Here demographic equilibrium is assumed, hence the population  $H_i$  of each node is assumed to be constant, which implies that the number of susceptible individuals at time  $t$  is  $S_i(t) = H_i - I_i(t) - R_i(t) - V_i^S(t) - V_i^I(t) - V_i^R(t)$ , but it is also possible to model  $\frac{dS_i}{dt}$  dynamically.

The force of infection represents the rate at which an individual enters in contact with the disease. It is written as:

$$F_i(t) = \beta \left[ (1-m) \frac{B_i}{K+B_i} + m \sum_{j=1}^n Q_{ij} \frac{B_j}{K+B_j} \right]. \quad (1.7)$$

The parameter  $\beta$  represents the maximum exposure rate (and is proportional to the basic reproduction number  $R_0$ , see CHAPTER 5), which may vary in time due to non-pharmaceutical interventions or the awareness of the population<sup>50</sup>. The fraction  $B_i/(K+B_i)$ <sup>51</sup> is the probability of becoming infected due to the exposure to a concentration  $B_i$  of *V. cholerae*,  $K$  being the half-saturation constant<sup>52</sup>. The force of infection in a given node depends for a fraction  $(1-m)$  to the local concentration of *V. cholerae*  $B_i$ . The remaining fraction  $m$ , the community-level probability that individuals travel outside their node, are exposed to the concentration  $B_j$  of the remote communities.

The human mobility is encoded in matrix  $Q$ , where  $Q_{ij}$  represents the probability that a mobile individual living in node  $i$  reaches  $j$  as a destination. Because of human mobility, a susceptible individual residing at node  $i$  can be exposed to pathogens in the destination community  $j$ . In the lack of detailed mobility data, the probabilities  $Q_{ij}$  can be estimated through a gravity model of human mobility<sup>53</sup>:

$$Q_{ij} = \frac{H_j e^{-d_{ij}/D}}{\sum_{k \neq i}^n H_k e^{-d_{ik}/D}}, \quad (1.8)$$

where the attractiveness of node  $j$  depends on its population size  $H_j$ , while the deterrence factor is assumed to be dependent on the distance  $d_{ij}$  between the two communities via an exponential kernel (with shape factor  $D$ ).

Individuals are removed from the susceptible compartment  $S$  at rate  $F(t)$ . A fraction  $\sigma$  of the infected individuals develops symptoms, passing from compartment  $S$  to  $I$  and shed *V. cholerae* into the reservoir at rate  $\theta$ . The remaining fraction  $(1-\sigma)$  does not develop symptoms and does not contribute to the disease transmission<sup>54</sup>, and is considered temporally immune, thus passing from compartment  $S$  to  $R$ . Symptomatic infected individuals recover at a rate  $\gamma^{55}$ , or die due to cholera or other causes at rates  $\alpha$  or  $\mu$ , respectively. Recovered individuals lose their immunity at rate  $\rho$ , the average rate of loss of immunity for individuals that previously had been asymptomatic or symptomatic, or die at a rate  $\mu$ . In both cases, they return to susceptible (death is associated with the birth of a susceptible to keep the population constant).

The environmental concentration of *V. cholerae*  $B_i$  may increase due to infected individuals shedding *Vibrios* and hydrologic travel. As for the force of infection, bacterial shedding is proportional for a fraction  $1-m$  on the local infected individuals and for a fraction  $m$  on the infected individuals moving according to the mobility network. The increase in bacteria concentration is modeled with the rate  $p/W_i$ , where  $p$  is the rate at which bacteria excreted by an infected individual reach and contaminate the local water reservoir of volume  $W_i$  (assumed to be proportional to population size, i.e.,  $W_i = cH_i$ <sup>56</sup>). Rainfall-induced runoff might

<sup>50</sup> Bertuzzo, Finger, et al., “On the Probability of Extinction of the Haiti Cholera Epidemic” in: *Stochastic Environmental Research and Risk Assessment* (2016).

<sup>51</sup> In the classical SIR model, this fraction would be  $\frac{I}{H}$ , a mass action reaction on the proportion of infected in the population.

<sup>52</sup> Codeço, “Endemic and Epidemic Dynamics of Cholera: The Role of the Aquatic Reservoir” in: *BMC Infectious Diseases* (2001).

<sup>53</sup> Erlander et al., *The Gravity Model in Transportation Analysis: Theory and Extensions* (1990).

<sup>54</sup> In the model presented in CHAPTER 3, asymptomatic individuals are assumed to contribute to disease transmission, and are explicitly modeled in a compartment  $A$ .

<sup>55</sup> Hence the mean infectious period is  $\frac{1}{\gamma}$  [duration]. Correspondingly the life expectancy without cholera would be  $\frac{1}{\mu}$ .

<sup>56</sup> Rinaldo, Bertuzzo, Mari, et al., “Reassessment of the 2010–2011 Haiti Cholera Outbreak and Rainfall-Driven Multiseason Projections” in: *Proceedings of the National Academy of Sciences* (2012).

cause additional pathogen loads to enter the water reservoir due to effects such as the overflow of pit latrines and washout of open-air defecation sites. The contamination rate  $p$  is thus intensified by the rainfall intensity  $J_i(t)$  via a coefficient  $\lambda^{57}$ . By introducing the dimensionless bacterial concentrations  $B_i^* = B_i/K$ , it is possible to group three model parameters into a single shedding intensity ratio  $\theta = p/(cK)^{58}$ . Bacteria undergo hydrologic dispersal at a rate  $l$ : pathogens travel from node  $i$  to  $j$  with probability  $P_{ij}$ , which is assumed to be one if node  $j$  is the downstream nearest neighborhood  $i$ , and zero otherwise. *V. cholerae* decays at rate  $\mu_B$  in the water reservoir.

The estimation of the local incidence is computed integrating over the new symptomatic individuals,

$$\frac{dC_i}{dt} = \sigma F_i S_i, \quad (1.9)$$

During the vaccination campaign, oral cholera vaccines doses are distributed at rate  $\nu_i(t)$  to susceptible and recovered individuals, which enter the compartments  $V^S$  and  $V^R$ . As the OCV provides a partial immunity having efficacy  $\eta$ ,  $0 \leq \eta \leq 1$ , vaccinated susceptibles ( $V^S$ ) can become infected ( $V^I$ ) through a decreased force of infection of a factor  $(1 - \eta)$  with respect to non-vaccinated individuals. Vaccinated infected individuals behave exactly like infected ones, but are placed in a different compartment to exclude them from future vaccination campaigns. After recovering at rate  $\gamma$ , they lose their vaccine protection at rate  $\rho_v$ .

Derivations of this model are used in the next 2 chapters. Hydrological transport is always neglected ( $P_{ij} = 0, \forall i, j$ ), and other differences are:

- in CHAPTER 2: the model is very similar, with only one spatial node. Compartment for the incubating individuals  $E$  and  $V_E$  are added because of the higher temporal resolution of the date. Moreover, a precise non-linear formulation of the effect of rainfall is developed. The force of infection is further generalized to encompass non-linear rainfall effects and human-to-human transmission. Finally, alongside the present ordinary differential equations formulation, a version of the model with stochastic transitions is developed to better deal with low incidence data.
- in CHAPTER 3: the model is also a very similar stochastic translation of the present model, with mobility (but no hydrological transport) implemented as human-to-human transmission. Shedding (infectious) asymptomatic are modeled in an additional compartment  $A$ . The effect of vaccination is described very precisely with vaccine efficacy waning over time. As the effect of synchronized immunity is sought after, the time spent recovered is more accurately represented as an erlang distribution with parameter  $3$ . The rainfall uses the same non-linear effect introduced in CHAPTER 2.

The COVID-19 models proposed later are very different but share some features e.g. the mobility formulation is close in CHAPTER 4 and CHAPTER 6, and more generally the general structure is similar across all compartmental models.

<sup>57</sup> Rinaldo, Bertuzzo, Mari, et al., “Reassessment of the 2010–2011 Haiti Cholera Outbreak and Rainfall-Driven Multiseason Projections” in: *Proceedings of the National Academy of Sciences* (2012); Righetto et al., “Rainfall Mediations in the Spreading of Epidemic Cholera” in: *Advances in Water Resources* (2013).

<sup>58</sup> Bertuzzo, Azalee, et al., “On the Space-Time Evolution of a Cholera Epidemic” in: *Water Resources Research* (2008).

## CHAPTER 2

# RAINFALL AS A DRIVER OF EPIDEMIC CHOLERA: COMPARATIVE ASSESSMENTS OF THE EFFECT OF INTRA-SEASONAL PRECIPITATION EVENTS

The correlation between cholera epidemics and climatic drivers, in particular seasonal tropical rainfall, has been studied in a variety of contexts. Several mechanistic models have included rainfall as a driver of cholera transmission by focusing on two possible pathways: either by increasing exposure to contaminated water (*e.g.* due to worsening sanitary conditions during water excess) or by increased water contamination with freshly excreted bacteria (*e.g.* due to washout of open-air defecation sites or overflows). In this chapter, the explanatory power of these different model structures is assessed by formal model comparison using deterministic and stochastic models of the type susceptible-infected-recovered-bacteria (SIRB). The incorporation of rainfall effects is generalized using a nonlinear function that can increase or decrease the relative importance of large precipitation events. The modeling framework is applied to the daily epidemiological data collected during the 2015 cholera outbreak in Juba, South Sudan. This epidemic is characterized by a particular intra-seasonal double peak of incidence in apparent relation with particularly strong rainfall events. Results suggest that rainfall-based models in both their deterministic and stochastic formulations outperform models that do not account for rainfall. In fact, classical SIRB models are not able to reproduce the second peak, suggesting it was rainfall-driven. Moreover stronger support is found for rainfall acting on increased exposure rather than on exacerbated water contamination. Although these results are context-specific, they stress the importance of a systematic and comprehensive appraisal of transmission pathways and their environmental forcings when embarking on the modeling of epidemic cholera.

This section is adapted from: Joseph Lemaitre, Damiano Pasetto, Javier Perez-Saez, Carla Sciarra, Joseph Francis Wamala, and Andrea Rinaldo. “Rainfall as a Driver of Epidemic Cholera: Comparative Model Assessments of the Effect of Intra-Seasonal Precipitation Events”. In: *Acta Tropica* 190 (Feb. 1, 2019), pp. 235–243, referred in the following as the postprint, and its supplementary information as SI.

### RAINFALL AND THE TRANSMISSION OF CHOLERA

Two main exposure pathways fuel cholera transmission across endemic and epidemic settings. First, an *indirect* exposure occurs from consumption of water contaminated by untreated sewage; rainfall and the ensuing hydrologic transport processes might play a role in water contamination, for instance through the washout of open-air defecation sites and sewage circulation in the environment. Alternatively, *direct*, or human-to-human exposure occurs when the bacteria is transmitted from an infectious individual directly to a susceptible one, for example via contaminated food or fomite. In this case, environmental factors do not play a major role, except for transmission changes due to behavior modifications. The mechanism behind the transmission of cholera has been postulated to be a

combination of environmentally-mediated and direct exposures<sup>1</sup>.

**CHOLERA AND RAINFALL** The relationship between different climatic and environmental factors and the transmission of cholera has long been studied. Works linking cholera outbreaks to anomalies in the El Niño Southern Oscillation<sup>2</sup> have paved the way for a new field in epidemiological research.

Previous studies have highlighted the role of climatic drivers on cholera dynamics, mostly focusing on climate change effects on disease spread<sup>3</sup> or on the impacts of spatial and temporal heterogeneities<sup>4</sup>. The effect of temperature on cholera transmission has been described (though mainly through the lense of *Vibrios* survival and ecology in natural environments), but for rainfall, it remains to be fully elucidated, possibly due to the multiple ways in which it can influence transmission at the local and regional scales<sup>5</sup>. Indeed, intense rainfall events have been shown to alter infection risk through a variety of potential mechanisms, including: flooding, leading to sewage contamination of water sources<sup>6</sup>; increased hydrologic transport-driven iron availability in environmental waters that enhances pathogen survival and the expression of toxins<sup>7</sup>; dry spells inducing persistent low water levels leading to increased use of unsafe water sources<sup>8</sup>; and crowding during strong flood events<sup>9</sup>.

Most countries where associations between rainfall and cholera risk have been studied experience endemic cholera transmission. Empirical studies have shown a range of correlations, both positive and negative, endowed with time lags ranging from weeks to months<sup>10</sup>. In general, rainfall has been found to increase cholera transmission, but there is evidence of the inverse, possibly due to pathogen dilution<sup>11</sup>. Such variability reflects the variety of potential mechanisms whereby rainfall may alter infection risk. Similarly, a clear empirical correlation between intense rainfall and enhanced transmission is found in several regions hit by cholera epidemics<sup>12</sup> (fig. 2.1). Results therein showed that at all spatial scales and locations examined, tropical storms were correlated with increased cholera incidence with lags of the order of a few days. As a consequence, accounting for the related forcing of dynamic models resulted in improved fits of reported incidence.

Properly incorporating the effects of rainfall in mathematical models of cholera transmission is thus paramount to discriminate among the above-mentioned alternative transmission pathways, thus unlocking a predictive framework to evaluate the potentially rainfall-sensitive efficacy of available intervention strategies in endemic and epidemic settings. This becomes critically important when evaluating the number of averted infections by deploying vaccines, as was done in the aftermath of the passage of Hurricane Matthew<sup>13</sup>, or considering optimal deployment in space and time.

The effect of rainfall on indirect cholera transmission has been accounted for in two main fashions in recent mathematical models. On one side, a *contamination*-centered approach suggesting that bursts of infections could be linked to increased contamination of the water compartment<sup>14</sup>, as in the ECHO model presented in CHAPTER 2. This process conceptualizes the washout of open-air defecation sites by hydrologic transport. The same ‘transport’ effect may be realized by sewer collectors’ overflows. In fact, both mechanisms have the net effect of charging progressively the bacterial concentration in the water reservoir<sup>15</sup>. Pathogens’ loads are washed out from a hydrologic catchment enclosing human settlements and their infective individuals shedding bacteria. Therein, pathogen survival and thus

<sup>1</sup> Sugimoto et al. 2014; Bi, Azman, et al. 2016; Lessler, Salje, et al. 2016; Rinaldo, Bertuzzo, Blokesch, et al. 2017.

<sup>2</sup> Colwell 1996; Pascual et al. 2000; Hashizume, Chaves, et al. 2013.

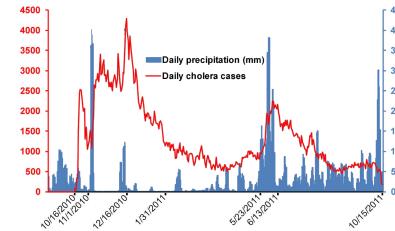


fig. 2.1: Daily cholera cases (red) and daily rainfall (blue) in Haiti from September 15, 2010 to October 16, 2011. It highlights the visual correlation between heavy rainfall event and cholera outbreaks. Adapted from Gaudart et al., “Spatio-Temporal Dynamics of Cholera during the First Year of the Epidemic in Haiti” in: *PLOS Neglected Tropical Diseases* (2013).

<sup>3</sup> Rinaldo, Bertuzzo, Blokesch, et al. 2017; Hashizume, Armstrong, et al. 2008; Magny, Thiaw, et al. 2012; Rodó et al. 2013; Ramírez et al. 2016; Vezzulli, Grande, et al. 2016.

<sup>4</sup> Reiner et al. 2012; Baker-Austin et al. 2013; Vezzulli, Colwell, et al. 2013; Cash et al. 2014; Escobar et al. 2015; Vezzulli, Pezzati, et al. 2015; Perez-Saez, King, et al. 2017.

<sup>5</sup> Rinaldo, Bertuzzo, Mari, et al. 2012; Eisenberg, Kujbida, et al. 2013; Baracchini et al. 2017.

<sup>6</sup> Ruiz-Moreno et al. 2007; Hashizume, Armstrong, et al. 2008.

<sup>7</sup> Lipp et al. 2002; Faruque et al. 2005; Hill et al. 2011.

<sup>8</sup> Rebaudet, Sudre, et al. 2013b.

<sup>9</sup> Reiner et al. 2012.

<sup>10</sup> Ruiz-Moreno et al. 2007; Emch et al. 2008; Magny, Thiaw, et al. 2012.

<sup>11</sup> Ruiz-Moreno et al. 2007.

<sup>12</sup> Magny, Murtugudde, et al. 2008; Magny, Thiaw, et al. 2012; Rebaudet, Sudre, et al. 2013b; Rebaudet, Sudre, et al. 2013a; Jutla et al. 2013, but Haiti is treated specifically in the next chapter.

<sup>13</sup> Pasetto, Finger, Rinaldo, et al. 2017.

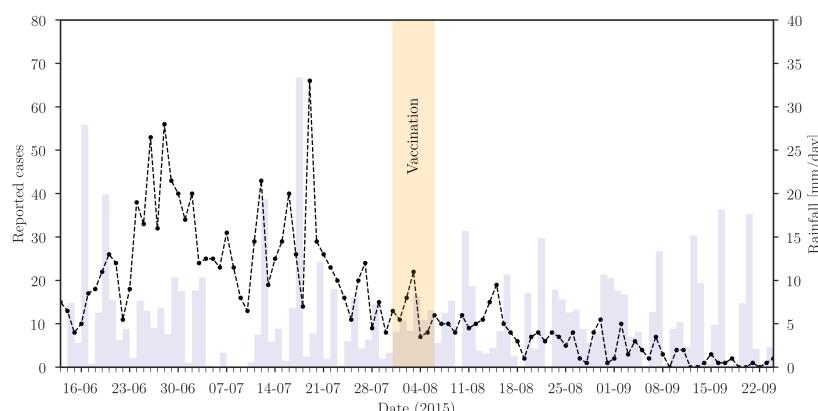
<sup>14</sup> Rinaldo, Bertuzzo, Mari, et al. 2012.

<sup>15</sup> Codeço 2001.

the toxicity of their loads depend on hydrologic residence time distributions<sup>16</sup>. Such loads increase as a function of rainfall, which acts as a proxy of runoff volumes. The second approach is *exposure*-centered and employs a rainfall-dependent exposure rate subsuming both pathogen availability and the probability of the ingestion of contaminated water during wet spells<sup>17</sup>. Although both approaches are physically plausible, they have not been compared directly on the same datasets within a formal statistical framework, which would allow to highlight their respective merits and further recommendations for their use in different settings.

In this chapter, the explanatory power of these different types of rainfall-driven mechanistic models applied to a cholera outbreak in South Sudan is compared. The link between rainfall and cholera during the outbreak recorded in Juba in 2015 when an intra-seasonal peak of cholera cases was recorded possibly in correspondence to intense precipitation events is quantitatively examined. The analysis of the lagged relationship between rainfall rates and revamped cholera incidence is addressed via dynamical compartmental models considered both in deterministic and stochastic versions incorporating both direct (human-to-human) and indirect (water-to-human) disease transmission, and rainfall effects on both contamination and exposure.

## CASE STUDY: THE 2015 CHOLERA OUTBREAK IN JUBA, SOUTH SUDAN



<sup>16</sup> Rinaldo, Bertuzzo, Mari, et al. 2012; Rinaldo, Bertuzzo, Blokesch, et al. 2017.

<sup>17</sup> Eisenberg, Kujbida, et al. 2013.

In the past years, South Sudan had been struck by several cholera outbreaks<sup>18</sup>. Here the analysis focuses on the 2015 outbreak in Juba, when a particular double peak of cholera cases was associated with a strong intra-seasonal precipitation event (fig. 2.2). Access was granted to epidemiological records for the 2014 and 2015 cholera epidemics that include daily cholera incident cases and hospitalization at the second-lowest administrative level (Payams), reported by the Ministry of Health of South Sudan. The cases in the 7 Payams that constitute the administrative area of Juba have been aggregated to obtain the reported time series for the county level. The population data for Juba county is taken from the South Sudan National Bureau of Statistics<sup>19</sup>. Daily rainfall estimates for years 2014 and 2015 were obtained from the Climate Data Library<sup>20</sup>. In 2015, 167'377 OCV doses were distributed in the county of Juba during 6 days of a mass vaccination campaign started on July 31, 2015<sup>21</sup>, and are taken into account in the model.

fig. 2.2: Reported cholera cases (dots) and precipitation (bars) during the 2015 cholera epidemic in Juba. The timing of the vaccination campaign is highlighted in yellow.

<sup>18</sup> More details on the history of cholera in South Sudan are found in: Sciarri et al., *Mathematical Modeling of Cholera Epidemics in South Sudan* (2018)

<sup>19</sup> SSNBS, *Population Projections for South Sudan by County 2015 - 2020* (2015).

<sup>20</sup> IRI/LDEO, "Climate Data Library" (2016), with a resolution of 0.1°, rainfall was averaged over the study area.

<sup>21</sup> Abubakar et al., "The First Use of the Global Oral Cholera Vaccine Emergency Stockpile: Lessons from South Sudan" in: *PLOS Medicine* (2015); Azman, Parker, et al., "Effectiveness of One Dose of Oral Cholera Vaccine in Response to an Outbreak: A Case-Cohort Study" in: *The Lancet Global Health* (2016); Parker et al., "Adapting to the Global Shortage of Cholera Vaccines: Targeted Single Dose Cholera Vaccine in Response to an Outbreak in South Sudan" in: *The Lancet Infectious Diseases* (2017).

## CHOLERA MODELS

### GENERALIZED CHOLERA MODEL

The proposed model is inspired by the model presented in the previous chapter: we have the susceptible  $S$ , infected  $I$ , and recovered  $R$  compartments for individuals, with an additional variable  $B$  describing the concentration of the bacteria in the environment. Previous modelling exercises had considered rainfall intensity  $J(t)$  either to i) multiplicatively increase water *contamination* with bacteria shed by infected individuals<sup>22</sup>, or ii) assumed that the rainfall multiplicatively increases the *exposure* to contaminated water<sup>23</sup>. A generalized formulation of these cholera-forced models, wherein both formulations are nested, is proposed. It enables the systematic comparison of the effect of rainfall through these two different transmission pathways.

The model described in CHAPTER I was designed to deal with weekly data. Here, given the daily temporal resolution at which incidence data was available for the 2015 Juba epidemic, a compartment of exposed individuals  $E$  is added to the model structure. It describes the incubation period: the lag between the time of exposure and the onset of the symptoms. Moreover, to account for the vaccination campaigns that were deployed in Juba in August 2015, four compartments ( $V^S$ ,  $V^E$ ,  $V^I$ , and  $V^R$ ) are added to describe the dynamics of vaccinated individuals and their removal from the pool of susceptibles.

The proposed generalized cholera model is described in fig. 2.3, and formulated as:

$$S(t) = H - I(t) - E(t) - R(t) - V^S(t) - V^E(t) - V^I(t) - V^R(t) \quad (2.1)$$

$$\frac{dE}{dt} = \sigma F(t)S - (\phi + \mu + \nu)E \quad (2.2)$$

$$\frac{dI}{dt} = \phi E - (\gamma + \mu + \alpha)I \quad (2.3)$$

$$\frac{dR}{dt} = (1 - \sigma)F(t)S + \gamma I - (\rho + \mu + \nu)R \quad (2.4)$$

$$\frac{dB}{dt} = -\mu_B B + \theta [1 + f_c(J(t))] (I + V^I) \quad (2.5)$$

$$\frac{dV^S}{dt} = \nu S - \mu V^S + \rho_v V^R - (1 - \eta)F(t)V^S \quad (2.6)$$

$$\frac{dV^E}{dt} = \nu E + \sigma(1 - \eta)F(t)V^S - (\phi + \mu)V^E \quad (2.7)$$

$$\frac{dV^I}{dt} = \phi V^E - (\gamma + \alpha + \mu)V^I \quad (2.8)$$

$$\frac{dV^R}{dt} = \nu R - (\mu + \rho_v)V^R + \gamma V^I + (1 - \sigma)(1 - \eta)F(t)V^S, \quad (2.9)$$

where  $F(t)$  takes into account both human-to-human transmission and nonlinear water-to-human transmission:

$$F(t) = \underbrace{\beta_B \left[ \frac{B}{K + B} \right] \left( 1 + f_e(J(t)) \right)}_{\text{indirect transmission}} + \underbrace{\beta_I \frac{(I + V^I)}{H}}_{\text{direct transmission}}. \quad (2.10)$$

The notation is consistent with CHAPTER I and we refer the reader to the description of the processes provided on p. 28, and reminded in the margin. In

<sup>22</sup> Bertuzzo, Finger, et al., “On the Probability of Extinction of the Haiti Cholera Epidemic” in: *Stochastic Environmental Research and Risk Assessment* (2016); Pasetto, Finger, Rinaldo, et al., “Real-Time Projections of Cholera Outbreaks through Data Assimilation and Rainfall Forecasting” in: *Advances in Water Resources* (2017).

<sup>23</sup> Eisenberg, Kujbida, et al., “Examining Rainfall and Cholera Dynamics in Haiti Using Statistical and Dynamic Modeling Approaches” in: *Epidemics* (2013).

symbol	signification
$\alpha$	cholera induced mortality
$\mu$	natural mortality
$\sigma$	symptomatic fraction
$\gamma$	infectious period
$\rho$	duration of immunity
$\rho_v$	duration of vaccine protection
$\eta$	vaccine efficacy
$\theta$	shedding rate
$\mu_B$	bacteria decay in water
$J(t)$	rainfall

Reminder of the parameters already described in CHAPTER I (p. 28).

addition to the previously described dynamics, exposed individuals become symptomatic infected at a rate  $\phi$ , which corresponds to an average incubation period of  $1/\phi \approx 1.5$  days<sup>24</sup>. As few reliable data on changes in Juba's population are available for the years of interest, the total population  $H$  is assumed to be constant, as in (2.1).

Moreover, the effect of rainfall is more precisely described. Functions  $f_c(J(t))$  and  $f_e(J(t))$  account for the rainfall effect respectively by increasing the bacteria contamination in the water reservoir or directly through amplifying the exposure in the force of infection. With the objective of assessing the importance of rain-

<sup>24</sup> Azman, Rudolph, et al., "The Incubation Period of Cholera: A Systematic Review" in: *Journal of Infection* (2013).

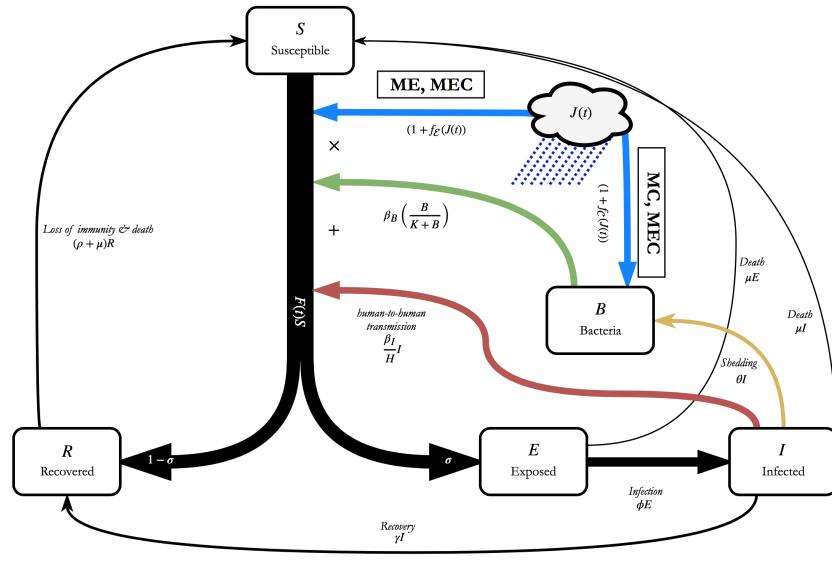


fig. 2.3: Transition diagram for the different cholera models considered, with the different variations ME, MR, and MEC indicated.

fall on cholera transmission, a generalization of the linear relation found in the litterature<sup>25</sup> is proposed, by using a nonlinear function form for  $f_{c,e}(J(t))$ :

$$f_{c,e}(J(t)) = \lambda_{c,e} \left( \frac{J(t)}{\max_t J(t)} \right)^{\alpha_{c,e}}, \quad (2.11)$$

where the subscripts  $c,e$  respectively denote the effect of rainfall on exposure and contamination,  $\max_t J(t)$  is the maximum recorded rainfall intensity during the epidemic, and the  $\alpha_{c,e} \geq 0$  controls for the relative importance of different rainfall intensities in their effect on the force of infection. Indeed, since the ratio  $\frac{J(t)}{\max_t J(t)} \in [0, 1]$ , for  $\alpha_{c,e} \gg 1$  the ratio will tend to 0 for all small precipitation events, leaving only the effect of the strongest events, whereas for  $\alpha_{c,e} < 1$  all precipitation events will be assigned a similar weight in the force of infection (fig. 2.4). The formulations found in litterature are recovered by setting  $\alpha_{c,e} = 1$ . The flexibility allowed by eqn. (2.11) allows to discriminate between rainfall effects along a continuum from acting on disease transmission regardless of intensity to a threshold-like effect for the largest events which could be associated to severe flooding causing damages to the city's water and sanitary system, for instance leading to sewer overflow.

<sup>25</sup> Specifically the reference exposure (Eisenberg, Kujibida, et al. 2013) and contamination (Rinaldo, Bertuzzo, Mari, et al. 2012) models.

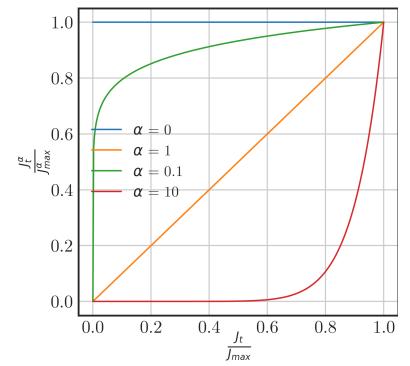


fig. 2.4:  $\left( \frac{J_t}{J_{\max}} \right)^\alpha$  for different values of  $\alpha$ . For  $\alpha = 1$ , the linear model described in previous works is obtained, and larger  $\alpha$  corresponds to an increased relative importance of heavy rainfall events.

## COMPETING TRANSMISSION MODELS

The relevance of the two rainfall-driven transmission pathways is assessed here by comparing the following models:

- **MN** SIRB model without rainfall:  $\lambda_c = \lambda_e = 0, \beta_I = 0$ , as the null hypothesis for the importance of rainfall.
- **MC** SIRB model where rainfall enhances the *contamination* of the water reservoir:  $\lambda_c = 0, \beta_I = 0$ .
- **ME** SIRB model where rainfall increases the *exposure* to bacteria:  $\lambda_c = 0, \beta_I = 0$ .
- **MEC** SIRB model combining both approaches **MC** and **ME** ( $\beta_I = 0$ ). Both ways of accounting rainfall play a role simultaneously.

Model	$\lambda_c$	$\alpha_c$	$\lambda_e$	$\alpha_e$	$\beta_I$
MN	–	–	–	–	–
MNH	–	–	–	–	×
MC	×	×	–	–	–
ME	–	–	×	×	–
MCH	×	×	–	–	×
MEH	–	–	×	×	×
MEC	×	×	×	×	–
MECH	×	×	×	×	×

For each model is explored the possibility of adding explicitly direct, human-to-human transmission ( $\beta_I > 0$ ), which is indicated with an H at the end of the model name: **MNH**, **MCH**, **MEH**, and **MECH**. The different parameters associated with the considered models are summarized in tab. 2.1.

The eight models are compared on the basis of their ability to match the time series of daily reported cases during the cholera epidemic in Juba of 2015. The postprint **SI**, tab. 1 summarizes which parameters are calibrated for each model and their prior distribution. The degrees of freedom of the models,  $n_p$ , vary from  $n_p = 7$  for **MN** to  $n_p = 12$  for **MECH**. Given the low number of daily reported cases and their ensuing variability, a stochastic equivalent is also implemented<sup>26</sup> of the deterministic ODE system, eqns. (2.1)–(2.9), formulated as a continuous-time partially observed Markov process model, accounting for both demographic and disease transmission stochasticities<sup>27</sup>. The generalized stochastic model is structured exactly like the deterministic one and a similar stochastic model is presented in the next chapter. Hence stochastic model equations are left in the postprint **SI**.

## MODELS CALIBRATION

**INITIAL CONDITIONS** The history of cholera epidemics in South Sudan, and particularly in Juba, plays an important role in the determination of the size of susceptible and recovered compartments at the beginning of the 2015 epidemic. These two compartments are also largely impacted by the rate of immunity loss  $\rho$  of the recovered individuals and by the probability of asymptomatic infection  $1 - \sigma$ <sup>28</sup>. The initial conditions must therefore be estimated for each parameter set considered during calibration.

To take into account the uncertainty associated with the immunity landscape of Juba, the initial number of recovered individuals in April 2014 is calibrated for each model. Then, the detailed daily data of suspected cases during 2014 is used

**ME** models uses a similar transmission pathway as in Eisenberg, Kujbida, et al. (2013), whereas **MC** models were described in Rinaldo, Bertuzzo, Mari, et al. (2012). These models were adapted to the generalized configuration, e.g. in eqn. (2.10) precipitation enters in the term  $(1 + f_E(J(t)))$  which entails water-to-human transmission also when  $J(t) = 0$ , unlike the published version. The steps take to adapt these models are left in the postprint **SI**, section 1.

tab. 2.1: Parameters considered in the eight compared models.  $\lambda$  and  $\alpha$  characterize the functional forms considering the precipitation eqn. (2.11).  $\beta_I$  is the exposure for human-to-human transmission. A dash – indicate that the parameter is set to zero, whereas a cross × indicate that the parameter is used.

<sup>26</sup> The sole difference in formulation is the introduction of overdispersion in the infection process. The force of infection is multiplied by a time-continuous white noise process, as detailed in the eqns. (3.1)–(3.2).

<sup>27</sup> Bretó, He, et al., “Time Series Analysis via Mechanistic Models” in: *The Annals of Applied Statistics* (2009).

For details about fixed and calibrated parameters, along with prior and posterior distributions, the reader is referred to the supplementary information of Lemaitre, Pasetto, Perez-Saez, et al., “Rainfall as a Driver of Epidemic Cholera: Comparative Model Assessments of the Effect of Intra-Seasonal Precipitation Events” in: *Acta Tropica* (2019).

<sup>28</sup> in the literature, values range from  $\sigma = 0.5$ , meaning one asymptomatic per each symptomatic infected, to less than  $\sigma = 0.01$ , corresponding to more than 99 asymptomatic infected per each symptomatic infected (Fung 2014).

to estimate the associated number of recovered individuals at the start of the 2015 epidemic. Suspected cases in 2014 are scaled to the reporting fraction and undergo an exponential decay with an average time of immunity loss  $1/\rho^{29}$ . Simulations are then initialized on June 5, 2015 considering two exposed individuals, two infected, and the associated steady-state bacteria concentration.

**CALIBRATION** Calibration of the deterministic model is performed using a Markov Chain Monte Carlo based algorithm<sup>30</sup>, which draws samples from the posterior distribution of the parameters. Inference on the stochastic model is performed using a frequentist iterated filtering algorithm<sup>31</sup>. Both models were fit against the daily reported cases accounting for over- or under-reporting, and assuming a Poisson likelihood<sup>32</sup>. Each datapoint at reporting date  $t_i$ ,  $y_{t_i}$ , is assumed to belong to a Poisson distribution centered on the modeled incidence  $C_{t_i}$ <sup>33</sup> and its associated parameter vector  $\theta$  altered by the reporting rate  $\varepsilon$ , as:

$$y_{t_i} \sim \text{Poisson}(\varepsilon C_{t_i}(\theta)), \quad \varepsilon > 0, \quad (2.12)$$

Models were then compared using the Bayesian Information Criterion (BIC), Bayes factors, and the likelihood ratio test for the nested models.

## RESULTS

### MODEL SELECTION OF RAINFALL EFFECTS ON CHOLERA TRANSMISSION

The summary statistics of the deterministic and stochastic models considered in the study are given in tab. 2.2. Overall, the stochastic models outperform their deterministic counterparts for all model structures by  $\approx 40$  log-likelihood units. Both model types agree on the significance of rainfall in explaining the time series of daily reported cases, in particular through the increased exposure pathway, although the specific ordering of the models differs between model types. The BFs for the deterministic models suggest a strong support for model **M****E**, followed by model **M****E** ( $BF_{ME,MEH}^{-1} = 0.16$ ), with very little support for all other models ( $BF_{ME,MNH}^{-1} < 10^{-2}$  for all other models than **M****E**). For the stochastic model, the BFs estimated with the BIC suggest the strongest support for model **M****E**, with the basic SIRB model coming in second with 5 times less evidence ( $BF_{MN,ME}^{-1} \approx 0.15$ ). When considering only the BIC, model **M****E** ranks first for both the deterministic and the stochastic formulations. Interestingly, all models that include human-to-human transmission present smaller or equal log-likelihoods than their counterparts with only the bacteria compartment, which suggests that the data does not support both environmental and human-to-human transmission within the set of the models considered here.

The results of the nested LR-tests confirm the statistical significance of including rainfall in the cholera transmission models, with the effect on exposure better supported by the data in both model types than the effect on contamination. In the deterministic case, the extension of the basic SIRB (model **M****N**) with rainfall effects were significant for all direct comparisons (fig. 2.5,A). The addition of human-to-human transmission was not significant mostly due to the above-mentioned lower estimate of the log-likelihood in these models. When considering only a single effect of rainfall (either increasing exposure or contamination), **M****E** outperforms **M****C** in terms of likelihood for the same number of parameters.

<sup>29</sup> Something similar has been done in (Pasetto, Finger, Rinaldo, et al. 2017).

<sup>30</sup> Differential Evolution Adaptive Metropolis, **DREAM** (Vrugt 2016), was developed to explore high-dimensional parameter spaces. Given the parameters' prior distribution, **DREAM** searches and selects new samples in the posterior distribution by using multiple **MCMC** chains that run in parallel and that jointly contribute to the computation of the proposal parameter samples. **MCMC** chains converge toward the posterior probability distribution based on the log-likelihood function of the data given the model output.

<sup>31</sup> MIF (Ionides et al. 2015) is a frequentist-based approach for identifying the maximum likelihood estimation. It has proved successful even for a range of complex models of cholera dynamics (King, Ionides, et al. 2008; Baracchini et al. 2017). The MIF<sub>2</sub> algorithm, which employs iterated Bayes maps, is implemented in the R package **pomp** (King, Nguyen, et al. 2015).

<sup>32</sup> Camacho et al. 2018.

<sup>33</sup>  $C_{t_i}$  is computed as in eqn. (1.9), but for the  $E \rightarrow I$  transition. For the deterministic model:  $C_{t_i} = \int_{t_i}^{t_{i+1}} \phi(E(t) + V^E(t)) dt$ . For the stochastic model the reporting process is described as  $C_{t_i} = [N_{EI}(t_{i+1}) - N_{EI}(t_i)] + [N_{VEV^I}(t_{i+1}) - N_{VEV^I}(t_i)]$ , where  $N_{AB}$  denotes the stochastic counting process of transitions between classes  $A$  and  $B$ .

Model	Deterministic				Stochastic			
	$n$	$\hat{\ell}$	BIC	$BF^{-1}$	$n$	$\hat{\ell}$ (s.e.)	BIC	$BF^{-1}$
MN	7	-368.62	770.27	$3.1 \times 10^{-5}$	8	-326.45 (0.105)	690.65	$1.5 \times 10^{-1}$
MNH	8	-368.95	775.64	$1.1 \times 10^{-9}$	9	-327.52 (0.052)	697.51	$4.7 \times 10^{-3}$
MC	9	-358.32	759.11	$5.5 \times 10^{-3}$	10	-323.50 (0.037)	696.01	$2.5 \times 10^{-2}$
MCH	10	-359.06	765.30	$1.7 \times 10^{-4}$	11	-324.89 (0.041)	701.68	$5.9 \times 10^{-4}$
ME	9	-356.96	756.40	$1.6 \times 10^{-1}$	10	-319.81 (0.035)	687.38	1
MEH	10	-358.06	763.30	$6.3 \times 10^{-4}$	11	-320.64 (0.030)	693.18	$4.1 \times 10^{-2}$
MEC	11	-356.87	765.64	1	12	-320.17 (0.031)	696.96	$6.2 \times 10^{-3}$
MECH	12	-357.55	771.73	$2.4 \times 10^{-6}$	13	-320.38 (0.024)	702.09	$4.8 \times 10^{-4}$

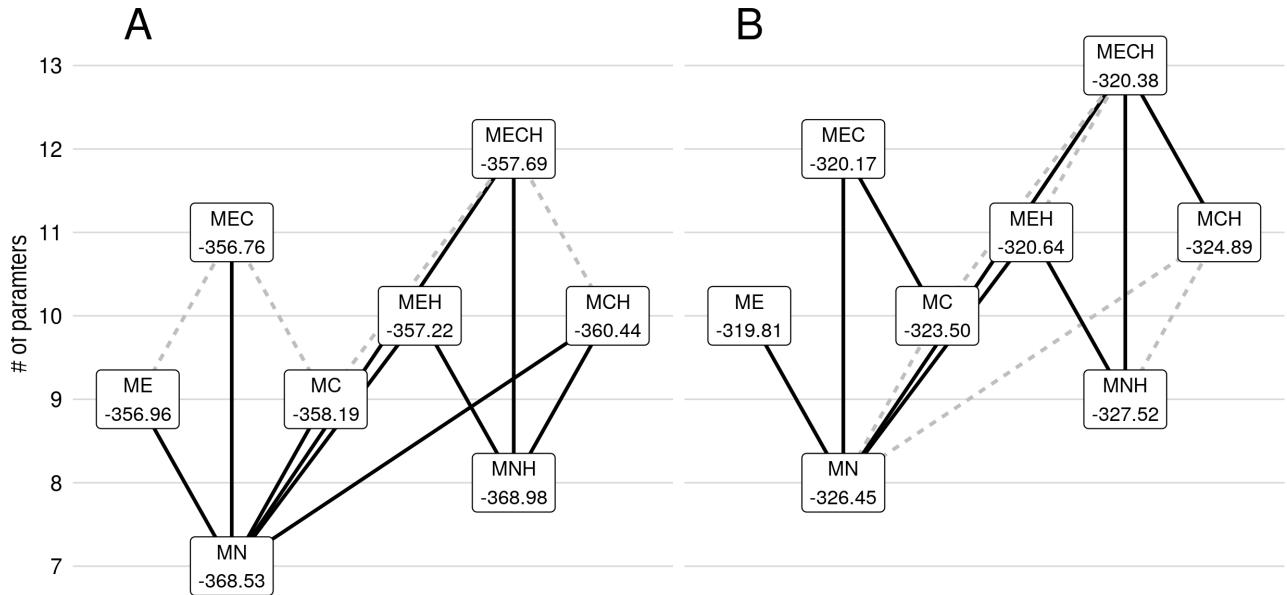
Interestingly, the inclusion of rainfall-induced contamination in model ME is rejected due to the very limited increase of the estimated log-likelihood of MEC, in contrast with the BFs favoring the latter. Model ME is thus the one retained by the LR-tests in the deterministic set of models. In the case of the stochastic models, the LR-tests also highlight the importance of the effect of rainfall on exposure rather than on contamination (fig. 2.5,B). In fact, the much stronger performance of MN in comparison with its deterministic counterpart relative to all other model structures imposes a stronger condition for retaining additional transmission processes. Indeed, both models MC and MCH were rejected when compared to MN, thus only models with rainfall-driven exposure were retained. As in the deterministic case, model ME is the one finally retained due to the lack of significance of the inclusion of additional transmission processes. Note that the conclusion based on the LR-tests for the deterministic models should be taken with caution because the MCMC algorithm used for calibration does not directly aim at maximizing the likelihood, but rather at sampling from the posterior distribution of the parameters given the data and the model. Moreover, the best likelihood visited by the chains when sampling the posteriors that are used in the LR-tests is not a formal estimate of the models' likelihood. However, the fact that the LR-tests applied to both model types agree with the selection of ME supports their use in both cases.

Both statistical methods for model comparison therefore agree about the importance of the effect of intra-seasonal rainfall on the exposure to transmission during the 2015 cholera epidemic in Juba. For the deterministic type of models, the BFs suggest stronger support for model MEC, and the LR-tests for ME, whereas for the stochastic models both the BIC-based estimates of BFs and the LR-tests favor ME.

#### INTRASEASONAL RAINFALL EVENTS AND THE 2015 JUBA EPIDEMIC

The comparison between the estimated output cases computed by the basic SIRB model (MN) and the most significant rainfall-based processes (MEC and ME for the deterministic and stochastic types, respectively) highlight the importance of rainfall in retrieving the second epidemiological peak (fig. 2.6). Both deterministic

tab. 2.2: Model comparison statistics. For each model is reported its number of parameters  $n$ , the associated estimated log-likelihood  $\hat{\ell}$  (and its Monte Carlo standard error for the stochastic model), and the inverse of the Bayes Factor ( $BF^{-1}$ ) with respect to the model with the largest evidence. The BFs for the deterministic models were computed directly from the parameters posteriors, whereas for the stochastic models they were estimated with the Bayesian Information Criterion (BIC) as  $BF_i \approx e^{\frac{1}{2}(BIC_i - BIC_{min})}$ . The BIC for the deterministic models was computed using the maximum log-likelihood value visited with the MCMC algorithm across chains.



and stochastic models fit well the general trend of the data, but they underestimate the large number of reported cases on July 19 (65 cases). Instead, the more complex models **ME** and **M**E**C** follow the SIRB dynamics and then are forced by the precipitation that occurred on July 18 (33 mm/d) toward the epidemiological peak.

Model calibration results suggest that precipitations with smaller intensities did not have a strong impact on cholera transmission during the 2015 epidemic in Juba. The exponents  $\alpha_c$  and  $\alpha_e$  were found to be systematically larger than 1. Thus, in the considered epidemic, the nonlinear function used to account for rainfall in the model, eqn. (2.11), helps to isolate the contribution of the largest rainfall.

The best measures of fit computed for the stochastic **ME** (see tab. 2.2) are thus explained by a larger sensitivity to precipitation, which causes the match between the mean of the simulated cases and the data during the second peak.

Comparing the two model types, stochastic results have a larger 95% confidence interval, which better encompasses most of the data. In particular, both epidemiological peaks are well captured by the stochastic models, while the deterministic results systematically underestimate them. Two factors contribute to this result: the intrinsic stochastic nature of the model, which requires the simulation of various model runs for the same set of parameters, and the noise that necessarily perturbs the force of the infection yielding an overdispersion in infections. The standard deviation of such (assumed white) noise is estimated in each stochastic model, and it is interesting to note that the MLE obtained for **ME** is slightly smaller than in **MN** (0.028 versus 0.022), highlighting again that the data are retrieved with a lower uncertainty when rainfall is included in the model. This is evident in fig. 2.6, where the width of the 95% confidence interval of models **ME** and **M**E**C** is smaller with respect to **MN**.

Finally, despite having different BFs, the deterministic models **ME** and **M**E**C** are qualitatively similar in terms of output response, indicating that the recorded changes in the log-likelihood function do not correspond to qualitative changes in the output.

fig. 2.5: Likelihood ratio tests of model nesting. The LL-tests were computed for each nested pair of models  $\{\mathcal{M}_0, \mathcal{M}_1\}$ , with parameter vectors  $\theta^0, \theta^1$ , for which at least one of the parameters that is not null is  $\theta^1$  is equal to 0 in  $\theta^0$ . Each model is labeled with its associated estimated maximum log-likelihood value,  $\hat{\ell}$ , for the deterministic (A) and the stochastic (B) models, and linked based on whether the likelihood ratio is significantly (full black lines) or not (dashed gray lines) at the 5% level. The absence of lines indicates a lower  $\hat{\ell}$  for the more complex model.

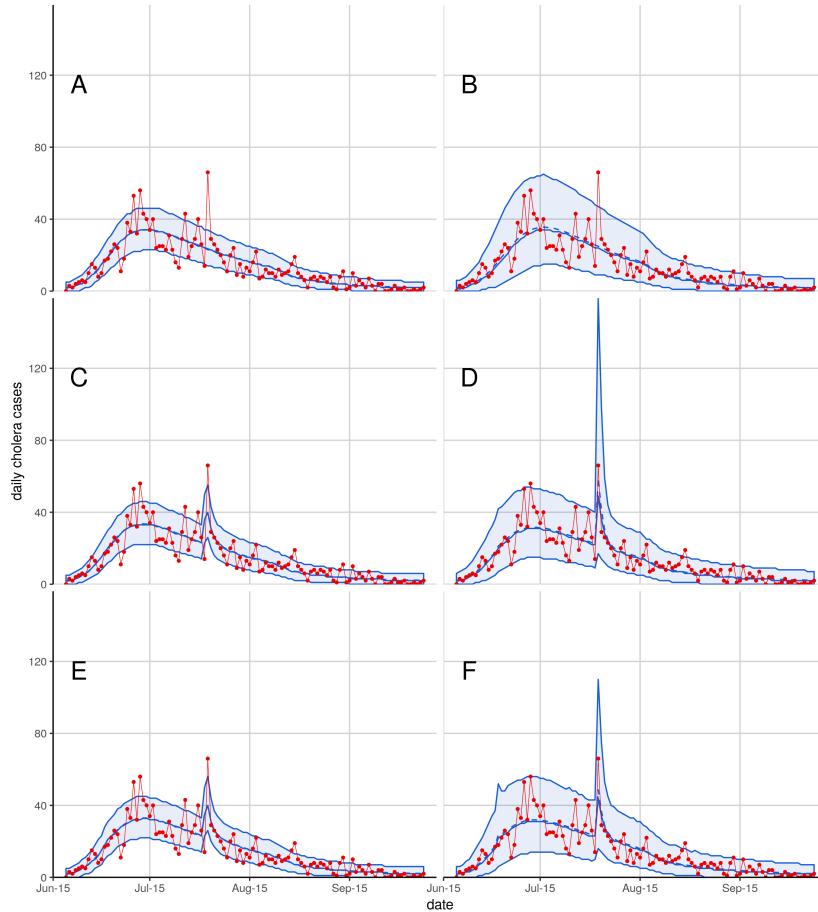


fig. 2.6: Simulations of the MN (A-B), ME (C-D) and MEC (E-F) models. Simulations for the deterministic versions (A,C,E) are given by the mean (blue dashed line), median (blue full line) and 95% simulation envelopes (blue ribbon) of 100 simulations of the measurement model for each trajectory from 100 samples from the posteriors of model parameters against reported daily cholera cases (red line and dots). Simulations from the stochastic models (B, D, F) are given for 10'000 simulations of the stochastic process and measurement models.

## DISCUSSION

In this chapter, a general mechanistic SIRB-based epidemiological model is developed to evaluate the relevance of rainfall in the amplification of cholera transmission, focusing on the 2015 Juba outbreak. Two rainfall-based transmission processes were compared: the direct increase of the exposure to the contaminated water (model ME)<sup>34</sup>, and the increase of water contamination by flooded open defecation sites (model MC)<sup>35</sup>. In addition, human-to-human transmission is considered (models' name with H).

Regarding the epidemiological model, this study introduced two innovations with respect to previous modeling attempts of cholera epidemics<sup>36</sup>. First, the focus on daily incidence data, as opposed to weekly epidemiological reports commonly used in cholera studies, motivated the introduction of a compartment of exposed individuals, eqn. (2.2), to account for the incubation period of the disease and, thus, the lag between the possibly rainfall-driven infection process and the manifestation of the symptoms resulting in the time-series of daily reported cases at our disposal. Second, a non-linear version of rainfall driver, in the form of a power-law controlled by a single parameter, was introduced to generalize the previous linear dependence. Such formulation has the flexibility to either emphasize the impact of the largest rainfall events or give equal weight to all non-zero rainfall intensities.

All model assumptions were compared for both deterministic and stochastic

<sup>34</sup> Eisenberg, Kujibida, et al., "Examining Rainfall and Cholera Dynamics in Haiti Using Statistical and Dynamic Modeling Approaches" in: *Epidemics* (2013).

<sup>35</sup> Rinaldo, Bertuzzo, Mari, et al., "Reassessment of the 2010–2011 Haiti Cholera Outbreak and Rainfall-Driven Multiseason Projections" in: *Proceedings of the National Academy of Sciences* (2012).

<sup>36</sup> Bertuzzo, Finger, et al., "On the Probability of Extinction of the Haiti Cholera Epidemic" in: *Stochastic Environmental Research and Risk Assessment* (2016); Pasetto, Finger, Rinaldo, et al., "Real-Time Projections of Cholera Outbreaks through Data Assimilation and Rainfall Forecasting" in: *Advances in Water Resources* (2017).

model types, to draw more general conclusions. The statistics and tests used to compare the model results (tab. 2.2 and fig. 2.5) supported the significance of rainfall effects during the 2015 epidemic in Juba. In fact, results showed that for both model types there exists a significant positive effect of including rainfall drivers, in particular because standard SIRB models were not able to reproduce the second epidemiological peak of reported cases that occurred in July during the recession period. All models considering rainfall, instead, showed an increase of the number of cases in correspondence of the second peak, which was due to the large rainfall rates that occurred during the previous days (fig. 2.6). This difference in the simulated responses of models considering (or not) rainfall lead to stronger support for rainfall-based models. Due to the small variations among the likelihoods of rainfall-based models, however, it is not straightforward to conclude the best way to include the rainfall effect (tab. 2.2). Models with the minimum BIC were those considering the increase in exposure (model ME) for both the stochastic and the deterministic model types. For the deterministic models, the computation of the Bayes Factors (BFs), which should provide a direct estimation of the model probability, suggested the selection of the model combining exposure and contamination processes (model MEC). However, this information criterion might be unstable due to numerical issues and oscillations in the MCMC used for calibration<sup>37</sup>. Because the models' outputs were similar for MEC and ME (fig. 2.6), it is advised to select the approach endowed with less parameters<sup>38</sup>, in this case ME, as indicated by the BIC. Note that the inclusion of human-to-human transmission was not statistically relevant in this modeling exercise.

The comparison between the likelihoods of the two models' types (deterministic and stochastic) showed that considering the stochasticity of the processes improves the model results (tab. 2.2). This suggests that also deterministic models should include a stochastic term in the computation of the force of infection, eqn. (2.10), which might increase the flexibility of the outputs.

Several limitations should be considered when analyzing the present results. The calibration exercise attempted in this study considered daily rainfall and cholera reported cases, which are characterized by significant random fluctuations that might partly cloud the description of the underlying infection processes. Small random delays in reporting could change the infection curve and thus the effect of rainfall. This issue was partially addressed by considering the exposed compartment, eqn. (2.2), for simulation of the incubation period, and unknown reporting rate  $\varepsilon$  for the observed cases.

Here, it is aimed to reproduce the epidemic by modeling epidemiological transmission processes. While non-linear rainfall effects and possible over-reporting are taken into account, human mobility effect<sup>39</sup> is not considered, which could help modeling the importation of infected individual. Moreover, in this model asymptomatic infected individuals do not contribute transmission. From a modeling viewpoint, these unaccounted processes were compensated by the calibration procedure, at the loss of predictive power.

The prior bounds to be assigned to parameters are typically wide<sup>40</sup> because the rates governing transmission processes are highly dependent on the specific epidemiological context so that somewhat contradictory values had been estimated in the literature. These considerations, together with the intrinsic noise affecting recorded cases, underlie the possibility that some of the model parameters might

<sup>37</sup> Raftery et al., "Estimating the Integrated Likelihood via Posterior Simulation Using the Harmonic Mean Identity" (2007).

<sup>38</sup> in an attempt to maximize the model predictive accuracy we want to reduce potential overfitting.

It is also possible that the stochastic model has benefitted from improved identifiability. It has been shown that stochastic models can often extract more information about parameters than deterministic counterparts, see: Browning et al., "Identifiability Analysis for Stochastic Differential Equation Models in Systems Biology" in: *Journal of The Royal Society Interface* (2020).

<sup>39</sup> Gatto, Mari, et al., "Generalized Reproduction Numbers and the Prediction of Patterns in Waterborne Disease" in: *Proceedings of the National Academy of Sciences* (2012); Bertuzzo, Casagrandi, et al., "On Spatially Explicit Models of Cholera Epidemics" in: *Journal of The Royal Society Interface* (2010); Mari, Bertuzzo, Finger, et al., "On the Predictive Ability of Mechanistic Models for the Haitian Cholera Epidemic" in: *Journal of The Royal Society Interface* (2015); Perez-Saez, King, et al., "Climate-Driven Endemic Cholera Is Modulated by Human Mobility in a Megacity" in: *Advances in Water Resources* (2017).

<sup>40</sup> Akman et al., "Examination of Models for Cholera: Insights into Model Comparison Methods" in: *Letters in Biomathematics* (2016).

be unidentifiable<sup>41</sup>, in the sense that different parameter combinations would yield the same model output (also called equifinality).

The lack of available data prevented us to include the effects of the overall efforts towards WaSH improvement in this study, but vaccination is implemented in order to eliminate one possible covariate of the rainfall effect. Despite these limitations, the proposed model comparison using both a deterministic and a stochastic model gave coherent results. The agreement of the two modeling types strengthened the results regarding the importance of rainfall patterns to significantly affect the development of cholera cases in time.

Overall, the findings of the study are consistent with the lessons learned in South Sudan with most of the transmission starting with the onset of the rainy season. In 2016 and 2017, cases in the dry season were observed and associated to the overexploitation of scarce water resources by nomadic herdsmen. This suggests that, as already observed, a general assessment of the relationship between precipitation and general waterborne or water-based disease infections is far from obvious and surely case-dependent. It has been argued, for example, that in the domain of water-based parasitic infections rainfall could not only boost disease transmission but also reduce it substantially<sup>42</sup>, e.g. by increasing water flow (which in turn decreases habitat suitability in water). Rainfall patterns may also drastically affect human activities related to water contacts, thus potentially altering exposure and transmission risk<sup>43</sup>. To that end, a hydrology-driven assessment cannot ignore certain characteristics, in particular the ephemeral or permanent nature of the waterways fostering contacts among pathogens and hosts<sup>44</sup>. Also, temporal fluctuations of rainfall patterns may be particularly important in determining the seasonality of transmission<sup>45</sup>.

<sup>41</sup> Eisenberg, Robertson, et al., “Identifiability and Estimation of Multiple Transmission Pathways in Cholera and Waterborne Disease” in: *Journal of Theoretical Biology* (2013).

<sup>42</sup> McCreesh and Booth, “Challenges in Predicting the Effects of Climate Change on Schistosoma Mansoni and Schistosoma Haematobium Transmission Potential.” in: *Trends in parasitology*. (2013).

<sup>43</sup> Lai, Biedermann, et al., “Spatial Distribution of Schistosomiasis and Treatment Needs in Sub-Saharan Africa: A Systematic Review and Geostatistical Analysis” in: *The Lancet. Infectious Diseases* (2015).

<sup>44</sup> Perez-Saez, Mande, et al., “Hydrology and Density Feedbacks Control the Ecology of Intermediate Hosts of Schistosomiasis across Habitats in Seasonal Climates” in: *Proceedings of the National Academy of Sciences* (2016).

<sup>45</sup> Bertuzzo, Mari, Righetto, et al., “Hydroclimatology of Dual-Peak Annual Cholera Incidence: Insights from a Spatially Explicit Model” in: *Geophysical Research Letters* (2012); Bertuzzo, Mari L., et al., “Prediction of the Spatial Evolution and Effects of Control Measures for the Unfolding Haiti Cholera Outbreak” in: *Geophysical Research Letters* (2011); McCreesh, Nikulin, et al., “Predicting the Effects of Climate Change on Schistosoma Mansoni Transmission in Eastern Africa” in: *Parasites & Vectors* (2015); Perez-Saez, Mande, et al., “Hydrology and Density Feedbacks Control the Ecology of Intermediate Hosts of Schistosomiasis across Habitats in Seasonal Climates” in: *Proceedings of the National Academy of Sciences* (2016).

# CHAPTER 3

## ESTIMATING THE PROBABILITY OF ELIMINATING CHOLERA FROM HAITI THROUGH A MASS VACCINATION CAMPAIGN

This chapter describes a model developed for a multi-modeling study. Four teams with existing experience in modeling cholera in Haiti were tasked to model scenarios of mass vaccination administration towards cholera immunity. The timing was meant to take advantage of the low incidence to affect the endemicity of cholera in Haiti to achieve elimination. The multi-model study was led by Elizabeth C. Lee and has been published as:

Elizabeth C Lee, Dennis L Chao, Joseph C Lemaitre, Laura Matrajt, Damiano Pasetto, Javier Perez-Saez, Flavio Finger, Andrea Rinaldo, Jonathan D Sugimoto, M Elizabeth Halloran, Ira M Longini, Ralph Ternier, Kenia Vissieres, Andrew S Azman, Justin Lessler, and Louise C Ivers. “Achieving Coordinated National Immunity and Cholera Elimination in Haiti through Vaccination: A Modelling Study”. In: *The Lancet Global Health* 8.8 (Aug. 1, 2020), e1081–e1089

For this collaborative effort, a hidden Markov model of cholera transmission is developed, with specific modeling choices to account for vaccination and to gather information on elimination timing and probability. The present chapter is adapted from Supplement 3 of the aforementioned publication, which focuses on the model designed by our team. The reader is referred to the publication by Lee et al. for the modeling exercise results and the comparison between models predictions.

**CHOLERA IN HAITI** In January 2010, an earthquake hit Haiti, disrupting healthcare and water infrastructure and displacing a million persons. Ten months later, Cholera (*V. Cholerae* of serogroup O1, serotype Ogawa and biotype El Tor) was introduced in Haiti by United Nation peacekeeper soldiers<sup>1</sup>. This introduction in a naive population caused a major outbreak, totaling 820'000 reported cases and 10'000 death, most of them occurring in the first two years<sup>2</sup>. Since 2016, cases decreased steadily and finally, the last cholera (culture) confirmed case occurred in early 2019<sup>3</sup>. In Haiti, cholera exhibited a seasonal pattern with two annual peaks and its burden was spread among every department (fig. 3.1).

### DESCRIPTION OF THE MODEL

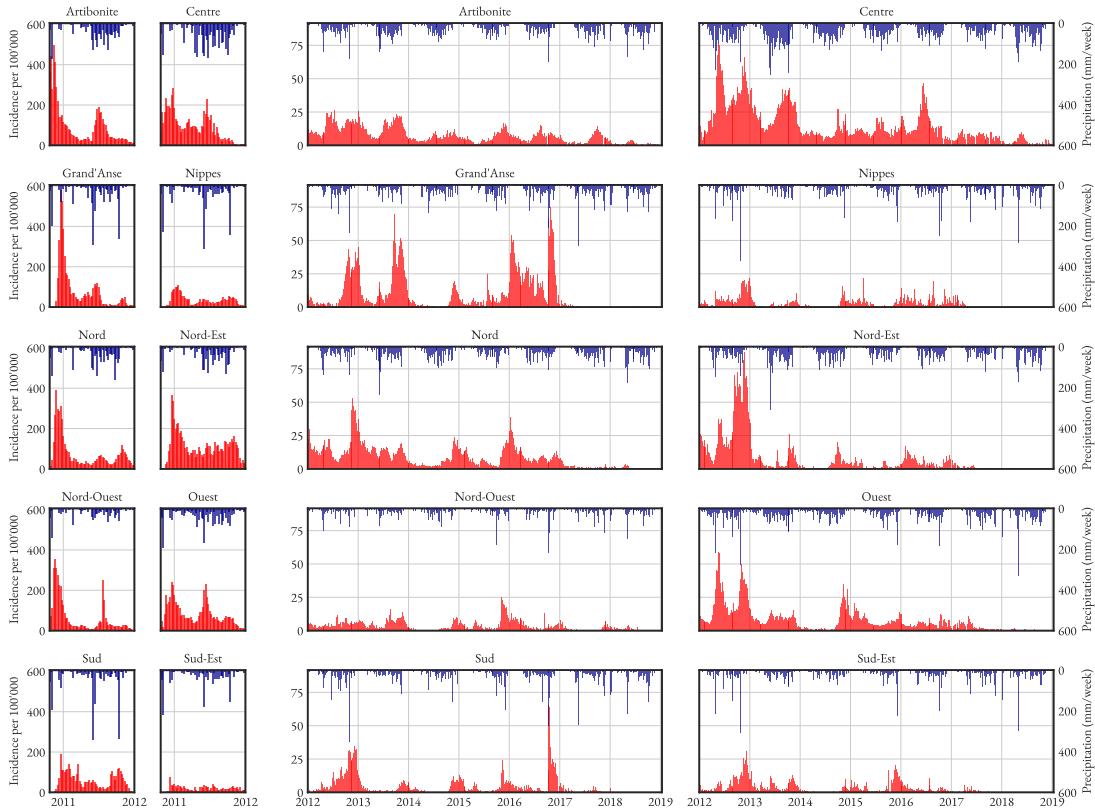
**GENERAL PRINCIPLES** The cholera model adopted to study the Haitian epidemic is a stochastic compartmental model applied at the level of the ten Haitian departments. The model is based on a Partially-Observed Markov Process (POMP), simulating the stochastic transitions between compartments as discrete events. It is the stochastic translation of a deterministic SIRB model based on ordinary differential equations which has been extensively used to simulate the Haitian cholera epidemic in previous studies<sup>4</sup> and is described in CHAPTER 1.

<sup>1</sup> Frerichs et al., “Nepalese Origin of Cholera Epidemic in Haiti” in: *Clinical Microbiology and Infection* (2012); Piarroux et al., “Understanding the Cholera Epidemic, Haiti” in: *Emerging Infectious Diseases* (2011).

<sup>2</sup> Barzilay et al., “Cholera Surveillance during the Haiti Epidemic — The First 2 Years” in: *New England Journal of Medicine* (2013).

<sup>3</sup> Mitchell et al., *PAHO/WHO | Haiti Reaches One-Year Free of Cholera* (2020).

<sup>4</sup> Rinaldo, Bertuzzo, Mari, et al., “Reassessment of the 2010–2011 Haiti Cholera Outbreak and Rainfall-Driven Multiseason Projections” in: *Proceedings of the National Academy of Sciences* (2012); Bertuzzo, Finger, et al., “On the Probability of Extinction of the Haiti Cholera Epidemic” in: *Stochastic Environmental Research and Risk Assessment* (2016); Pasetto, Finger, Camacho, et al., “Near Real-Time Forecasting for Cholera Decision Making in Haiti after Hurricane Matthew” in: *PLOS Computational Biology* (2018).



The model subdivides the population of each department<sup>5</sup> into compartments counting the number of individuals at the different stages of the disease. In addition to susceptible  $S$ , symptomatic infected  $I$ , and recovered individuals  $R$ , a compartment for infectious asymptomatic  $A$  is added for this exercise. As in previous chapters, an environmental compartment describing the bacterial concentration  $B$  in the local environment is used to estimate the force of infection. Precipitation is an important environmental driver of cholera transmission<sup>6</sup>, especially in Haiti<sup>7</sup>. As done in the contamination formulation in CHAPTER 2, rainfall increases the rate at which bacteria shed by infected individuals enter the environmental reservoir and thus increases the bacterial concentration and finally the force of infection. A diagram of the model is given in fig. 3.2.

## MODEL DYNAMICS

The following cholera dynamics are included in the model:

*Force of Infection and mobility* The force of infection in each department contains as an additional term the number of cholera cases in the rest of Haiti. This allows for a possible introduction of cholera due to human mobility between departments. The force of infection in each department is composed of two parts. The first is related to the local bacterial concentration of the department. The second is related to case importation from other departments through human-to-human transmission. The corresponding equation for the  $i^{th}$  department reads:

$$F_0^i(t) = \beta^i \frac{B_i(t)}{1 + B_i(t)} + c^i \sum_{j \neq i} (I_j(t) + A_j(t)).$$

fig. 3.1: Weekly incident cases and rainfall in the ten departments of Haiti from 2010 to 2019. Note that the first two years are shown with a different scale. The resurgence in 2016 is linked to hurricane Matthew (Pasetto, Finger, Camacho, et al. 2018; Khan et al. 2017). The seasonality pattern is evident especially in Artibonite.

<sup>5</sup> The case data was provided for the ten departments of Haiti, so the model spatial unit was chosen to be the department. Other teams decided either the same scale, coarser (national for Lee et al.) or finer (1km grid for Chao et al.).

<sup>6</sup> Camacho et al., “Cholera Epidemic in Yemen, 2016–18: An Analysis of Surveillance Data” in: *The Lancet Global Health* (2018).

<sup>7</sup> Rinaldo, Bertuzzo, Mari, et al. 2012.

The force of infection uses indirect (water-mediated) local transmission but direct (human-to-human) transmission for mobility exchanges. One of the reasons behind this discrepancy is technical: scalability issues linked to particle depletion hinder iterated filtering performance. Thus, it becomes very hard to calibrate a spatial model. To alleviate this issue, a custom procedure has been developed (see below) that was only tractable with a simple mobility link. Instead today, the choice would be to use the improved methods that have been developed for spatial inference on pomp models (Asfaw et al. 2021; Park and Ionides 2020).

The first term in the sum represents local transmission governed by the department-specific exposure parameter  $\beta^i$  which multiplies the logistic dose-response of the re-scaled local bacterial concentration  $B = B^*/K$ , where  $B^*$  is the unscaled concentration of vibrios and  $K$  the half-saturation constant of the logistic function  $\frac{B_i^*(t)}{K+B_i^*(t)}$ . Case importation from other departments is given by the sum of the asymptotically and symptomatically infected in department  $j$ , modulated by a parameter  $c^j$  which represents the intensity of case introduction from other departments in Haiti to department  $i$ .

*Rainfall* We account for rainfall through the contamination pathway described in the previous chapter, using the same non-linear relationship. Rainfall increases the amount of *Vibrios* per infected individual that contaminates the water reservoir. In Haiti, rainfall was empirically associated with resurgence of cholera infections in many research works<sup>8</sup>

*Symptoms* A proportion  $\sigma$  of infected individuals become symptomatic, and  $1 - \sigma$  remain (shedding, hence infectious) asymptomatic and transitions into compartment  $A$  (instead of  $R$  in the previous chapters, because past model formulation did not account for shedding asymptomatic).

*Shedding* Both symptomatic and asymptomatic infected individuals shed bacteria. The shedding rate of asymptomatics,  $\theta_A$ , is modeled as a fraction of the shedding rate of symptomatic individuals  $\theta_I$ <sup>9</sup>.

*Recovery rate* The recovery rate is the same for both asymptomatic and symptomatic individuals ( $\gamma_I = \gamma_A = 0.2 \text{ d}^{-1}$ )<sup>10</sup>.

*Acquired immunity* Individuals acquire natural immunity and remain in the recovered compartment ( $R$ ) for a period that lasts for  $1/\rho = 8$  years on average, before reintegrating the susceptible compartment<sup>11</sup>.

*Erlang-distributed immunity loss* To model extinction through coordinated immunity, it is needed to better approximate distribution that typically characterizes the duration of immunity<sup>12</sup>. Instead of a exponential distribution, recovered individuals pass through a succession of three separate recovered compartments ( $R_1, R_2, R_3$ ) characterized by the same transition rate  $\rho_1 = \rho_2 = \rho_3 = 3\rho$ . Hence the duration of immunity is Erlang-distributed.

*Bacterial dynamics* The size of the bacterial reservoir is proportional to the population density  $D_i$  of the department. This is another difference with respect to past models which used population. The goal is to better capture the risk posed by cholera in slums and dense urban areas such as Port-au-Prince. Bacteria die at rate  $\mu_B$  (with a mean persistence, calibrated, of less than three days). Rainfall influences the bacteria concentration by increasing the rate at which bacteria enter the environmental reservoir.

*Reporting process* The reported cases are modeled by a negative binomial distribution with dispersion parameter  $p$ . Over- or under-reporting is accounted for through the reporting parameter  $\varepsilon$ .

**STOCHASTICITY** Overdispersion in the infection process is introduced by multiplying the force of infection  $F_0$  by a time-continuous white noise process

<sup>8</sup> Gaudart et al. 2013; Bertuzzo, Finger, et al. 2016; Adams 2012; Periago et al. 2012; Adams 2013; Rinaldo, Bertuzzo, Mari, et al. 2012; Eisenberg, Kujbida, et al. 2013; Kirpich, Weppelmann, Yang, Ali, et al. 2015.

<sup>9</sup> Kühn et al., “Glucose- but Not Rice-Based Oral Rehydration Therapy Enhances the Production of Virulence Determinants in the Human Pathogen *Vibrio Cholerae*” in: *PLOS Neglected Tropical Diseases* (2014).

<sup>10</sup> Kaper et al., “Cholera” in: *Clinical Microbiology Reviews* (1995); Codeço, “Endemic and Epidemic Dynamics of Cholera: The Role of the Aquatic Reservoir” in: *BMC Infectious Diseases* (2001).

<sup>11</sup> This was surprisingly consistent across teams, with values ranging from 5 to 8 years.

<sup>12</sup> King, Ionides, et al., “Inapparent Infections and Cholera Dynamics” in: *Nature* (2008).

$\xi(t)$  defined as the differentiation of an integrated noise process  $\xi(t) = \frac{d}{dt}\Gamma(t)$ , here taken to be have a Gamma distribution with mean  $\Delta t$  and variance  $\sigma^2\Delta t^{13}$ :

$$\xi(t) = \Gamma(t + \Delta t) - \Gamma(t) \sim \text{Gamma} \left( \frac{\Delta t}{\sigma^2}, \sigma^2 \right). \quad (3.1)$$

Since  $\xi(t)$  is non-negative it can serve as a multiplicative noise on the force of infection:

$$F_i(t) = F_0^i(t)\xi(t), \quad (3.2)$$

which yields to over-dispersion in the transitions.

<sup>13</sup> Bretó and Ionides, “Compound Markov Counting Processes and Their Applications to Modeling Infinitesimally Over-Dispersed Systems” in: *Stochastic Processes and their Applications* (2011).

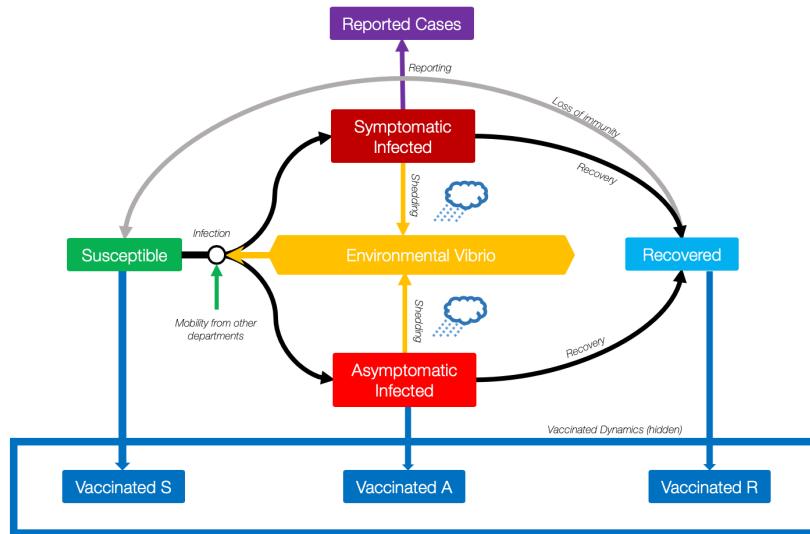


fig. 3.2: Schematic diagram of the cholera transmission processes in a single department. Dynamics of vaccinated compartments are not shown.

**VACCINATION DYNAMICS** The estimated vaccine efficacy was waning from 76% over 60 months for adults, and conservatively assumed to not protect after 5 years. These assumptions were shared across modeling teams. At each vaccination campaign, the available vaccine doses are uniformly distributed among susceptible  $S$ , asymptomatic infected  $A$  and recovered  $R_1, R_2, R_3$  individuals. The rate of vaccination is denoted with  $r_V$ . Individuals can receive either one or two doses of OCV, which yield efficacies of  $\eta_{1d}(t)$  and  $\eta_{2d}(t)$ , respectively. These parameters are common across modeling teams and the reader is referred to the manuscript for their time-varying values. There is no age structure but the efficacy is set to be the population-weighted average of estimated efficacy for those under 5 years old and those over 5 years old. The model considers ten additional compartments for each vaccination campaign, in order to distinguish among individuals who received one (compartments  $V_{1d}^S, V_{1d}^A, V_{1d}^{R_k}, k = 1, 2, 3$ ) or two (compartments  $V_{2d}^S, V_{2d}^A, V_{2d}^{R_k}, k = 1, 2, 3$ ) doses of OCV. Vaccinated susceptible individuals ( $V_{1d}^S$  and  $V_{2d}^S$ ) have a lower probability to become infected (and thus entering classes  $I$  or  $A$ ) than non-vaccinated susceptibles. This is modeled through the multiplicative reduction of the force of infection by a factor  $(1 - \eta_{1d}(t))$  or  $(1 - \eta_{2d}(t))$  respectively. The vaccination campaign window is split equally between departments (i.e for a vaccination campaign of 5 years duration, each department will be vaccinated during a 6 month period). Vaccine efficacy starts waning after the first

half of the duration of the department's vaccination campaign. For example, if for department  $i$  the vaccination campaign  $j$  spans from  $t_a^{ij}$  to  $t_b^{ij}$ , then:

$$\eta^{ij}(t) = \begin{cases} \eta_0(0) & \text{if } t < t_a^{ij} + \frac{t_b^{ij} - t_a^{ij}}{2} \\ \eta_0(t - (t_a^{ij} + \frac{t_b^{ij} - t_a^{ij}}{2})) & \text{if } t > t_a^{ij} + \frac{t_b^{ij} - t_a^{ij}}{2}, \end{cases} \quad (3.3)$$

where  $\eta_0(t)$  is the scenario dependent vaccine efficacy as defined in the meta-supplement. The rates at which individuals leave compartments  $V^A$  and  $V^{R_k}$  ( $k = 1, 2, 3$ ) are equivalent to  $A$  and  $R_i$ . Individuals enter the compartment  $V^S$  with a vaccine efficacy reduced according to the amount of time they spent in  $V^A$  and  $V^{R_k}$ . The actual deployment of the vaccine doses is shown in fig. 3.4.

**OTHER INTERVENTIONS** WaSH and other intervention efforts are not explicitly considered in the model, but their impact is implicitly taken into account by calibrating the exposure rates  $\beta^i$  to disease incidence that occurred while interventions were taking place.  $\beta^i$  is modeled to be constant in time, meaning that changes in number or type of interventions or population behavior over time are not taken into account.

## MODEL EQUATIONS

The model is implemented as a stochastic counting process<sup>14</sup>. Let  $N_{AB}(t)$  be the number of individuals transiting between compartments  $A, B \in \mathcal{X}$  in the time interval  $[0, t]$  where  $\mathcal{X}$  is the state vector,

$$\mathcal{X} = \{S, I, A, R_k, V_{j,1d}^S, V_{j,1d}^A, V_{j,1d}^{R_k}, V_{j,2d}^S, V_{j,2d}^A, V_{j,2d}^{R_k}\} \text{ for } j = 1, \dots, J, \text{ and } k = 1, 2, 3,$$

and  $J$  is the number of vaccination campaigns in the department. The number of transitions during a time-step  $\Delta t$  is  $\Delta N_{AB}(t) = N_{AB}(t + \Delta t) - N_{AB}(t)$ . Given the state of the system at time  $t$ ,  $\mathcal{X}_t$ , and a force of infection  $F_j(t)$  the transition rates read (transitions are written only for one dose of OCV and one vaccination campaign to shorten the equations):

$$\mathbb{P} [\Delta N_{SI}(t) = 1 | \mathcal{X}_t] = \sigma F_j(t) S(t) \Delta t + o(\Delta t) \quad (3.4)$$

$$\mathbb{P} [\Delta N_{SA}(t) = 1 | \mathcal{X}_t] = (1 - \sigma) F_j(t) S(t) \Delta t + o(\Delta t) \quad (3.5)$$

$$\mathbb{P} [\Delta N_{SV_{1d}^S}(t) = 1 | \mathcal{X}_t] = r_{V_{1d}}(t) S(t) \Delta t + o(\Delta t) \quad (3.6)$$

$$\mathbb{P} [\Delta N_{S\bullet}(t) = 1 | \mathcal{X}_t] = \mu S(t) \Delta t + o(\Delta t) \quad (3.7)$$

$$\mathbb{P} [\Delta N_{IR_1}(t) = 1 | \mathcal{X}_t] = \gamma I(t) \Delta t + o(\Delta t) \quad (3.8)$$

$$\mathbb{P} [\Delta N_{I\bullet}(t) = 1 | \mathcal{X}_t] = (\mu + \alpha) I(t) \Delta t + o(\Delta t) \quad (3.9)$$

$$\mathbb{P} [\Delta N_{AR_1}(t) = 1 | \mathcal{X}_t] = \gamma A(t) \Delta t + o(\Delta t) \quad (3.10)$$

$$\mathbb{P} [\Delta N_{AV_{1d}^A}(t) = 1 | \mathcal{X}_t] = r_{V_{1d}}(t) A(t) \Delta t + o(\Delta t) \quad (3.11)$$

$$\mathbb{P} [\Delta N_{A\bullet}(t) = 1 | \mathcal{X}_t] = \mu A(t) \Delta t + o(\Delta t) \quad (3.12)$$

$$\mathbb{P} [\Delta N_{R_k R_{k+1}}(t) = 1 | \mathcal{X}_t] = 3\rho R_k(t) \Delta t + o(\Delta t), \quad k = 1, 2 \quad (3.13)$$

$$\mathbb{P} [\Delta N_{R_3 S}(t) = 1 | \mathcal{X}_t] = 3\rho R_3(t) \Delta t + o(\Delta t) \quad (3.14)$$

$$\mathbb{P} [\Delta N_{R_k V_{1d}^{R_k}}(t) = 1 | \mathcal{X}_t] = r_{V_{1d}}(t) R_k(t) \Delta t + o(\Delta t), \quad k = 1, 2, 3 \quad (3.15)$$

$$\mathbb{P} [\Delta N_{R_k \bullet}(t) = 1 | \mathcal{X}_t] = \mu R_k(t) \Delta t + o(\Delta t), \quad k = 1, 2, 3 \quad (3.16)$$

<sup>14</sup> Bretó, He, et al., "Time Series Analysis via Mechanistic Models" in: *The Annals of Applied Statistics* (2009).

$$\mathbb{P} \left[ \Delta N_{V_{1d}^S I}(t) = 1 \mid \mathcal{X}_t \right] = \sigma(1 - \eta_{1d}^{ij}(t)) F_j(t) V_{1d}^S(t) \Delta t + o(\Delta t) \quad (3.17)$$

$$\mathbb{P} \left[ \Delta N_{V_{1d}^A}(t) = 1 \mid \mathcal{X}_t \right] = (1 - \sigma)(1 - \eta_{1d}^{ij}(t)) F_j(t) V_{1d}^S(t) \Delta t + o(\Delta t) \quad (3.18)$$

$$\mathbb{P} \left[ \Delta N_{V_{1d}^{\bullet}}(t) = 1 \mid \mathcal{X}_t \right] = \mu V_{1d}^S(t) \Delta t + o(\Delta t) \quad (3.19)$$

$$\mathbb{P} \left[ \Delta N_{V_{1d}^A V_{1d}^{R_1}}(t) = 1 \mid \mathcal{X}_t \right] = \gamma V_{1d}^A(t) \Delta t + o(\Delta t) \quad (3.20)$$

$$\mathbb{P} \left[ \Delta N_{V_{1d}^A \bullet}(t) = 1 \mid \mathcal{X}_t \right] = \mu V_{1d}^A(t) \Delta t + o(\Delta t) \quad (3.21)$$

$$\mathbb{P} \left[ \Delta N_{V_{1d}^{R_k} V_{k+1}^{R_k}}(t) = 1 \mid \mathcal{X}_t \right] = 3\rho V^{R_k}(t) \Delta t + o(\Delta t), \quad k = 1, 2 \quad (3.22)$$

$$\mathbb{P} \left[ \Delta N_{V_{1d}^{R_3} V_{1d}^S}(t) = 1 \mid \mathcal{X}_t \right] = 3\rho V_{1d}^{R_3}(t) \Delta t + o(\Delta t) \quad (3.23)$$

$$\mathbb{P} \left[ \Delta N_{V_{1d}^{R_k} \bullet}(t) = 1 \mid \mathcal{X}_t \right] = \mu V^{R_k}(t) \Delta t + o(\Delta t), \quad k = 1, 2, 3, \quad (3.24)$$

assuming that  $\mathbb{P}[\Delta N_{XY} > 1 \mid \mathcal{X}_t] = o(\Delta t) \forall X, Y \in \mathcal{X}$  and  $\mathbb{P}[\Delta N_{X\bullet} > 1 \mid \mathcal{X}_t] = o(\Delta t) \forall X \in \mathcal{X}$ . Note that  $\mathbb{P}[\Delta N_{X\bullet}(t) = 1 \mid \mathcal{X}_t]$  denotes probability that individuals die and it is governed by the same parameter  $\mu$  for all compartments except  $I$ .

The ensuing stochastic variations of the state variables are:

$$\Delta I(t) = \Delta N_{SI}(t) - \Delta N_{IR_1}(t) - \Delta N_{I\bullet}(t) \quad (3.25)$$

$$\Delta A(t) = \Delta N_{SA}(t) - \Delta N_{AR_1}(t) - \Delta N_{AV^A}(t) - \Delta N_{A\bullet}(t) \quad (3.26)$$

$$\Delta R_1(t) = \Delta N_{IR_1}(t) + \Delta N_{AR_1}(t) - \Delta N_{R_1 R_2}(t) - \Delta N_{R_1 V^{R_1}}(t) - \Delta N_{R_1 \bullet}(t) \quad (3.27)$$

$$\Delta R_2(t) = \Delta N_{R_1 R_2}(t) - \Delta N_{R_2 R_3}(t) - \Delta N_{R_2 V^{R_2}}(t) - \Delta N_{R_2 \bullet}(t) \quad (3.28)$$

$$\Delta R_3(t) = \Delta N_{R_2 R_3}(t) - \Delta N_{R_3 S}(t) - \Delta N_{R_3 V^{R_3}}(t) - \Delta N_{R_3 \bullet}(t) \quad (3.29)$$

$$\Delta V^S(t) = \Delta N_{SV^S}(t) - \Delta N_{V^S I}(t) - \Delta N_{V^S A}(t) - \Delta N_{V^S \bullet}(t) \quad (3.30)$$

$$\Delta V^A(t) = \Delta N_{AV^A}(t) - \Delta N_{V^A V^{R_1}}(t) - \Delta N_{V^A \bullet}(t) \quad (3.31)$$

$$\Delta V^{R_1}(t) = \Delta N_{R_1 V^{R_1}}(t) + \Delta N_{V^A V^{R_1}}(t) - \Delta N_{V^{R_1} V^{R_2}}(t) - \Delta N_{V^{R_1} \bullet}(t) \quad (3.32)$$

$$\Delta V^{R_2}(t) = \Delta N_{R_2 V^{R_2}}(t) + \Delta N_{V^{R_1} V^{R_2}}(t) - \Delta N_{V^{R_2} V^{R_3}}(t) - \Delta N_{R_2 \bullet}(t) \quad (3.33)$$

$$\Delta V^{R_3}(t) = \Delta N_{R_3 V^{R_3}}(t) + \Delta N_{V^{R_2} V^{R_3}}(t) - \Delta N_{V^{R_3} V^S}(t) - \Delta N_{V^{R_3} \bullet}(t) \quad (3.34)$$

$$S(t) = H_i - \sum_{X \in \mathcal{X} \setminus \{S\}} X(t), \quad (3.35)$$

where the equation for  $S(t)$  enforces a constant total population. The rescaled bacterial concentration  $B$  is necessary to estimate the force of infection and is computed using the following differential equation:

$$\frac{dB}{dt} = -\mu_B B + (1 + \lambda (J(t))^r) D_i [\theta_I I + \theta_A A], \quad (3.36)$$

where  $J(t)$  is the precipitation over time and  $D_i$  is the average population density of the department<sup>15</sup>. Parameter  $\mu_B$  expresses the mortality rate of the bacteria in the environment,  $\theta_I$  and  $\theta_A$  are the shedding rates of symptomatic and asymptomatic infected individuals, and  $\lambda$  and  $r$  are the parameters of the power-law that controls the non-linear impact of precipitation, as in CHAPTER 2. The model was simulated with a constant time-step of 4.8h, and the differential equation for the bacterial concentration was integrated using a Runge-Kutta 4 scheme.

Let  $C(t_j)$  denote the number of people that develop symptoms and seek healthcare during the observation interval  $[t_j, t_{j+1})$  (i.e. the true incidence). Thus:

$$C(t_j) = [N_{SI}(t_{j+1}) - N_{SI}(t_j)] + [N_{V^S I}(t_{j+1}) - N_{V^S I}(t_j)]. \quad (3.37)$$

<sup>15</sup> total department population over department area. It is assumed that density is more important than population size, a change from the historical model described in CHAPTER 1

Our Markov process formulation requires a measurement model linking the time series of the reported incidence to  $C(t_j)$  as the observed variable to the process model in eqn. (3.4)–(3.35). A negative-binomial measurement model accounting for over- or under-reporting of cholera incidence is chosen, i.e.

$$\text{cases}(t_j) \sim \text{NB}(\varepsilon(t)C(t_j), p).$$

where  $\varepsilon(t) > 0$  represents the proportion of cases reported. To account for the change of the case definition that occurred on January 1, 2018, the reporting rate changes over time:

$$\varepsilon(t) = \begin{cases} \varepsilon_1 & \text{if } t < \text{Jan 1st, 2018} \\ \varepsilon_2 & \text{otherwise.} \end{cases} \quad (3.38)$$

The parameters of the model are shown in tab. 3.1.

Parameter	Calibration	Value or bound	Unit	description
$\beta^i$ ( $\times 10$ dept.)	yes	$[0, \infty]$	–	Exposure
$c^i$ ( $\times 10$ dept.)	yes	$[0, \infty]$	–	Force of infection in dept. $i$ from cases in other depts.
$\varepsilon_1$	yes	$[0, 2]$	–	reporting fraction before January 1, 2018
$\varepsilon_2$	yes	$[0, 2]$	–	reporting fraction after January 1, 2018
$\sigma_w$	yes	$[0, 0.1]$	–	std-dev of the perturbation of $F(t)$
$p$	yes	$[0, \infty]$	–	dispersion parameter of reporting
$\theta_I$	yes	$[0, \infty]$	–	Shedding sympt.
$\theta_A$	yes	$[0, \theta_I]$	–	Shedding asympt.
$\mu_B$	yes	$[0, \infty]$	$d^{-1}$	Bacterial mortality in environment
$r$	yes	$[0, \infty]$	–	Exponent rainfall
$\lambda$	yes	$[0, \infty]$	–	Coef. rainfall
$\rho$	no	$1/(8 \cdot 365)$	$d^{-1}$	Loss of immunity (Levine et al., 1981 and Koelle et al, 2004)
$\sigma$	no	0.25	–	Symptomatic/exposed
$\alpha$	no	0.004	$d^{-1}$	Mortality due to cholera (Bertuzzo et al., 2008)
$\gamma$	no	1/5	$d^{-1}$	Recovery rate sympt. (Kaper, 1995)
$k$	no	3	–	number of recovered compartments
$\mu$	no	$1/(63.6 \cdot 365)$	$d^{-1}$	mortality rate from life expectancy (World bank, 2017)
$\eta_{1d}(t), \eta_{2d}(t)$	no	as in spec.	–	Vaccine efficacy for 1 and 2 doses

tab. 3.1: Model parameters

## CALIBRATION OF THE MODEL

**DATA** Remote-sensed precipitation estimates from NASA's TRMM and GPM missions are used<sup>16</sup>. Rainfall measurements are provided on a regular grid. To get a value for each department, the measurements are averaged over the extent of the department. The time series of future rainfall up to the year 2030 is constructed by sampling from the past 20 years of data with replacement blocks of 15 days. To keep the correct seasonality the day of the year of each block is preserved.

<sup>16</sup> TRMM 3B42 RT Derived Daily Product (Huffman et al. 2007) is used from October 2010 to March 2015, and GPM from April 2015 to December 2019.

**FITTING PROCEDURE** The model is calibrated separately for each department on the weekly reported cases from March 1, 2014 to January 1, 2019. The calibration procedure is based on a frequentist multiple iterated filtering algorithm (MIF2)<sup>17</sup>. The initial conditions on March 1, 2014 are derived by enforcing the model dynamics on the reported cases from the start of the epidemic in 2010.

<sup>17</sup> Ionides et al., "Inference for Dynamic and Latent Variable Models via Iterated, Perturbed Bayes Maps" in: *Proceedings of the National Academy of Sciences* (2015).

The MIF<sub>2</sub> algorithm performance deteriorates quickly with the spatial dimension of the model as the number of particles needed for calibration increases exponentially<sup>18,19</sup>. To address this problem each department is first calibrated independently. In a second step, using the departmental calibration as a starting point, the entire spatial model is calibrated. The department-specific calibration procedure is as follow:

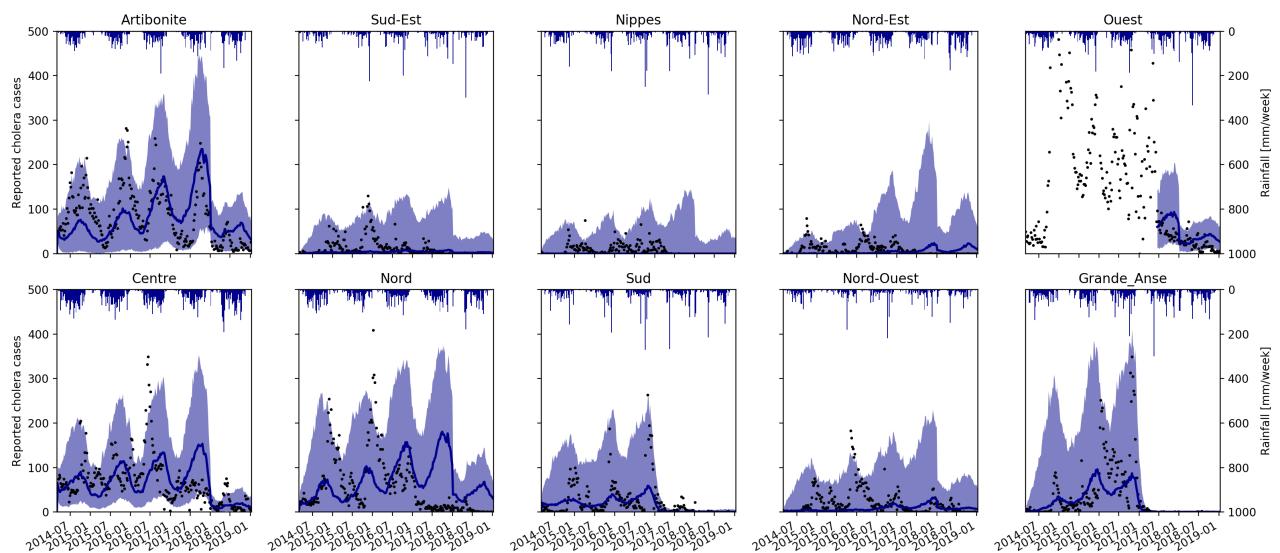
1. All unknown parameters are calibrated on the reported cases of Artibonite, where the epidemic had a clear seasonal dynamic from 2014 to 2018 with a sufficiently large number of cases, thus providing a good signal for the model. This allows to calibrate the unknown epidemiological and rainfall-related parameters on the most informative time series available.
2. For the other nine departments, the most sensitive parameters are calibrated: the exposure  $\beta^i$  and the mobility parameter  $c^i$  only, while fixing the remaining parameters to their best fit found for Artibonite.
3. The large, rainfall unrelated cholera outbreak in the Ouest department in 2015–2016 (mainly Port-au-Prince)<sup>20</sup> is excluded from the calibration since it is not part of the endemic dynamics that are the focus of this study.

During this phase, the mobility coefficients  $c_i$  are calibrated using the reported cholera cases (data) from the other departments (appropriately scaled with the reporting rate and symptomatic fraction). After visual convergence is reached

<sup>18</sup> Park and Ionides, *A Guided Intermediate Resampling Particle Filter for Inference on High Dimensional Systems* (2017).

<sup>19</sup> As mentioned, new methods have addressed these limitations (Asfaw et al. 2021; Park and Ionides 2020)

<sup>20</sup> unpublished field investigations speak of the notable contribution of the vandalism of main water pipes by gangs (Rebaudet, Bult, Gaudart, Michel, Gazin, Evers, Beaulieu, Abedi, Osei, Barrais, Pierre, Moore, Boncy, Adrien, Beigbeder, et al. 2018).



in each department, the departmental best fits are taken as starting points for a country-wide calibration. This mainly affects the mobility parameter  $c_i$ , which governs the departmental interdependence, as it is now calibrated on the actual simulated incidence from other departments. The resulting model fit is shown in fig. 3.3.

fig. 3.3: Visual assessment of model fit for the selected parameters set. The median (blue line) and the q025 and q975 quantiles (shaded area) over 1000 realization of the stochastic model are shown. Weekly reported cholera cases are shown as black dots.

## RESULTS AND DISCUSSION

The common outputs for the exercise are the distribution of time to elimination and the probability of elimination in different scenarios. The results are presented below for the six following scenarios, also shared across modeling teams:

- **No vaccination** status quo.
- **NATIONAL OVER 2 YEARS** a nation-wide OCV campaign over 2 years, to reach a coverage of 70% of the population with two doses, 10% with one dose, leaving 20% unvaccinated.
- **NATIONAL OVER 5 YEARS** a nation-wide OCV campaign reaching the same coverage, but with a slower deployment over 5 years.
- **2 DEPARTMENTS OVER 2 YEARS** Campaign focusing only on Centre and Artibonite, the two departments with the highest recent incidence of cholera. Vaccines are deployed to reach the same coverage as above, over 2 years.
- **3 DEPARTMENTS OVER 2 YEARS** Same as above, but with the addition of Ouest department.
- **NATIONAL OVER 2 YEARS, HIGH COVERAGE** same as the National over 2 years, but reaching a much higher coverage of 95% two doses, 1.67% one dose, and 3.33% unvaccinated individuals.

These scenarios are summarized in fig. 3.4 in terms of the cumulative number of doses deployed.

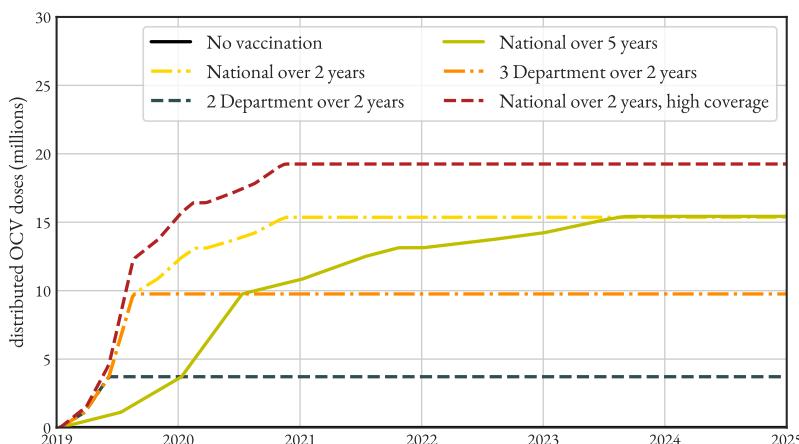
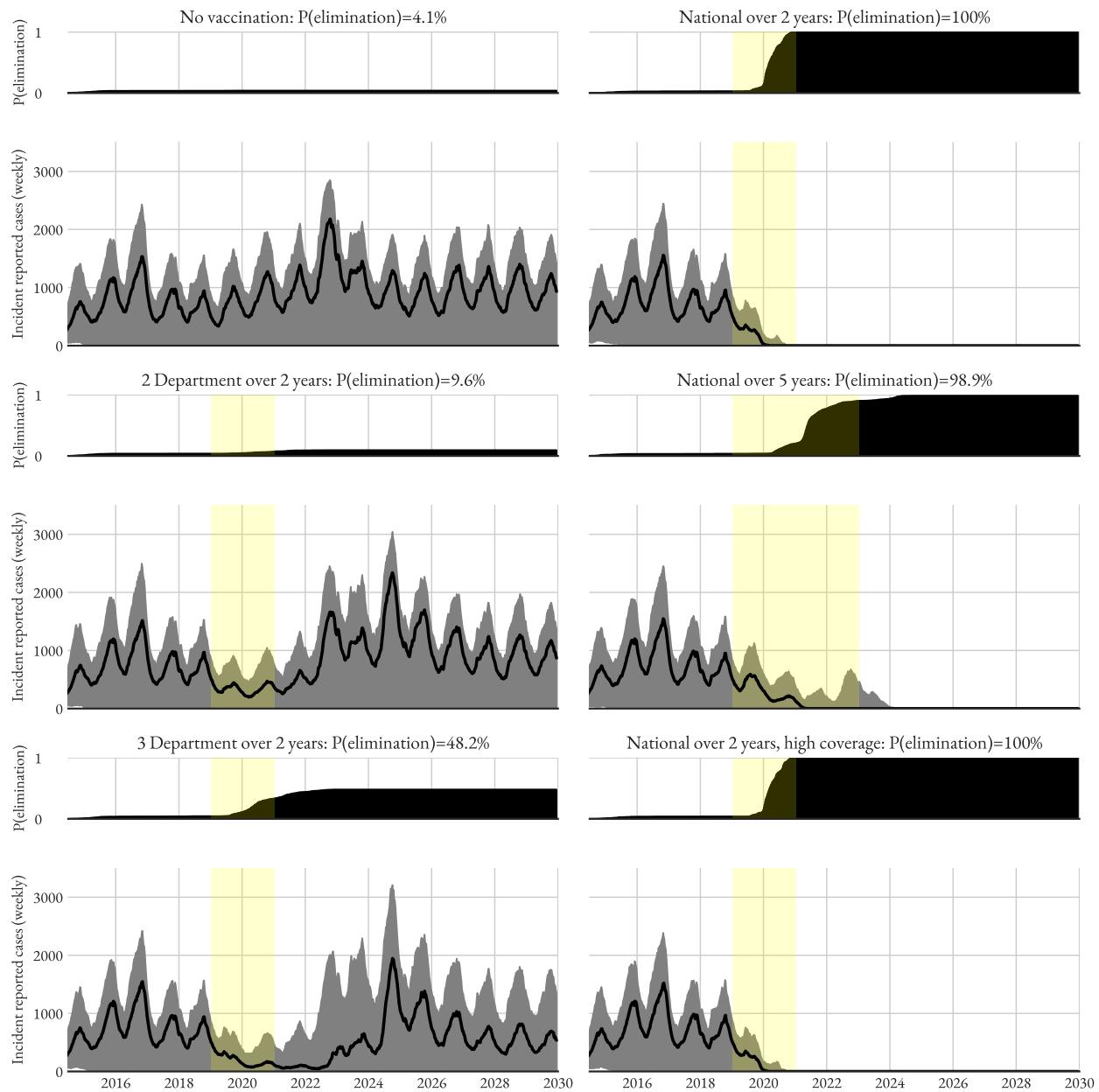


fig. 3.4: Rate of vaccine doses deployment in each considered scenarios.

The results in terms of probability of elimination and incidence are shown in fig. 3.5. Only a limited long-term impact of the **2 DEPARTMENTS** scenarios is projected. However, all **NATIONAL** scenarios considered were projected to lead to elimination, despite the limited vaccine efficacy and coverage. The **2 DEPARTMENTS** and **3 DEPARTMENTS** scenarios exhibit a probability of elimination of about 9.6% and 48% respectively (compared to 4.1% in the status quo scenario). The important impact of adding Ouest to the campaign is due to the department's population, around 4M, which represents about 40% of the Haitian population. While a slower timing decreases slightly the probability of elimination,

Only results from our model are presented here. Due to the difficulty of the exercise, the results across teams were quite different, and the reader is referred to Lee et al. (2020) and its supplement to compare the results with the other modeling teams.

it is noted that coverage is far more important than speed for these vaccination campaigns.



Haiti has seen no laboratory-confirmed cases of cholera since February 2019. This has sparked claims of cholera elimination, which would mean that the Americas would be free from Cholera. However, in the model results, the probability of elimination in the no vaccination scenario is very low (4.1%). It is always very difficult to predict or explain disease transmissions, and it is even harder for elimination. Possible explanations for the discrepancy observed between what is projected by the model and what happened are (i) issues in the model design, especially concerning the mobility acting as a constant additive pressure for the introduction of

fig. 3.5: Modeling results for the considered mass vaccination campaign scenarios, with the weekly reported cases (median and 95% confidence interval), and cumulative probability of elimination over time. The timing of the vaccine distribution in each scenario is highlighted in yellow.

cholera cases, hence lowering the probability of extinction; this was done purposefully to sustain transmission as it improved model fit and it was not believed Haiti to be close to elimination at that time, (ii) transmission parameters values error might have biased the estimate of the susceptible population on March 1, 2014, as the absence of serosurveys makes the correct estimation of this quantity critical, (iii) not giving enough weight to the decrease in reported cases from 2017<sup>21</sup>, (iv) WaSH interventions and rapid response teams that have been put in place in Haiti and whose activities have intensified in the last few years<sup>22</sup> might have played a role in disease elimination.

Further work must be performed to study the reasons behind the surprising (to us and other modeling groups, one year before) elimination of cholera from Haiti, and to assess all factors, including WaSH interventions, environmental drivers, and socio-economic changes in a unified modeling framework. The effort to investigate this (using the same methods presented for COVID-19 in CHAPTER 5) have been started, but the project has been disrupted owing to COVID-19 pandemic, which is the focus of the three remaining chapters.

<sup>21</sup> a 90% decrease of reporting is necessary for the model to replicate the decline in cases. However, it is highly unlikely that the observed decrease in cases is solely due to changes in the reporting process.

<sup>22</sup> Rebaudet, Bulit, Gaudart, Michel, Gazin, Evers, Beaulieu, Abedi, Osei, Barrais, Pierre, Moore, Boncy, Adrien, Guillaume, et al., “The Case-Area Targeted Rapid Response Strategy to Control Cholera in Haiti: A Four-Year Implementation Study” in: *PLOS Neglected Tropical Diseases* (2019).



# CHAPTER 4

## A SCENARIO MODELING PIPELINE FOR COVID-19 EMERGENCY PLANNING

Coronavirus disease 2019 (COVID-19) has caused strain on health systems worldwide due to its high mortality rate and the large portion of cases requiring critical care and mechanical ventilation. During these uncertain times, public health decision-makers, from city health departments to federal agencies, sought the use of epidemiological models for decision support in allocating resources, developing non-pharmaceutical interventions, and characterizing the dynamics of COVID-19 in their jurisdictions. In response, a flexible scenario modeling pipeline was developed, that could quickly tailor models for decision-makers seeking to compare projections of epidemic trajectories and healthcare impacts from multiple intervention scenarios in different locations. Here, the components and configurable features of the COVID Scenario Pipeline are presented. Model limitations and active areas of development to meet ever-changing decision-maker needs are also presented.

This chapter presents a modeling pipeline tailored for the response to COVID-19. The COVID Scenario Pipeline is an ongoing project, still actively developed, and many improvements have been added since the version presented in this thesis, where a perspective from July 2020 is taken. The state of knowledge on the SARS-CoV-2 and COVID-19 was rapidly expanding, and the present chapter highlights the challenges of dealing with such uncertainties. This chapter is based on: Joseph C. Lemaitre, Kyra H. Grantz, Joshua Kaminsky, Hannah R. Meredith, Shaun A. Truelove, Stephen A. Lauer, Lindsay T. Keegan, Sam Shah, Josh Wills, Kathryn Kaminsky, Javier Perez-Saez, Justin Lessler, and Elizabeth C. Lee. “A Scenario Modeling Pipeline for COVID-19 Emergency Planning”. In: *Scientific Reports* 11.1 (1 Apr. 6, 2021), p. 7534, where Kyra H. Grantz, Joshua Kaminsky, Hannah R. Meredith, Shaun A. Truelove contributed equally. It is referred in the following as the postprint (and its supplementary information as SI)

### INTRODUCTION

In late 2019, the virus responsible for coronavirus disease 2019 (COVID-19) was detected in Wuhan, China<sup>1</sup>. Since its emergence, SARS-CoV-2 has spread rapidly, causing significant morbidity and mortality and prompting the World Health Organization to declare a pandemic on March 11, 2020<sup>2</sup>. With 4.07M confirmed deaths and 190M confirmed cases as of July 2021, it is one of the deadliest pandemics in history. In addition to its significant individual health impacts, COVID-19 has put considerable strain on health systems, as a large fraction of cases requires mechanical ventilation or critical care<sup>3</sup>. In every stage of the pandemic thus far, there has been a need for flexible decision support tools that can be used to model and compare critical planning scenarios.

Epidemiological models have played an important role in shaping public health policy and interventions throughout the pandemic. The methods used have ranged widely—from agent-based modeling approaches that simulate the global movement of individuals and their contacts in household, workplace, and

<sup>1</sup> Zhu et al., “A Novel Coronavirus from Patients with Pneumonia in China, 2019” in: *New England Journal of Medicine* (2020).

<sup>2</sup> WHO, *WHO Director-General’s Opening Remarks at the Media Briefing on COVID-19* (2020).

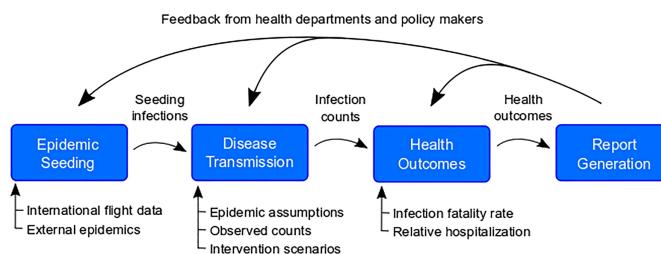
<sup>3</sup> Huang et al., “Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan, China” in: *The Lancet* (2020).

leisure settings<sup>4</sup>, to population-level models that incorporate features like age-specific transmission, asymptomatic and pre-symptomatic transmission, and metapopulation structure<sup>5</sup>, to curve fitting approaches that use data from early in the COVID-19 pandemic to project future burden<sup>6</sup>. Likewise, the goals of these models have varied widely, from assessing importation risk, estimating the fraction of cases attributable to transmission from unobserved infections, projecting the impact of non-pharmaceutical interventions that target different populations, and forecasting the needs of the healthcare system.

Within this space, there was a need for a modeling pipeline that could provide flexible but sophisticated epidemiological models to decision-makers who needed to plan and compare specific interventions. Here a scenario modeling pipeline is detailed: a modular framework that projects epidemic trajectories and health care impacts under different suites of interventions in order to aid in scenario planning. The flexibility of this approach has allowed us to provide rapid support to multiple organizations at the same time while customizing the pipeline models to situation-specific questions and data. This framework has been used to provide tailored estimates of the relative impacts of different scenarios of disease transmission, severity, and control, thus guiding intervention policies in several states, countries, and humanitarian aid settings.

## METHODS

### Pipeline at a glance



The pipeline consists of multiple modular components designed to run in sequence to produce results and reports focused on policy-relevant outcomes in any set of geographic locations (i.e., multiple countries, a single country, a set of subnational administrative units, or a single sub-national administrative unit) (fig. 4.1). While the pipeline was developed to be extended, the current core components are (1) epidemic seeding, (2) the transmission model, (3) health outcome generation engine, and (4) report generation.

These modular components of the pipeline fit together because each is composed of multiple pieces: an input format, an output format, one or more code libraries (where applicable), and a runner script. The standardized input and output formats ensure that components may be switched out according to user preference without impacting other phases.

The pipeline runs these components in sequence, according to the specifications outlined in a configuration file. This makes it easy to add or modify a component. To add a component, its input format is specified, and incorporate its dependencies. When appropriate, entire components may be substituted with data from outside of the pipeline, provided that the data meet the input formats

<sup>4</sup> Ferguson et al., *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand in: 20* (2020).

<sup>5</sup> Chinazzi et al., “The Effect of Travel Restrictions on the Spread of the 2019 Novel Coronavirus (COVID-19) Outbreak” in: *Science* (2020); Branas et al., “Flattening the Curve before It Flattens Us: Hospital Critical Care Capacity Limits and Mortality from Novel Coronavirus (SARS-CoV2) Cases in US Counties” in: *medRxiv* (2020); Moghadas et al., “Projecting Hospital Utilization during the COVID-19 Outbreaks in the United States” in: *Proceedings of the National Academy of Sciences* (2020); Davies et al., “Age-Dependent Effects in the Transmission and Control of COVID-19 Epidemics” in: *Nature Medicine* (2020).

<sup>6</sup> IHME et al., “Forecasting the Impact of the First Wave of the COVID-19 Pandemic on Hospital Demand and Deaths for the USA and European Economic Area Countries” in: *medRxiv* (2020).

fig. 4.1: Overview of the pipeline. The pipeline has four modules, each with specific inputs that can be specified by the user. First, it identifies when and in which model locations epidemics are seeded using an air importation model or confirmed case data. Second, the epidemic seeding events are used to initiate the disease transmission model, which is informed by epidemiological assumptions and intervention scenarios. The disease transmission model produces daily incident infection counts and infection prevalence. Next, health outcomes like hospitalizations and ICU admissions are calculated from these infection counts according to assumptions about health outcome risks and infection fatality ratios. Finally, these health outcomes may be summarized using templates and functions from the report generation component of the pipeline.

required by the next pipeline phase.

#### MODULE 1: EPIDEMIC SEEDING AND INITIALIZATION.

“Epidemic seeding” refers to how the disease transmission module is initialized with infected individuals. A seeding module must produce one or more seeding files that specify an added number of incident cases occurring due to “seeding” at particular dates and locations. The pipeline currently contains two epidemic seeding options: (1) seeding according to first case appearance in data, and (2) seeding according to an air travel importation model.

**SEEDING ACCORDING TO EARLIEST IDENTIFIED CASES** This seeding option enables users to seed the model according to COVID-19 case data. It currently supports user-supplied data and downloads from two commonly used public sources, the Johns Hopkins University Center for Systems Science and Engineering (JHU-CSSE) COVID-19 Dashboard<sup>7</sup> and USAFacts, a database that collates data from US state health departments<sup>8</sup>. Drawing from the user-specified data source, this option identifies the first five days that cases were reported in each modeled location. It is assumed that confirmed cases were infected a user-specified number of days prior to when they were reported, and that there is a user-specified ratio of infections to confirmed cases. Seed infections are, hence, created in each modeled location on the estimated days of infection for the first 5 days with reported cases; they are drawn stochastically from a Poisson distribution where the mean is the product of the number of reported cases and the user-specified ratio.

**SEEDING ACCORDING TO AN AIR IMPORTATION MODEL** A previously published model of measles importation was adapted to model the rate of COVID-19 importation to specific locations due to air travel<sup>9</sup>. This seeding option, available in the Github repository “HopkinsIDD/covidImportation”, uses complete itinerary (origin to final destination) air travel volume data from OAG<sup>10</sup> for all airports in the world, source location populations, and source location incidence data, to inform a model with which absolute counts of importation are estimated on a daily basis into airports<sup>11</sup>.

Geographic areas surrounding airports are classified spatially into “airport catchment areas” with a Voronoi tessellation of space in reference to the latitude and longitude coordinates of the airport<sup>12</sup> (fig. 4.2). When there are multiple airports within close proximity, the user may specify a threshold distance under which airports may be grouped into a single cluster that is defined by its centroid. The probability of importation is assigned to each intersection of a Voronoi tile and an administrative unit boundary. This probability is calculated as the proportion of the airport catchment area population that lives in that intersection, assuming that population is distributed evenly by area. All air importations on a given day are then aggregated to the administrative unit level and seeded into the epidemic model as newly infected individuals.

#### MODULE 2: TRANSMISSION MODEL AND INTERVENTION SCENARIOS.

The disease transmission module takes in seeding information and produces an epidemic model output file that contains, at minimum for subsequent module compatibility, daily counts of incident infections, indexed to their time of symp-

<sup>7</sup> Dong et al., “An Interactive Web-Based Dashboard to Track COVID-19 in Real Time” in: *The Lancet Infectious Diseases* (2020).

<sup>8</sup> USAFacts, *US COVID-19 Cases and Deaths by State* (2021).

<sup>9</sup> Truelove, Mier-y-Teran-Romero, et al., “Epidemics, Air Travel, and Elimination in a Globalized World: The Case of Measles” in: *medRxiv* (2020).

<sup>10</sup> OAG, *Flight Data in OAG* (2020).

<sup>11</sup> Truelove, Lauer, et al., *HopkinsIDD/covidImportation: Initial Release of covidImportation R Package* (2020).

<sup>12</sup> Balcan et al., “Modeling the Spatial Spread of Infectious Diseases: The GLobal Epidemic and Mobility Computational Model” in: *Journal of Computational Science* (2010).



fig. 4.2: A possible tessellation of airports in China, where the country is divided in regions depending on the closest airport.

tom onset. The currently implemented default transmission module comprises a metapopulation model with stochastic Susceptible-Exposed-Infected-Recovered (SEIR) disease dynamics.

**DISEASE DYNAMICS** The core model is a modified SEIR compartmental model where the time in the “Infected” compartment follows an Erlang distribution (i.e., the infected compartment is split into  $k$  compartments) to produce more realistic infectious periods where the chance of recovery depends on the time since infection<sup>13</sup>, and a coefficient ( $\alpha$ ) can be set to help the model approximate non-homogeneous mixing between susceptible and infected individuals and non-exponential growth<sup>14</sup>. By default  $k$  is set to 3 compartments. The transition of individuals between disease compartments is simulated stochastically with binomial random draws:

$$N_{S \rightarrow E}(t) = \text{Binom}\left(S, 1 - e^{-\Delta t \cdot \text{FOI}(t)}\right) \quad (4.2)$$

$$N_{E \rightarrow I^{(1)}}(t) = \text{Binom}\left(E, 1 - e^{-\Delta t \cdot \sigma}\right) \quad (4.3)$$

$$N_{I^{(1)} \rightarrow I^{(2)}}(t) = \text{Binom}\left(I^{(1)}, 1 - e^{-\Delta t \cdot \gamma'}\right) \quad (4.4)$$

$$N_{I^{(2)} \rightarrow I^{(3)}}(t) = \text{Binom}\left(I^{(2)}, 1 - e^{-\Delta t \cdot \gamma'}\right) \quad (4.5)$$

$$N_{I^{(3)} \rightarrow R}(t) = \text{Binom}\left(I^{(3)}, 1 - e^{-\Delta t \cdot \gamma'}\right) \quad (4.6)$$

$$\gamma' = \gamma \cdot k, \quad (4.7)$$

where  $S, E, I^{(1)}, I^{(2)}, I^{(3)}$ , and  $R$  represent the number of individuals in those respective compartments,  $\text{FOI}(t)$  is the force of infection from the infected population on the susceptible population,  $\frac{1}{\sigma}$  is the latent period,  $\frac{1}{\gamma}$  is the infectious period, and  $k$  is the number of  $I$  compartments. The force of infection, which modulates transition of individuals from the  $S$  to  $E$  compartments:

$$\text{FOI}(t) = \beta \cdot \frac{I(t)^\alpha}{H}, \quad (4.8)$$

$$(4.9)$$

where  $H$  is the total population,  $\beta$  is the daily transmission probability as defined by  $R_0$  and the infectious period, and  $\alpha$  is the mixing coefficient and:

$$I(t) = \sum_{j=1}^k I^{(j)} \quad (4.10)$$

$$H = S(t) + E(t) + \sum_{j=1}^k I^{(j)}(t) + R(t) \quad (4.11)$$

$$\beta(t) = R_0 \cdot \gamma. \quad (4.12)$$

**METAPOPULATION DYNAMICS** The model is capable of simulating disease spread in multiple locations jointly according to assumptions about population mobility between individual model locations (e.g., administrative units). The SEIR disease dynamics described above are simulated in each model location with a modification to the force of infection term that accounts for this impact of mobility on disease spread.

The force of infection in a given location  $i$  is calculated from a combination of local infections and infections in locations that are connected to it according to the mobility matrix, as follows:

The model for SARS-CoV-2 transmission is now entirely configurable to any compartmental model following an arbitrary transition graph. It allows for multi-strain, age-stratified models or models with vaccinated compartments.

<sup>13</sup> Yan and Chowell, *Quantitative Methods for Investigating Infectious Disease Outbreaks* (2019).

<sup>14</sup> Finkenstädt et al., “A Stochastic Model for Extinction and Recurrence of Epidemics: Estimation and Inference for Measles Outbreaks” in: *Biostatistics* (2002).

Early 2020, COVID Scenario Pipeline runs where parametrized based on the serial interval and the basic reproductive number of COVID-19 as follow:

**SERIAL INTERVAL** The serial interval (SI) represents the interval between two subsequent infections. Here  $\sigma$  and  $\gamma$  are parameterized with estimates of the range of the serial interval or generation time, such that:

$$SI = \frac{1}{2} \left( \frac{1}{\gamma} \right) + \frac{1}{\sigma}, \quad (4.1)$$

where it is assumed that the average infection occurs halfway through an index case’s infectious period. For SARS-CoV-2, the serial interval was estimated to be in range 6.5 – 8.2, from: Bi, Wu, et al., “Epidemiology and Transmission of COVID-19 in 391 Cases and 1286 of Their Close Contacts in Shenzhen, China: A Retrospective Cohort Study” in: *The Lancet Infectious Diseases* (2020), tab. S4.

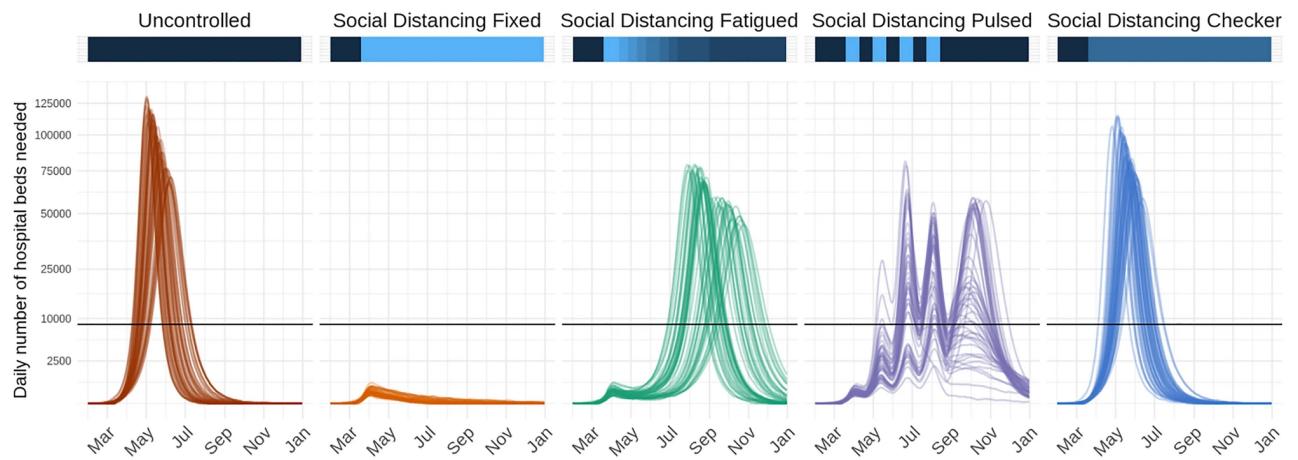
**THE BASIC REPRODUCTIVE NUMBER**  $R_0$  is the number of newly infected caused by an infected in a fully susceptible population. It has been estimated in the range 2 – 3 from early modeling works Riou et al., “Pattern of Early Human-to-Human Transmission of Wuhan 2019 Novel Coronavirus (2019-nCoV), December 2019 to January 2020” in: *Eurosurveillance* (2020) (this work also characterizes the dispersion of the number of secondary cases).

$$\text{FOI}_i = \left(1 - \sum_{j \neq i} p_{\text{away}} \frac{M_{i,j}}{H_i}\right) \cdot \beta_i(t) \frac{\left(I_i^{(1)} + I_i^{(2)} + I_i^{(3)}\right)^{\alpha}}{H_i} + \sum_{j \neq i} \left(p_{\text{away}} \frac{M_{i,j}}{H_i} \cdot \beta_j(t) \frac{\left(I_j^{(1)} + I_j^{(2)} + I_j^{(3)}\right)^{\alpha}}{H_j}\right), \quad (4.13)$$

with  $p_{\text{away}}$  the percent of the time individuals that move spend away ( $p_{\text{away}} \approx 0.5$  in the case of commuting).  $H_i$  is the population of node  $i$ .  $M$  is a mobility matrix such that  $M_{i,j}$  represents the daily movement of individuals (e.g., commuting) from origin  $i$  to destination  $j$ . The transition of individuals between disease compartments may be modified to index by location  $i$ , for example:

$$N_{S_i \rightarrow E_i}(t) = \text{Binom}\left(S_i, 1 - e^{-\Delta t \cdot \text{FOI}_i(t)}\right). \quad (4.14)$$

Users may provide a symmetric or asymmetric wide-form mobility matrix for all model locations or a long-form sparse mobility matrix that indicates only pairs of model locations with connectivity.



**APPLICATION OF TRANSMISSION MODIFIERS** In the absence of vaccines and other preventive treatments, non-pharmaceutical interventions, such as school closures, social distancing, stay-at-home directives, and testing and isolation are critical strategies for reducing disease transmission. Disease transmission may also change across space or time as the result of exogenous factors like seasonality or spatial heterogeneity in contact patterns. The model enables users to specify changes to the basic reproductive number ( $R_0$ ) and the inverse of the infectious period ( $\gamma$ ), for pre-specified periods of time to all or subsets of model locations independently. These interventions or exogenous changes can be implemented with fixed or distributional effectiveness. In addition, users may specify a rate of fatiguing effectiveness (e.g., declining adherence to a policy) over a certain number of days. This format enables flexibility in scenario planning; for instance, the model can be used to examine the effects of chaining multiple interventions together over time (e.g., school closure then stay-at-home), gradual declining adherence of the population to an intervention, switching interventions on and off over time, spatially heterogeneous interventions, or innate spatiotemporal heterogeneities (fig. 4.3).

fig. 4.3: Time series of the daily number of hospital beds needed across five possible intervention scenarios in a fictional location with nine counties. Lines represent results from 50 stochastic model simulations. Horizontal black lines represent the total hospital bed capacity in the fictional location. The colored horizontal bars along the top visualize the effectiveness of interventions at a given time point along with a dark blue to light blue spectrum; dark blue indicates a period with no reductions to transmission, while light blue indicates a period with more restrictive action (i.e., low transmissibility).

Transmission modifiers like non-pharmaceutical interventions or other exogenous factors modulate the daily transmission term below:

$$\beta'_i(t) = (1 - r_i(t)) \cdot \beta_i(t), \quad (4.15)$$

where  $\beta'_i(t)$  is the daily transmission rate after accounting for transmission modifier  $r_i(t)$  at the specified location  $i$  at time  $t$ . When transmission modifiers are in effect  $\beta'_i(t)$  replaces  $\beta_i(t)$  in the force of infection term  $\text{FOI}_i(t)$ . Note that several effects at the same time in the same place have a compounded effect on transmission of reductions. The effect of the  $k$  intervention in place at time  $t$  and node  $i$  is:

$$\beta'_i(t) = \beta_i(t) \cdot \prod_k (1 - r_k(t)), \quad (4.16)$$

but the compounding effect may be chosen to be additive (*e.g.* for vaccination rate) instead of multiplicative.

The specification of transmission modifiers is completely user-specified. A set of common non-pharmaceutical intervention scenarios that can be applied for user-specified dates and locations are included in the tab. s1 of the postprint s1. These intervention scenarios have been compiled according to a review of the literature on the potential impact of non-pharmaceutical interventions on respiratory virus transmission<sup>15</sup>.

### MODULE 3: CALCULATION OF HEALTH OUTCOMES

This pipeline module translates outputs from the transmission model into health outcomes such as hospitalizations and deaths. It takes in counts of daily incident infections and produces daily counts for specific health outcomes at appropriate time delays.

The current default implementation<sup>16</sup> produces hospital and intensive care unit (ICU) admissions, current hospital and ICU occupancy, ventilators needs, and the number of deaths. The modeling of health outcomes assumes that there is some transition probability from infection to death, infection to hospitalization, hospitalization to ICU admission, and ICU admission to ventilator use. When modeling health outcomes, is considered the probability of its occurrence (*e.g.*, the probability that infections are hospitalized), the time delay relative to its disease course (*e.g.*, the time between hospital admission and ICU admission), and where applicable, the duration in a given state (*e.g.*, how long a patient remains ventilated)<sup>17</sup>.

For settings where hospitalization, ICU admission, and ventilator use are infrequent, it may be more appropriate to think of these health outcome projections as different levels of disease severity.

The user may specify health outcome probabilities and delays, conditional on the flows described above. For use as default values, tables of parameter values derived from a literature review of COVID-19 health outcomes are provided in the postprint s1.

The pipeline currently contains two versions of this module with different approaches to the specification of health outcome risks: (1) unadjusted, uniform risk and (2) location-specific risks, adjusted by key demographic or health factors in each location.

<sup>15</sup> Which shows how research works conducted for past epidemics allowed to estimate the range of expectations for NPIs effectiveness well before evidence for COVID-19 became available. This table includes *e.g.* the outcomes of different responses to the 1918 influenza as input for social distancing policies and estimates on how influenza transmission was impacted when schools were closed due to holidays or inclement weather for the impact of school closures. As for current estimates, the next chapter deals with evaluating the impact of NPIs against SARS-CoV-2 transmission, and these estimates were used to parametrize the COVID Scenario Pipeline on Switzerland.

<sup>16</sup> at the time of submission. Similar to the transmission module, the outcome module now allows arbitrary transition graphs between outcomes, allowing it to accurately reflect any facet of the healthcare and reporting system. Among other improvements, note that transition probabilities (such as reporting fraction) maybe be modified in time and space with modifiers very similar to these for NPIs.

<sup>17</sup> Despite an added complexity, this explicit approach to model the healthcare system has some advantages with respect to using compartments (as done in the following chapters); the principal being that it is possible to use arbitrary distribution, such as log-normal, to model the residence time in each stage.

Column name (notation)	Description
$p_{\text{hosp} \text{inf}}$ ( $p_{\text{hosp} \text{inf}}$ )	Probability of hospitalization among infected individuals
$p_{\text{icu} \text{hosp}}$ ( $p_{\text{icu} \text{hosp}}$ )	Probability of ICU admission among hospitalized individuals
$p_{\text{vent} \text{icu}}$ ( $p_{\text{vent} \text{icu}}$ )	Probability of ventilation among individuals in the ICU
$p_{\text{death} \text{inf}}$ ( $p_{\text{death} \text{inf}}$ )	Probability of death among infected individuals
$\text{rr}_{\text{hosp} \text{inf}}$	Relative risk of hospitalization (given infection) relative to the average across all geoids
$\text{rr}_{\text{death} \text{inf}}$	Relative risk of death (given infection) relative to the average across all geoids

tab. 4.1: Health outcome risk parameters.

**UNADJUSTED, POPULATION-WIDE HEALTH OUTCOME RISKS** This option generates health outcome estimates with unadjusted risk across all locations, assuming fixed values for all health outcome probabilities, delays, and durations.

It is assumed that the number of infections admitted to the hospital is a draw from the Binomial distribution, lagged by a fixed time from symptom onset to hospital admission:

$$n_{t_{\text{inf}}+t_{\text{inf}} \rightarrow \text{hosp}}^{\text{hosp}} \sim \text{Binom}\left(n_{t_{\text{inf}}}^{\text{inf}}, p_{\text{hosp}|\text{inf}}\right), \quad (4.17)$$

where  $n^{\text{hosp}}$  is the number of hospital admissions,  $n^{\text{inf}}$  is the number of infections,  $t_{\text{inf}}$  is the time of infection,  $t_{\text{inf}} \rightarrow \text{hosp}$  is the mean time delay between infection and hospital admission, and  $p_{\text{hosp}|\text{inf}}$  is the probability of hospitalization given infection (tab 4.1).

A similar assumptions is made for the transitions between other outcomes (hospitalization admission to ICU admission, ICU admission to ventilator use, infection to death):

$$n_{t_{\text{hosp}}+t_{\text{hosp}} \rightarrow \text{icu}}^{\text{icu}} \sim \text{Binom}\left(n_{t_{\text{hosp}}}^{\text{hosp}}, p_{\text{icu}|\text{hosp}}\right) \quad (4.18)$$

$$n_{t_{\text{icu}}+t_{\text{icu}} \rightarrow \text{vent}}^{\text{vent}} \sim \text{Binom}\left(n_{t_{\text{icu}}}^{\text{icu}}, p_{\text{vent}|\text{icu}}\right) \quad (4.19)$$

$$n_{t_{\text{inf}}+t_{\text{inf}} \rightarrow \text{death}}^{\text{death}} \sim \text{Binom}\left(n_{t_{\text{inf}}}^{\text{inf}}, p_{\text{death}|\text{inf}}\right). \quad (4.20)$$

The number of patients currently hospitalized, admitted to ICU, and ventilated are generated from the incident number of events and fixed (non-distributional) user-defined durations for each event.

As information about the death and hospitalization rates was scarce early on in the pandemic, decision-makers wanted to consider how things would unfold over different scenarios of health burden. To facilitate these needs, the pipeline separates hospitalization and death rates from the other outcomes, and allows users to consider multiple scenarios for different rates (*e.g.* scenarios for different COVID-19's IFR before evidence on disease severity precised the estimates).

**LOCATION-SPECIFIC HEALTH OUTCOME RISKS** The severity of disease from SARS-CoV-2 infection can vary greatly between locations due to differences between populations; individuals that are older, with limited access to health care, or with certain pre-existing health conditions are at greater risk for severe disease and death. For this reason, the hospitalization module also supports the specification of location-specific relative risks, which can be used, for example, to standardize hospitalization and mortality rates according to the age distribution of the population. Users may provide a wide format data file with standardized variable names for the transition probabilities and relative risks by geoid. These transition probabilities are conditional on previous states (*e.g.*, probability of hospitalization

given infection is named `p_hosp_inf`). The probability of hospitalization and death given infection are specified as relative values compared to population-wide averages specified in the configuration file. The location-specific standardizations apply only to the health outcomes, making a critical assumption that all individuals are at equal risk of infection.

To facilitate the construction of such files a companion package, `covidSeverity`<sup>18</sup>, is provided. It produces location-specific relative death and hospitalization rates based on the age distribution of the local population. The package generates these outputs for US counties based on data from the US Census Bureau, and this is provided as a model input file as part of the main pipeline implementation (see `COVIDScenarioPipeline/sample_data/geoid-params.csv`). The package also includes built-in functionality to pull data from WorldPop<sup>19</sup> and generate adjustments for any location of interest.

The `covidSeverity` package applies a logistic generalized additive model (GAM) with a penalized cubic spline for age and a random effect for age-specific estimates of the risk of each health outcome from the literature, thus producing estimates of risk for 10-year, aggregated age categories. These age-specific estimates are then applied to the population age distribution in a given location.

The epidemic transmission model, the intervention module, and the outcome calculation are implemented in python, just-in-time compiled to machine code using Numba<sup>20</sup> for performance. Since this description, the team working on the pipeline has brought many improvements, both computational and conceptual, that have enabled the pipeline to remain useful in 2021.

#### MODULE 4: SUMMARIZATION OF MODEL OUTPUTS

This component of the pipeline provides wrapper functions for the lightweight summarization of model outputs into quantiles, plotting functions for common figures, and R Markdown templates to facilitate the rapid generation of technical reports. This module is available in the R package `report.generation` in the Github repository “`HopkinsIDD/COVIDScenarioPipeline`.”

Two key functions that read and process individual transmission and health outcome model output files are provided. In managing individual files with these functions, the processing time and memory load is reduced. Both of these functions take processing functions as arguments, thus enabling aggregation and filtering to occur at the level of individual files.

The package contains technical report R Markdown templates for US states, US counties, and individual countries, a diagnostic report template, and a template that is maintained solely for integration testing. When `report.generation` is installed and loaded, the templates become available to the user. Many parameters are drawn from the configuration file automatically and pre-written R Markdown chunks about the module options and methods can be referenced within the package.

Common figures include summary tables and time courses for estimated health outcomes under different interventions, maps that portray cumulative cases at the county level, comparisons between model estimates and observed cases and deaths, and visualizations for when ICU and ventilator capacity for each county is exceeded. The vignette in the postprint `s1` walks through example report outputs in more detail and a template-generated report example is provided.

<sup>18</sup> Lauer, Truelove, et al., *HopkinsIDD/covid-Severity: Initial Release of covidSeverity R Package* (2020).

<sup>19</sup> [worldpop.org](http://worldpop.org)

<sup>20</sup> Lam et al., “Numba: A LLVM-Based Python JIT Compiler” (2015).

## MODEL SPECIFICATION

All components and settings for simulations from the COVID Scenario Pipeline model are specified in an easily modifiable YAML configuration file. Different options are described in detail in the postprint and in the GitHub repository wiki.

## RESULTS

The COVID Scenario Pipeline has been used to support several partners, including the state of California and the national US response with regular reports tailored to their needs. While these are private, it is however possible to see the current projections from the pipeline: regular outputs are shared as part of the two following multi-modeling experiments<sup>21</sup>:

**COVID-19 FORECAST HUB** Along with dozens of other modeling teams, the Johns Hopkins ID Dynamics COVID-19 Working Group forecasts a few weeks ahead the COVID-19 pandemic in the US. In this case, no common assumptions are made: teams strive to accurately project expected dynamics. The current forecasts are visible on [covid19forecasthub.org](https://covid19forecasthub.org)<sup>22</sup>.

**SCENARIO MODELING HUB** proposes longer-term projections under different scenarios where *e.g.* variant transmission, immune escape, or vaccination hesitancy differs. Each round consists of 4 scenarios whose assumptions are shared across teams, while the rest is left to the choice of the modeling teams. The scenarios projections for current and past rounds are visible at [covid19scenariomodelinghub.org](https://covid19scenariomodelinghub.org)<sup>23</sup>.

## EXAMPLE OF SCENARIO PLANNING REPORT FOR CANTON DE VAUD

Moreover, a report produced on April 9, 2020 for Canton de Vaud main hospital, CHUV, is presented in the next two pages. This report is part of a series of reports produced during the first wave of the COVID-19 pandemic in collaboration with Canton de Vaud authorities. While different from typical pipeline reports, it shows how the assumptions, projections, and uncertainties are communicated to decision-makers.

After summarizing the assumptions taken in this report, short-term forecasts are presented, both graphically and in terms of quantities of interests such as peak ICUs and hospital occupancy. As the inference module of the pipeline did not exist at this time, these projections were generated using a simple filtering method to select parameter values that were reasonable given the observed epidemiological dynamics. Longer-term projections are presented for three transmission scenarios, with an additional visualization of the probability of ICU capacity exceedance under these scenarios. Finally, assumptions and limitations which characterize the projections are summarized at the end of the report.

In the report appendix, along with additional projections, a live estimation of the hospitalized case-fatality ratio is provided. In fact, access to individual-level healthcare data was provided as part of the interactions with public-health officials of Canton de Vaud. This dataset improved considerably scenario reports estimates that were subsequently tailored for local specificities of the health system and provided the foundations for the study presented in CHAPTER 5<sup>24</sup>.

**MODEL ACCESS** The project is open-source under the GNU General Public License v3.0 license, and code is available at [github.com/HopkinsIDD/COVIDScenarioPipeline](https://github.com/HopkinsIDD/COVIDScenarioPipeline). The master branch of this repository consists of a Python package “SEIR” and two R packages “hospitalization” and “report.generation” which correspond to the second, third, and fourth modules of the model pipeline. Air importation-based seeding is implemented in the covidImportation package ([github.com/HopkinsIDD/covidImportation](https://github.com/HopkinsIDD/covidImportation)), while seeding according to the earliest identified cases is performed in scripts within the COVID Scenario Pipeline repository.

<sup>21</sup> While it is possible to single out the COVID Scenario Pipeline output on these websites, the interest of these studies lies in the ensemble trajectory that combines projections from many models, mitigating issues of opinionated model design and assumptions.

<sup>22</sup> Cramer et al., “Evaluation of Individual and Ensemble Probabilistic Forecasts of COVID-19 Mortality in the US” in: *medRxiv* (2021).

<sup>23</sup> Borcherding, “Modeling of Future COVID-19 Cases, Hospitalizations, and Deaths, by Vaccination Rates and Nonpharmaceutical Intervention Scenarios — United States, April–September 2021” in: *MMWR. Morbidity and Mortality Weekly Report* (2021), presents round 4 results.

<sup>24</sup> It is interesting to compare how the real-time data analysis in this report compare with the detailed estimates presented in the APPENDIX TO CHAPTER 5 that uses two additional weeks of data.

# Planning Scenarios for COVID19 in Vaud

Updated 2020-04-09

The aim of this model is to provide estimates about the course of the epidemic under a specific set of assumptions for planning purposes. These are not direct forecasts of the epidemic.

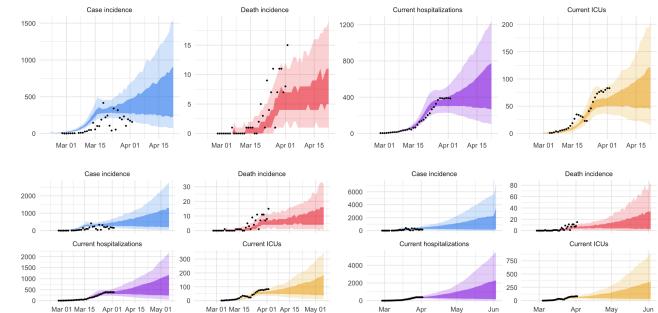
## Summary

This report compares the health impact and surge capacity needs for the COVID-19 epidemic in Canton de Vaud. The underlying model is based on our best understanding of SARS-CoV-2 natural history and transmission, and uses current demographic and epidemiologic data from Vaud and other locations (Table 1). We project the number of hospitalizations, ICU admissions, and deaths in both the short and long term under different public-health intervention scenarios.

**Table 1.** Key parameters and assumptions used in the Vaud model. Hospitalization values derived from Canton de Vaud reflects our current estimate under the available data. As so, it underestimate the durations as it's biased towards short hospital stays. As more data become available, we'll provide better estimates.

Parameter	Value	Source
Population size of Canton VD	793'129	Federal Statistical Office
Generation time	6.5 to 8.2 days	IDD@JHU (mailto:IDD@JHU) analysis of reported cases
Basic reproduction number R0	1.8 to 3.3	Multiple
Reproduction number with current measures	0.45 to 1.71	Multiple
Duration of hospitalization	6.8 days	VD data
Duration of ICU stay	6.3 days	VD data
Proportion of ICU cases among hospitalized patients	20%	VD data
Ratio death / hospitalization	10%	VD data
Ratio hospital death / total death	0.66	VD data

## Short-term cantonal estimates with current measures in place



**Figure 1.** Modelled incidence of cases and death, along with current hospital and ICU occupancy at horizons 2 weeks (top line), 4 weeks (bottom left) and 8 weeks (bottom right). The black dot represents cantonal data, the dark colored region the 50% confidence interval and the lightly colored region the 95% confidence interval.

**Table 2.** Short term projections. In all tables, we display the mean across all simulations, along with the 95% interquartile range.

	2 weeks		4 weeks		8 weeks	
	median	95%CI	median	95%CI	median	95%CI
Peak current hospitalizations	500	(290-1,200)	600	(290-2,200)	800	(300-5,000)
Peak current ICUs	90	(50-200)	100	(50-300)	140	(50-900)
Cumulative cases	16,000	(8,000-30,000)	23,000	(9,000-60,000)	40,000	(10,000-180,000)
Cumulative ICUs	500	(270-800)	700	(310-1,600)	1,300	(400-5,000)
Cumulative hospitalizations	2,500	(1,400-5,000)	4,000	(1,600-9,000)	7,000	(1,800-28,000)
Cumulative deaths	190	(120-300)	300	(150-600)	600	(170-2,100)

## Long term planning under different scenarios

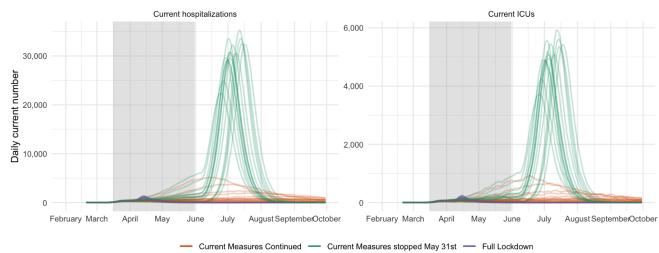
We consider three planning scenarios:

- **Current Measures**, which assumes measures as they were instituted in canton de Vaud:
- School closure on March 13,
- Physical distancing measures progressively established from March 16 to March 20.

We assume these moderately restrictive non-pharmaceutical interventions (NPIs) to reduce transmission by 48 - 75% ( $R_0$  from 0.45 to 1.71) based on a city with similar social distancing measures in place during the 1918 influenza pandemic and on the most recent  $R_0$  estimates in Switzerland. In this scenario, these measures are maintained for the foreseeable future.

- **Current Measures Stopped May 31st**, the measures described in the Current Measures scenario are held up until May 31st, then we assume a return to high transmissibility ( $R_0$  from 1.8 to 3.3) for the foreseeable future.
- **Full lockdown**, which assumes that very aggressive non-pharmaceutical interventions that reduce transmission by 81 - 88% ( $R_0$  from 0.25 to 0.63) are put in place from April 5 on. This scenario is based on the Wuhan lockdown that started on January 23, 2020.

## Hospital and ICU usage



**Figure 2.** Time series for 15 random example simulations for three planning scenarios of SARS-CoV-2 spread in Vaud. Interventions in the Current Measures Discontinued Scenario are applied from March 16 to May 31 (grey box), while in other scenarios measures are continued until October.

**Table 3. Current Measures Continued**

	May 1		June 1		September 1	
	median	95%CI	median	95%CI	median	95%CI
Peak current hospitalizations	500	(270-1,900)	700	(270-5,000)	900	(280-5,000)
Peak current ICUs	90	(50-280)	120	(50-700)	170	(50-900)

**Table 4. Current Measures Stopped May 31st**

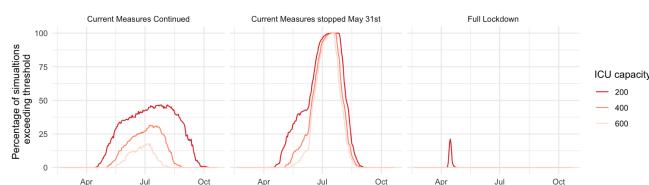
	May 1		June 1		September 1	
	median	95%CI	median	95%CI	median	95%CI
Peak current hospitalizations	500	(290-1,800)	800	(300-5,000)	30,000	(24,000-40,000)
Peak current ICUs	90	(50-270)	140	(50-800)	5,000	(4,000-6,000)
Cumulative cases	20,000	(9,000-50,000)	40,000	(10,000-160,000)	700,000	(700,000-800,000)
Cumulative ICUs	600	(300-1,300)	1,200	(300-5,000)	27,000	(26,000-27,000)
Cumulative hospitalizations	3,100	(1,600-7,000)	6,000	(1,800-25,000)	130,000	(130,000-140,000)
Cumulative deaths	250	(140-500)	500	(170-1,800)	13,000	(13,000-14,000)

**Table 5. Full Lockdown**

	May 1		June 1		September 1	
	median	95%CI	median	95%CI	median	95%CI
Peak current hospitalizations	800	(400-1,600)	800	(400-1,600)	800	(400-1,600)
Peak current ICUs	150	(80-280)	150	(80-280)	150	(80-280)
Cumulative cases	19,000	(11,000-30,000)	21,000	(12,000-30,000)	21,000	(12,000-30,000)
Cumulative ICUs	700	(400-1,100)	700	(400-1,300)	700	(400-1,300)

Cumulative hospitalizations	3,000	(2,000-6,000)	4,000	(2,100-6,000)	4,000	(2,100-6,000)
Cumulative deaths	290	(180-500)	400	(200-600)	400	(200-600)

## ICU Capacity exceedence



**Figure 3. Fraction of simulations exceeding ICU capacity thresholds.** In each scenario, we show the percentage of simulations exceeding the 200, 400, 600 ICU beds threshold.

## Contributors

Prepared by Joseph Lemaitre, Javier Perez-Saez, Andrew Azman, Andrea Rinaldo, Jacques Fellay.

**Pipeline from Johns Hopkins ID Dynamics Working Group (<http://www.iddynamics.jhsph.edu/>):** Kyra H. Grantz, Joshua Kaminsky, Lindsay T. Keegan, Steve Lauer, Elizabeth C. Lee, Joseph Lemaitre, Justin Lessler, Hannah R. Meredith, Shaun A. Truelove.

**Contact:** joseph.lemaitre@epfl.ch (mailto:joseph.lemaitre@epfl.ch?subject=Hyperlinks), jacques.fellay@epfl.ch (mailto:jacques.fellay@epfl.ch?subject=Hyperlinks)

## Limitations

We note several limitations to our work, among which:

There remains considerable uncertainty around some of the key epidemiologic features of COVID-19, including the average duration of infectiousness and time to recovery or death. We have used commonly accepted and well supported estimates and believe that they are appropriate for planning purposes.

We assume equal risk of infection and progression to hospitalization or death among all individuals in the canton at a given time point. There is evidence of age-specific differences in clinical burden and perhaps susceptibility to infection that are not (yet) considered here.

We assume  $R_e$ , the effective reproductive number, to be constant in each scenario, other than changes due to onset of non-pharmaceutical interventions. As capacity for surveillance, contact tracing, and testing improve, and as the general public becomes increasingly aware of the outbreak and modifies their own

behavior, we might expect that  $R_e$  will decrease dynamically, perhaps even as a function of the perceived COVID19 burden. As the outbreak continues, it will be possible to refine the scenarios considered here to better reflect the actual epidemic situation.

We do not explicitly model the role of asymptomatic infection when calculating the number of expected hospitalizations. All infectious individuals are considered at risk of hospitalization, though some may recover or die prior to hospitalization. A substantial asymptomatic burden may reduce the number of hospitalized cases.

## Model assumptions

We built a stochastic, Susceptible-Exposed-Infected-Recovered (SEIR) model of SARS-CoV-2 transmission in canton Vaud.

**Population.** The total population and age distribution were collected from the website of the Swiss Federal Statistical Office. The entire population of the canton is assumed to be susceptible at the beginning of the model. We assume equal risk of infection for all individuals at a given time.

**Initial conditions.** We used the first confirmed cases recorded in Vaud on February 25 to inform the importations that seeded our epidemic model. We assumed that there was a reporting rate of 20% during this period using Poisson draws with a rate parameter five times the cumulative number of cases. Baseline transmissibility in the model is similar to the early days of the Wuhan epidemic ( $R_0 = 2-3$ ).

**Reproductive Number.** The reproductive number  $R_0$ , or the average number of secondary cases caused by a single infected individual in a susceptible population, varies by scenario. Note that the reproductive number is highly-context specific. There is still uncertainty in the range of possible  $R_0$  values for SARS-CoV-2, and varying effectiveness of non-pharmaceutical interventions, including social distancing, improved hand hygiene, and case detection and isolation, may further reduce  $R_0$ .

**COVID-19 Natural History.** From analysis of 181 confirmed SARS-CoV-2 cases among travelers and other publicly reported cases, we estimate the incubation period, or the time from exposure to symptom onset, to follow an exponential distribution with mean 5.2 days (IQR 1.5, 7.21 days) (Fig. S1). The average duration of infectiousness following symptom onset is between 1.3 days and 3, following an Erlang distribution with 3 compartments. We thus sample across a range of possible mean serial intervals (time from symptom onset of an index case to symptom onset of a secondary case infected by the index) from 6.5 to 8.2 days.

**Rates of Death and Hospitalization.** From our analysis of 391 confirmed SARS-CoV-2 infections in Shenzhen, China, we estimate the average time to hospitalization from symptom onset follows a log-normal distribution with median 3.42 days (IQR 2.01, 5.83) (Fig. 3). Other rates where derived from CHUV and Vaud hospitalization data. We fitted a log normal distribution to the observed distribution of time hospitalized, time in ICU, time from hospitalization to ICU and time to death. We estimate from current data that, among those hospitalized, 20% will be admitted to the ICU.

**Filtering.** Simulated time series of incidence and hospitalizations are based on 30'000 random draws from reasonable bounds of the parameter values (Table 1). Among these draws, we resample to narrow the set to a simulation that matches the observed reality in hospitalization. Namely, we assign to each simulation a weight that is proportional to the likelihood w.r.t. the incident hospitalization.

We use simple statistical models using the cumulative distribution of times to hospitalization, ICU admission, and death, as well as the durations of hospitalization and ICU stay to calculate the number of incident and cumulative hospitalizations and ICU admissions and deaths per day, accounting for

appropriate delays since infection and symptom onset.

Uncertain parameters include the length of the serial interval,  $R_0$  and the probability of hospitalization. Simulated time series of incidence and hospitalizations are based on 30'000 random draws from reasonable bounds of the parameter values (Table 1). Simulated trajectories are then resampled based on the degree to which they match the observed time series of cumulative hospitalizations, which gives a set of simulations which represent well the observed data.

## Key Sources

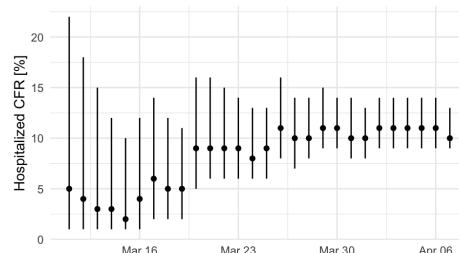
- Impact of broad scale non-pharmaceutical interventions is based on the observed impact of such programs in 1918 as reported in (Bootsma and Ferguson 2007). We took parameter estimates on the effectiveness of interventions from Milwaukee in this study.
- School closure impacts based on data from (Jackson et al. 2020), (Cauchemez et al. 2008) and (Litvinova et al. 2019)
- Approximate generation time is based on data from (Bi et al. 2020)

## Appendix

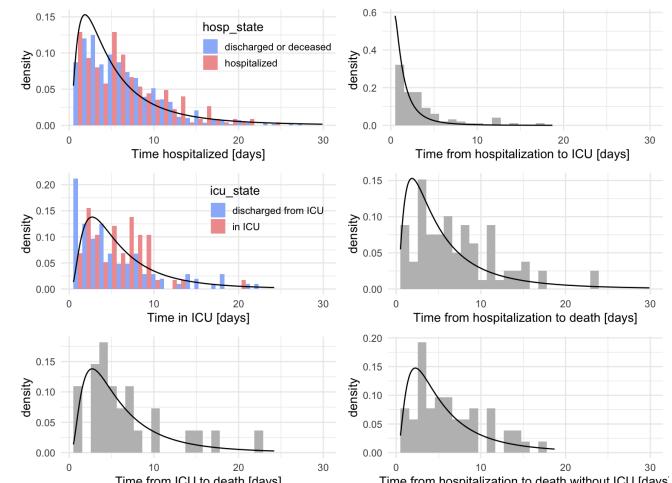
### Hospitalization data analysis

#### Key points:

- 956 have been hospitalized
- The mean hospitalized CFR in the past 3 days is of 11%
- 20% of patients were treated in an ICU
- The mean hospitalization time is of 6.4 days
- The mean ICU time is of 5.6 days

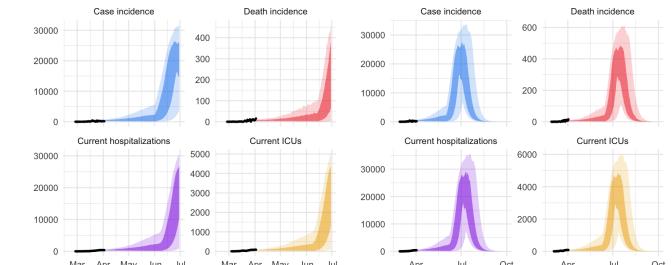


**Figure A.1. Hospitalized case fatality ratio (hCFR).** Estimates of hCFRs where based on the methods of Ruan et al. (2020) (<https://doi.org/10.1371/journal.pone.0006852>) (points are the Maximum Likelihood estimates and bars give 95% CIs).



**Figure A.2. Time distributions of modelled hospitalization processes.** Bars represent data from all VD hospitals and lines estimated probability distribution functions as used in the model.

### Epidemic dynamics under different scenarios



## DISCUSSION

<sup>25</sup> The COVID scenario pipeline is presented as an open-source modeling framework that aims to balance epidemiological rigor with the flexibility and urgency required by public health policymaking. The modularity of the proposed framework has enabled us to adapt the assumptions about COVID-19 epidemiology, transmission, and health outcome risks in response to emerging information and to different settings. The pipeline implementation of non-pharmaceutical interventions is highly adaptable for policymakers desiring to compare the impact of different potential scenarios.

<sup>25</sup> This section discusses the COVID Scenario Pipeline as presented in this chapter, *i.e.* going back to the July 2020 perspective.

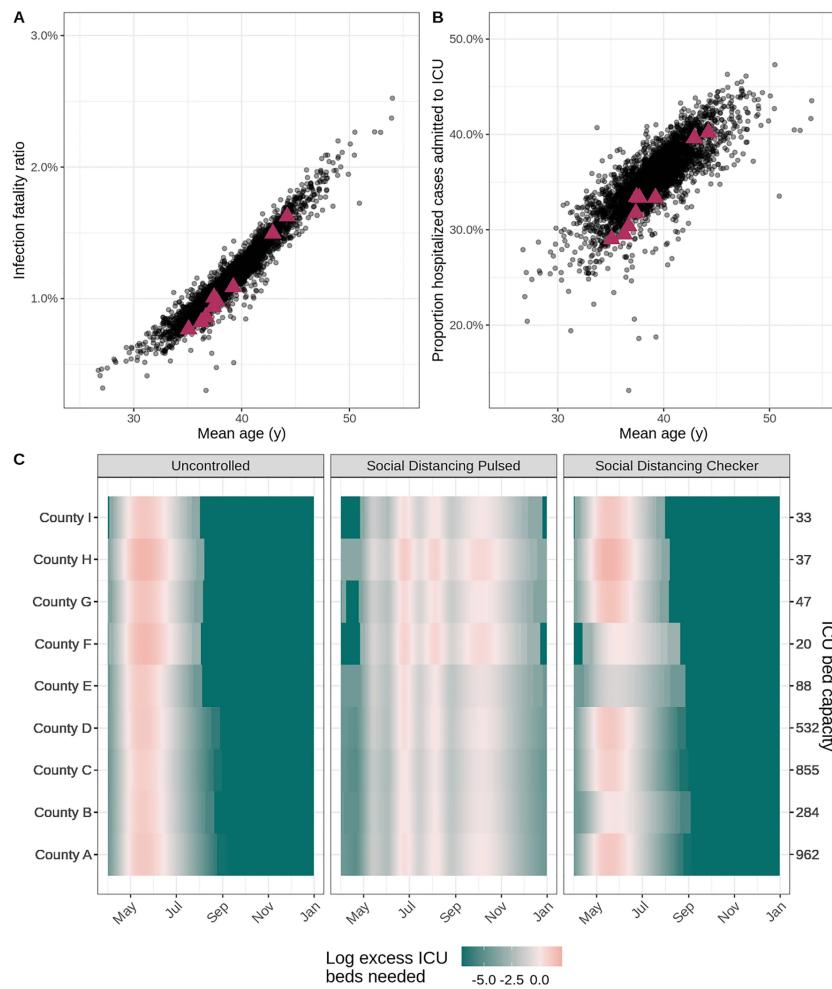
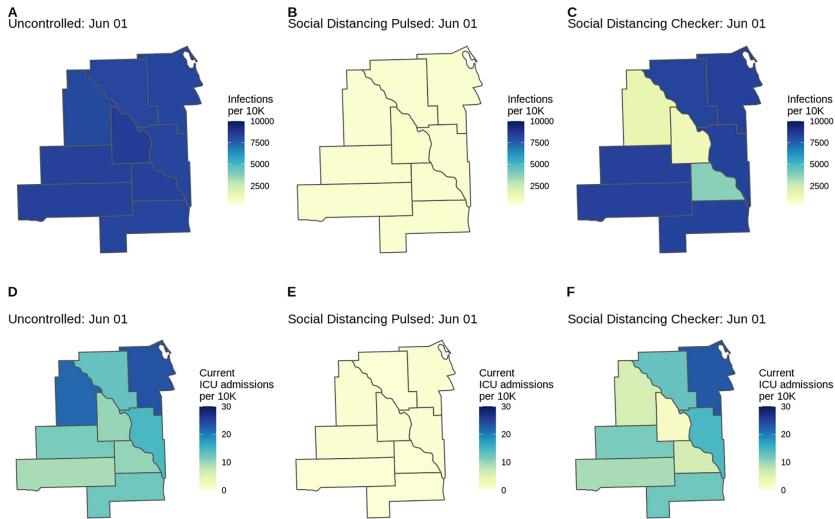


fig. 4.4: Health outcome risks and logistical needs for fictional counties. In the scatterplots, each point indicates (A) the age-adjusted infection fatality ratio and (B) the risk of ICU admission given hospitalization by mean age for a county within the United States. Data for nine fictional counties is marked by magenta triangles. (C) The heat maps display county-level ICU bed needs, shaded according to the log-ratio above or below the assumed ICU bed capacity (secondary y-axis) in each county (primary y-axis) for three example intervention scenarios (panels). The salmon pink shading indicates periods of time where ICU bed needs exceed capacity in the fictional counties.

Throughout the course of the pandemic, the default settings of the pipeline have been adapted in response to the changing needs and questions of our collaborators. At the beginning of the pandemic, air importation seeding was a critical determinant of epidemic onset in specific locations. Now that cases of COVID-19 are present worldwide, the development shifted towards more empirical methods of epidemic seeding that better match trajectories of confirmed cases in specific locations as policy questions have shifted to more operational needs. As new data emerged, it was decided to move from calculating unadjusted health outcomes to health outcomes based on the age-standardized risk of hospitalization, ICU admission, and death according to emerging case-study data.



As the COVID-19 pandemic continues, the plan is to continue to expand the scope of the COVID Scenario Pipeline to changing needs and questions. New model releases include a health outcomes model expansion that will enable a multiplicity of pathways to ICU occupancy, ventilator usage, and death. As questions have shifted to near-term operational needs, inference has been incorporated into the pipeline, thus enabling the calibration of model trajectories to deaths and confirmed case counts, short-term forecast of health outcomes, and estimation of location-specific transmission parameters and NPI effectiveness. Moreover, the epidemic transmission model has been re-written to allow for arbitrary transition graphs between compartments, which allow one to explicitly model vaccination campaigns, variants, immune escape and multiple age classes. While the pipeline's generic structure means that different modules may be readily replaced, the implementation of the model as described here has several limitations with regard to the epidemiology of COVID-19. The disease transmission model does not explicitly incorporate age-specific contact or transmission rates or asymptomatic transmission, nor does it consider how factors like testing rates may lead to time-varying biases in reporting. Data to inform these factors was scarce when this model was initially developed, but these processes can change the dynamics displayed by scenario projections. Moreover, this model structure means that age-targeted interventions (e.g., cocooning of high-risk age groups) cannot be modeled, nor are interventions targeting asymptomatic individuals (e.g., systematic age-specific testing) in a mechanistic manner. However, it is possible, and done in real settings, to adjust for these types of interventions through overall population-level reductions in disease transmission. Additionally, the current model cannot tune changes in population mobility or connectivity over time, although it is known that travel and movement restrictions played a role in changing the spread of COVID-19.

The health outcomes module as described in this chapter also has several structural limitations; it assumes that only one progression in health outcome severity exists (infection to hospitalization to ICU to ventilator use), although it is known that many disease course progressions are possible. In addition, the delays and durations involved in the modeled health outcomes progression are fixed in time for a given location, although stochastic variation may be incorporated across

fig. 4.5: County-level COVID-19 risk for three scenarios (Uncontrolled, Pulsed, Checker) in the fictional Location X. Choropleths for model outcomes on June 1, 2020 of the (A–C) cumulative infection rate per 10,000 population and (D–F) number of patients currently admitted to the ICU per 10,000 population for the Uncontrolled, Social Distancing Pulsed, and Social Distancing Checker intervention scenarios. County-level variation in attack rates can arise from differences in risk of importation, mobility patterns connecting subdivisions, and differences in non-pharmaceutical interventions applied in each location. If location-specific health outcome risks are specified (as are age-standardized health outcome risks in this example), this may serve as another source of county-level variation.

**RECENT DEVELOPMENTS** The present chapter solely focuses on a perspective from July 2020, but the current state of the project has seen numerous improvements since then. The pipeline transmission and outcomes model, as presented, were hard-coded simple models. These models allowed to successfully capture the early COVID-19 dynamics and to present reports to decision-makers. These simple models, focusing on a few important aspects of disease transmission, captured accurately epidemiological dynamics and uncertainties. Provided uncertain estimates of the serial interval and  $R_0$  (or other quantities), these models allow one to derive insights on projected future impact of the epidemic. Simplicity is a strength as it prevents overconfidence and the reliance on too many unknowns while highlighting the key aspects in transmission. However, as the pandemic progressed it brought more information and the need for additional flexibility. As months of evidence are now available, the pipeline benefits from an additional inference module (based on MCMC with adaptation for large-scale computations). And as mentioned in margin notes, most of the assumptions described in this chapter have now been relaxed while keeping the framework principle similar (independent module and single configuration file). The transmission module allows age-stratified, multi-strain models with vaccinated compartments and arbitrary transition rates in space and time. The health outcome allows arbitrary transition graphs and probabilities can change in space and time. Many other developments, such as helpers to write configuration files (a USA configuration file is about 30'000 lines), performance improvements, and report features were (and still are) added.

simulations.

Nevertheless, the modular approach taken is meant to allow for easy substitution of models with improvement in any of these areas while still taking advantage of other pipeline components. This feature has been leveraged throughout the course of the COVID-19 pandemic, and individual modules continue to develop. This flexibility does come at a cost, as the modular pipeline approach requires us to write and read files at the end and beginning of each phase, respectively. This procedure requires more disk space and input/output steps than other modeling approaches that can hold all of the necessary data in memory until a single output is produced at the end. Still, these slowdowns are not critically limiting; it has been possible to run 1000 county-level simulations of the United States in less than 10 min on a 96 core server.

These limitations point to a broader need to consider the totality of evidence generated by epidemiological models. While the proposed approach is well-suited to answering policy questions about interventions, it is critical for policymakers to explore projections from multiple models in order to understand the range of possible trajectories and the sensitivity of results to different assumptions. Models that incorporate individual-level behaviors may be better for considering the impact of specific contact tracing strategies or location-specific measures like workplace occupancy or symptom screening policies that are not well captured in compartmental models such as ours<sup>26</sup>. Other individual-based models are better suited for addressing heterogeneities due to differences in household or social structures<sup>27</sup>. Models incorporating real-time mobility data can best characterize the impact of movement-related restrictions<sup>28</sup> models with age-specific transmission may provide more detail on the impact of age-specific interventions like “cocooning”<sup>29</sup> or closing and opening schools<sup>30</sup>. Still other models are particularly suited to address questions about health systems burden and forecast operational needs<sup>31</sup> and to consider the economic impacts of transmission and interventions<sup>32</sup>. Integrating knowledge from multiple models, where appropriate, with careful consideration of the assumptions and appropriate applications of each model, will strengthen response and preparedness<sup>33</sup>.

Our flexible modeling pipeline brings an important voice to this “conversation” of models, by allowing rapid and flexible specification and simulation of even very complex intervention scenarios, and providing flexibility to rapidly update models as our understanding of a disease changes. This approach only reaches its full potential when parameters are based on careful and ongoing consideration of the literature and available data. But, when appropriately used as part of an iterative approach to decision making, this pipeline can be a valuable tool for public health decision-making.

<sup>26</sup> Kucharski, Klepac, et al., “Effectiveness of Isolation, Testing, Contact Tracing, and Physical Distancing on Reducing Transmission of SARS-CoV-2 in Different Settings: A Mathematical Modelling Study” in: *The Lancet Infectious Diseases* (2020); Firth et al., “Using a Real-World Network to Model Localized COVID-19 Control Strategies” in: *Nature Medicine* (2020); Hinch et al., “OpenABM-Covid19 - an Agent-Based Model for Non-Pharmaceutical Interventions against COVID-19 Including Contact Tracing” in: *medRxiv* (2020).

<sup>27</sup> Wilder et al., “Modeling Between-Population Variation in COVID-19 Dynamics in Hubei, Lombardy, and New York City” in: *Proceedings of the National Academy of Sciences* (2020); Kerr et al., “Covasim: An Agent-Based Model of COVID-19 Dynamics and Interventions” in: *medRxiv* (2021).

<sup>28</sup> Lai, Ruktanonchai, et al., “Effect of Non-Pharmaceutical Interventions to Contain COVID-19 in China” in: *Nature* (2020).

<sup>29</sup> Duque et al., “COVID-19: How to Relax Social Distancing If You Must” in: *medRxiv* (2020).

<sup>30</sup> Ferguson et al., *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand* in: *20* (2020).

<sup>31</sup> Branas et al., “Flattening the Curve before It Flattens Us: Hospital Critical Care Capacity Limits and Mortality from Novel Coronavirus (SARS-CoV2) Cases in US Counties” in: *medRxiv* (2020); Laboratory, *COVID-19 Cases and Deaths Forecasts* (); Weissman et al., “Locally Informed Simulation to Predict Hospital Capacity Needs During the COVID-19 Pandemic” in: *Annals of Internal Medicine* (2020).

<sup>32</sup> Acemoglu, Chernozhukov, et al., *Optimal Targeted Lockdowns in a Multi-Group SIR Model* (2020); Silva et al., “COVID-ABS: An Agent-Based Model of COVID-19 Epidemic to Simulate Health and Economic Effects of Social Distancing Interventions” in: *Chaos, Solitons & Fractals* (2020).

<sup>33</sup> Shea et al., “Harnessing Multiple Models for Outbreak Management” in: *Science* (2020).

# CHAPTER 5

## ASSESSING THE IMPACT OF NON-PHARMACEUTICAL INTERVENTIONS ON SARS-CoV-2 TRANSMISSION IN SWITZERLAND

Following the rapid dissemination of COVID-19 in Switzerland, large-scale non-pharmaceutical interventions (NPIs) were implemented by the cantons and the federal government between February 28 and March 20, 2020. Estimates of the impact of these interventions on SARS-CoV-2 transmission are critical for decision making in the present and future outbreaks. This chapter aims to assess the impact of these NPIs on disease transmission by estimating changes in the basic reproduction number  $R_0$  at national and cantonal levels in relation to the timing of these NPIs. For the whole country and eleven cantons, the time-varying  $R_0$  is estimated by fitting a stochastic transmission model explicitly simulating within-hospital dynamics. The dataset includes individual-level data from more than 1000 hospitalized patients in Switzerland and public daily reports of hospitalizations and deaths. The national  $R_0$  is estimated to be 2.8 (95% confidence interval 2.1–3.8) at the beginning of the epidemic. Starting from around March 7, a strong reduction of the time-varying  $R_0$  is found, with an 86% median decrease (95% quantile range [QR] 79–90%) to a value of 0.40 (95% QR 0.3–0.58) in the period of March 29 to April 5. At the cantonal level,  $R_0$  decreased by between 53% and 92% over the course of the epidemic. Reductions in time-varying  $R_0$  were synchronous with changes in mobility patterns as estimated through smartphone activity, which started before the official implementation of NPIs. Most of the reduction of transmission is inferred to be attributable to behavioral changes as opposed to natural immunity, the latter accounting for only about 4% of the total reduction in effective transmission. As Switzerland considers relaxing some of the restrictions of social mixing, current estimates of time-varying  $R_0$  well below one are promising. However, as of April 24, 2020, at least 96% (95% QR 95.7–96.4%) of the Swiss population remains susceptible to SARS-CoV-2. These results warrant a cautious relaxation of social distance practices and close monitoring of changes in both the basic and effective reproduction numbers.

This chapter is based on: Joseph C. Lemaitre, Javier Perez-Saez, Andrew S. Azman, Andrea Rinaldo, and Jacques Fellay. “Assessing the Impact of Non-Pharmaceutical Interventions on SARS-CoV-2 Transmission in Switzerland”. In: *Swiss Medical Weekly* 150.2122 (May 30, 2020), and J. Perez-Saez shares co-authorship of the work. It is referred in the following as the postprint (and its supplementary information as s1).

### INTRODUCTION

As of May 13, 2020, the ongoing coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has resulted in more than 4.1 million cases and 280,000 deaths globally<sup>1</sup>. Independent estimates of the basic reproduction number  $R_0$  for SARS-CoV-2 from the initial phases of the epidemic in China, Europe, and the United States have generally ranged from 2–3<sup>2</sup> with doubling times on the order of 2–4 days. In response to the rapid increase in reported cases and hospitalizations, most countries have

<sup>1</sup> WHO, *WHO Situation Report 90* (2020).

<sup>2</sup> Riou et al., “Pattern of Early Human-to-Human Transmission of Wuhan 2019 Novel Coronavirus (2019-nCoV), December 2019 to January 2020” in: *Eurosurveillance* (2020).

implemented non-pharmaceutical interventions (NPIs), including compulsory face mask use, border and school closures, quarantine of suspected and confirmed cases, up to population-wide home isolation<sup>3</sup>. An observation study conducted in Hong Kong during this pandemic estimated that social distancing measures and school closures reduced COVID-19 transmission, as characterized by the effective reproductive number, by 44%<sup>4</sup>. Comparable reductions have been observed in many different settings<sup>5</sup>. Making decisions around relaxing NPIs requires both a careful assessment of the level of pre-relaxation transmission (e.g.,  $R_0$ ) and quantification of the expected increase in transmission from the relaxation of different NPIs.

From the initial reported case on February 24 to April 29, Switzerland reported more than 24'400 laboratory confirmed COVID-19 cases and 1'408 official deaths affecting all 26 cantons<sup>6</sup>. The federal government issued a series of special decrees from February 28 banning gatherings of more than 1'000 people culminating on March 20 recommended home isolation (fig. 5.1). One month after the first NPI, daily confirmed case incidence had decreased from a peak of more than 1000 to a daily average of under 170 in the week of April 20–26 (fig. 5.1). Initial reports have suggested a basic reproduction number ( $R_0$ ) of 3.5 at the start of the epidemic, with a decrease of 85% by March 20<sup>7</sup>. However, these estimates, part of

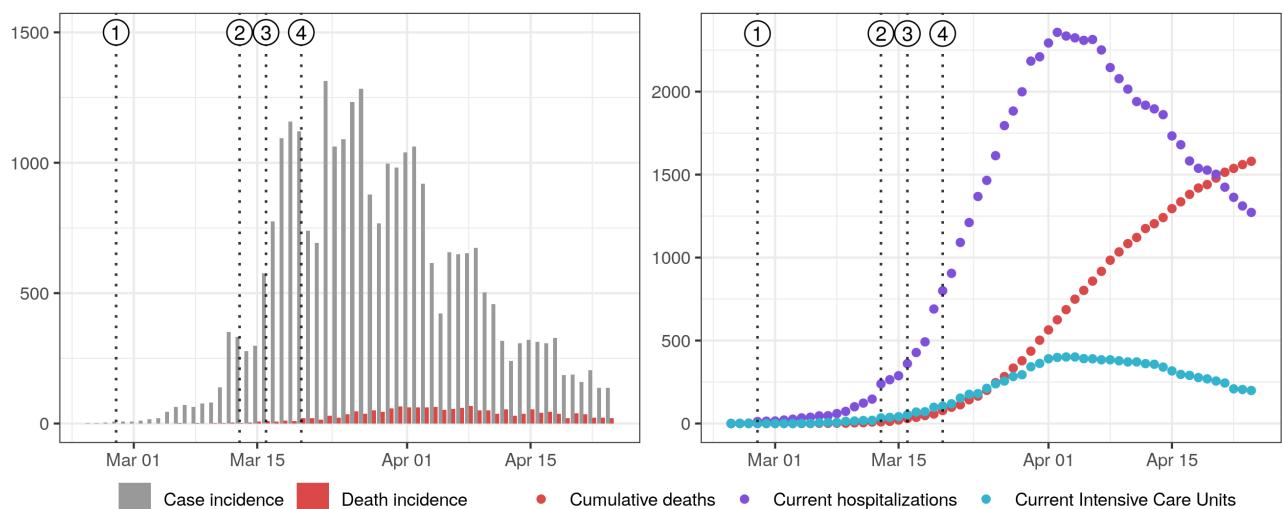
<sup>3</sup> HIT COVID Team, *Health Interventions Tracking for COVID-19 (HIT-COVID)* (2020).

<sup>4</sup> Cowling et al., “Impact Assessment of Non-Pharmaceutical Interventions against Coronavirus Disease 2019 and Influenza in Hong Kong: An Observational Study” in: *The Lancet Public Health* (2020).

<sup>5</sup> Flaxman et al., *Report 13: Estimating the Number of Infections and the Impact of Non-Pharmaceutical Interventions on COVID-19 in 11 European Countries* in: *35* (2020).

<sup>6</sup> FSP, *Rapport sur la situation épidémiologique en Suisse et dans la Principauté de Liechtenstein* (2020).

<sup>7</sup> Flaxman et al., *Report 13: Estimating the Number of Infections and the Impact of Non-Pharmaceutical Interventions on COVID-19 in 11 European Countries* in: *35* (2020).



a multicountry analysis of NPIs, relied on death incidence and did not account for specifics of the hospitalization processes in Switzerland. Moreover, changes in  $R_0$  were assumed to happen on the date of a NPI implementation, thus not allowing for the exploration of the relative timing between NPIs and changes in transmission. Furthermore, such delays likely bias the estimate  $R_0$ .

NPIs affecting daily activities such as school closures and gathering bans aim at having a direct impact on mobility patterns to reduce potentially infectious social contacts. In other terms, the causal pathway from NPIs to transmission reduction is mediated by changes in mobility. Recent releases of mobility data from smartphone software providers give the possibility to study the associations between the implementation of movement-limiting measures, behavioral change, and the related changes in  $R_0$ . Context-specific data on the degree and speed of compliance with these types of NPIs and associations with the observed decreases

fig. 5.1: COVID-19 epidemic curve in Switzerland and timing of non-pharmaceutical interventions. Dotted lines indicate the issuing of NPIs: (1) ban on gatherings of more than 1'000 people, (2) school closure, (3) closure of non-essential activities, and (4) ban on gatherings of more than five people. Left: Daily case incidence along with death incidence. Right: Current hospitalizations, intensive care units (ICUs), and cumulative deaths. Data from Probst (2020), and may therefore present inconsistencies with official reports from the Federal Office of Public Health.

in  $R_0$  could inform scenario-building should the future reinstatement of measures become necessary. Given that  $R_0$  directly represents transmission potential, its quantification enables us to estimate the proportion of reduction in transmission attributable to behavioral changes. In this sense, tracking of  $R_0$  is more suited to study the impact of NPIS than the effective reproduction number  $R_{eff}$ , which is an aggregate measure of transmission capturing aspects of both infectious contacts and of population susceptibility.

Here, the goal is to estimate the changes in  $R_0$  throughout the epidemic at both the national and cantonal levels using detailed data on hospitalizations and deaths from Switzerland between February 24 and April 24. In order to understand the estimated changes in  $R_0$ , its relationship with the timing of NPIS and human mobility estimates derived from cell phone data is explored.

## METHODS

The setting for the present chapter arose while providing scenario modeling reports, as presented in CHAPTER 4. CHUV, the hospital coordinating Canton de Vaud's response, provided access to individual hospitalization data. The dataset and pre-processing steps to account for right-censoring are presented in the APPENDIX TO CHAPTER 5. This dataset allowed for refined estimates of stays in hospital and death processes, which in turn were used as assumptions in a COVID-19 model; giving said model sufficient stiffness to estimate  $R_0$  in time. Model equations and fixed and inferred parameter values are left in the postprint SI.

## MODEL AND ASSUMPTIONS

**MODEL STRUCTURE** A stochastic compartmental model of the COVID-19 epidemic and hospitalization processes is developed for each canton of Switzerland. The model is structured around the classical S, E, I, and R compartments<sup>8</sup>, and the model diagram is shown in fig. 5.2. A susceptible individual  $S$  might be exposed after contact with infectious  $I$  individuals. Upon exposure, a formerly susceptible individual goes through an incubation period  $E$  before becoming infectious  $I$ . The individual then recovers  $R$  and does not participate in transmission anymore. In addition to these dynamics, infected individuals have some probability of developing severe symptoms. Estimates derived from data from the Canton de Vaud, as shown in the APPENDIX, show a high proportion of deaths outside of hospitals ( $\approx 50\%$ ) hence two pathways are modeled depending if the individual seek or has access to hospital care. Some severely infected will be treated in hospitals after a delay from symptom onset  $I_b$ . In this case, hospitalization leads to discharge (recovery) or death, either through normal hospitalization ( $H_s$  and  $H_d$  respectively) or passing through Intensive Care Units (ICUs, compartments  $U_s$  and  $U_d$  respectively). Otherwise, the severely infected may recover or die without passing through hospitalization, going into compartment  $I_d$ . There is no need to explicitly model a latent stage where individuals are still asymptomatic but infectious<sup>9</sup> as only hospitalisation and death data is used. The model is implemented as a hidden Markov model using the pomp R package<sup>10</sup>.

**ASSUMPTIONS** The reader is referred to the postprint SI for the exhaustive description of model transitions and parameters. As parameter identifiability is needed to capture the dynamics of  $R_0$ , most of the parameters were fixed to

<sup>8</sup> Kermack et al., “A Contribution to the Mathematical Theory of Epidemics” in: *Proceedings of the Royal Society A* (1927).

<sup>9</sup> Ganyani et al., “Estimating the Generation Interval for COVID-19 Based on Symptom Onset Data” in: *medRxiv* (2020); He et al., “Temporal Dynamics in Viral Shedding and Transmissibility of COVID-19” in: *Nature Medicine* (2020); Liu et al., “The Contribution of Pre-Symptomatic Infection to the Transmission Dynamics of COVID-2019” (2020).

<sup>10</sup> King, Nguyen, et al., *Statistical Inference for Partially Observed Markov Processes via the R Package Pomp* (2015).

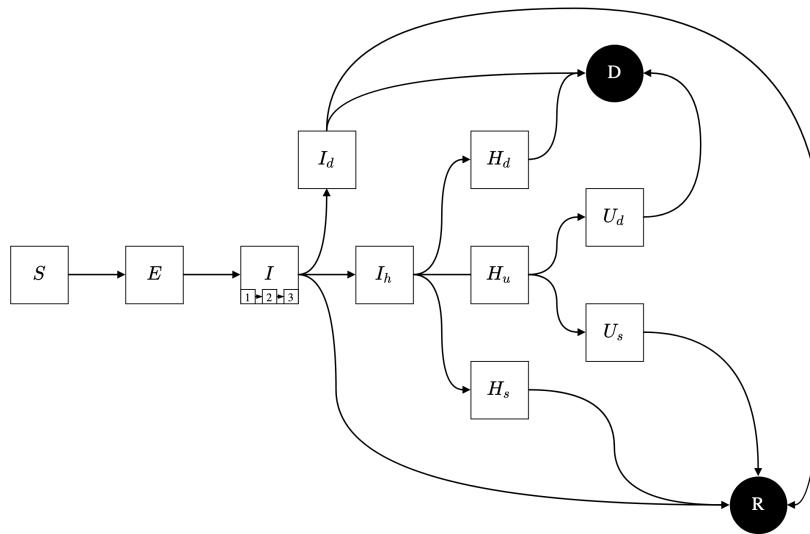


fig. 5.2: Schematic diagram of COVID-19 transmission and hospitalization processes. There are two sinks: Death  $D$  and recovered  $R$ . Each stage with regard to the disease may be implemented with several compartments (subscript numbered boxes) to better represent the time distribution spent in that stage.

values from the literature or obtained analyzing the data from Canton de Vaud<sup>11</sup>. The transmission model is parameterized assuming a mean generation time of 5.2 days<sup>12</sup>, and an exposed and non-infectious duration of 2.9 days<sup>13</sup>, yielding a mean duration of 4.6 days in the infectious compartments. It is assumed that 7.5% of infections were severe and would require hospitalization, that 50% of deaths happened outside of hospitals (from data on Canton de Vaud described in the APPENDIX, and data from Geneva collected from OpenZH), that 16% of those hospitalized would die (data from Vaud, see APPENDIX), and that the infection fatality ratio (IFR) was 0.75%, which is in the range of published estimates<sup>14</sup>. Individual-level data on hospitalised cases from the canton of Vaud is used to estimate the distribution of time spent in the hospital and in the intensive care unit (ICU). Times to discharge and death are estimated using survival models that account for right-censoring of observations (see APPENDIX). The number of compartments for the linear-chain trick (*i.e.* the shape parameter for the erlang distributed residence time) for each observable hospitalization state was obtained by fitting Erlang distributions to the data of Canton de Vaud. To account for right-censoring the model is fitted to the estimated log-normal distributions described in the APPENDIX instead directly to observed times to events but rather. The rate parameter of the Erlang distributions is calibrated for shape parameters between 1 and 10 by minimizing the Kullback-Leibler (KL) divergence between the Erlang and estimated log-normal distributions. The final fit was taken to be the one with the smallest KL-divergence. We found that all hospitalization processes were best represented with exponential distributions (Erlang with a shape parameter of 1)<sup>15</sup>.

## DATA AND INFERENCE

**DATASET** Curated data from OpenZH<sup>16</sup> up to April 24 is used. This dataset included, by canton, the number of hospitalised COVID patients, and the cumulative numbers of deaths, cases and hospital discharges; the latter not available for all cantons. The cantonal estimate is focused on cantons that had enough cases and data to obtain meaningful results, keeping 11 of the 26 cantons (Bern,

<sup>11</sup> As noted by Gostic et al., the accuracy of  $R_0$  estimates obtained with the method presented here is sensible to assumptions on model structure and parameters; see Gostic et al., “Practical Considerations for Measuring the Effective Reproductive Number,  $R_t$ ” in: *PLOS Computational Biology* (2020).

<sup>12</sup> Ganyani et al., “Estimating the Generation Interval for COVID-19 Based on Symptom Onset Data” in: *medRxiv* (2020).

<sup>13</sup> He et al., “Temporal Dynamics in Viral Shedding and Transmissibility of COVID-19” in: *Nature Medicine* (2020).

<sup>14</sup> Verity et al., “Estimates of the Severity of Coronavirus Disease 2019: A Model-Based Analysis” in: *The Lancet Infectious Diseases* (2020); Russell et al., “Estimating the Infection and Case Fatality Ratio for Coronavirus Disease (COVID-19) Using Age-Adjusted Data from the Outbreak on the Diamond Princess Cruise Ship, February 2020” in: *Eurosurveillance* (2020).

<sup>15</sup> In the first preprinted version without the survival analysis, the shape parameters were estimated to be of 2 for many compartments. It highlights the importance of reasoning about potential biases and of careful data pre-processing.

<sup>16</sup> openZH, *openZH/Covid\_19* (2020).

Basel-Landschaft, Basel-Stadt, Fribourg, Geneva, Jura, Neuchâtel, Ticino, Vaud, Valais and Zurich). These cantons account for 66% of the Swiss population. The national estimate uses curated national aggregate data and thus encompassed all cantons<sup>17</sup>. Unknown parameters of the model are fitted using maximum likelihood inference through iterated filtering<sup>18</sup>. No attempt to include confirmed case data into the observation model were made because of the heterogeneous testing strategies adopted across cantons and over time. The model is therefore fitted to death incidence and changes in current hospitalisations using appropriate likelihood functions:

$$\begin{aligned} \text{deaths}(t) &\sim \text{Poisson}(\Delta D(t)) \\ \Delta \text{hosp}(t) &\sim \text{Skellam}(\Delta H(t), \Delta D_H(t) + \Delta R_H(t)), \end{aligned} \quad (5.1)$$

where,  $\Delta D(t)$ ,  $\Delta H(t)$ ,  $\Delta D_H(t)$ ,  $\Delta R_H(t)$  are respectively the number of new deaths, hospitalized, and deaths and discharged from hospitals at time  $t$ , and  $\Delta \text{hosp}(t)$  is the difference between the number of current hospitalizations at times  $t$  and  $t - 1$ , for which a Skellam distribution<sup>19</sup> is chosen. The full log-likelihood of the observation model was taken as the sum of the individual log-likelihoods of the  $\Delta \text{hosp}(t)$  and of the  $\text{deaths}(t)$ . The model is calibrated separately for each canton on the daily death and hospitalization until April 24. Hospitalization incidence data at the cantonal level would have provided more information, but this data was not accessible. Fitting to hospitalization incidence data, which access was granted in the canton of Vaud, yielded similar results to those when changes in current hospitalizations were used. The majority of model parameters were either derived from Vaud data or the literature (see postprint s1), which enables the identifiability of the reproduction number  $R_0$ .

**REPRODUCTION NUMBER ESTIMATION** Time-varying basic reproduction numbers  $R_0$  were estimated following a similar approach to Cazelles et al.<sup>20</sup>, recently applied to COVID-19 transmission in Hubei<sup>21</sup>. The method aims at inferring the time series of  $R_0$ <sup>22</sup> that yields model dynamics that are in best agreement with the whole set of available observations. As such, the value of  $R_0$  at a given point in time is informed by the whole data, and therefore does not have the limitations of being either “forward-looking” or “backwards-looking” as it is the case of commonly used statistical methods applied for this purpose<sup>23</sup>.

The model equations are similar to the previously presented partially-observed Markov Processes models, and are left in the postprint s1. The difference lies in the force of infection. Given the state of the system at time  $t$ ,  $\mathcal{X}_t$ , and using the same notations as in CHAPTER 3, the transition  $S \rightarrow E$  reads:

$$\mathbb{P} [\Delta N_{SE}(t) = 1 \mid \mathcal{X}_t] = \underbrace{\beta(t) \frac{I_1(t) + I_2(t) + I_3(t)}{P}}_{\text{Force of infection}} S(t) \Delta t + o(\Delta t). \quad (5.2)$$

Time-varying  $R_0(t) = \beta(t)/(3r_I)$  is modelled as a geometric random walk defined by its calibrated variance, where  $\beta$  is the transmission parameter and  $1/(3r_I)$  is the mean duration spent in the infectious compartments  $I_1$  to  $I_3$ . Once the time series is inferred, the timing and slope of changes in  $R_0$  are evaluated by using linear changepoint models<sup>24</sup>. The null model corresponded to a linear decrease between two plateaus corresponding to the baseline value at the start of the epidemic and a low value after the implementation of NPIs. To allow for different slopes in

<sup>17</sup> Probst, Daenuprobst/Covid19-Cases-Switzerland (2020).

<sup>18</sup> Ionides et al., “Inference for Dynamic and Latent Variable Models via Iterated, Perturbed Bayes Maps” in: *Proceedings of the National Academy of Sciences* (2015).

<sup>19</sup> which represents the difference between two independent random variables, each Poisson-distributed. See Skellam, “The Frequency Distribution of the Difference between Two Poisson Variates Belonging to Different Populations” in: *Journal of the Royal Statistical Society. Series A (General)* (1946).

<sup>20</sup> Cazelles et al., “Accounting for Non-Stationarity in Epidemiology by Embedding Time-Varying Parameters in Stochastic Models” in: *PLOS Computational Biology* (2018).

<sup>21</sup> Kucharski, Russell, et al., “Early Dynamics of Transmission and Control of COVID-19: A Mathematical Modelling Study” in: *The Lancet Infectious Diseases* (2020).

<sup>22</sup> As mentioned,  $R_0$ , not  $R_{eff}$  is computed as the presented method estimates the basic reproduction number, *i.e.* the expected number of infections generated by one infected individual in a fully susceptible population. To do so, the susceptible and recovered populations are explicitly modeled while other methods usually compute the effective reproduction number in the population by deconvolution of the observed incidence.

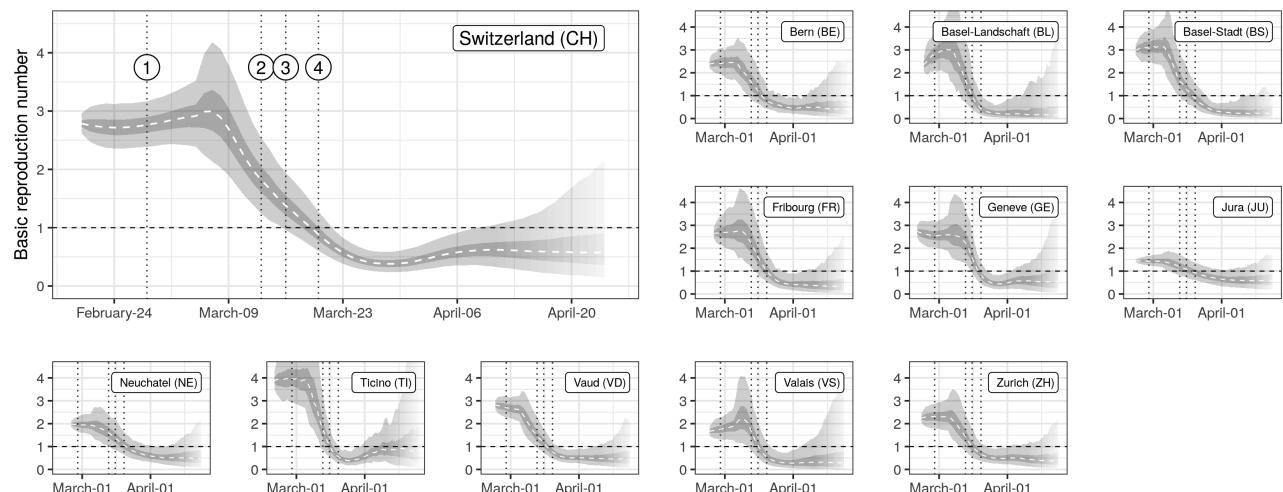
<sup>23</sup> Wallinga et al., “Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures” in: *American Journal of Epidemiology* (2004); Cori et al., “A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics” in: *American Journal of Epidemiology* (2013); Lipsitch, Joshi, et al., “Comment on Pan A, Liu L, Wang C, et al. Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China.” (2020).

<sup>24</sup> Lindeløv, *Mcp: An R Package for Regression With Multiple Change Points* (2020).

the decreasing phase of  $R_0$ , models with one and two additional breakpoints (corresponding to two and three different slopes) are also fitted, and the best model was selected using Bayesian model selection based on leave-one-out cross-validation (details in section 6 of the postprint s1).

The estimated changes in  $R_0$  were contrasted with changes in activity-related mobility data produced by Google<sup>25</sup>. Changes in activity were expressed as relative changes with respect to a baseline computed as the median over a 5-week period from January 3 to February 6. Mobility changes were computed for different categories: grocery and pharmacy, parks, transit stations, retail and recreation, residential, and workplace. Mobility estimates were based on smartphone-based geo-location location data (GPS, WiFi connections, Bluetooth) from users who activated Location History for their Google Account. These data were used to determine changes in the number of visits to and length of stays in locations categorized into the above-mentioned types. The dataset therefore only covers a sample of the Swiss population who use smartphones, the latter representing around 80% of the total population in 2020<sup>26</sup>. Gaps in the dataset were filled by linear interpolation and a 7-day moving average was applied to smooth out weekly seasonality in activities. The cross-correlation between changes in  $R_0$  and the averaged changes in each type of activity was computed with lags up to 10-days. Changes in  $R_0$  were computed based on location-specific baselines taken as the mean value of  $R_0$  from the beginning of the simulations, 5 days before the first reported case in each canton, until March 8. Changepoint models are employed to identify dates of change in mobility patterns and in  $R_0$ .

## RESULTS



Over the study period,  $R_0$  trends follow a common trajectory nationally and across cantons, starting with a high plateau ( $R_0 > 2$ ) in the early stage of the epidemic followed by a rapid reduction starting at the beginning of March, and reaching a low and stable value ( $R_0 < 1$ ) from the end of March onwards (fig. 5.3).

At the beginning of the epidemic,  $R_0$  is estimated at 2.8 (95% confidence interval [CI] 2.06–3.83) at the national level, with cantonal-level values ranging from 2.5 to 3.1 (postprint s1 tab. 5). The onset of the reduction was estimated to be be-

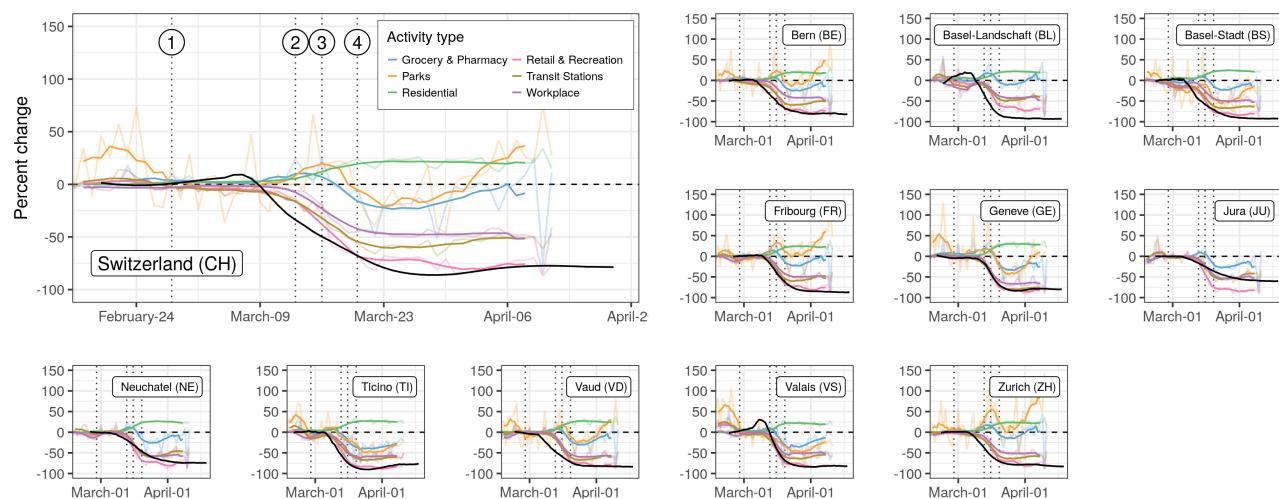
<sup>25</sup> Google LLC, *Google COVID-19 Community Mobility Report* (2020).

<sup>26</sup> O'Dea, *Smartphone Users in Switzerland 2018–2024* (2020).

All data and code except for individual hospitalization data from the canton of Vaud have been deposited on Zenodo ([doi.org/10.5281/zenodo.3862075](https://doi.org/10.5281/zenodo.3862075)). The change point analysis, model fit and additional results are omitted from this thesis and may be found in the supplementary information of (Lemaitre, Perez-Saez, et al. 2020).

fig. 5.3: Estimates of changes in the basic reproduction number  $R_0$ . Median (dashed line), IQR (dark gray), and the 95% QR (light gray) of the estimated time series of  $R_0$  are shown for each canton. Vertical dotted lines indicate the issuing of NPIs as described in fig. 5.1. Transparency at the end of the time series indicates increasing uncertainty (style inspired by the reports of CMMID).

tween March 4 (Basel-Stadt and Vaud) and March 11 (Geneva and Valais) at the cantonal level and on March 7 at the national level (postprint s1 fig. II). Overall strong support is found for the reduction in  $R_0$  starting before school closures on March 13 (probability 0.99 at the national level, postprint s1 fig. II). Once started, a strong decrease in  $R_0$  is estimated at the national (reduction of 0.16/day) and cantonal levels (between 0.14/day in Jura to 0.18/day in Basel-Landschaft) (fig. 5.3). No strong support was found for changes of the slope during the decrease phase at either at the national or cantonal levels except for Bern, Basel-Stadt, and Vaud, for which additional changes in slopes were inferred towards the stabilization of  $R_0$  at low values (postprint s1 tab. 7).  $R_0$  in Switzerland has been estimated to drop below 1 on March 19 (95% CI March 16–22) with individual cantons meeting this threshold between March 16 (Basel-Stadt) and March 20 (Neuchâtel) (postprint s1 fig. 10). The probability that  $R_0$  had already fallen below one was low when schools closed on March 13 (national 0.006, cantonal from 0 in Geneva to 0.23 in Basel-Landschaft), and high by the time gatherings of five people or more were banned on March 20 (national 0.92, cantonal from 0.52 in Neuchâtel to 0.99 in Ticino) (postprint s1 fig. 9). The estimated plateau value of  $R_0$  after the reduction, that is, from March 29 to April 10, was 0.4 (95% quantile range [QR] 0.3–0.6) at the national level, with median values at the cantonal level ranging from 0.2–0.7 (postprint s1 tab. 5). At the national level,  $R_0$  was reduced by 86% (95% QR 79–90%), with median reductions ranging from 53% (Jura) to 92% (Basel-Stadt) at the cantonal level. A gradual reduction in  $R_0$  leading to values below one around the third week of March is consistent with the observed reduction of confirmed case incidence in early April, when taking into consideration the delays due to the incubation period, with a median of 5.2 days<sup>27</sup>, and between symptom onset and reporting<sup>28</sup>. Similarly, the inflection in the number of current hospitalizations and ICU usage in early April also supports  $R_0$  dropping below one in mid-March.



Activity-related mobility patterns changed markedly in all cantons since the beginning of the epidemic (fig. 5.4). Mobility related to work, retail and recreation, and transit stations dropped by 50% to 75% at the national level, with cantonal-level reductions ranging from 30% to 80% depending on activities. Residential-related mobility increased across cantons between 20% and 30%. Strong support

Additional results may be found in the supplementary information of Lemaitre, Perez-Saez, et al., “Assessing the Impact of Non-Pharmaceutical Interventions on SARS-CoV-2 Transmission in Switzerland” in: *Swiss Medical Weekly* (2020).

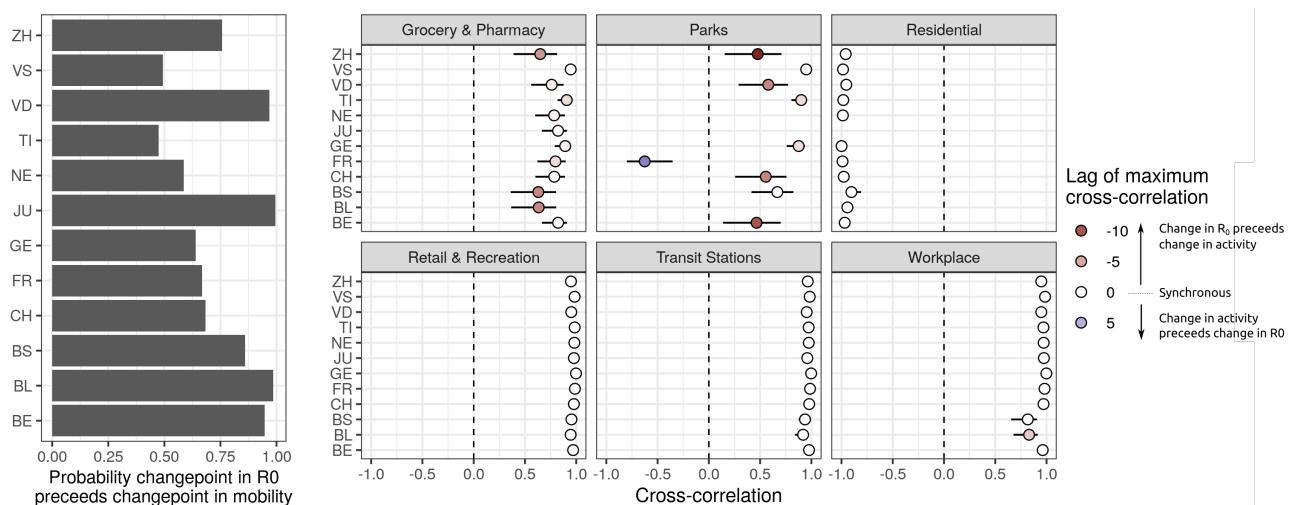
<sup>27</sup> Lauer, Grantz, et al., “The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application” in: *Annals of Internal Medicine* (2020).

<sup>28</sup> Bi, Wu, et al., “Epidemiology and Transmission of COVID-19 in 391 Cases and 1286 of Their Close Contacts in Shenzhen, China: A Retrospective Cohort Study” in: *The Lancet Infectious Diseases* (2020).

fig. 5.4: Changes in mobility patterns and  $R_0$ . Changes in mobility with respect to baseline are shown by activity type in terms of the daily values (transparent lines) and 7-day rolling mean (full lines), against the median estimate of  $R_0$  (black line). Vertical dotted lines indicate the issuing of NPIs as described in fig. 5.1.

is found for mobility changes starting simultaneously for all activity types within each canton. Changes in mobility are estimated to have started between March 6 to 14 for all cantons (postprint s1 fig. 12), thus finding strong support for changes starting before school closure on March 13 (national-level mean probability across activities 0.70, cantonal range 0.55–0.99).

Based on the changepoint models, reductions in  $R_0$  likely started (probability 0.76) before observed reductions in mobility at the national level and across cantons (fig. 5.4). Changes in  $R_0$  were highly correlated with changes in mobility, the strongest associations being with mobility related to work, transit stations, retail and recreation, and residential (cross-correlations >0.9 in all cantons and nationally, fig. 5.5). In the majority of cases, the correlation between mobility and



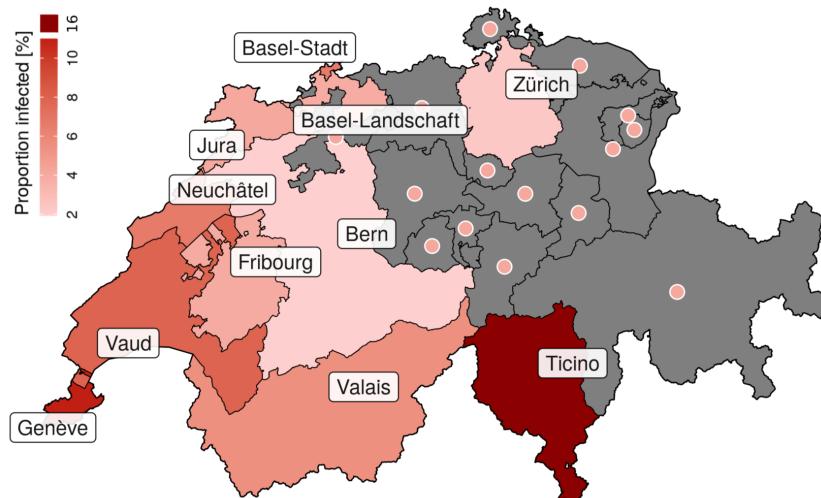
$R_0$  was strongest with no lag between the two. However, changes in mobility to workplaces lagged behind changes in  $R_0$  in Basel-Stadt and Basel-Landschaft. Correlations between changes in  $R_0$  and grocery and pharmacy mobility were less marked (national level 0.65), with changes in mobility occurring after changes in  $R_0$  (negative lags in fig. 5.5). In most cantons, a strong increase in park mobility after March 25 resulted in a positive correlation with changes in  $R_0$ , but with negative lags (change in activity after change in  $R_0$ , fig. 5.5). The linear associations between the level of reduction in mobility and maximum reduction in  $R_0$  across cantons were found to be not significant, except for a small effect size for reduced park mobility (regression coefficient of 0.15, 95% CI 0.02–0.25) (postprint s1 fig. 8).

The effective reproduction number  $R_{eff}$ , a common measure which is an aggregate measure of transmission capturing aspects of both infectious contacts and of population susceptibility was estimated. Across cantons,  $R_{eff}$  was extremely close to  $R_0$ , indicating that a small fraction of the population is expected to have a natural immunity to SARS-CoV-2. As of April 24, an estimated 3.9% (95% QR 3.6–4.3%) of the population nationally had been infected, with median estimates ranging from 1.9% (Bern) to 16% (Ticino) (fig. 5.6). Modeled estimates of the proportion infected of people infected in the canton of Geneva are in agreement with preliminary results from ongoing serological studies, which have estimated the seroprevalence to be 9.7% (95% CrI 6.1–13.1%) in the third week of April<sup>29</sup>, compared with modeled estimates of 8.9% (95% QR 7.8–10.1%) after accounting for

fig. 5.5: Timing between changes in  $R_0$  and mobility. Left: the probability that the first changepoint in  $R_0$  occurred before the first changepoint in mobility-related activity. Right: Maximum cross-correlations between time series of changes in  $R_0$  and changes in mobility (bars 95% CI). Lags refer to the delay between changes in mobility-related activity and changes in  $R_0$  (positive lag  $k$  indicates that current changes in mobility have maximal cross-correlation with changes in  $R_0$   $k$  days in the past).

<sup>29</sup> Stringhini et al., “Repeated Seroprevalence of Anti-SARS-CoV-2 IgG Antibodies in a Population-Based Sample from Geneva, Switzerland” in: *medRxiv* (2020).

the time from infection to seroconversion<sup>30</sup>(postprint s1 fig. 7).



<sup>30</sup> Wölfel et al., “Virological Assessment of Hospitalized Patients with COVID-2019” in: *Nature* (2020).

fig. 5.6: Modelled proportion of people infected with SARS-CoV-2 in Switzerland up to April 24. Estimates were produced for 11 of 26 cantons for which enough data were available (unmodelled cantons are shown in gray with points indicating the national-level estimated incidence proportion of 3%). Values are reported in the postprint s1 tab. 6.

## DISCUSSION

Our results suggest a strong reduction of  $R_0$  across Switzerland since the start of the epidemic. The reduction in  $R_0$  started around March 7, thus about 1 week before the implementation of lockdown-type NPIs. Analysis of activity-related mobility data also showed strong support for changes in mobility starting before the implementation of most NPIs. Estimated reductions of viral transmission were strongly correlated and mostly synchronous with observed changes in mobility patterns, although the initiation of changes in transmission preceded measurable changes in activity-related mobility.

The methods used to infer the time series of  $R_0$  do not rely on assumptions on the shape of how it changed in time, nor on the dates at which changes started. Alternative methods that rely on fixed dates (such as that of the Imperial College COVID-19 Response Team<sup>31</sup>) might be biased as changes in transmission are not synchronous with policy changes. Distribution based methods such as provided by R package EpiEstim<sup>32</sup> are flexible but subject to bias when misused<sup>33</sup>. In addition, the present approach enables the estimation of  $R_0$ , which is a direct quantification of transmission potential, as opposed to the effective reproduction number  $R_{eff}$ , which also accounts for the effect of susceptible depletion as done in the above-mentioned statistical approaches. This enabled us to estimate the proportion of reduction in transmission attributable to behavioral changes, which is therefore more suited to study the impact of NPIs. Aside from these methodological differences, the present estimates are in line with other estimates in Switzerland: Althaus et al.<sup>34</sup> estimated a reduction of 89 % (83–94%) from a baseline of 2.78 (2.51–3.11), Scire et al.<sup>35</sup> estimated a reduction of 76 % (70–82%) from a baseline of 1.88 (1.80–1.98) and Imperial College estimated a reduction of 60% (50–80%) from a baseline of 3.5 (2.8–4.3)<sup>36</sup>.

The presented results provide strong support for a reduction of transmission starting about 1 week before school closure, the first national-level NPI targeting daily activities, which was ordered on March 13. Moreover, initiation of transmission reduction was found to precede changes in mobility patterns as detectable

<sup>31</sup> Flaxman et al., *Report 13: Estimating the Number of Infections and the Impact of Non-Pharmaceutical Interventions on COVID-19 in 11 European Countries* in: *35* (2020).

<sup>32</sup> Wallinga et al., “Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures” in: *American Journal of Epidemiology* (2004); Cori et al., “A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics” in: *American Journal of Epidemiology* (2013).

<sup>33</sup> Lipsitch, Joshi, et al., “Comment on Pan A, Liu L, Wang C, et al. Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China.” (2020).

<sup>34</sup> Althaus, *Real-Time Modeling and Projections of the COVID-19 Epidemic in Switzerland* (2020).

<sup>35</sup> Scire et al., “Reproductive Number of the COVID-19 Epidemic in Switzerland with a Focus on the Cantons of Basel-Stadt and Basel-Landschaft” in: *Swiss Medical Weekly* (2020).

<sup>36</sup> Flaxman et al., *Report 13: Estimating the Number of Infections and the Impact of Non-Pharmaceutical Interventions on COVID-19 in 11 European Countries* in: *35* (2020).

from the Google dataset. A possible explanation for this initial decrease in transmission could be linked to the strong increase in public interest in COVID-19 in February as measured by Google searches for COVID-19-related keywords (fig. 5.7).

In fact, a second sharp rise in Google searches is estimated to have started on March 7 (95% CrI March 3–9), which overlaps with the estimated start of the national level decrease in  $R_0$  on March 7 (probability that changepoints coincide 0.76). The Federal Office for Public Health issued an information campaign on COVID-19 on February 28, which was updated on March 2 to stress basic hygiene rules<sup>37</sup>. This may have resulted in voluntary social distancing as well as increased hygiene early on without noticeable changes in mobility patterns. This type of proactive change in behavior would be in line with early changes in mobility patterns, which were estimated to precede school closures and subsequent measures. The present results suggest that the value of  $R_0$  was likely already below one on March 20, when the federal government banned gatherings of more than five people and recommended voluntary home isolation for the whole population. This result should however be taken within context, as the announcement was anticipated on social networks earlier that week, and so was probably already impacting social distancing behavior. Therefore caution is recommended in any causal interpretation of these results on the role of this last NPI on driving  $R_0$  below one.

Despite the strong association between the changes in mobility and reductions in  $R_0$  within each canton, the lack of cross-cantonal associations between the level of reduction in mobility types and the level of reduction in  $R_0$  suggests context-specific pathways between COVID-19 transmission and mobility intensity. These warrants caution in attempting to apply general relations between mobility and transmission reduction. Investigation of general associations will require more in-depth studies controlling for other factors such as population density, economic activities and social mixing patterns, and inter-cantonal mobility patterns, in addition to the incorporation of potential environmental drivers of transmission such as temperature and relative humidity<sup>38</sup>.

Several limitations to this work are noted. First, due to the relatively recent introduction of SARS-CoV-2 in Switzerland compared with the length of hospital and ICU stays, the time distribution of hospital in- and out-patients is biased towards shorter duration (see APPENDIX), which is addressed by accounting with right-censoring using survival models. In addition, because of the limited data available in some places, it was only possible to fit the model for 11 of the 26 cantons. Modeling results presented in this work are subject to hypothesis on yet uncertain parameters of COVID-19, including the infection fatality rate and the proportion of severe infections requiring hospitalization. An important uncertainty is the fraction of asymptomatic infections and their relative contribution to disease transmission. It is assumed that all infected individuals contribute equally to transmission, which means the estimation of the proportion of people infected would under-estimate true cumulative incidence if there were a large fraction of asymptomatic infections with a relatively low contribution to transmission. Evidence from South Korea, however, suggests that only a small fraction (2%) of confirmed COVID-19 infections are asymptomatic, and none of the household members of these asymptomatic carriers were infected<sup>39</sup>. Moreover, model results are in agreement with preliminary results from ongoing serological studies in Switzerland<sup>40</sup>. The presented estimates of time-varying basic reproduction num-

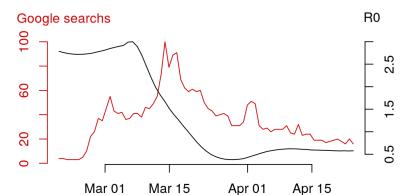


fig. 5.7: Google trends for COVID-19 and changes in  $R_0$  in Switzerland. Trends corresponds to the amount of searches for the keyword “coronavirus” (red line) between February 15 and April 30 and are given as a percent of the maximum number of searches in the period, the time evolution of  $R_0$ .

<sup>37</sup> OFSP, *Nouvelles Règles d’hygiène et de Conduite Pour Se Protéger Contre Le Nouveau Coronavirus* (2020).

<sup>38</sup> Neher et al., “Potential Impact of Seasonal Forcing on a SARS-CoV-2 Pandemic” in: *Swiss Medical Weekly* (2020); Kissler et al., “Projecting the Transmission Dynamics of SARS-CoV-2 through the Postpandemic Period” in: *Science* (2020).

<sup>39</sup> Park, Kim, et al., “Early Release - Coronavirus Disease Outbreak in Call Center, South Korea - Volume 26, Number 8—August 2020 - Emerging Infectious Diseases Journal - CDC” (2020).

<sup>40</sup> Stringhini et al., “Repeated Seroprevalence of Anti-SARS-CoV-2 IgG Antibodies in a Population-Based Sample from Geneva, Switzerland” in: *medRxiv* (2020).

bers assume that the generation interval for COVID-19 in Switzerland remained unchanged, thus potentially ignoring the joint role of  $R_0$ , the infectious period, and contact rates in determining the disease's intrinsic growth rate<sup>41</sup>. If the generation interval increased with the reduction of social contact then these estimates are conservative overestimates of the “true” value of  $R_0$ , which is encouraging from a public health perspective. Inferred disease dynamics and estimated time-varying  $R_0$  also depend on the values of the incubation period, which is set to the estimates currently available in the literature. In the present modeling framework, the initial conditions were estimated along with changes in  $R_0$ , which could, however, be influenced by the role of imported cases in driving disease dynamics, especially in cantons bordering regions with strong COVID-19 transmission in early February (Eastern France for Basel-Stadt and Basel-Landschaft and Northern Italy for Ticino). Since importations were not modeled, this could yield an overestimation of the initial value of  $R_0$ , which warrants caution in interpreting specific values of  $R_0$  in these cantons. This potential overestimation would, however, not affect the strong inferred reduction in  $R_0$ . Another limitation of this study is that it was not possible to disentangle the individual contribution of each NPI on  $R_0$  in this analysis owing to the early onset of changes in  $R_0$  and in mobility patterns, as well as the very close spacing between the different types of NPI. This information would, however, be extremely valuable in supporting decisions on NPI strategies against COVID-19. Efforts to constitute a global database of NPIs will provide the opportunity to extend this type of analysis to other settings and produce evidence for the effect of different types of NPIs<sup>42</sup>.

As the Swiss government plans to gradually lift restrictions, close monitoring of changes in  $R_0$  is critical, given that the reductions in transmission appear to be almost entirely driven by changes in behavior, not through herd immunity. Near real-time estimates of  $R_0$  may serve as a critical tool for public health and political decision makers in the months to come, and efforts should be made to refine models like ours using new data, including those from population-based serological studies, mobility data, and more detailed individual-level data on COVID-19 cases across the spectrum of severity.

<sup>41</sup> Yan, “Separate Roles of the Latent and Infectious Periods in Shaping the Relation between the Basic Reproduction Number and the Intrinsic Growth Rate of Infectious Disease Outbreaks” in: *Journal of Theoretical Biology* (2008).

<sup>42</sup> HIT COVID Team, *Health Interventions Tracking for COVID-19 (HIT-COVID)* (2020).

While relying on hospitalization and death allowed for an early robust identification of the basic reproduction number, it would be necessary to add a reporting process and case data to update this estimate through 2020–2021. Methods based on observed data are easier to maintain, and reliable continuous updates of the reproduction number in Switzerland are available on the Swiss National COVID-19 Task Force website [scientifictaskforce.ch/en/current-situation/](http://scientifictaskforce.ch/en/current-situation/). It is provided by the ETHZ, with the method described in Huisman et al., “Estimation and Worldwide Monitoring of the Effective Reproductive Number of SARS-CoV-2” in: *medRxiv* (2021). The modeled incidence from the present CHAPTER has been later used as “truth” to compare waste-water data with reported cases, see: Fernandez-Cassi et al., “Wastewater Monitoring Outperforms Case Numbers as a Tool to Track COVID-19 Incidence Dynamics When Test Positivity Rates Are High” in: *Water Research* (2021).



# CHAPTER 6

## OPTIMIZING THE SPATIO-TEMPORAL ALLOCATION OF COVID-19 VACCINES: ITALY AS A CASE STUDY

The development of vaccines has sparked high hopes towards the control of SARS-CoV-2 transmission without resorting to extensive community-wide restrictions. While vaccine distribution campaigns are underway across the world, communities face the challenge of a fair and effective distribution of limited supplies. One may wonder whether suitable spatial allocation strategies might significantly improve a campaign's efficacy in averting damaging outcomes. To that end, the problem of optimal control of COVID-19 vaccinations is addressed in a country-wide geographic and epidemiological context characterized by strong spatial heterogeneities in transmission rate and disease history. The vaccine allocation strategies in space and time that minimize the number of infections in a prescribed time horizon are searched for. Scenarios of unfolding disease transmission across the 107 provinces of Italy, from January to April 2021, are generated by a spatially explicit compartmental COVID-19 model tailored to the Italian geographic and epidemiological context (Bertuzzo, Mari, Pasetto, et al. 2020; Gatto, Bertuzzo, et al. 2020). A novel optimal control framework is developed to derive optimal vaccination strategies given the epidemiological projections and constraints on vaccine supply and distribution logistics. Optimal schemes significantly outperform simple alternative allocation strategies based on incidence, population distribution, or prevalence of susceptibles in each province. Results suggest that the complex interplay between the mobility network and the spatial heterogeneities imply highly non-trivial prioritization of local vaccination campaigns. By accounting for spatial heterogeneities and human mobility networks, the presented approach complements currently used allocation methods based on criteria such as age or risk. The extent of the improvement grants further inquiry aimed at refining other possibly relevant factors so far neglected. This CHAPTER thus provides a proof-of-concept of the potential of optimal control for complex and heterogeneous epidemiological contexts at country, and possibly global, scales.

This chapter is based on the following preprint: Joseph C. Lemaitre, Damiano Pasetto, Mario Zanon, Enrico Bertuzzo, Lorenzo Mari, Stefano Miccoli, Renato Casagrandi, Marino Gatto, and Andrea Rinaldo. "Optimizing the Spatio-Temporal Allocation of COVID-19 Vaccines: Italy as a Case Study". In: *medRxiv* (May 13, 2021), p. 2021.05.06.21256732

### INTRODUCTION

The deployment of SARS-CoV-2 vaccines is limited in many countries by the available stock and the logistics to distribute the doses<sup>1</sup>. Current prioritization approaches typically target groups at higher risk of severe outcomes<sup>2</sup>, or their indirect protection by vaccinating those with higher disease transmission<sup>3</sup>. The main hypothesis of this chapter is that taking into account spatial heterogeneities in disease transmission when designing prioritization strategies will significantly improve the effectiveness of vaccination campaigns. The distribution of doses inside each country is limited by the logistic capabilities of the healthcare network and the rate at which the vaccine stock is replenished. Decisions concerning the best allocation strategies are to be taken under these constraints. Moreover,

<sup>1</sup> Khamsi, "If a Coronavirus Vaccine Arrives, Can the World Make Enough?" In: *Nature* (2020); National Academies of Sciences, *Framework for Equitable Allocation of COVID-19 Vaccine* (2020).

<sup>2</sup> Spassiani et al., "Vaccination Criteria Based on Factors Influencing COVID-19 Diffusion and Mortality" in: *Vaccines* (2020); Matrajt et al., "Vaccine Optimization for COVID-19: Who to Vaccinate First?" In: *Science Advances* (2020).

<sup>3</sup> Gallagher et al., "Indirect Benefits Are a Crucial Consideration When Evaluating SARS-CoV-2 Vaccine Candidates" in: *Nature Medicine* (2021); Tuite, Zhu, et al., "Alternative Dose Allocation Strategies to Increase Benefits from Constrained COVID-19 Vaccine Supply" in: *Annals of Internal Medicine* (2021).

the complex coupling between regions due to human mobility and the spatial heterogeneities in disease history make the discovery of such optimal allocation strategies an arduous task.

An optimal control framework to explore COVID-19 vaccine distribution in space and time is proposed. It is applied to the SARS-CoV-2 epidemic in Italy where strong spatial effects arise from the geography of the disease, heterogeneous lockdown exit strategies, and post-lockdown control measures<sup>4</sup>.

The problem of vaccine allocation is of primary importance for public-health officials, epidemiologists, and economists<sup>5</sup>. Roll-out strategies are conventionally based on the prioritization of individuals at risk, such as health workers and elderly people<sup>6</sup>. However, the heterogeneous ways in which different regions may be affected by each successive wave raise questions about spatial prioritization strategies. What is the best feasible spatial allocation, given supply and logistic constraints? Would that differ significantly from current non geographically optimized plans? Should vaccines be distributed on the basis of demography or would it be better to prioritize areas currently subject to an outbreak? How relevant are the susceptibility profile and future transmission in each region?

Epidemiological modeling has long been used to answer questions about the impact of vaccination campaigns, often by comparing outcomes under different scenarios<sup>7</sup>. Optimization, i.e., the search for the best possible course of action that maximizes or minimizes an objective metric, has been carried out theoretically since the seventies<sup>8</sup>. Recent dramatic improvements of both algorithms<sup>9</sup> and computational power prompted applied studies using different methods to rigorously find optimal mitigation strategies: most of the time through iterative parameter search<sup>10</sup>, but also using genetic algorithms<sup>11</sup> or solving the Hamilton-Jacobi-Bellman equations<sup>12</sup>.

Interesting developments have recently arose during the ongoing SARS-CoV-2 pandemic<sup>13</sup>. The urgency of effective vaccination campaigns led to the development of modeling frameworks for the optimization of vaccine allocation, based on age or risk<sup>14</sup>, dose timing<sup>15</sup>, and the deployment of testing resources, using optimal control<sup>16</sup> or Bayesian experimental design<sup>17</sup>, along with prioritization based on social contact networks<sup>18</sup>.

To our knowledge, optimal spatial allocation of COVID-19 vaccines at a country scale has never been performed yet. This question is distinct from, and complementary to, risk-based prioritization. Spatial heterogeneities in disease transmission are complex, as seen during the initial outbreaks, supporting the significance of the posed problem towards effective control of the epidemic. However, the connectivity network underlying spatial epidemiological models may generate complex large-scale control problems whose solution requires tailored formulations and efficient algorithms.

This work aims to find optimal strategies for this problem through modern optimization methods based on distributed direct multiple shooting, automatic differentiation, and large-scale nonlinear programming<sup>19</sup>. This allows us to solve the large-scale optimization problems arising from epidemiological models, even when considering hundreds of spatial nodes.

<sup>4</sup> Marziano et al., “Retrospective Analysis of the Italian Exit Strategy from COVID-19 Lockdown” in: *Proceedings of the National Academy of Sciences* (2021).

<sup>5</sup> Emanuel et al., “An Ethical Framework for Global Vaccine Allocation” in: *Science* (2020); Lipsitch and Dean, “Understanding COVID-19 Vaccine Efficacy” in: *Science* (2020).

<sup>6</sup> Bubar et al., “Model-Informed COVID-19 Vaccine Prioritization Strategies by Age and Serostatus” in: *Science* (2021); Fitzpatrick et al., “Optimizing Age-Specific Vaccination” in: *Science* (2021); Baden et al., “Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine” in: *New England Journal of Medicine* (2020); Yang et al., “Who Should Be Prioritized for COVID-19 Vaccination in China? A Descriptive Study” in: *BMC Medicine* (2021).

<sup>7</sup> such as the study by Lee et al. (2020) presented in CHAPTER 3.

<sup>8</sup> Morton et al. 1974; Sethi et al. 1978; Greenhalgh 1988.

<sup>9</sup> Quirynen et al., “Multiple Shooting in a Microsecond” (2015).

<sup>10</sup> Sah et al. 2018; Medlock et al. 2009.

<sup>11</sup> Patel et al. 2005.

<sup>12</sup> Zakary et al. 2017; Miller Neilan et al. 2011.

<sup>13</sup> Fitzpatrick et al., “Optimizing Age-Specific Vaccination” in: *Science* (2021); Thul et al., *Stochastic Optimization for Vaccine and Testing Kit Allocation for the COVID-19 Pandemic* (2021); Moore et al., “Vaccination and Non-Pharmaceutical Interventions: When Can the UK Relax about COVID-19?” In: *medRxiv* (2021).

<sup>14</sup> Matraji et al. 2020; Spassiani et al. 2020; Fitzpatrick et al. 2021; Bubar et al. 2021.

<sup>15</sup> Saad-Roy et al., “Epidemiological and Evolutionary Considerations of SARS-CoV-2 Vaccine Dosing Regimes” in: *medRxiv* (2021); Kadire et al., “Delayed Second Dose versus Standard Regimen for Covid-19 Vaccination” in: *New England Journal of Medicine* (2021).

<sup>16</sup> Acemoglu, Fallah, et al., *Optimal Adaptive Testing for Epidemic Control: Combining Molecular and Serology Tests* (2021).

<sup>17</sup> Chatzimanolakis et al., “Optimal Allocation of Limited Test Resources for the Quantification of COVID-19 Infections” in: *Swiss Medical Weekly* (2020).

<sup>18</sup> Chen et al., “Prioritizing Allocation of COVID-19 Vaccines Based on Social Contacts Increases Vaccination Effectiveness” in: *medRxiv* (2021).

<sup>19</sup> Bock et al., “A Multiple Shooting Algorithm for Direct Solution of Optimal Control Problems” (1984); Savorgnan et al., “Multiple Shooting for Distributed Systems with Applications in Hydro Electricity Production” in: *Journal of Process Control* (2011); Andersson et al., “CasADi – A Software Framework for Nonlinear Optimization and Optimal Control” in: *Mathematical Programming Computation* (In Press, 2018); Wächter et al., “On the Implementation of an Interior-Point Filter Line-Search Algorithm for Large-Scale Nonlinear Programming” in: *Mathematical Programming* (2006).

## MATERIALS AND METHODS

The formulation of the optimal control problem has three main components: 1) an objective function to be minimized, here the total incidence in Italy from January 11, 2021 to April 11, 2021; 2) the set of constraints that the control must satisfy, in this case the limitations on vaccine administration rate in each province and the total vaccine stock in Italy; and 3) the spatial epidemiological model<sup>20</sup> governing the transmission dynamics with the daily vaccination rates in each province as control variables. The objective, the model, and the constraints may be tailored to specific applications within the proposed framework.

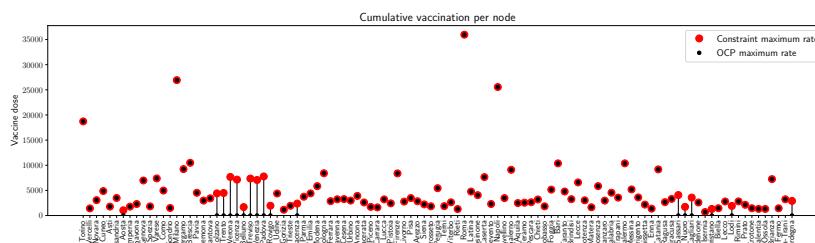
### OBJECTIVE FUNCTION

Optimizing calls for a metric, whose selection is critical in determining the optimal solution and its outcome. The choice of an objective function relates to health, economy, and ethics. Possible candidates are the minimization of e.g. DALYs (the Disability-Adjusted Life Years), the number of deaths, and economic loss<sup>21</sup>. All these objectives are linked and may be combined together. As the model considered for this work does not have risk classes, without loss of generality it is optimized for the minimization of the incidence on Italy from January 11 to April 11, 2021. Minimization of the deaths would yield the same results with the present model.

### CONSTRAINTS

Two types of constraints are defined: supply constraints, which determine the weekly delivery to the national stockpile; and logistic constraints, which limit the maximum rate of local vaccine distribution in each province.

The supply constraints ensure that the model does not distribute more vaccines than what is actually available in stock. It is assumed that the national supply of vaccine doses is empty on January 11, 2021 and is replenished every Monday. Four scenarios are considered with weekly deliveries of 479'700 (realistic, baseline value), 1M, 1.5M and 2M vaccine doses.



From the national stockpile, doses may be allocated to any of the 107 Italian provinces, but the logistic constraints limit the rate at which it is possible to distribute the vaccines in each province. The maximum number of individuals who can be vaccinated in a province per day is assumed to be proportional to the province's population, such that the national maximum distribution capacity equals 500,000 doses per day, i.e., 3.5M per week if every province vaccinates at its maximum rate (which in retrospect is close to Italy's vaccination rate as of May 1, 2021) (fig. 6.1).

<sup>20</sup> Gatto, Bertuzzo, et al., "Spread and Dynamics of the COVID-19 Epidemic in Italy: Effects of Emergency Containment Measures" in: *Proceedings of the National Academy of Sciences* (2020); Bertuzzo, Mari, Pasetto, et al., "The Geography of COVID-19 Spread in Italy and Implications for the Relaxation of Confine-ment Measures" in: *Nature Communications* (2020).

<sup>21</sup> Du et al., "Comparative Cost-Effectiveness of SARS-CoV-2 Testing Strategies in the USA: A Modelling Study" in: *The Lancet Public Health* (2021).

fig. 6.1: Local maximum vaccination rate constraint  $v_i^{\max}$  for each province. This logistic constraint bounds the maximum number of vaccines to 0.5M of doses per day, with a local rate that is proportional to the node population. Here the maximum vaccination rate for each province (the constraint the solution has to comply with) is shown in red, and the maximum rate prescribed by the optimal solution while simulating the pessimistic scenario with a stock-pile delivery of 479'700 doses, in black. The optimal solution uses the maximal capacity of the logistic net-work while respecting the constraint defined.

### COVID-19 TRANSMISSION MODEL

The optimal control framework may be used with any compartmental SARS-CoV-2 transmission model that can be approximated by ordinary differential equations. To demonstrate its usefulness, it is applied a complex model based on previous work that was aimed to describe the first wave of COVID-19 infections in Italy<sup>22</sup>.

The proposed framework is constituted of two disease transmission models, one “true” model and a simplified one used for control:

- The full model is a COVID-19 model, as designed in Gatto, Bertuzzo, et al. (2020) and Bertuzzo, Mari, Pasetto, et al. (2020). This model is ODE-based, includes full connectivity based on mobility data, and is implemented in MATLAB using the adaptive step `ode45` integration scheme. Using data assimilation, the joint-posterior distribution is obtained for all parameters of this model.
- The simplified model used for optimal control is an approximation of the above model, integrated using an explicit Runge-Kutta 4 method with fixed step size. The problem is simplified by limiting the connectivity to the largest mobility fluxes (fig. 6.3) and optimizing only one (median) realization of the posterior. This model is implemented in Python with the CasADi library.

**MODEL DESCRIPTION** The model subdivides the Italian population into its 107 provinces represented as a network of connected nodes. Each province has local dynamics describing the number of individuals present in each of the model compartments: susceptible  $S$ , exposed  $E$ , pre-symptomatic  $P$  (incubating infectious), symptomatic infectious  $I$ , asymptomatic infectious  $A$ , hospitalized  $H$ , quarantined  $Q$ , recovered  $R$ , and dead  $D$ . A tenth compartment, vaccinated individuals  $V$ , is added to the original nine, as shown in fig. 6.2.

Apart from hospitalized  $H$ , quarantined  $Q$ , dead  $D$ , and symptomatic individuals  $I$ , a fraction of the other individuals commutes between provinces along the mobility network, thus node-to-node disease transmission is introduced along the network shown in fig. 6.3. The COVID-19 transmission dynamics are described by the following set of ordinary differential equations in each province  $i$  with population  $N_i$ :

$$\begin{aligned} \dot{S}_i &= -\lambda_i(t)S_i - r_i^v(t)S_i \\ \dot{E}_i &= \lambda_i(t)S_i - (\delta^E + r_i^v(t))E_i \\ \dot{P}_i &= \delta^E E_i - (\delta^P + r_i^v(t))P_i \\ \dot{I}_i &= \sigma\delta^P P_i - (\gamma^I + \eta)I_i \\ \dot{A}_i &= (1 - \sigma)\delta^P P_i - (\gamma^A + r_i^v(t))A_i \\ \dot{Q}_i &= \zeta\eta I_i - \gamma^Q Q_i \\ \dot{H}_i &= (1 - \zeta)\eta I_i - (\gamma^H + \alpha^H)H \\ \dot{R}_i &= \gamma^I I_i + \gamma^A A_i + \gamma^H H_i + \gamma^Q Q_i - r_i^v(t)R_i \\ \dot{V}_i &= r_i^v(t) \cdot (S_i + E_i + P_i + A_i + R_i). \end{aligned} \quad (6.1)$$

Susceptible individuals get exposed to the pathogen at rate  $\lambda_i(t)$ , corresponding to the force of infection for community  $i$ , thus becoming latently infected (but not

<sup>22</sup> Gatto, Bertuzzo, et al., “Spread and Dynamics of the COVID-19 Epidemic in Italy: Effects of Emergency Containment Measures” in: *Proceedings of the National Academy of Sciences* (2020); Bertuzzo, Mari, Pasetto, et al., “The Geography of COVID-19 Spread in Italy and Implications for the Relaxation of Confine-ment Measures” in: *Nature Communications* (2020).

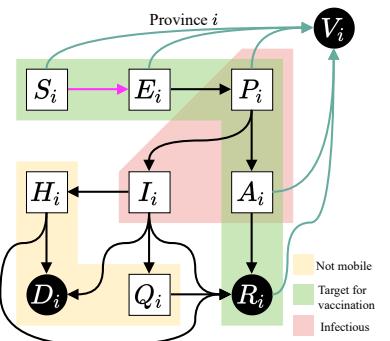


fig. 6.2: Diagram representing the compartments of the epidemiological model and the possible transitions in a single province. The control is the vaccination rate (teal arrows), aiming at minimizing incident infections (pink arrow). Individuals in compartments outside of the yellow block are able to move along the mobility network.

infectious yet). Exposed individuals transition to the post-latent, infectious stage at rate  $\delta^E$ . Post-latent individuals progress to the next infectious classes at rate  $\delta^P$ , developing an infection that can be either symptomatic—with probability  $\sigma$ —or asymptomatic—with probability  $1 - \sigma$ . Symptomatic infectious individuals recover from infection at rate  $\gamma^I$  and may seek treatment at rate  $\gamma$ . Asymptomatic individuals recover at rate  $\gamma^A$ . Infected individuals who sought treatment are either hospitalized (rate  $1 - \zeta$ ) or quarantined (rate  $\zeta$ ) at home and are considered to be effectively removed from the community, thus not contributing to disease transmission. Individuals who recover from the infection are assumed to have long-lasting immunity to reinfection at the timescale studied, but possible loss of immunity can be easily included in the model. Hospitalized individuals die at rate  $\alpha_H$  and recover at rate  $\gamma^H$ .

Individuals in compartments  $S, E, P, A, R$  might receive vaccine doses. If the chosen strategy allocates  $v_i(t)$  doses in node  $i$  at time  $t$ , the vaccination rate is:

$$r_i^v(t) = \frac{v_i(t)}{S_i(t) + E_i(t) + P_i(t) + A_i(t) + R_i(t)}. \quad (6.2)$$

Vaccinated individuals are moved at rate  $r_i^v(t)$  from their original compartments to compartment  $V$ , where they do not contribute to the infection anymore.

The estimated vaccine efficacy and immunity duration depend on the vaccine type. Because the focus is on spatial patterns and differences among vaccination strategies, the vaccination process is simplified by assuming a one-dose vaccine with instantaneous 100% efficacy. Moreover, vaccine protection is assumed to persist during the three months of projection considered.

**FORCE OF INFECTION AND SIMPLIFICATIONS** In addition to the province's local dynamics, it also considers that local susceptibles may enter in contact with infected individuals that are traveling, and oppositely, susceptible commuters may become infected through contact with local infected. Compartments  $P$ ,  $A$ , and  $I$  have different degrees of infectiousness and contribute to the force of infection (eqn. (6.3) and (6.4)), which represents the rate at which susceptibles  $S$  become infected and, thus, enter the exposed compartment  $E$ . The force of infection in each province is divided (for computational reasons) in a local and a mobility component. The local component describes transmission among the local individuals. The mobility component considers that local susceptibles may enter in contact with infected individuals that are traveling, and oppositely, susceptible commuters may become infected through contact with local infected. Connected provinces contribute to this process depending on the strength of the mobility fluxes from and to the node of interest.

Namely the force of infection  $\lambda_i(t)$  is split between the sum of the local force of infection  $\lambda_i^L(t)$ , from infected in node  $i$  and a mobility-driven force of infection from the network  $\lambda_i^N(t)$ , hence  $\lambda_i(t) = \lambda_i^L(t) + \lambda_i^M(t)$ . While running the model, it became apparent that  $\lambda_i^M(t) \ll \lambda_i^L(t)$ . Hence this artificial separation will be exploited when simplifying the model for optimal control. As described below,  $\lambda_i^M(t)$  is updated every day whereas  $\lambda_i^L(t)$  is updated at each integration step.

As for the formulations of the force of infection it reads:

$$\lambda_i^L(t) = C_{i,i} \beta_0 \beta_i(t) \cdot \frac{C_{i,i} (P_i + \varepsilon_A A_i) + \varepsilon_I I_i}{C_{i,i} \cdot (S_i + E_i + P_i + R_i + A_i + V_i) + I_i}, \quad (6.3)$$

and the influence of other provinces on province  $i$  is written as:

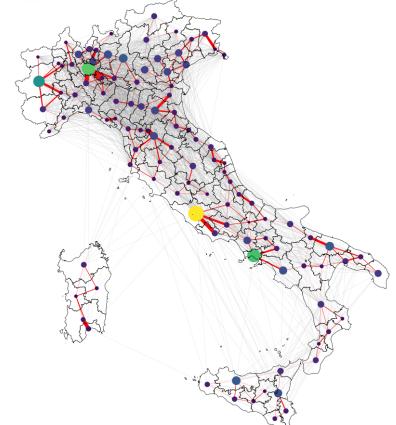
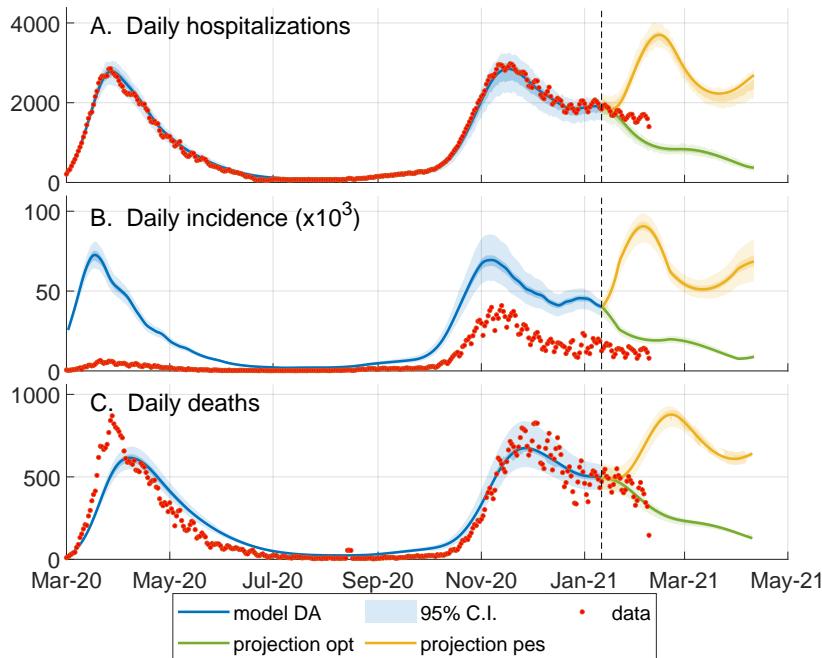


fig. 6.3: Mobility network: the force of infection in a province is coupled to the dynamics of other connected provinces. To reduce the problem to a tractable size, only the most important connections (red edges) are considered when optimizing, but the full network (red and grey edges) is used to assess the different strategies. Nodes size and color display each province's population, and edges width shows the strength of the coupling between each pair of provinces.

$$\lambda_i^M(t) = \sum_{m,m \neq i} C_{i,m} \cdot \frac{\sum_{n,n \neq m} [C_{n,m} \cdot \beta_0 \beta_n(t)(P_n + \varepsilon_A A_n)] + \varepsilon_I \beta_0 \beta_m(t) I_m}{\sum_{l,l \neq m} [C_{l,m} \cdot (S_l + E_l + P_l + R_l + A_l + V_l)] + I_m}, \quad (6.4)$$

where  $\beta_0$  is the baseline transmission rate, while  $\beta_i(t)$  is a spatially distributed and time-varying parameter describing site- and time-specific variations in transmissibility due to non-pharmaceutical interventions or other exogenous factors like variants. The parameters  $\varepsilon_A$  and  $\varepsilon_I$  represent the reduction of transmission respectively for asymptomatic and symptomatic individuals with respect to pre-symptomatic individual transmissions. Matrix  $C$  accounts for mobility: each element  $C_{i,j}$  of the matrix ( $i \neq j$ ) represents the proportion of individuals moving from  $i$  to  $j$ , while the diagonal elements  $C_{i,i}$  are the proportions of individual who do not move in each node  $i$ .

The objective for this modeling work is to minimize the total incidence of infections, i.e.,  $\int_{t_i}^{t_f} \sum_i \lambda_i(t) S_i$ . Note that for the present model, this is equivalent to optimizing the total deaths or hospital admissions, as without risk classes the sizes of these two compartments are proportional to each other.



**CALIBRATION AND SCENARIOS** The epidemiological model, previously calibrated during the first wave of COVID-19 in Italy, is updated up to January 11, 2021 using an iterative particle filtering. This data assimilation scheme allows us to capture the second wave of infections that hit Italy in the Fall of 2020, a necessary requirement to generate model projections that take into account the whole epidemic history, as shown in fig. 6.4. In this approach, model projections are described by an ensemble of a thousand trajectories associated with different parameters, whose distributions quantify the model uncertainty. Two projection scenarios are considered, each characterized by a possible rate of epidemic transmission, see fig. 6.4. The “Optimistic” scenario assumes a constant lowering of transmission from January 11, 2021 to April 11, 2021; the “Pessimistic” scenario

fig. 6.4: Data assimilation and projected scenarios for optimization. Comparison between the model outputs (95% confidence interval (CI) of the ensemble, blue shaded area) and the corresponding epidemiological data from March 2020 to January 2021. The orange and green shaded areas respectively show the ensemble dynamics (95% CI) of what is called pessimistic and optimistic transmission scenarios from January to April. The optimal vaccination strategy in the optimistic (or pessimistic) scenario is computed with respect to the continuous green (or orange) line, representing the model trajectory obtained using the median of each ensemble parameter. A. The data on the daily hospitalizations is estimated as described in (Bertuzzo, Mari, Pasetto, et al. 2020); this data at the regional level is assimilated on a moving window of 14 days to update the model parameters describing the local transmission rates. B. Daily number of newly exposed individuals versus the reported positive cases. Note that, the model assumes the presence of 90% asymptomatics among the exposed individuals, who are possibly not detected by the surveillance system. C. Daily number of deaths.

The detailed methods for the data assimilation scheme is provided in the supplementary information of (Lemaitre, Pasetto, Zanon, et al. 2021), and this thesis focuses on the optimal control problem. However, data assimilation is a key component allowing to update models states and parameters as new data becomes available, making it necessary for Model Predictive Control, see section DISCUSSION.

considers a gradual increase in transmission until mid-February 2021, which results in a third wave.

For each scenario, the optimal control problem is solved for one reference model trajectory, whose parameters and state on January 11, 2021 are obtained as the median values of the 1'000 model realizations. In this way, the reference trajectory approximately represents the ensemble median in each province. Then, the effectiveness of the optimal allocation is assessed on the full ensemble of trajectories.

This division is necessary in order to solve the optimal control problem in a reasonable time. To adapt the presented framework to another model/country, one would need to update the “true” model to a suitable candidate (which could be a stochastic model, a Hidden Markov model, or any other kind) and design a tractable approximation of this new model to be solved by optimal control.

The optimal vaccination course that minimizes the objective is computed based on the simplified model. Then, this strategy and the alternative ones are evaluated on the full model, for different posterior realizations. If the simplified model is sufficiently accurate, the performance loss is small and the proposed strategy outperforms simpler strategies, as shown in the simulation results.

#### FORMULATION OF THE OPTIMAL CONTROL PROBLEM

**OPTIMAL CONTROL PROBLEM DEFINITION** Let  $n$  the number of spatial nodes ( $n = 107$  provinces in Italy) and  $m$  the number of epidemic states in the model ( $m = 9$  states). The states of the system are noted  $x(t) \in \mathbb{R}_+^{n \times m}$ , i.e.,  $x(t)$  is a vector containing the epidemic variables  $S_i(t), E_i(t), P_i(t), I_i(t), A_i(t), Q_i(t), H_i(t), R_i(t), V_i(t)$  for every province  $i = 1, \dots, n$ . The rate of vaccine rollout for every node at time  $t$  is written  $v(t) = (v_1(t), \dots, v_n(t)) \in \mathbb{R}_+^n$ , representing the control variable.

The dynamics of the epidemiological model, in eqn. (6.1), are expressed as an ordinary differential equation in each province  $i$ :

$$\dot{x}_i(t) = F_i(x_i(t), v_i(t), m_i(t), t), \quad (6.5)$$

where  $m_i(t)$  carries the contribution of other provinces to the force of infection of node  $i$ , i.e. eqn. (6.4). The epidemiological model can be noted as following system of ordinary differential equations coupling disease transmission among all provinces:

$$\dot{x}(t) = F(x(t), v(t), m(t), t). \quad (6.6)$$

For simplicity, the time dependence is dropped in the equations below, and the state and control variables for the full system are written as:

$$x = (x_1, \dots, x_n), \quad v = (v_1, \dots, v_n),$$

with the global dynamics for all provinces then denoted:

$$F(x, v) = (F_1(x_1, v_1, m_1), \dots, F_n(x_n, v_n, m_n)).$$

The coupled force of infection in node  $i$  is denoted  $\lambda_i$ . The cost function is defined as the national incidence, i.e., the sum of new infections (transitions  $S_i \rightarrow E_i$ ) in all provinces at time  $t$ , for every node  $i$ , i.e.,

$$L(x, v) = \sum_{i=1}^n \lambda_i S_i.$$

For the sake of generality the terminal cost  $M$  is introduced, which can be used to ensure that the system is left in a proper state instead of optimizing for short-term gain. Since properly designing the terminal cost could require a long analysis, for simplicity it is not used in this work, hence  $M(\cdot) = 0$ .

Given the dynamical system with states  $x$ , controls  $v$ , and dynamics  $F$ , the optimal control problem is:

$$\min_{v(\cdot)} \int_0^T L(x(t), v(t)) dt + M(x(T)) \quad (6.7a)$$

$$\text{s.t. } x(0) = \hat{x}_0, \quad (6.7b)$$

$$\dot{x}(t) = F(x(t), v(t)), \quad \forall t \in [0, T], \quad (6.7c)$$

$$H(x(t), v(t)) \leq 0, \quad \forall t \in [0, T], \quad (6.7d)$$

where the aim is to minimize the cost function over the control horizon  $T$ , while enforcing the modeled SARS-CoV-2 transmission dynamics (eqns. (6.7b) and (6.7c)). Moreover, the constraints imposed by vaccine availability and the maximum vaccination rate are lumped in function  $H$  that expands to

$$v_i(t) \geq 0, \quad i \in \mathbb{I}_1^n, \quad (6.8a)$$

$$\int_{t_d}^{t_{d+1}} v_i(t) dt \leq v_i^{\max} \propto N_i, \quad i \in \mathbb{I}_1^n, t_d \in \mathbb{I}_0^T, \quad (6.8b)$$

$$\int_0^t \sum_{i=1}^n v_i(t) dt \leq D(t), \quad \forall t \in [0, T], \quad (6.8c)$$

where time is measured in days, and  $\mathbb{I}_a^b$  is the set of all integers  $a \leq k \leq b$ . eqn. (6.8a) enforces that one can only distribute a non-negative amount of vaccine doses. Eqn. (6.8b) states the logistic constraints, which limit to  $v_i^{\max}$  the amount of individuals that can be vaccinated each day in each node; here  $t_d$  is the time at which each day starts. The daily vaccination capacity of each province is set to be proportional to its population size  $N_i$ , assuming a fair distribution of the sanitary infrastructure among provinces, as shown in fig. 6.1. The constraint on the national stockpile is materialized by eqn (6.8c), which ensures that the total vaccine allocation across all nodes does not exceed the stockpile  $D(t)$ . The stockpile is replenished every Monday by the delivery of new vaccines, hence  $D(t)$  is a staircase function.

**TRANSFORMING THE OPTIMAL CONTROL PROBLEM INTO A NON-LINEAR PROGRAMMING PROBLEM** The optimal control problem in eqn. (6.7) is solved by a direct method, also called *first discretize, then optimize*, which transforms the control problem into a nonlinear programming problem. Within direct methods, direct multiple shooting is employed.

Our time window  $[0, T]$  is divided into a uniform time grid  $t_0, \dots, t_N$ , with  $N + 1$  points and  $t_0 = 0, t_N = T$ . Hence there are  $N$  intervals  $[t_k, t_{k+1}]$ , and the states at time  $t_k$  is denoted  $x_k = x(t_k)$ , and  $v_k$  represent the controls in interval  $[t_k, t_{k+1}]$ . The control function is parameterized using basis functions with local support.

The continuous-time dynamics  $F$  in eqn. (6.6) are transformed by numerical integration into the discrete-time operator  $f$ . The system dynamics are then discretized to obtain a discrete-time system:

$$x_{k+1} = f(x_k, v_k),$$

For an overview of the possible solution approaches for optimal control problems the interested reader is referred to Betts (2010) and Biegler (2010). In particular, in this work, a variant of direct multiple shooting Bock et al. (1984) tailored to distributed systems Savorgnan et al. (2011) is used.

For discretization, a uniform time grid is chosen here, i.e.,  $t_{k+1} = t_k + \delta_t$  and a piecewise constant control function, i.e.,  $v(t) = v_k, t \in [t_k, t_{k+1}]$ . This is a common choice but other possibilities exist. Here  $\delta t = 1$  day.

satisfying  $x_k = x(t_k)$  for all  $k = 0, \dots, N$ . Moreover, the cost function is also discretized, to obtain:

$$l(x_k, v_k) = \int_{t_k}^{t_{k+1}} L(x(t), v(t)).$$

The discretization is performed using numerical integration techniques (here a fourth-order Runge-Kutta scheme, with 50 steps per day) to obtain a good approximation of the true trajectory and cost. Finally, the path constraints  $H$  are relaxed and imposed at a finite amount of time instants, here coinciding with the time grid  $t_0, \dots, t_N$ . Since in the present case the constraints only involve the controls, no approximation is introduced by enforcing these constraints only on this uniform grid. The optimal control problem (6.7) is then approximated by following the nonlinear programming problem:

$$\min_{x, v} M(x_N) + \sum_{k=0}^{N-1} l(x_k, v_k) \quad (6.9a)$$

$$\text{s.t. } x_0 = \hat{x}_0 \quad (6.9b)$$

$$x_{k+1} = f(x_k, v_k), \quad k \in \mathbb{I}_0^{N-1}, \quad (6.9c)$$

$$H(x_k, v_k) \leq 0, \quad k \in \mathbb{I}_0^{N-1}. \quad (6.9d)$$

Here both the states  $x = (x_0, \dots, x_N)$  and the controls  $v = (v_0, \dots, v_{N-1})$  are defined as optimization variables, which is a distinguishing trait of multiple shooting as opposed to single shooting.

The force of infection associated with mobility is assumed constant over each day. Thus, each node dynamics can be made independent of the other nodes dynamics by introducing an auxiliary control variable  $z$  that is constrained to match the force of infection due to the other nodes at the beginning of each time interval. Then, the dynamics of the decoupled system in each node can be written as:

$$\begin{aligned} \dot{x}_i(t) &= F_i(x_i(t), z_{i,k}), & t \in [t_k, t_{k+1}] \\ x_i(t_k) &= x_{i,k} + g_i(v_{i,k}), & z_{i,k} = e_i(x). \end{aligned}$$

**SOLVING THE NON-LINEAR PROGRAMMING PROBLEM** Nonlinear programming problems may be solved by readily available solvers using the primal-dual interior-point method. The main difficulty in solving the proposed nonlinear programming problem (6.9) is the large dimension of the system and the nonlinearities of the model. In order to bring the problem to a tractable form, three simplifications are introduced: (a) vaccines are administered instantaneously at the beginning of each day, rather than with a constant rate over the whole day; (b) the component of the force of infection taking into consideration the mobility of individuals across provinces is evaluated at the beginning of each day and remains constant through the day; and, (c) the mobility network is simplified, by keeping only the most important connections (see fig. 6.3), thus increasing the sparsity of the underlying spatial connectivity matrix. These simplifications deliver a significant computational advantage, and the impact on the model accuracy was verified to be limited. The nonlinear programming problem arising from the simplified epidemiological model is non-convex and involves approximately  $10^5$  variables and  $10^5$  constraints. The problem is formulated within the automatic differentiation framework CasADi<sup>23</sup> and solve it using Ipopt<sup>24</sup> with the HSL ma86 large sparse symmetric indefinite solver<sup>25</sup>, which exploit the sparsity of the problem. Solving

Note that, even if the optimal strategy is computed using the simplified model, its impact in terms of averted cases and deaths is evaluated using the full epidemiological model without any of these simplifications. A more detailed discussion on this subject, and on the impact of the simplification is given in the APPENDIX TO CHAPTER 6.

The full framework and analysis code is available here: [https://github.com/jcblemai/COVID-19\\_italy-vaccination-oc](https://github.com/jcblemai/COVID-19_italy-vaccination-oc).

<sup>23</sup> Andersson et al., “CasADi – A Software Framework for Nonlinear Optimization and Optimal Control” in: *Mathematical Programming Computation* (In Press, 2018).

<sup>24</sup> Wächter et al., “On the Implementation of an Interior-Point Filter Line-Search Algorithm for Large-Scale Nonlinear Programming” in: *Mathematical Programming* (2006).

<sup>25</sup> HSL. *A Collection of Fortran Codes for Large Scale Scientific Computation.* () .

the optimal control problem is both CPU and RAM-intensive. Numerical computations are performed on the Helvetios cluster in the EPFL HPC facility (one problem per computing node, each equipped with 36 2.3 GHz cores and 192 GB RAM). On this cluster, it takes approximately four days to solve the large-scale problem presented here. It should be possible to solve even larger problems with more RAM available.

## RESULTS

Using state-of-the-art linear algebra solvers and automatic differentiation, each scenario is solved for the optimal vaccination allocation from January 11, 2021 to April 11, 2021. These scenarios are a combination of two projection scenarios (pessimistic vs optimistic) and four assumptions on the weekly stockpile delivery (479'700, 1M, 1.5M or 2M doses delivered per week). In each scenario, the optimal solution is a spatially explicit vaccine roll-out policy, i.e. an indication of the number of vaccine doses to be deployed in each province each day.

### COMPARISON OF VACCINATION STRATEGIES

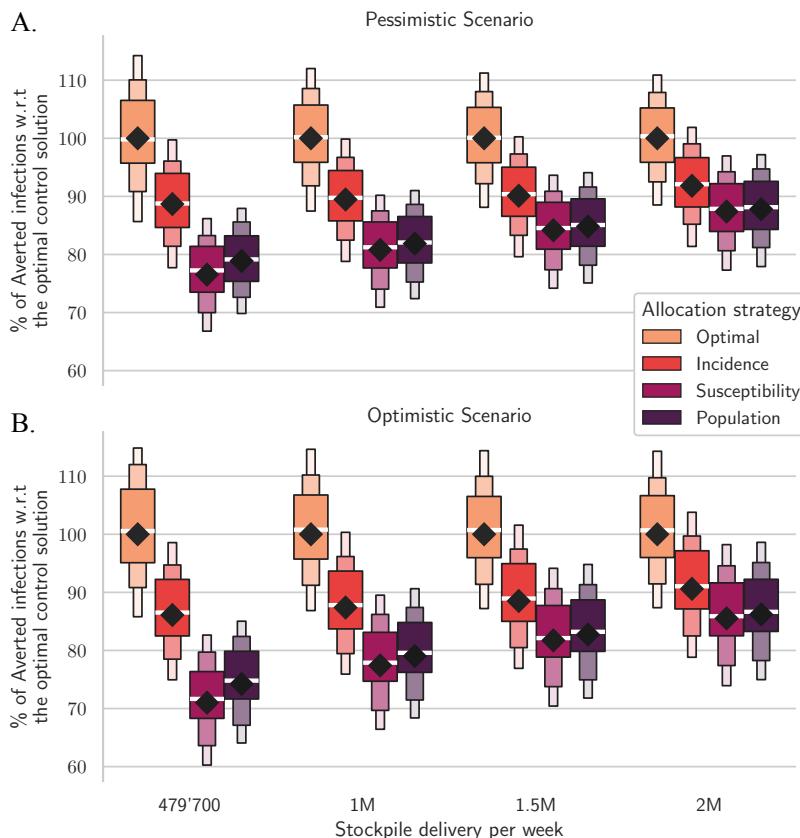


fig. 6.5: Comparison between different vaccine allocation strategies. Percentage of averted infections per vaccine doses from January 11, 2021 to April 11, 2021 resulting from province-scale vaccine allocation strategies for both the pessimistic (panel A) and the optimistic (panel B) scenarios based on: the optimal solution, population, proportion of susceptible individuals, and projected incidence (see color codes in the legend). The vaccine allocation is optimized for a median reference trajectory (diamonds), and the performance of the computed vaccination strategy is assessed over the whole posterior (box plots). For each projection scenario, results are normalized by the number of averted infections in the reference solution (see tab. 6.1 for absolute values). Results for alternative scenarios and vaccination strategies are shown in fig. A5.

In order to measure the potential impact of the optimal allocation strategy, it is compared against three non-optimized, yet reasonable alternative approaches: vaccinate proportionally to the future incidence as projected by the epidemiological model; vaccinate proportionally to the number of susceptible individuals in each province; vaccinate proportionally to the province's population. Comparisons

with additional alternative strategies are presented in the APPENDIX. Spatial prioritization based on epidemiological criteria, such as past<sup>26</sup> or future<sup>27</sup> incidence, has been thoroughly used in real campaigns and prospective studies.

For each of the eight scenarios considered, the number of averted infections is computed with respect to a zero-vaccination baseline, alongside with the number of averted infections per vaccination dose (see tab. 6.1). In the optimistic transmission scenario, characterized by a recess of the epidemic, the vaccination campaign has a lower impact on the averted infections per dose as only a small percentage of the vaccinated individuals would have been at risk of transmission. Obviously, the impact of the vaccination campaign is more evident in the pessimistic scenario where, for all strategies, the averted infections are larger than the vaccines deployed due to indirect protection effects. By virtue of the law of diminishing returns, the number of averted infections per dose decreases when increasing the stockpile. The optimal solution always outperforms all alternative strategies in

<sup>26</sup> Lee et al., “Achieving Coordinated National Immunity and Cholera Elimination in Haiti through Vaccination: A Modelling Study” in: *The Lancet Global Health* (2020).

<sup>27</sup> Pasetto, Finger, Camacho, et al., “Near Real-Time Forecasting for Cholera Decision Making in Haiti after Hurricane Matthew” in: *PLOS Computational Biology* (2018).

Weekly stockpile delivery	Method	Averted Infections (Millions)		Averted Infections per doses	
		Optimistic	Pessimistic	Optimistic	Pessimistic
2 millions	OPTIMAL	6.98	30.6	0.268	1.18
	Incidence	6.32	28.1	0.243	1.08
	Population	6.02	26.8	0.231	1.03
	Susceptibility	5.97	26.7	0.229	1.02
1.5 millions	OPTIMAL	5.52	24.1	0.283	1.24
	Incidence	4.89	21.7	0.25	1.11
	Population	4.57	20.4	0.234	1.05
	Susceptibility	4.51	20.3	0.231	1.04
1 million	OPTIMAL	3.9	16.9	0.3	1.3
	Incidence	3.41	15.1	0.262	1.16
	Population	3.08	13.8	0.237	1.06
	Susceptibility	3.02	13.7	0.232	1.05
479'700	OPTIMAL	2.01	8.54	0.322	1.37
	Incidence	1.73	7.57	0.278	1.21
	Population	1.49	6.74	0.239	1.08
	Susceptibility	1.43	6.54	0.229	1.05

terms of the number of averted infections and in terms of averted infections per dose (see again tab. A4). Incidence-based allocation comes second, while population and susceptibility-based allocation are distant third and fourth respectively. The improvement between optimal and incidence-based allocation is significant, ranging from 8.8% (pessimistic, 2M doses/week) to 16.2% (optimistic, 479'000 doses/week). In fig. 6.5, the black diamonds represent the percentage of averted infections obtained using each strategy for the reference trajectory, with respect to the averted infections resulting from the optimal strategy. The optimal strategy is observed to have the largest relative benefits for the smallest stockpile in both the optimistic and pessimistic scenarios.

In the pessimistic scenario (see fig. 6.5.A), when 479'700 doses are made available each week, the averted infections associated with the optimal strategy in the reference projection are 1.37 per dose: 25 % more compared to the strategies based on population or susceptible distributions (1.08 averted infections per dose), and

tab. 6.1: Absolute number of averted infections and averted infections per dose during the first three months of 2021 as evaluated for the reference trajectory (see fig. 6.4, and fig. 6.5 for an assessment on a set of trajectories drawn for the posterior distribution). The first column represents the considered scenarios of weekly stockpile replenishment, i.e. the number of doses delivered to Italy every week, ranging from 479'700 to 2M.

more than 10 % higher compared to the strategy based on the projected infections (1.21 averted infections per dose). These differences are smaller but still significant when increasing the weekly stockpile deliveries up to 2M doses; similar results are obtained also for the optimistic transmission scenario (fig. 6.5.B).

The optimal control strategy considered here is computed for a reference model trajectory, which is the median of an ensemble of 1000 realizations. To further investigate the effectiveness of the optimal solution, it is applied to a subset of trajectories randomly sampled from the ensemble. The box plots in fig. 6.5 display the main quantiles of the averted infections computed for the ensemble of trajectories. The optimal strategy still yields better results than the ensemble of projections related to the other strategies, thus suggesting that the computed solution is robust even under the presence of perturbations in the forecasts of the epidemic dynamics. More importantly, for each realization of the ensemble and for each projection scenario, the optimal strategy systematically averts more infections than any of the other control strategies. Our results suggest

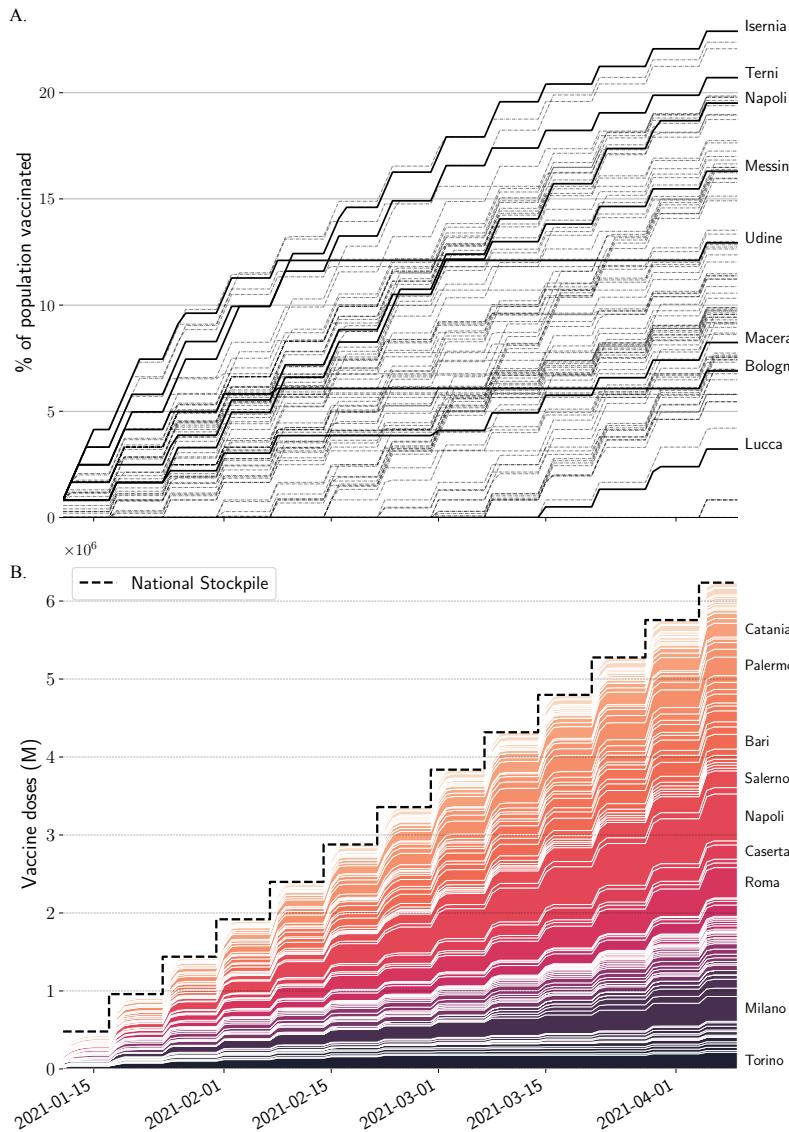


fig. 6.6: Optimal vaccine allocation for the baseline, pessimistic scenario. A. Cumulative proportion of vaccine doses administered in the 107 provinces, (some of which are highlighted). The local distribution rate is limited by a rate that is proportional to population. This logistic constraint is visualized here as the maximum slope, equal for every province. B. Stacked cumulative absolute number of vaccines in the 107 provinces of Italy. The national stockpile is shown in black and is replenished every week (say on Mondays) with 479'700 doses. The name of the provinces with a final allocation of more than 150'000 doses is shown on the right.

that it is possible to considerably increase the impact of vaccination campaigns by optimizing the vaccine allocation in space and time. For this task, optimal control provides the best possible strategy and sets a benchmark for the theoretical potential of a vaccination campaign.

#### ANALYSIS OF THE OPTIMAL VACCINE ALLOCATION

The optimal vaccine allocation obtained as the solution of the optimal control framework is complex to analyze. The strategy must obey the imposed logistic and supply constraints: 1) The vaccine stockpile is replenished every Monday by a fixed amount of doses (e.g., 479'700 doses in the baseline scenario), and 2) the maximum possible distribution capacity per province is limited, proportionally to the node population, so that the number of doses distributed across the country can be of 0.5M per day at maximum (more details in fig. 6.1).

The optimal vaccination course in time for the pessimistic, 479'700 doses/week scenario is shown in fig. 6.6. The optimal allocation respects is observed the two constraints on distribution (fig. 6.6.A) and supply (fig. 6.6.B). Moreover, no province is vaccinated at the maximum possible rate during the whole campaign, and provinces display a variety of vaccination patterns. Note that the vaccines received every Monday are always distributed during the following week, but that the rate of delivery on a national level increases with time (fig. 6.6.B). Surprisingly, the optimal solution favors precise targeting over the speed of delivery, in order to allocate more doses to provinces where the impact of vaccines on the whole system is projected to be higher. Hence, in order to control infections, precise targeting may trump delivery speed.

Furthermore, in the optimal solution every time a province is vaccinated, the rate of vaccination is equal to the maximum rate allowed by the local logistic constraint. This property is common to the alternative vaccination strategies, hence the difference in performance is due to the spatial allocation patterns. Another outlook on the temporal allocation is shown in fig. 6.7: for every week (*i.e.* for every stockpile delivery), the percentage of doses allocated to each province is shown as a stacked plot, highlighting how main spatial focus of the optimal strategy changes over the course of the epidemic.

In fig. 6.8, one can already see by visual inspection that the optimal allocation distributes most of the available doses on a few provinces with high incidence. These provinces are neither the most connected nor the most populous in Italy. The optimal strategy makes then use of the information on the network connectivity to fine-tune the allocation and deploys the vaccination on more provinces than the incidence-based strategy.

To further investigate these patterns, fig. 6.9.A is shown the number of administered doses vs the incidence projected without vaccines (the proxy variables leading to the second-best control performance), both normalized according to the resident population in each province. An allocation pattern whereby provinces with a higher incidence receive more vaccines is observed. However, the allocation is non-linear with respect to the projected incidence, suggesting that to better control the epidemic, the optimal allocation strategy takes into account other factors such as the importance of each province within the mobility network, as well as the proportion of susceptibles. When the weekly stockpile delivery is increased, as shown in fig. 6.9.B, this pattern shifts to the right while remaining qualitatively consistent. Hence, the optimal allocation strategy is robust with respect to the

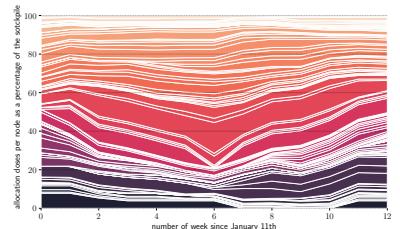
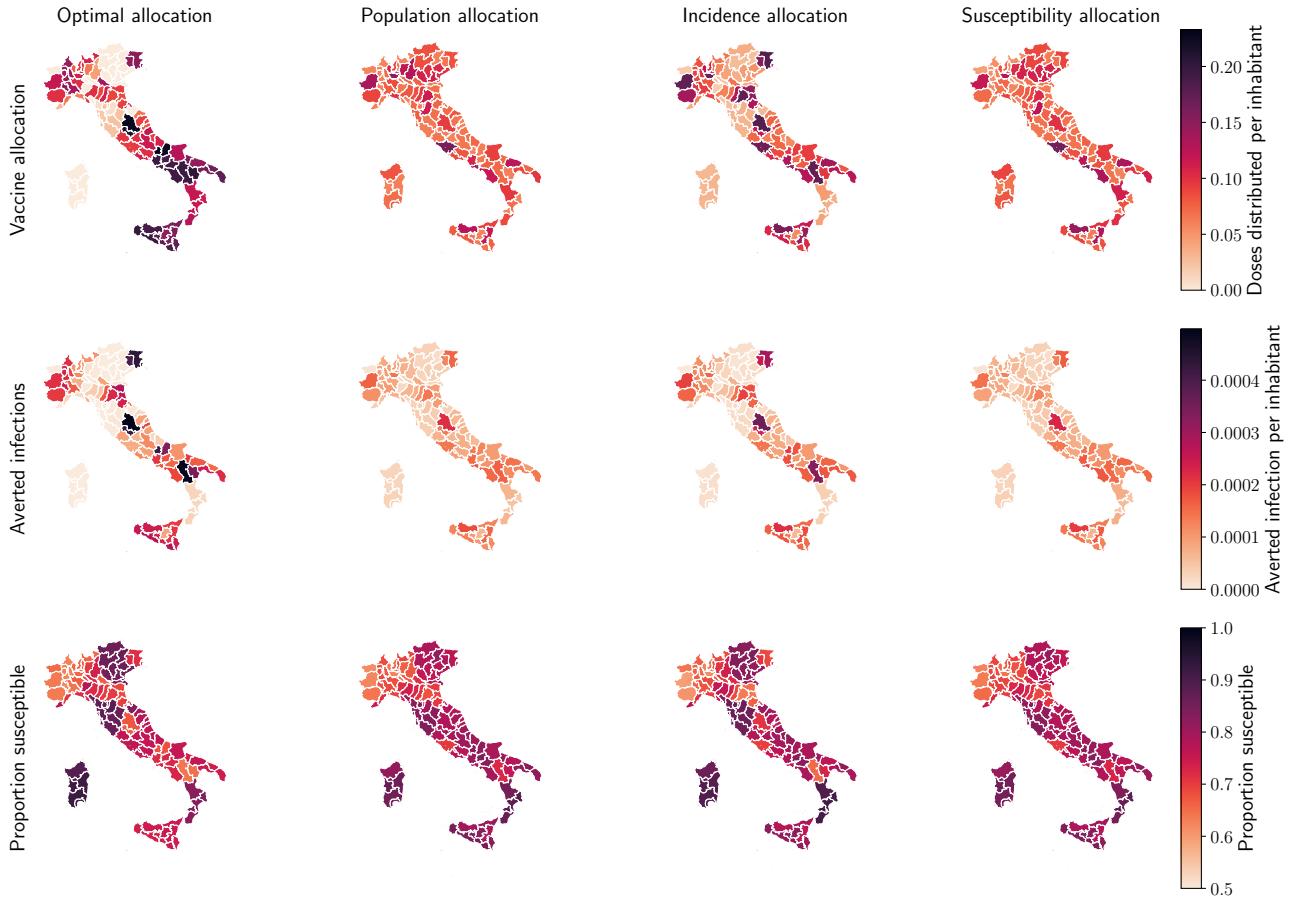


fig. 6.7: Temporal vaccine allocation for the pessimistic scenario with a stockpile delivery of 479'700 doses per week. This view unravels the temporal pattern in the allocation by showing the allocated doses per node as a percentage of the total weekly stockpile delivery.



overall vaccine availability constraint, and the same nodes are prioritized.

## DISCUSSION

Without any constraint on supply, each country would vaccinate its population as fast as possible according to the available infrastructure. However limitations in vaccine supply and rate of delivery are a reality for every country, hence the available doses should be deployed in space and time following a fair and effective strategy. In stockpile-limited settings, like most current vaccination campaigns worldwide, careful allocation may significantly increase the number of averted infections and deaths. The goal is to distribute the vaccines where they have the strongest beneficial impact on the dynamics of the epidemic. However, deriving an algorithm capable of computing spatially optimal allocation strategies in real, heterogeneous settings is far from trivial and the presented approach is, to the best of our knowledge, the first attempt in this direction.

A novel optimal control framework has been developed, that delivers the best vaccination strategy under constraints on supply and logistics. This allows us to compute the allocation strategy that maximizes the number of averted infections during a projection of the COVID-19 epidemic in Italy from January 11, 2021 to April 11, 2021. Results show that the optimal strategy has a complex structure that mainly reflects the projected incidence of each province, but also

fig. 6.8: Spatial vaccine distribution patterns for the optimal allocation (left) and alternative strategies based on population, incidence and susceptibility (additional alternative strategies are presented in the APPENDIX). For each province and strategy is shown: the proportion of vaccinated individuals (top), the number of averted infections per inhabitant with respect to a no vaccination baseline (middle), and the proportion of individuals who are still susceptible at the end of the control horizon (bottom).

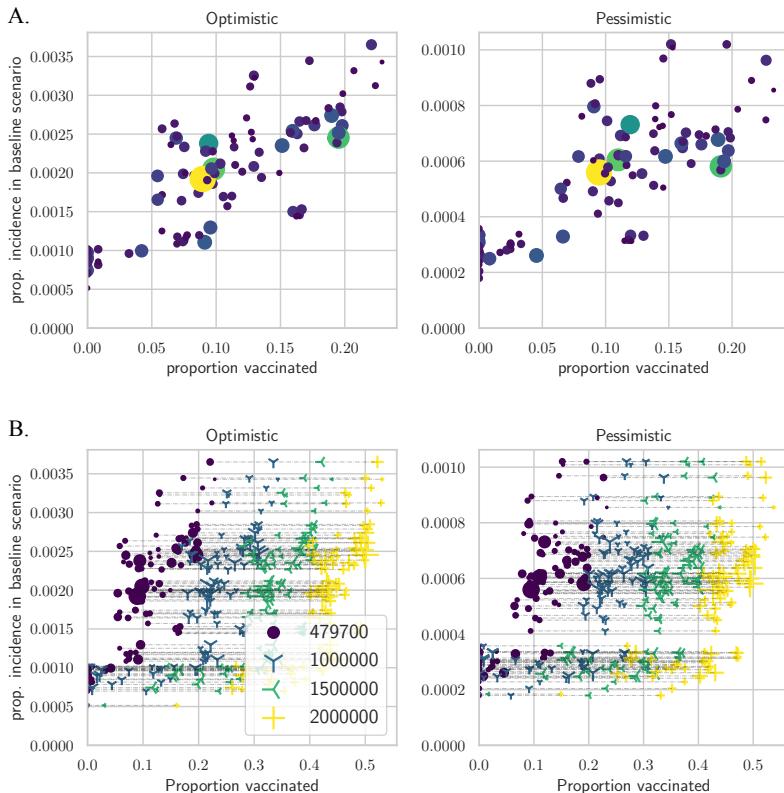


fig. 6.9: Analysis of the optimal solution. A. Vaccinated individuals according to the optimal strategy against the projected incidence without vaccination, both normalized by the province population. The scenario with a weekly stockpile delivery of 479'700 doses is considered. B. Same as in A, but considering all four scenarios of weekly stockpile replenishment. Each dot represents a province, with dot size proportional to its population.

takes into account the spatial connectivity provided by the mobility network and the landscape of acquired population immunity. Although the reason why this strategy is optimal is not immediately intuitive, simulations clearly outline its better overall performances over other, more straightforward strategies. This comparison suggests that the simplicity underlying intuitive vaccination strategies may undermine their effectiveness, and calls for complementing these simple approaches with rigorous and objective mathematical tools, like optimal control, that allow a full account of the complexity of the problem.

The present work demonstrates that it is possible to solve optimal control problems for spatially explicit dynamical models of infectious diseases at a national scale, thus overcoming the computational limitations that, up to now, precluded this kind of application. The proposed framework can account for any compartmental epidemic model, with up to hundreds of connected spatial nodes. Supply and logistic constraints can be adapted to the actual landscape of decisions faced by the stakeholders, such as no/reduced vaccine delivery on weekends, or the need for fairness in vaccine distribution, e.g. by ensuring that each region receives at least a fixed fraction of the available vaccines. This is especially important as in the optimal allocation scenarios, some provinces receive no vaccine at all. Moreover, the optimization can be carried for single-dose vaccines, as done here, or for two-dose vaccines, where one could potentially optimize the time between the first and second dose (and if a second dose should be administered at all), clearly also considering the intervals recommended by the health authorities.

The method developed is obviously not devoid of limitations. The main one is that the optimal vaccination strategy strongly depends on the projection of

the underlying epidemiological model. These projections are subject to several sources of uncertainty, especially for long term horizons, for example due to model design and calibration<sup>28</sup>, the generation of baseline transmission scenarios, and unforeseen events that may change the course of the epidemic (such as the importation of cases, the emergence of new virus variants, changes in disease awareness or social distancing policies). The optimal vaccination strategy is thus reliable only if the projections given by the underlying model dynamics are sufficiently accurate. A successful approach developed by the automatic control community to tackle that issue, named Model Predictive Control<sup>29</sup>, consists in compensating the performance losses expected over long horizons by constantly adapting the optimal strategy. In this context, Model Predictive Control might be implemented using the following steps: (a) at the beginning of each week, the state of the system is estimated by using newly acquired epidemiological data; (b) the optimization problem is solved over a fixed prediction horizon using the estimated state as an initial condition; (c) the optimal strategy for the first week is applied and, as soon as the next week starts, these steps are repeated starting from (a). This method corrects the model inaccuracies by constantly resetting the initial state to the estimated one. Moreover, constraints may be updated to account for unexpected deliveries or new orders. Future work will aim at further evaluating the benefits of implementing this scheme for the design of optimal vaccination strategies.

Moreover, the epidemiological model underlying the optimal control problem has known validity and limitations<sup>30</sup>. An additional limitation of the model for the specific scopes of this work is that it does not account explicitly for risk-based classes, and thus does not account for the heterogeneities that may result from the demography of the population, as well as from the age-related transmission and clinical characteristics associated with COVID-19. While surely limiting for operational use of the tools, the scope of this CHAPTER is to provide a proof-of-concept of the relevance of spatial effects, which have not been addressed so far in the literature. To that end, the presented results support the relevance of the research question posed and the proposed framework can be extended to optimize across both spatial and risk heterogeneities. A counter-factual assumption in this work is that a one-dose vaccine with full and instantaneous efficacy against transmission is considered. At the time of development, the details about COVID-19 vaccines were not released, and this hypothesis allowed us to demonstrate the framework in a simple setting. This framework could be further extended to account also for the simultaneous deployment of different vaccine types, some of which may require the administration of two doses. This extension too is the subject of ongoing research, in particular to extend the modeling tools described here to accommodate the peculiarities of each authorized vaccine candidate while designing effective spatiotemporal deployment strategies.

In conclusion, in this work the vaccine allocation at a country scale has been optimized on different scenarios of epidemic transmission and vaccine availability. A spatially explicit compartmental model that had already been successfully used to describe the COVID-19 pandemic in Italy is updated using a data assimilation scheme. Then, the model is discretized, transformed, and simplified and a framework is constructed to perform large-scale nonlinear optimization on vaccine allocation, subject to stockpile and logistic constraints. Solving this problem yielded a complex solution that outperforms other strategies by a significant margin and proves robust across posterior realizations of the underlying model. As such, besides inherent limitations, it provides a benchmark against which other, possibly simpler, vaccine rollout strategies can be compared.

<sup>28</sup> Cramer et al., “Evaluation of Individual and Ensemble Probabilistic Forecasts of COVID-19 Mortality in the US” in: *medRxiv* (2021).

<sup>29</sup> Rawlings et al., *Model Predictive Control: Theory, Computation, and Design* (2017).

<sup>30</sup> Gatto, Bertuzzo, et al., “Spread and Dynamics of the COVID-19 Epidemic in Italy: Effects of Emergency Containment Measures” in: *Proceedings of the National Academy of Sciences* (2020); Bertuzzo, Mari, Pasetto, et al., “The Geography of COVID-19 Spread in Italy and Implications for the Relaxation of Confine-ment Measures” in: *Nature Communications* (2020).





# CONCLUSION AND PERSPECTIVES

Each of the five chapters of this thesis introduces an infectious disease model tailored to tackle a public-health policy related question, whether concerning the identification of transmission pathways or the effects of past or planned interventions. Naturally, the angle of approach to the complex phenomena that are cholera and COVID-19 transmission varies depending on the research question and the specificities of the study setting. Hence each modeling work takes a different path within the interactive framework to model infectious disease dynamics<sup>31</sup>. Let's recall here the main takeaways from the different modeling exercises.

First, as in classical hypothesis testing, a model might be considered as a simplified representation of reality. Among different hypotheses, the aim is to find the “true” model underlying the observed dynamics. This perspective is taken in CHAPTER 2, where the explanatory power of different pathways for rainfall-mediated cholera transmission is compared<sup>32</sup>. Results stress the importance of rainfall as a covariate for cholera transmission while highlighting the complexity of the mechanistic pathways considered. While context-dependent, it is nevertheless interesting to observe how the different transmission routes proposed in the literature explain the intra-seasonal rainfall events in Juba, South Sudan.

The perspective on models is shifted in the policy related questions proposed in CHAPTERS 3 and 4: here, a sufficient predictive accuracy with respect to the interventions and transmission dynamics of interest is desired. Models are designed to most adequately reproduce an unknown, highly complex reality in order to perform experiments<sup>33</sup>. Building on ECHO's experience on cholera modeling<sup>34</sup> a spatial stochastic model of cholera transmission in Haiti is proposed in CHAPTER 3. It is used as part of a larger multi-modeling study<sup>35</sup>, to study the probability of cholera elimination from Haiti under different scenarios of mass vaccination campaigns. A retrospective analysis reveals that while the proposed model fits past dynamics, it under-estimates the probability of elimination of cholera in Haiti without any vaccination campaign. This discrepancy is a reminder that modeling is no silver bullet, further stressing the importance of careful model design aligned with the study objective.

Non-withstanding the COVID-19 pandemic, cholera would have been the sole focus of the present thesis. This regretful interruption left unanswered some research questions on this ancient disease. There have been no confirmed cholera cases since early 2019 in Haiti, which is surprising with regard to the aforementioned projections. What is the cause of the elimination of cholera from Haiti? What role did climate, herd immunity, and the WaSH interventions carried out in the past years<sup>36</sup> played in this elimination? The answer would provide additional insights on the path towards the elimination of cholera by 2030, as sought after by WHO and GTFCC. Toward this goal, the oral cholera vaccine stockpile put

<sup>31</sup> see INTRODUCTION, fig. 1 and Heesterbeek et al., “Modeling Infectious Disease Dynamics in the Complex Landscape of Global Health” in: *Science* (2015).

<sup>32</sup> Rinaldo, Bertuzzo, Mari, et al., “Reassessment of the 2010–2011 Haiti Cholera Outbreak and Rainfall-Driven Multiseason Projections” in: *Proceedings of the National Academy of Sciences* (2012); Eisenberg, Kujibida, et al., “Examining Rainfall and Cholera Dynamics in Haiti Using Statistical and Dynamic Modeling Approaches” in: *Epidemics* (2013).

<sup>33</sup> Indeed, accurate mechanistic modeling is still required for the processes of interest, and these two perspectives are related more than opposed.

<sup>34</sup> Rinaldo, Gatto, et al., *River Networks as Ecological Corridors: Species, Populations, Pathogens* (2020).

<sup>35</sup> Lee et al., “Achieving Coordinated National Immunity and Cholera Elimination in Haiti through Vaccination: A Modelling Study” in: *The Lancet Global Health* (2020).

<sup>36</sup> Rebaudet, Bulit, Gaudart, Michel, Gazin, Evers, Beaulieu, Abedi, Osei, Barrais, Pierre, Moore, Boncy, Adrien, Guillaume, et al., “The Case-Area Targeted Rapid Response Strategy to Control Cholera in Haiti: A Four-Year Implementation Study” in: *PLOS Neglected Tropical Diseases* (2019).

in place at the disposal of countries is brought to play a strategic role<sup>37</sup>. However the population at risk vastly exceeds the vaccine supply and production<sup>38</sup>, and effective targeting of the existing doses remains an important problem<sup>39</sup> on the road towards elimination, a possible application of the optimal control framework developed here for COVID-19.

The rapid spread of SARS-CoV-2 mobilized infectious disease epidemiologists across the world. In this thesis, most of the work undertaken for the response to the COVID-19 pandemic took place within or around the COVID Scenario Pipeline, a configurable framework to project epidemic trajectories and healthcare impacts under different suites of interventions. The pipeline is used to support several partners including the state of California and the national US response. It is still being actively developed to address the ever-changing needs of decision-makers. Since the description given in CHAPTER 4, a year of historical data brought the need for high-dimensional inference algorithms, and the challenges posed by SARS-CoV-2 necessitated more flexible disease transmission and health outcome modeling. These developments allowed the capture of the dynamics of competing strains, immune escape, and the successive vaccination campaigns<sup>40</sup>. Hence the pipeline is now an operational forecasting platform that provides a unified framework to project and forecast disease dynamics from emergence to endemicity<sup>41</sup>.

Assessing the effectiveness of past policies is paramount to project disease dynamics and to inform future decisions. In Switzerland, the pipeline has been used to inform CHUV, the main hospital of Canton de Vaud. The dialogue with policy-makers triggered the exchange of a epidemiological dataset to improve the accuracy of scenario planning reports. In turn, this data has enabled a research study on the estimation of SARS-CoV-2 reproduction number,  $R_0$ , in Switzerland. Using stochastic models and iterated particle filtering<sup>42</sup> the effectiveness of non-pharmaceutical interventions against COVID-19 in Switzerland is uncovered. Among the numerous takeaways from this early COVID-19 work presented in CHAPTER 5, the timing of the decrease in transmission preceding NPIs implementation is especially interesting.

Finally, provided an accurate model of transmission dynamics and interventions effectiveness, optimal control is the ultimate stage of infectious disease modeling towards informed decisions: policies that minimize the burden of diseases are programmatically designed. To date, both the demanding prerequisites and the difficulty of controlling epidemiological models at scales that are relevant to decision-makers were limiting the application of optimal control on practical problems. In CHAPTER 6, a large-scale optimal control framework coupled to an existing model of COVID-19 transmission<sup>43</sup> is presented. Using automatic differentiation and non-linear programming, the proposed solver designs the most effective vaccination strategy for a given objective, feasible under operational constraints. This proof of concept is performed on spatial vaccines allocation against COVID-19 in Italy. While limitations of the model, such as the absence of age-stratification, restrict the scope of the results, this first attempt at country scale optimization of a compartmental epidemiological model is a promising proof-of-concept for this method that provides a benchmark of what is possible to achieve with the available resources.

Taken as a whole, the present thesis demonstrates the relevance of modeling as a tool to inform public-health decisions through five studies. Yet, it also shows that

<sup>37</sup> Global Task Force on Cholera Control, *Ending Cholera - A Global Roadmap to 2030* (2017).

<sup>38</sup> Pezzoli, "Global Oral Cholera Vaccine Use, 2013–2018" in: *Vaccine* (2019).

<sup>39</sup> Lessler, Moore, et al., "Mapping the Burden of Cholera in Sub-Saharan Africa and Implications for Control: An Analysis of Data across Geographical Scales" in: *The Lancet* (2018).

<sup>40</sup> Borchering, "Modeling of Future COVID-19 Cases, Hospitalizations, and Deaths, by Vaccination Rates and Nonpharmaceutical Intervention Scenarios — United States, April–September 2021" in: *MMWR. Morbidity and Mortality Weekly Report* (2021).

<sup>41</sup> Updated outputs of the pipeline are visible on the COVID-19 Scenario Modeling Hub and the COVID-19 Forecast Hub ([covid19scenariomodelinghub.org](https://covid19scenariomodelinghub.org) and [covid19forecasthub.org](https://covid19forecasthub.org)).

<sup>42</sup> an alternative procedure to the predominant  $R_0$  estimation methods, see Gostic et al., "Practical Considerations for Measuring the Effective Reproductive Number,  $R_t$ " in: *PLOS Computational Biology* (2020); Pasetto, Lemaitre, et al., "Range of Reproduction Number Estimates for COVID-19 Spread" in: *Biochemical and Biophysical Research Communications* (2020).

<sup>43</sup> Gatto, Bertuzzo, et al., "Spread and Dynamics of the COVID-19 Epidemic in Italy: Effects of Emergency Containment Measures" in: *Proceedings of the National Academy of Sciences* (2020); Bertuzzo, Mari, Pasetto, et al., "The Geography of COVID-19 Spread in Italy and Implications for the Relaxation of Confine-ment Measures" in: *Nature Communications* (2020).

there is no one-size-fit-all approach to infectious disease modeling: each research question and transmission setting requires numerous adjustments to capture and project the disease dynamics. This takeaway might seem discouraging, but past efforts are not lost as each model builds on the previous works, borrowing conceptual breakthroughs and methods. With these new tools, the scientific and public-health communities are left better armed to face existing and upcoming threats<sup>44</sup>. Hopefully, the methods developed as part of the present thesis, whether at their infancy for optimal control or mature like the COVID Scenario Pipeline, will strengthen the arsenal of methods to guide decisions against infectious disease transmission.

The sustaining transmission of cholera and COVID-19 is the product of complex interactions between the environment, individuals, pathogens, and societies. Despite tremendous advances, especially for COVID-19, the data available about infectious diseases will stay noisy, missing, and biased in the foreseeable future. Modeling helps to design effective policies, allowing to properly weight different courses of action and to convey the associated outcomes and uncertainties. As such, it has been a key instrument in the fight against infectious diseases, a point made more evident during the COVID-19 pandemic.

Using compartmental models and computer-age statistical inference methods, the present thesis demonstrates the relevance of modeling to inform the control measures against epidemics. The scientific understanding of transmission pathways and intervention effectiveness is unraveled in specific contexts. Moreover, the (ongoing) issue of scenario planning reports to inform COVID-19 to decision-makers provide a concrete example of the influence of modeling on public-health policies. While the solutions to fight infectious diseases exist, resources are notoriously lacking, even more so because outbreaks piles upon impoverished communities, or in regions plagued by conflicts, natural disasters, and instability. Computational epidemiology allows one to communicate the trade-offs on policies and to ensure the best allocation of the limited available resources, so that *e.g.* the effect of every vaccine dose is maximized.

Indeed, there is only so much modeling can do, and for many communities, the elephant in the room is the brutal lack of resources to fight the infectious diseases that pose a constant threat to children and adults. The COVID-19 pandemic has pointed out what is possible in terms of mobilization and collaboration from partners around the world. Hopefully, it will set a precedent, and other diseases that plague impoverished and stigmatized communities will also benefit from the involvement and resources necessary to lighten this unfair burden<sup>45</sup>.

<sup>44</sup> In the past, outbreaks and pandemics have sparked rapid advances on various topics of infectious disease modeling. The response to the COVID-19 pandemic has benefitted extensively from conceptual and concrete tools developed in the past, such as the  $R_0$  package that was developed after HiNi (Obadia et al. 2012), but also *e.g.* real-time forecasting methods against Ebola (2014–2015), statistical approaches for cholera, multi-modeling studies for influenza, ... Indeed the COVID-19 pandemic has sparked the development of an unprecedented amount of novel tools.

<sup>45</sup> Even solely from a modeling perspective, the sudden availability of mobility datasets, something long desired for cholera, is an example of the paradigm shift brought by a pandemic that affected every country in the world, and hopefully lessons learned will facilitate collaboration against other diseases.



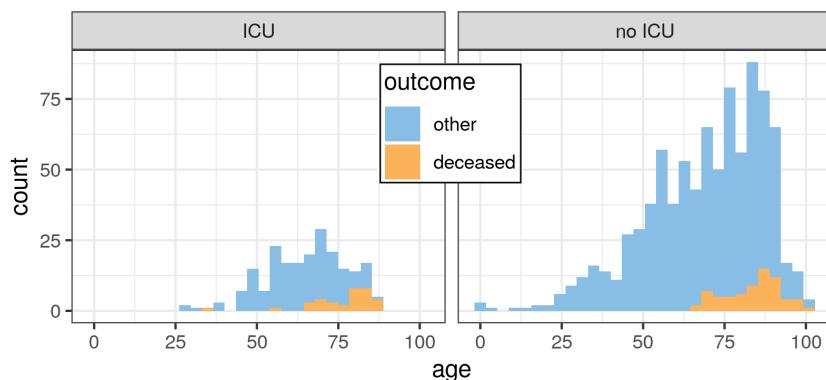
## APPENDICES



## APPENDIX TO CHAPTER 5

### CANTON DE VAUD HOSPITALIZATION DATA

The dataset contains individual-level data from 1093 patients, hospitalized in the canton of Vaud up to April 14, 2020. Of all patients, 41% (448/1093) were female and 59% were male (645/1093) with a median age of 70 years (fig. A1). Of all the hospitalized cases, 20% (214/1093) required use of an Intensive Care Unit (ICU).



Of 777 patients with a known outcome on April 14, 104 died (13%). The hospitalized Case Fatality Ratio (hCFR) is estimated by adjusting for the distribution of time hospitalization to death accounting for the fact that outcomes have not been yet observed for all patients (right-censoring<sup>46</sup>). To account for multiple outcomes (death and discharge), a parametric competing risk survival model is implemented<sup>47</sup>. A Bayesian approach is chosen<sup>48</sup>, that enables us to fit parametric distributions to times to events using accelerated failure models. This method allows for the joint estimation of the probability of each event type and the distributions of times to events. In this case the probability of death, i.e. the hCFR. A COVID-19 modeling study in France identified mixtures of probabilities of times to death, with a group dying faster with exponentially-distributed times and one dying slower<sup>49</sup>. The Bayesian survival framework is extended to test for mixture in times to death and recovery. Model patients being discharged from ICU and subsequently dying, which was the case for 4/138 patients with known outcome, are not taken into account. Neither are accounted for multiple ICU stays per patient since this information was not available. Both Gamma and Log-Normal distributions are fitted separately to patients that did not go into ICU, and patients that did. For the former, it is modeled times from hospitalization to death or discharge, and for the latter times from ICU entry to both outcomes. Models were fit with Stan<sup>50</sup>, and selection was done using Bayesian leave-one-out cross-validation<sup>51</sup>.

fig. A1: Age distribution of patients hospitalized for COVID-19 in the canton of Vaud up to April 14. Hospitalized individuals are divided in two subgroups depending on if they were treated in ICU (left) or not (right) during their stay. Moreover, only the 777 patients with known outcome are displayed here. The outcome is highlighted: either death (orange) or discharge/transfer to another hospital (blue).

<sup>46</sup> The dataset presented here is two weeks more recent than the one visible in the example scenario planning report for Canton de Vaud (see Results, CHAPTER 4). It takes into account 137 additional patients, and the comparison with fig. A1 strikingly shows how the distribution changes as the system is longer observed.

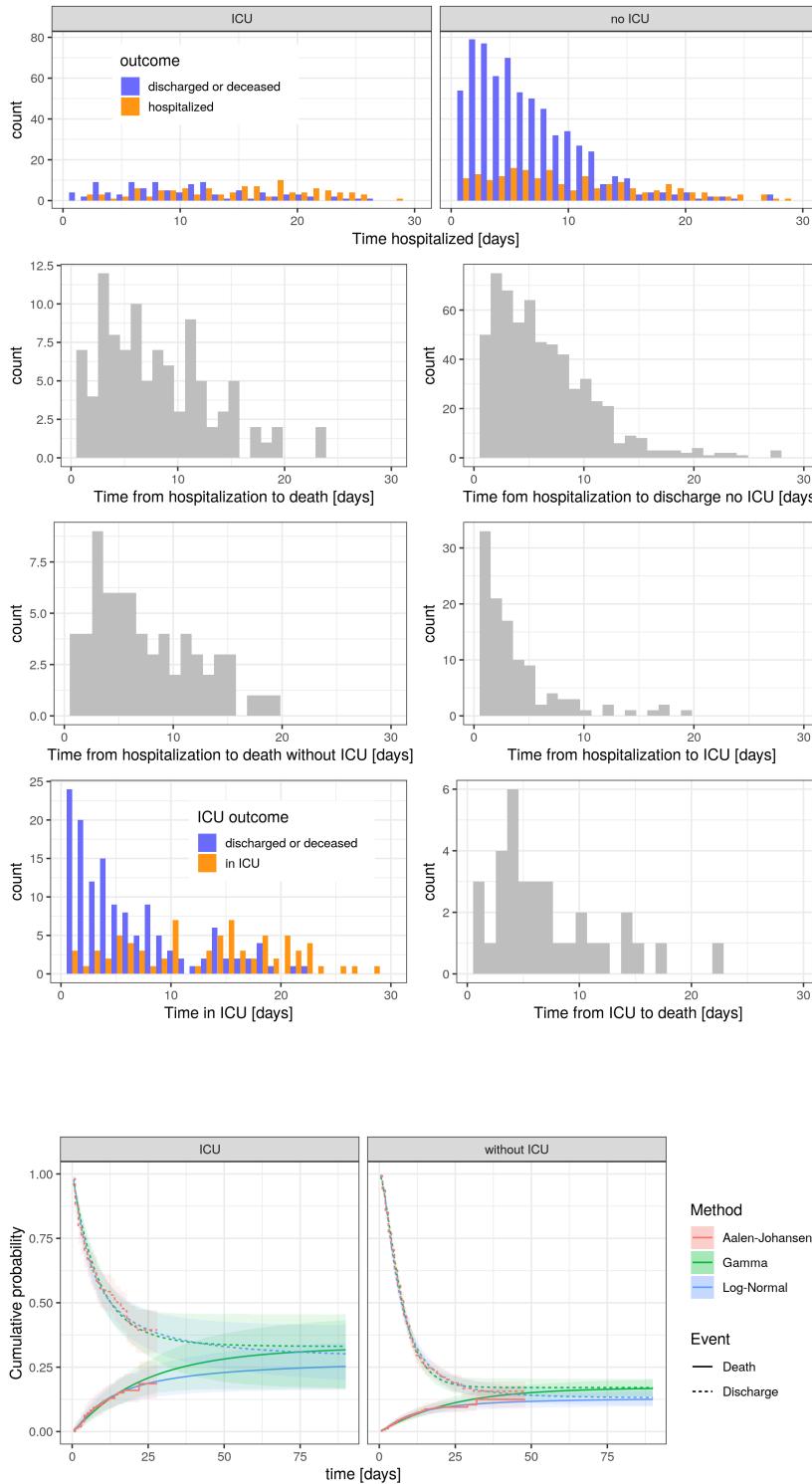
<sup>47</sup> Ghani et al., “Methods for Estimating the Case Fatality Ratio for a Novel, Emerging Infectious Disease” in: *American Journal of Epidemiology* (2005).

<sup>48</sup> Bellot et al., “Tree-Based Bayesian Mixture Model for Competing Risks” (2018).

<sup>49</sup> Salje et al., “Estimating the Burden of SARS-CoV-2 in France” in: *Science* (2020).

<sup>50</sup> Carpenter et al., “Stan: A Probabilistic Programming Language” in: *Journal of Statistical Software* (2017).

<sup>51</sup> Vehtari et al., “Practical Bayesian Model Evaluation Using Leave-One-out Cross-Validation and WAIC” in: *Statistics and Computing* (2017).



Times to death and discharge were best described by log-normal distributions with a single group both for patients having required ICU or not (fig. A3). When accounting for right-censoring and assuming log-normally distributed times to events, an overall hCFR is estimated at 16.0% (95% credible interval, CrI: 12.5-19.8), resulting from a hCFR of 28.1% (95% CrI: 16.4-40.9) for patients requiring ICU and 13.0% (95% CrI: 9.9-16.6) for patients that did not require it. Estimated hCFRs

fig. A2: Data from the canton of Vaud showing times to key hospitalization events. In order to perform this analysis, the patients are splitted in two categories: those who did not go through ICU during their stay and those who did. From left to right, top to bottom: total length of hospital stay for patients that went to ICU, then similarly for patient that did not. Then the time to death is shown for all patients, followed by both the time to discharge and to death for non-ICU patients. The last three graphs concerns ICU patients and detail ICU focused estimate: time from hospitalization to ICU, time in ICU and time from ICU to death. When meaningful, both currently hospitalized patient (orange) and already out-of-hospital patients (blue) are shown.

fig. A3: Survival functions of death and discharge for hospitalized patients and patients in ICU. The lines represent the mean estimated cumulative probability of dying (full) and 1 minus the cumulative probability of discharge (dotted) estimated with non-parametric (Aalen-Johansen estimator, shading gives the 95% CI) and parametric (assuming gamma and log-normal distributions, shading indicate the 95% CrI) methods. Time is in days from hospitalization for patients that did not require ICU, and time from ICU admission for those that did. The point at which the lines join represents the probability that the final outcome is death, which was estimated to be 28.1% (95% CrI: 16.4-40.9) for patients in ICU and 13.0% (95% CrI: 9.9-16.6) for patients not requiring ICU based on the log-normal distribution.

	mean	sd	mean (logscale)	sd (logscale)
Time hospitalized	8.49	6.58	1.81	0.87
Time to death	8.23	6.09	1.80	0.87
Time to discharge without ICU	6.29	4.66	1.56	0.80
Time hospitalized without ICU	7.35	5.79	1.68	0.85
Time to death without ICU	7.84	6.27	1.73	0.88
Time to ICU	2.35	3.79	0.18	1.05
Time hospitalized with ICU	13.14	7.50	2.37	0.72
Time in ICU	8.36	6.76	1.69	1.04
Time from ICU to discharge	8.68	6.99	1.71	1.07
Time from ICU to death	6.97	4.98	1.68	0.77

were slightly higher when assuming gamma-distributed times to events (overall hCFR of 20.3%, 95% CrI: 15.9–24.1). The distribution of times of hospitalization processes are shown in fig. A2, and fitted distribution parameters given in tab. A2.

Group	Event	Log-Normal			Gamma	
		median	mean-log	SD-log	scale	shape
Without ICU	Death	10 (7.3–16)	2.3 (2–2.8)	1.2 (0.95–1.5)	21 (21–22)	1.1 (0.83–1.4)
	Discharge	6.1 (5.6–6.6)	1.8 (1.7–1.9)	0.93 (0.87–0.99)	4.3 (4.4–4.2)	1.8 (1.6–2)
ICU	Death	13 (6.2–30)	2.6 (1.8–3.4)	1.3 (0.87–1.9)	21 (12–23)	1.2 (0.74–1.9)
	Discharge	6.4 (4.3–9.3)	1.8 (1.5–2.2)	1.3 (1.1–1.6)	9.1 (8.1–11)	1 (0.75–1.4)

tab. A3: Estimated parameters of hospitalization time distributions. These estimates differ from observed values given in tab. A2 by accounting for right-censoring of observations. Time from hospitalization to discharge or death, and from ICU admission to discharge or death are reported here. Estimates were obtained using competing risk survival model as described above. Parameters are given in terms of their posterior mean and 95% CrI (in parenthesis). For the log-normal distribution the parameters correspond to the mean and SD of the logarithm of the distribution.

tab. A2: Observed hospitalization time distributions. All times are in days and taken from the date of hospitalization if not specified otherwise. Note that these estimates are biased due to right-censoring of observations and probably under-estimate the true distributions. Estimates that account for right-censoring are shown in tab. A3.



# APPENDIX TO CHAPTER 6

## COMMENT ON THE SIMPLIFICATIONS

**DISCUSSION ON SIMPLIFICATION (A)** Realistically, vaccinations will occur at least eight hours per day. The assumption, while justified as a computationally convenient approximation of reality, is not a priori worse than assuming that vaccine administration takes place over the whole day. More refined approximations, while in principle possible, pose severe issues because of the nature of the system dynamics. While for most initial values the system dynamics can be easily simulated with time-continuous vaccinations, the system becomes stiff by construction once almost the entire population has been vaccinated. In this case, numerical integration errors can drive the size of some compartments to be negative, which violates the model assumptions and makes the result of the numerical integration meaningless. The main issue in this case is that the optimizer will exploit these inaccuracies in order to reduce the cost. Therefore, this issue is much more evident when solving optimal control problems than when simply simulating the system dynamics. Some simple approaches to tackle this issue are investigated, but no technique yielded satisfactory performances. It is our impression that ad-hoc integration strategies will be required in order to reliably simulate and optimize dynamics with continuous vaccination rates. While this will be the subject of future research, the results obtained with the current approximation have yielded sufficient accuracy.

**DISCUSSION ON SIMPLIFICATION (B)** This simplification has been proposed in Savorgnan et al.<sup>52</sup> as an approach to solve distributed optimal control problems by means of multiple shooting. In the original version, the coupling variable  $z$  is not necessarily piecewise constant, but rather piecewise polynomial. In simulations of this problem, the piecewise constant parametrization has been observed to yield sufficient accuracy.

The dynamics of each node are discretized using an explicit Runge-Kutta integrator of order four, with 50 integration steps per day. Alternative integrators such as explicit Euler, or implicit Runge-Kutta integrators, yielded similar results. Furthermore, in order to verify the accuracy of the integrator and the impact of the introduced simplifications on the solution accuracy, the system is simulated in open-loop, i.e. the optimal control trajectory is applied to the full model starting from the initial condition provided by the data assimilation scheme.

**DISCUSSION ON SIMPLIFICATION (C)** The mobility matrix is sparsified by pruning elements below a threshold (see fig. A4). This operation reduces the number of connection between nodes. Also in this case, the introduced simplification

<sup>52</sup> Savorgnan et al., “Multiple Shooting for Distributed Systems with Applications in Hydro Electricity Production” in: *Journal of Process Control* (2011).

has been verified through numerical simulations to have a small impact on the prediction and control accuracy.

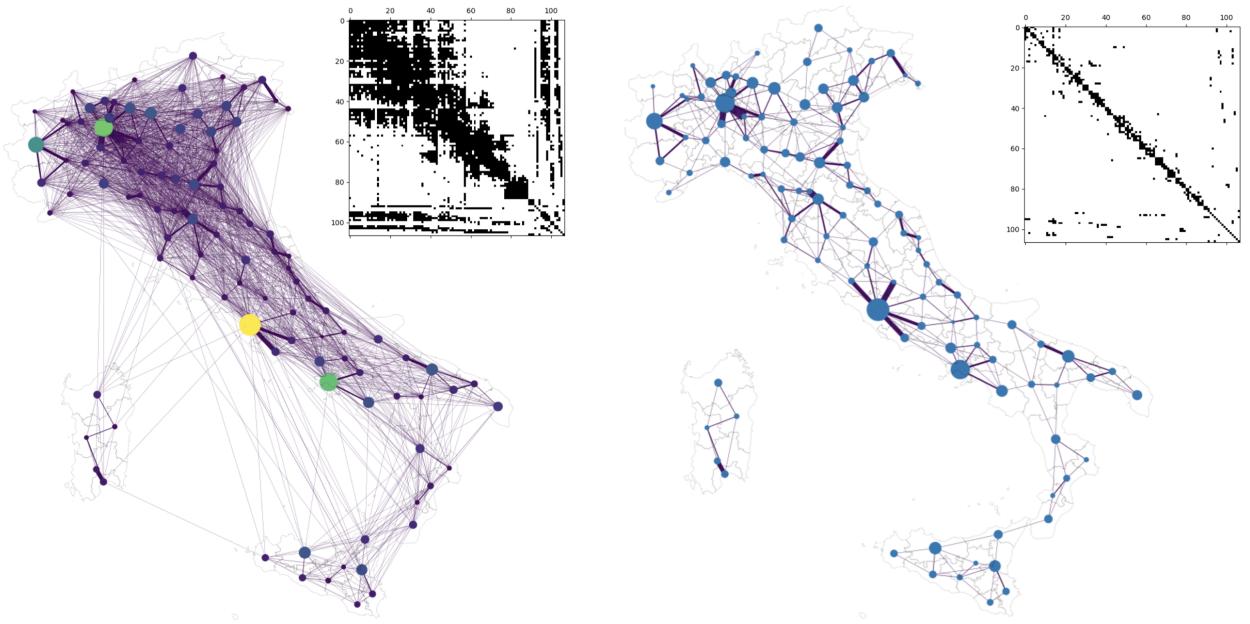


fig. A4: Simplification of the mobility matrix to obtain a sparse and tractable problem. After the optimization, the effectiveness of the optimal control strategy is assessed on the full model.

**POSSIBLE FURTHER IMPROVEMENTS** Applying optimal control in open loop, i.e., solving the optimization problem once and applying the control input over the whole time interval, may lead to poor performance due to model inaccuracy and external perturbations. A common remedy consists in closing the loop by repeatedly solving the optimal control problem by using the most updated information on the initial states. This is the principle behind Model Predictive Control (MPC)<sup>53</sup>. In this context, the state would be estimated on a daily, weekly, or monthly basis so as to solve again the optimal control problem and correct the optimal strategy.

#### DETAILS OF THE ALTERNATIVE STRATEGIES

Alternative strategies are created to be compared the optimal solutions. Each strategy uses a decision variable,  $\mathcal{U}_i$ , as a basis for the allocation of vaccines among provinces. The decision variable is one of:

- **MODELED FUTURE INCIDENCE, ABSOLUTE:** the modeled total future incidence in a no-vaccination scenario. This is equivalent to the objective of the optimal control problem with no control;
- **MODELED FUTURE INCIDENCE, PER POPULATION:** as above, but normalized by the resident population in each node;
- **MODELED INITIAL SUSCEPTIBILITY, ABSOLUTE:** the modeled number of susceptibles in each province at the start of the vaccination campaign;
- **MODELED INITIAL SUSCEPTIBILITY, PER POPULATION:** as above, but normalized by the resident population in each node;

<sup>53</sup> Rawlings et al., *Model Predictive Control: Theory, Computation, and Design* (2017).

- PROVINCE'S POPULATION.

Two strategies to distribute the doses are defined:

- **FOCUSED** Where every province is sorted (higher on top) according to its decision variable  $\mathcal{V}_i$ . The maximum local rate  $v_i^{max}$  is allocated to every province going down through the list, until the stockpile is empty. In other words, assuming an amount  $K$  of vaccines is available in the stockpile, the province index  $i$  that satisfy  $\max_i \mathcal{V}_i$  is searched for, and province  $i$  is assigned  $M_i = \min(v_i^{max}, K)$  vaccines. Then, the next province  $j$  that satisfy  $\max_{j,j \neq i} \mathcal{V}_j$  is searched for and assigned  $M_j = \min(v_j^{max}, K - M_i)$ . And so on, until no vaccines remains in the stockpile. This strategy will concentrate the allocation on nodes with the highest values of the considered decision variable.
- **PROPORTIONAL** In this case, assuming that on a given day there is a quantity of vaccine  $K$  in the stockpile, each province  $i$  receives an amount  $M_i = \min(v_i^{max}, K \cdot \frac{\mathcal{V}_i}{\sum_j \mathcal{V}_j})$ . This approach vaccinate each node proportionally to the value of its decision variable  $\mathcal{V}_i$ .

In the main text, the results for three alternative strategies are shown, namely *proportional absolute incidence*, *proportional population*, and *proportional susceptibility*—named respectively Incidence, Population, and Susceptibility. These strategy are good performers across scenarios, and show how different choices for the decision variables may affect the outcomes of the optimal control problem. In the next sections, the results for all these alternative strategies are shown.

## ADDITIONAL RESULTS

The results for all these strategies is presented in tab. A4, and are shown side-by-side in fig. A5. The optimal solutions outperforms all the others solution. In fact, for every given posterior realization, the optimal control solution always outperforms all other allocation strategies. Even if some scatter is observed when sampling the posterior, the performances of optimal strategies are clearly separated from the rest of the alternatives.

Scenario	Method	Averted Infections		Averted Infections per dose	
		Optimistic	Pessimistic	Optimistic	Pessimistic
2M	Optimal	6.98M	30.6M	0.268	1.18
	Incidence	6.32M	28.1M	0.243	1.08
	Proportional Incidence	6.23M	27.5M	0.239	1.06
	Focused Susceptibility	6.03M	26.9M	0.232	1.03
	Focused Proportional Susceptibility	6.03M	26.9M	0.232	1.03
	Focused Proportional Incidence	6.03M	26.9M	0.232	1.03
	Focused Population	6.03M	26.9M	0.232	1.03
	Focused Incidence	6.03M	26.9M	0.232	1.03
	Population	6.02M	26.8M	0.231	1.03
	Susceptibility	5.97M	26.7M	0.229	1.02
	Proportional Susceptibility	5.6M	25.3M	0.215	0.971
	Optimal	5.52M	24.1M	0.283	1.24
1.5M	Incidence	4.89M	21.7M	0.25	1.11
	Proportional Incidence	4.82M	21.3M	0.246	1.09
	Focused Population	4.58M	20.5M	0.235	1.05
	Focused Incidence	4.58M	20.5M	0.235	1.05
	Focused Proportional Incidence	4.58M	20.5M	0.235	1.05
	Focused Proportional Susceptibility	4.58M	20.5M	0.235	1.05
	Focused Susceptibility	4.58M	20.5M	0.235	1.05
	Population	4.57M	20.4M	0.234	1.05
	Susceptibility	4.51M	20.3M	0.231	1.04
	Proportional Susceptibility	4.18M	19.0M	0.214	0.975
	Optimal	3.9M	16.9M	0.3	1.3
	Incidence	3.41M	15.1M	0.262	1.16
1M	Proportional Incidence	3.34M	14.7M	0.257	1.13
	Focused Population	3.09M	13.9M	0.238	1.07
	Focused Susceptibility	3.09M	13.9M	0.238	1.07
	Focused Proportional Susceptibility	3.09M	13.9M	0.238	1.07
	Focused Incidence	3.09M	13.9M	0.238	1.07
	Focused Proportional Incidence	3.09M	13.9M	0.238	1.07
	Population	3.08M	13.8M	0.237	1.06
	Susceptibility	3.02M	13.7M	0.232	1.05
	Proportional Susceptibility	2.75M	12.6M	0.211	0.972
	Optimal	1.96M	8.39M	0.314	1.34
	Focused Proportional Incidence	1.95M	7.74M	0.312	1.24
	Proportional Incidence	1.69M	7.32M	0.271	1.17
479'700	Incidence	1.63M	7.21M	0.262	1.15
	Focused Incidence	1.59M	6.64M	0.254	1.06
	Focused Population	1.57M	6.85M	0.251	1.09
	Population	1.45M	6.57M	0.233	1.05
	Focused Susceptibility	1.45M	6.53M	0.232	1.04
	Susceptibility	1.41M	6.43M	0.225	1.03
	Focused Proportional Susceptibility	1.28M	6.09M	0.204	0.973
	Proportional Susceptibility	1.26M	5.89M	0.202	0.944

tab. A4: Absolute number of averted infections for each scenario.

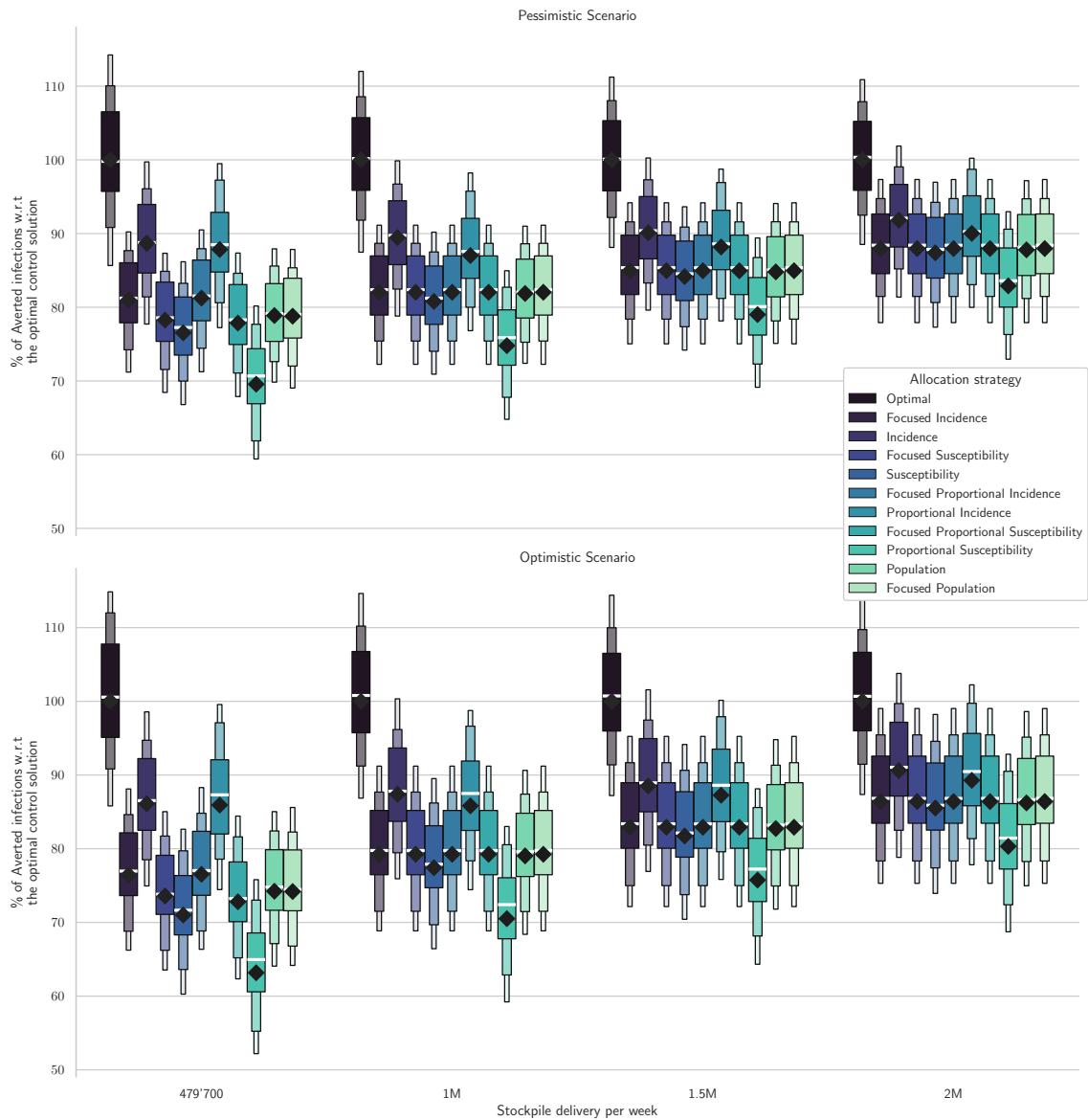


fig. A5: Comparison of different allocation strategies. Percentages of averted infections per vaccine dose from January 11, 2021 to April 11, 2021 using different vaccine distribution strategies for the pessimistic (top panel) and the optimistic (bottom panel) scenario based on: the optimal solution, the spatial distribution of the population, the amount of susceptible individuals at the beginning of the vaccination campaign, and the projected disease incidence in the absence of control. A median realization of the modeled posterior is optimized (diamonds), and the performance is assessed on the whole posterior (box plots). The results are normalized by the number of averted infections in the optimized solution (see tab. A4 for absolute values).



# BIBLIOGRAPHY

- Abubakar, Abdinasir, Andrew S. Azman, John Rumunu, Iza Ciglenecki, Trina Helderman, Haley West, Justin Lessler, David A. Sack, Stephen Martin, William Perea, Dominique Legros, and Francisco J. Luquero. "The First Use of the Global Oral Cholera Vaccine Emergency Stockpile: Lessons from South Sudan". In: *PLOS Medicine* 12.11 (Nov. 17, 2015), e1001901.
- Acemoglu, D., A. Fallah, A. Giometto, D. Huttenlocher, A. Ozdaglar, F. Parise, and S. Pattathil. *Optimal Adaptive Testing for Epidemic Control: Combining Molecular and Serology Tests*. Jan. 3, 2021. URL: <http://arxiv.org/abs/2101.00773> (visited on 03/08/2021).
- Acemoglu, Daron, Victor Chernozhukov, Iván Werning, and Michael D. Whinston. *Optimal Targeted Lockdowns in a Multi-Group SIR Model*. Working Paper 27102. National Bureau of Economic Research, May 2020.
- Adams, Patrick. "Cholera in Haiti Takes a Turn for the Worse". In: *The Lancet* 381.9874 (Apr. 13, 2013), p. 1264.
- . "Haiti Prepares for Cholera Vaccination but Concerns Remain". In: *The Lancet* 379.9810 (Jan. 7, 2012), p. 16.
- Akman, Olcay, Marina Romadan Corby, and Elsa Schaefer. "Examination of Models for Cholera: Insights into Model Comparison Methods". In: *Letters in Biomathematics* 3.1 (Jan. 1, 2016), pp. 93–118.
- Ali, Mohammad, Allyson R. Nelson, Anna Lena Lopez, and David A. Sack. "Updated Global Burden of Cholera in Endemic Countries". In: *PLoS Neglected Tropical Diseases* 9.6 (June 4, 2015).
- Althaus, Christian L. *Real-Time Modeling and Projections of the COVID-19 Epidemic in Switzerland*. 2020. URL: <https://ispmbern.github.io/covid-19/swiss-epidemic-model/> (visited on 04/26/2020).
- Anderson, Roy M. and Robert M. May. "Population Biology of Infectious Diseases: Part I". In: *Nature* 280.5721 (Aug. 1979), pp. 361–367.
- Andersson, Joel A E, Joris Gillis, Greg Horn, James B Rawlings, and Moritz Diehl. "CasADI – A Software Framework for Nonlinear Optimization and Optimal Control". In: *Mathematical Programming Computation* (In Press, 2018).
- Asfaw, Kidus, Joonha Park, Allister Ho, Aaron A. King, and Edward Ionides. *Partially Observed Markov Processes with Spatial Structure via the R Package spatPomp*. May 3, 2021. URL: <http://arxiv.org/abs/2101.01157> (visited on 06/26/2021).
- Azman, Andrew S, Lucy A Parker, John Rumunu, Fisseha Tadesse, Francesco Grandesso, Lul L Deng, Richard Laku Lino, Bior K Bior, Michael Lasuba, Anne-Laure Page, Lameck Ontweka, Augusto E Llosa, Sandra Cohuet, Lorenzo Pezzoli, Dossou Vincent Sodjinou, Abdinasir Abubakar, Amanda K Debes, Allan M Mpairwe, Joseph F Wamala, Christine Jamet, Justin Lessler, David A Sack, Marie-Laure Quilici, Iza Ciglenecki, and Francisco J Luquero. "Effectiveness of One Dose of Oral Cholera Vaccine in Response to an Outbreak: A Case-Cohort Study". In: *The Lancet Global Health* 4.11 (Nov. 1, 2016), e856–e863.
- Azman, Andrew S., Francisco J. Luquero, Iza Ciglenecki, Rebecca F. Grais, David A. Sack, and Justin Lessler. "The Impact of a One-Dose versus Two-Dose Oral Cholera Vaccine Regimen in Outbreak Settings: A Modeling Study". In: *PLOS Medicine* 12.8 (Aug. 25, 2015), e1001867.
- Azman, Andrew S., Francisco J. Luquero, Amabelia Rodrigues, Pedro Pablo Palma, Rebecca F. Grais, Cunhate Na Banga, Bryan T. Grenfell, and Justin Lessler. "Urban Cholera Transmission Hotspots and Their Implications for Reactive Vaccination: Evidence from Bissau City, Guinea Bissau". In: *PLOS Neglected Tropical Diseases* 6.11 (Nov. 8, 2012), e1901.
- Azman, Andrew S., Kara E. Rudolph, Derek A. T. Cummings, and Justin Lessler. "The Incubation Period of Cholera: A Systematic Review". In: *Journal of Infection* 66.5 (May 1, 2013), pp. 432–438.
- Azman, Andrew S., John Rumunu, Abdinasir Abubakar, Haley West, Iza Ciglenecki, Trina Helderman, Joseph Francis Wamala, Olimpia de la Rosa Vázquez, William Perea, David A. Sack, Dominique Legros, Stephen Martin, Justin Lessler, and

- Francisco J. Luquero. "Population-Level Effect of Cholera Vaccine on Displaced Populations, South Sudan, 2014". In: *Emerging Infectious Diseases* 22.6 (June 2016), pp. 1067–1070.
- Baden, L.R., H.M. El Sahly, B. Essink, and *et al.* "Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine". In: *New England Journal of Medicine* 384 (2020), pp. 403–416.
- Baker-Austin, Craig, Joaquin A. Trinanes, Nick G. H. Taylor, Rachel Hartnell, Anja Siiton, and Jaime Martinez-Urtaza. "Emerging Vibrio Risk at High Latitudes in Response to Ocean Warming". In: *Nature Climate Change* 3.1 (1 Jan. 2013), pp. 73–77.
- Balcan, Duygu, Bruno Gonçalves, Hao Hu, José J. Ramasco, Vittoria Colizza, and Alessandro Vespignani. "Modeling the Spatial Spread of Infectious Diseases: The GLobal Epidemic and Mobility Computational Model". In: *Journal of Computational Science* 1.3 (Aug. 1, 2010), pp. 132–145.
- Baracchini, Theo, Aaron A. King, Menno J. Bouma, Xavier Rodó, Enrico Bertuzzo, and Mercedes Pascual. "Seasonality in Cholera Dynamics: A Rainfall-Driven Model Explains the Wide Range of Patterns in Endemic Areas". In: *Advances in Water Resources* 108 (Oct. 1, 2017), pp. 357–366.
- Barzilay, Ezra J., Nicolas Schaad, Roc Magloire, Kam S. Mung, Jacques Boncy, Georges A. Dahourou, Eric D. Mintz, Maria W. Steenland, John F. Vertefeuille, and Jordan W. Tappero. "Cholera Surveillance during the Haiti Epidemic — The First 2 Years". In: *New England Journal of Medicine* 368.7 (Feb. 14, 2013), pp. 599–609.
- Bellot, Alexis and Mihaela Schaar. "Tree-Based Bayesian Mixture Model for Competing Risks". *International Conference on Artificial Intelligence and Statistics*. International Conference on Artificial Intelligence and Statistics. Mar. 31, 2018, pp. 910–918.
- Bertuzzo, E., S. Azaele, A. Maritan, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo. "On the Space-Time Evolution of a Cholera Epidemic". In: *Water Resources Research* 44.1 (2008).
- Bertuzzo, E., R. Casagrandi, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo. "On Spatially Explicit Models of Cholera Epidemics". In: *Journal of The Royal Society Interface* 7.43 (Feb. 6, 2010), pp. 321–333.
- Bertuzzo, E., L. Mari, L. Righetto, M. Gatto, R. Casagrandi, I. Rodriguez-Iturbe, and A. Rinaldo. "Hydroclimatology of Dual-Peak Annual Cholera Incidence: Insights from a Spatially Explicit Model". In: *Geophysical Research Letters* 39.5 (2012).
- Bertuzzo, Enrico, Flavio Finger, Lorenzo Mari, Marino Gatto, and Andrea Rinaldo. "On the Probability of Extinction of the Haiti Cholera Epidemic". In: *Stochastic Environmental Research and Risk Assessment* 30.8 (Dec. 1, 2016), pp. 2043–2055.
- Bertuzzo, Enrico, Lorenzo Mari, Damiano Pasetto, Stefano Miccoli, Renato Casagrandi, Marino Gatto, and Andrea Rinaldo. "The Geography of COVID-19 Spread in Italy and Implications for the Relaxation of Confinement Measures". In: *Nature Communications* 11.1 (1 Aug. 26, 2020), p. 4264.
- Bertuzzo, Mari L., Righetto L., Gatto M., Casagrandi R., Blokesch M., Rodriguez-Iturbe I., and Rinaldo A. "Prediction of the Spatial Evolution and Effects of Control Measures for the Unfolding Haiti Cholera Outbreak". In: *Geophysical Research Letters* 38.6 (Mar. 18, 2011).
- Betts, J.T. *Practical Methods for Optimal Control and Estimation Using Nonlinear Programming*. 2nd ed. SIAM, 2010.
- Bi, Qifang, Andrew S. Azman, Syed Moinuddin Satter, Azharul Islam Khan, Dilruba Ahmed, Altaf Ahmed Raj, Emily S. Gurley, and Justin Lessler. "Micro-Scale Spatial Clustering of Cholera Risk Factors in Urban Bangladesh". In: *PLOS Neglected Tropical Diseases* 10.2 (Feb. 11, 2016), e0004400.
- Bi, Qifang, Eva Ferreras, Lorenzo Pezzoli, Dominique Legros, Louise C Ivers, Kashmira Date, Firdausi Qadri, Laura Digilio, David A Sack, Mohammad Ali, Justin Lessler, Francisco J Luquero, and Andrew S Azman. "Protection against Cholera from Killed Whole-Cell Oral Cholera Vaccines: A Systematic Review and Meta-Analysis". In: *The Lancet Infectious Diseases* 17.10 (Oct. 1, 2017), pp. 1080–1088.
- Bi, Qifang, Yongsheng Wu, Shuijiang Mei, Chenfei Ye, Xuan Zou, Zhen Zhang, Xiaojian Liu, Lan Wei, Shaun A Truelove, Tong Zhang, Wei Gao, Cong Cheng, Xiujuan Tang, Xiaoliang Wu, Yu Wu, Binbin Sun, Suli Huang, Yu Sun, Juncen Zhang, Ting Ma, Justin Lessler, and Tiejian Feng. "Epidemiology and Transmission of COVID-19 in 391 Cases and 1286 of Their Close Contacts in Shenzhen, China: A Retrospective Cohort Study". In: *The Lancet Infectious Diseases* (Apr. 27, 2020).
- Biegler, Lorenz T. *Nonlinear Programming*. MOS-SIAM Series on Optimization. SIAM, 2010.

- Bock, H.G. and K.J. Plitt. "A Multiple Shooting Algorithm for Direct Solution of Optimal Control Problems". *Proceedings 9th IFAC World Congress Budapest*. Pergamon Press, 1984, pp. 242–247.
- Borcherding, Rebecca K. "Modeling of Future COVID-19 Cases, Hospitalizations, and Deaths, by Vaccination Rates and Non-pharmaceutical Intervention Scenarios — United States, April–September 2021". In: *MMWR. Morbidity and Mortality Weekly Report* 70 (2021).
- Box, George E. P. "Science and Statistics". In: *Journal of the American Statistical Association* 71.356 (Dec. 1, 1976), pp. 791–799.
- Branas, Charles C., Andrew Rundle, Sen Pei, Wan Yang, Brendan G. Carr, Sarah Sims, Alexis Zebrowski, Ronan Doorley, Neil Schluger, James W. Quinn, and Jeffrey Shaman. "Flattening the Curve before It Flattens Us: Hospital Critical Care Capacity Limits and Mortality from Novel Coronavirus (SARS-CoV2) Cases in US Counties". In: *medRxiv* (Apr. 6, 2020), p. 2020.04.01.20049759.
- Bretó, Carles, Daihai He, Edward L. Ionides, and Aaron A. King. "Time Series Analysis via Mechanistic Models". In: *The Annals of Applied Statistics* 3.1 (Mar. 2009), pp. 319–348.
- Bretó, Carles and Edward L. Ionides. "Compound Markov Counting Processes and Their Applications to Modeling Infinitely Over-Dispersed Systems". In: *Stochastic Processes and their Applications* 121.11 (Nov. 1, 2011), pp. 2571–2591.
- Brouwer, Andrew F., Mark H. Weir, Marisa C. Eisenberg, Rafael Meza, and Joseph N. S. Eisenberg. "Dose-Response Relationships for Environmentally Mediated Infectious Disease Transmission Models". In: *PLOS Computational Biology* 13.4 (Apr. 7, 2017), e1005481.
- Browning, Alexander P., David J. Warne, Kevin Burrage, Ruth E. Baker, and Matthew J. Simpson. "Identifiability Analysis for Stochastic Differential Equation Models in Systems Biology". In: *Journal of The Royal Society Interface* 17.173 (Dec. 23, 2020), p. 20200652.
- Bubar, Kate M., Kyle Reinholt, Stephen M. Kissler, Marc Lipsitch, Sarah Cobey, Yonatan H. Grad, and Daniel B. Larremore. "Model-Informed COVID-19 Vaccine Prioritization Strategies by Age and Serostatus". In: *Science* (Jan. 21, 2021).
- Buckee, Caroline, Abdisalan Noor, and Lisa Sattenspiel. "Thinking Clearly about Social Aspects of Infectious Disease Transmission". In: *Nature* 595.7866 (7866 July 2021), pp. 205–213.
- Butler, Susan M., Eric J. Nelson, Nitayana Chowdhury, Shah M. Faruque, Stephen B. Calderwood, and Andrew Camilli. "Cholera Stool Bacteria Repress Chemotaxis to Increase Infectivity". In: *Molecular Microbiology* 60.2 (2006), pp. 417–426.
- Byrne, Joseph Patrick. *Encyclopedia of Pestilence, Pandemics, and Plagues: A-M*. Greenwood Press, 2008. 872 pp.
- Camacho, Anton, Malika Bouhenia, Reema Alyusfi, Abdulhakeem Alkohlani, Munna Abdulla Mohammed Naji, Xavier de Radiguès, Abdinasir M Abubakar, Abdulkareem Almoalmi, Caroline Seguin, Maria Jose Sagrado, Marc Poncin, Melissa McRae, Mohammed Musoke, Ankur Rakesh, Klaudia Porten, Christopher Haskew, Katherine E Atkins, Rosalind M Eggo, Andrew S Azman, Marije Broekhuijsen, Mehmet Akif Saatcioglu, Lorenzo Pezzoli, Marie-Laure Quilici, Abdul Rahman Al-Mesbahy, Nevio Zagaria, and Francisco J Luquero. "Cholera Epidemic in Yemen, 2016–18: An Analysis of Surveillance Data". In: *The Lancet Global Health* 6.6 (June 1, 2018), e680–e690.
- Capasso, V. and S. L. Paveri-Fontana. "A Mathematical Model for the 1973 Cholera Epidemic in the European Mediterranean Region". In: *Revue D'épidémiologie Et De Santé Publique* 27.2 (Sept. 18, 1979), pp. 121–132.
- Carpenter, Bob, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. "Stan: A Probabilistic Programming Language". In: *Journal of Statistical Software* 76.1 (1 Jan. 11, 2017), pp. 1–32.
- Cash, Benjamin A., Xavier Rodó, Michael Emch, Md Yunus, Abu S. G. Faruque, and Mercedes Pascual. "Cholera and Shigellosis: Different Epidemiology but Similar Responses to Climate Variability". In: *PLOS ONE* 9.9 (Sept. 17, 2014), e107223.
- Cazelles, Bernard, Clara Champagne, and Joseph Dureau. "Accounting for Non-Stationarity in Epidemiology by Embedding Time-Varying Parameters in Stochastic Models". In: *PLOS Computational Biology* 14.8 (Aug. 15, 2018), e1006211.
- CDC. *Diagnosis and Detection | Cholera | CDC*. Dec. 13, 2018–. URL: <https://www.cdc.gov/cholera/diagnosis.html> (visited on 07/04/2021).
- Chao, Dennis L., M. Elizabeth Halloran, and Ira M. Longini. "Vaccination Strategies for Epidemic Cholera in Haiti with Implications for the Developing World". In: *Proceedings of the National Academy of Sciences* 108.17 (Apr. 26, 2011), pp. 7081–7085.

- Chatzimanolakis, Michail, Pascal Weber, Georgios Arampatzis, Daniel Wälchli, Ivica Kičić, Petr Karnakov, Costas Papadimitriou, and Petros Koumoutsakos. "Optimal Allocation of Limited Test Resources for the Quantification of COVID-19 Infections". In: *Swiss Medical Weekly* 150.5153 (Dec. 16, 2020).
- Chen, Jiangzhuo, Stefan Hoops, Achla Marathe, Henning Mortveit, Bryan Lewis, Srinivasan Venkatramanan, Arash Hadadian, Parantapa Bhattacharya, Abhijin Adiga, Anil Vullikanti, Aravind Srinivasan, Mandy L. Wilson, Gal Ehrlich, Maier Fenster, Stephen Eubank, Christopher Barrett, and Madhav Marathe. "Prioritizing Allocation of COVID-19 Vaccines Based on Social Contacts Increases Vaccination Effectiveness". In: *medRxiv* (Feb. 16, 2021), p. 2021.02.04.21251012.
- Chinazzi, Matteo, Jessica T. Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti, Kunpeng Mu, Luca Rossi, Kaiyuan Sun, Cécile Viboud, Xinyue Xiong, Hongjie Yu, M. Elizabeth Halloran, Ira M. Longini, and Alessandro Vespignani. "The Effect of Travel Restrictions on the Spread of the 2019 Novel Coronavirus (COVID-19) Outbreak". In: *Science* 368.6489 (Mar. 2020), pp. 395–400.
- Clemens, J. D., F. van Loon, D. A. Sack, M. R. Rao, F. Ahmed, J. Chakraborty, B. A. Kay, M. R. Khan, M. Yunus, J. R. Harris, J. D. Clemens, M. R. Rao, D. A. Sack, J. R. Harris, J. D. Clemens, M. R. Rao, A-M. Svennerholm, and J. Holmgren. "Biotype as Determinant of Natural Immunising Effect of Cholera". In: *The Lancet*. Originally Published as Volume 1, Issue 8746 337.8746 (Apr. 13, 1991), pp. 883–884.
- Codeço, Cláudia Torres. "Endemic and Epidemic Dynamics of Cholera: The Role of the Aquatic Reservoir". In: *BMC Infectious Diseases* 1 (Feb. 2, 2001), p. 1.
- Colwell, Rita R. "Global Climate and Infectious Disease: The Cholera Paradigm". In: *Science* 274.5295 (Dec. 20, 1996), pp. 2025–2031.
- Colwell, Rita R., Anwar Huq, M. Sirajul Islam, K. M. A. Aziz, M. Yunus, N. Huda Khan, A. Mahmud, R. Bradley Sack, G. B. Nair, J. Chakraborty, David A. Sack, and E. Russek-Cohen. "Reduction of Cholera in Bangladeshi Villages by Simple Filtration". In: *Proceedings of the National Academy of Sciences* 100.3 (Feb. 4, 2003), pp. 1051–1055.
- Cori, Anne, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. "A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics". In: *American Journal of Epidemiology* 178.9 (Nov. 1, 2013), pp. 1505–1512.
- COVID-19 Open Access Project. *Living Evidence on COVID-19*. 2020. URL: <https://ispbbern.github.io/covid-19/living-review/>.
- Cowling, Benjamin J, Sheikh Taslim Ali, Tiffany W Y Ng, Tim K Tsang, Julian C M Li, Min Whui Fong, Qiuyan Liao, Mike YW Kwan, So Lun Lee, Susan S Chiu, Joseph T Wu, Peng Wu, and Gabriel M Leung. "Impact Assessment of Non-Pharmaceutical Interventions against Coronavirus Disease 2019 and Influenza in Hong Kong: An Observational Study". In: *The Lancet Public Health* (Apr. 17, 2020).
- Cramer, Estee Y., Evan L. Ray, Velma K. Lopez, Johannes Bracher, Andrea Brennen, Alvaro J. Castro Rivadeneira, Aaron Gerding, Tilmann Gneiting, Katie H. House, Yuxin Huang, Dasuni Jayawardena, Abdul H. Kanji, Ayush Khandelwal, Khoa Le, Anja Mühlemann, Jarad Niemi, Apurv Shah, Ariane Stark, Yijin Wang, Nutcha Wattanachit, Martha W. Zorn, Youyang Gu, Sansiddh Jain, Nayana Bannur, Ayush Deva, Mihir Kulkarni, Srujana Merugu, Alpan Raval, Siddhant Shingi, Avtansh Tiwari, Jerome White, Spencer Woody, Maytal Dahan, Spencer Fox, Kelly Gaither, Michael Lachmann, Lauren Ancel Meyers, James G. Scott, Mauricio Tec, Ajitesh Srivastava, Glover E. George, Jeffrey C. Cegan, Ian D. Dettwiller, William P. England, Matthew W. Farthing, Robert H. Hunter, Brandon Lafferty, Igor Linkov, Michael L. Mayo, Matthew D. Parno, Michael A. Rowland, Benjamin D. Trump, Sabrina M. Corsetti, Thomas M. Baer, Marisa C. Eisenberg, Karl Falb, Yitao Huang, Emily T. Martin, Ella McCauley, Robert L. Myers, Tom Schwarz, Daniel Sheldon, Graham Casey Gibson, Rose Yu, Liyao Gao, Yian Ma, Dongxia Wu, Xifeng Yan, Xiaoyong Jin, Yu-Xiang Wang, YangQuan Chen, Lihong Guo, Yanting Zhao, Quanquan Gu, Jinghui Chen, Lingxiao Wang, Pan Xu, Weitong Zhang, Difan Zou, Hannah Biegel, Joceline Lega, Timothy L. Snyder, Davison D. Wilson, Steve McConnell, Robert Walraven, Yunfeng Shi, Xuegang Ban, Qi-Jun Hong, Stanley Kong, James A. Turtle, Michal Ben-Nun, Pete Riley, Steven Riley, Ugur Koyluoglu, David DesRoches, Bruce Hamory, Christina Kyriakides, Helen Leis, John Milliken, Michael Moloney, James Morgan, Gokce Ozcan, Chris Schrader, Elizabeth Shakhnovich, Daniel Siegel, Ryan Spatz, Chris Stiefeling, Barrie Wilkinson, Alexander Wong, Zhifeng Gao, Jiang Bian, Wei Cao, Juan Lavista Ferres, Chaozhuo Li, Tie-Yan Liu, Xing Xie, Shun Zhang, Shun Zheng, Alessandro Vespignani, Matteo Chinazzi, Jessica T. Davis, Kunpeng Mu, Ana Pastore y Piontti, Xinyue Xiong, Andrew Zheng, Jackie Baek, Vivek Farias, Andreea Georgescu, Retsef Levi, Deeksha Sinha, Joshua Wilde, Nicolas D. Penna, Leo A. Celi,

- Saketh Sundar, Sean Cavany, Guido España, Sean Moore, Rachel Oidtman, Alex Perkins, Dave Osthuis, Lauren Castro, Geoffrey Fairchild, Isaac Michaud, Dean Karlen, Elizabeth C. Lee, Juan Dent, Kyra H. Grantz, Joshua Kaminsky, Kathryn Kaminsky, Lindsay T. Keegan, Stephen A. Lauer, Joseph C. Lemaitre, Justin Lessler, Hannah R. Meredith, Javier Perez-Saez, Sam Shah, Claire P. Smith, Shaun A. Truelove, Josh Wills, Matt Kinsey, R. F. Obrecht, Katharine Tallaksen, John C. Burant, Lily Wang, Lei Gao, Zhiling Gu, Myungjin Kim, Xinyi Li, Guannan Wang, Yueying Wang, Shan Yu, Robert C. Reiner, Ryan Barber, Emmanuela Gaikedu, Simon Hay, Steve Lim, Chris Murray, David Pigott, B. Aditya Prakash, Bijaya Adhikari, Jiaming Cui, Alexander Rodríguez, Anika Tabassum, Jiajia Xie, Pinar Keskinocak, John Asplund, Arden Baxter, Buse Eylul Oruc, Nicoleta Serban, Sercan O. Arik, Mike Dusenberry, Arkady Epshteyn, Elli Kanal, Long T. Le, Chun-Liang Li, Tomas Pfister, Dario Sava, Rajarishi Sinha, Thomas Tsai, Nate Yoder, Jinsung Yoon, Leyou Zhang, Sam Abbott, Nikos I. Bosse, Sebastian Funk, Joel Hellewel, Sophie R. Meakin, James D. Munday, Katherine Sherratt, Mingyuan Zhou, Rahi Kalantari, Teresa K. Yamana, Sen Pei, Jeffrey Shaman, Turgay Ayer, Madeline Adee, Jagpreet Chhatwal, Ozden O. Dalgic, Mary A. Ladd, Benjamin P. Linas, Peter Mueller, Jade Xiao, Michael L. Li, Dimitris Bertsimas, Omar Skali Lami, Saksham Soni, Hamza Tazi Bouardi, Yuanjia Wang, Qinxia Wang, Shanghong Xie, Donglin Zeng, Alden Green, Jacob Bien, Addison J. Hu, Maria Jahja, Balasubramanian Narasimhan, Samyak Rajanala, Aaron Rumack, Noah Simon, Ryan Tibshirani, Rob Tibshirani, Valerie Ventura, Larry Wasserman, Eamon B. O'Dea, John M. Drake, Robert Pagano, Jo W. Walker, Rachel B. Slayton, Michael Johansson, Matthew Biggerstaff, and Nicholas G. Reich. "Evaluation of Individual and Ensemble Probabilistic Forecasts of COVID-19 Mortality in the US". In: *medRxiv* (Feb. 5, 2021), p. 2021.02.03.21250974.
- Davies, Nicholas G., Petra Klepac, Yang Liu, Kiesha Prem, Mark Jit, and Rosalind M. Eggo. "Age-Dependent Effects in the Transmission and Control of COVID-19 Epidemics". In: *Nature Medicine* 26.8 (8 Aug. 2020), pp. 1205–1211.
- De, S. N., J. K. Sarkar, and B. P. Tribedi. "An Experimental Study of the Action of Cholera Toxin". In: *The Journal of Pathology and Bacteriology* 63.4 (1951), pp. 707–717.
- Domman, Daryl, Fahima Chowdhury, Ashraful I. Khan, Matthew J. Dorman, Ankur Mutreja, Muhammad Ikhtear Uddin, Anik Paul, Yasmin A. Begum, Richelle C. Charles, Stephen B. Calderwood, Taufiqur R. Bhuiyan, Jason B. Harris, Regina C. LaRocque, Edward T. Ryan, Firdausi Qadri, and Nicholas R. Thomson. "Defining Endemic Cholera at Three Levels of Spatiotemporal Resolution within Bangladesh". In: *Nature Genetics* 50.7 (July 2018), pp. 951–955.
- Dong, Ensheng, Hongru Du, and Lauren Gardner. "An Interactive Web-Based Dashboard to Track COVID-19 in Real Time". In: *The Lancet Infectious Diseases* 20.5 (May 1, 2020), pp. 533–534.
- Du, Zhanwei, Abhishek Pandey, Yuan Bai, Meagan C. Fitzpatrick, Matteo Chinazzi, Ana Pastore y Piontti, Michael Lachmann, Alessandro Vespignani, Benjamin J. Cowling, Alison P. Galvani, and Lauren Ancel Meyers. "Comparative Cost-Effectiveness of SARS-CoV-2 Testing Strategies in the USA: A Modelling Study". In: *The Lancet Public Health* 6.3 (Mar. 1, 2021), e184–e191.
- Duque, Daniel, David P. Morton, Bismark Singh, Zhanwei Du, Remy Pasco, and Lauren Ancel Meyers. "COVID-19: How to Relax Social Distancing If You Must". In: *medRxiv* (May 5, 2020), p. 2020.04.29.20085134.
- Eisenberg, Marisa C., Gregory Kujbida, Ashleigh R. Tuite, David N. Fisman, and Joseph H. Tien. "Examining Rainfall and Cholera Dynamics in Haiti Using Statistical and Dynamic Modeling Approaches". In: *Epidemics* 5.4 (Dec. 1, 2013), pp. 197–207.
- Eisenberg, Marisa C., Suzanne L. Robertson, and Joseph H. Tien. "Identifiability and Estimation of Multiple Transmission Pathways in Cholera and Waterborne Disease". In: *Journal of Theoretical Biology* 324 (May 7, 2013), pp. 84–102.
- Emanuel, Ezekiel J., Govind Persad, Adam Kern, Allen Buchanan, Cécile Fabre, Daniel Halliday, Joseph Heath, Lisa Herzog, R. J. Leland, Ephrem T. Lemango, Florencia Luna, Matthew S. McCoy, Ole F. Norheim, Trygve Ottersen, G. Owen Schaefer, Kok-Chor Tan, Christopher Heath Wellman, Jonathan Wolff, and Henry S. Richardson. "An Ethical Framework for Global Vaccine Allocation". In: *Science* 369.6509 (Sept. 11, 2020), pp. 1309–1312.
- Emch, Michael, Caryl Feldacker, M. Sirajul Islam, and Mohammad Ali. "Seasonality of Cholera from 1974 to 2005: A Review of Global Patterns". In: *International Journal of Health Geographics* 7 (June 20, 2008), p. 31.
- Erlander, Sven and Neil F. Stewart. *The Gravity Model in Transportation Analysis: Theory and Extensions*. VSP, Dec. 1990. 248 pp.
- Escobar, Luis E., Sadie J. Ryan, Anna M. Stewart-Ibarra, Julia L. Finkelstein, Christine A. King, Huijie Qiao, and Mark E. Polhemus. "A Global Map of Suitability for Coastal Vibrio Cholerae under Current and Future Climate Conditions". In: *Acta Tropica* 149 (Sept. 1, 2015), pp. 202–211.

- Faruque, Shah M., Iftekhar Bin Naser, M. Johirul Islam, A. S. G. Faruque, A. N. Ghosh, G. Balakrish Nair, David A. Sack, and John J. Mekalanos. "Seasonal Epidemics of Cholera Inversely Correlate with the Prevalence of Environmental Cholera Phages". In: *Proceedings of the National Academy of Sciences* 102.5 (Feb. 1, 2005), pp. 1702–1707.
- Ferguson, N., D. Laydon, G. Nedjati Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunuba Perez, G. Cuomo-Dannenburg, A. Dighe, I. Dorigatti, H. Fu, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, L. Okell, S. Van Elsland, H. Thompson, R. Verity, E. Volz, H. Wang, Y. Wang, P. Walker, C. Walters, P. Winskill, C. Whittaker, C. Donnelly, S. Riley, and A. Ghani. *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand*. Report. Mar. 16, 2020.
- Fernandez-Cassi, Xavier, Andreas Scheidegger, Carola Bänziger, Federica Cariti, Alex Tuñas Corzon, Pravin Ganesanandamoorthy, Joseph C. Lemaitre, Christoph Ort, Timothy R. Julian, and Tamar Kohn. "Wastewater Monitoring Outperforms Case Numbers as a Tool to Track COVID-19 Incidence Dynamics When Test Positivity Rates Are High". In: *Water Research* 200 (July 15, 2021), p. 117252.
- Fewtrell, Lorna, Rachel B. Kaufmann, David Kay, Wayne Enanoria, Laurence Haller, and John M. Colford. "Water, Sanitation, and Hygiene Interventions to Reduce Diarrhoea in Less Developed Countries: A Systematic Review and Meta-Analysis". In: *The Lancet Infectious Diseases* 5.1 (Jan. 1, 2005), pp. 42–52.
- Finger, Flavio, Enrico Bertuzzo, Francisco J. Luquero, Nathan Naibei, Brahma Touré, Maya Allan, Klaudia Porten, Justin Lessler, Andrea Rinaldo, and Andrew S. Azman. "The Potential Impact of Case-Area Targeted Interventions in Response to Cholera Outbreaks: A Modeling Study". In: *PLOS Medicine* 15.2 (Feb. 27, 2018), e1002509.
- Finkenstädt, Bärbel F., Ottar N. Bjørnstad, and Bryan T. Grenfell. "A Stochastic Model for Extinction and Recurrence of Epidemics: Estimation and Inference for Measles Outbreaks". In: *Biostatistics* 3.4 (Dec. 1, 2002), pp. 493–510.
- Firth, Josh A., Joel Hellewell, Petra Klepac, Stephen Kissler, Adam J. Kucharski, and Lewis G. Spurgin. "Using a Real-World Network to Model Localized COVID-19 Control Strategies". In: *Nature Medicine* 26.10 (10 Oct. 2020), pp. 1616–1622.
- Fitzpatrick, Meagan C. and Alison P. Galvani. "Optimizing Age-Specific Vaccination". In: *Science* (Jan. 21, 2021).
- Flaxman, Seth, Swapnil Mishra, Axel Gandy, Juliette T Unwin, Helen Coupland, Thomas A. Mellan, Harisson Zhu, Tresnia Berah, Jeffrey Eaton, Pablo Perez Guzman, Nora Schmit, Lucia Cilloni, Kylie Ainslie, Marc Baguelin, Isobel Blake, Adhiratha Boonyasiri, Olivia Boyd, Lorenzo Cattarino, Constanze Ciavarella, Laura Cooper, Zulma Cucunuba Perez, Gina Cuomo-Dannenburg, Amy Dighe, Bimandra Djafara, Ilaria Dorigatti, Sabine Van Elsland, Rich Fitzjohn, Han Fu, Katy Gaythorpe, Lily Geidelberg, Nicholas Grassly, Will Green, Timothy Hallett, Arran Hamlet, Wes Hinsley, Ben Jeffrey, David Jorgensen, Edward Knock, Daniel Laydon, Gemma Nedjati Gilani, Pierre Nouvellet, Kris Parag, Igor Siveroni, Hayley Thompson, Robert Verity, Erik Volz, Caroline Walters, Haowei Wang, Yuanrong Wang, Olivier Watson, Peter Winskill, Xiaoyue Xi, Charles Whittaker, Patrick Walker, Azra Ghani, Cristl Donnelly, Steven Riley, Lucy Okell, Michaela Vollmer, Neil Ferguson, and Samir Bhatt. *Report 13: Estimating the Number of Infections and the Impact of Non-Pharmaceutical Interventions on COVID-19 in 11 European Countries*. Report. Mar. 30, 2020.
- Freedman, David. "From Association to Causation: Some Remarks on the History of Statistics". In: *Statistical Science* 14.3 (Aug. 1999), pp. 243–258.
- Frerichs, R. R., P. S. Keim, R. Barrais, and R. Piarroux. "Nepalese Origin of Cholera Epidemic in Haiti". In: *Clinical Microbiology and Infection* 18.6 (June 1, 2012), E158–E163.
- Fung, Isaac Chun-Hai. "Cholera Transmission Dynamic Models for Public Health Practitioners". In: *Emerging Themes in Epidemiology* 11.1 (Feb. 12, 2014), p. 1.
- Gallagher, Molly E., Andrew J. Sieben, Kristin N. Nelson, Alicia N. M. Kraay, Walter A. Orenstein, Ben Lopman, Andreas Handel, and Katia Koelle. "Indirect Benefits Are a Crucial Consideration When Evaluating SARS-CoV-2 Vaccine Candidates". In: *Nature Medicine* 27.1 (1 Jan. 2021), pp. 4–5.
- Ganyani, Tapiwa, Cecile Kremer, Dongxuan Chen, Andrea Torneri, Christel Faes, Jacco Wallinga, and Niel Hens. "Estimating the Generation Interval for COVID-19 Based on Symptom Onset Data". In: *medRxiv* (Mar. 8, 2020), p. 2020.03.05.20031815.
- Gatto, Marino, Enrico Bertuzzo, Lorenzo Mari, Stefano Miccoli, Luca Carraro, Renato Casagrandi, and Andrea Rinaldo. "Spread and Dynamics of the COVID-19 Epidemic in Italy: Effects of Emergency Containment Measures". In: *Proceedings of the National Academy of Sciences* (Apr. 23, 2020).

- Gatto, Marino, Lorenzo Mari, Enrico Bertuzzo, Renato Casagrandi, Lorenzo Righetto, Ignacio Rodriguez-Iturbe, and Andrea Rinaldo. "Generalized Reproduction Numbers and the Prediction of Patterns in Waterborne Disease". In: *Proceedings of the National Academy of Sciences* 109.48 (Nov. 27, 2012), pp. 19703–19708.
- Gaudart, Jean, Stanislas Rebaudet, Robert Barrais, Jacques Boncy, Benoit Faucher, Martine Piarroux, Roc Magloire, Gabriel Thimothe, and Renaud Piarroux. "Spatio-Temporal Dynamics of Cholera during the First Year of the Epidemic in Haiti". In: *PLOS Neglected Tropical Diseases* 7.4 (Apr. 4, 2013), e2145.
- Gelman, Andrew and Aki Vehtari. *What Are the Most Important Statistical Ideas of the Past 50 Years?* June 3, 2021. URL: <http://arxiv.org/abs/2012.00174> (visited on 07/20/2021).
- Ghani, A. C., C. A. Donnelly, D. R. Cox, J. T. Griffin, C. Fraser, T. H. Lam, L. M. Ho, W. S. Chan, R. M. Anderson, A. J. Hedley, and G. M. Leung. "Methods for Estimating the Case Fatality Ratio for a Novel, Emerging Infectious Disease". In: *American Journal of Epidemiology* 162.5 (Sept. 1, 2005), pp. 479–486.
- Glass, Roger I., Ann-Mari Svennerholm, M. R. Khan, Shamsul Huda, M. Imdadul Huq, and Jan Holmgren. "Seroepidemiological Studies of EI Tor Cholera in Bangladesh: Association of Serum Antibody Levels with Protection". In: *The Journal of Infectious Diseases* 151.2 (Feb. 1, 1985), pp. 236–242.
- Global Task Force on Cholera Control. *Ending Cholera - A Global Roadmap to 2030*. WHO, 2017, p. 32.
- Google LLC. *Google COVID-19 Community Mobility Report*. COVID-19 Community Mobility Report. 2020. URL: <https://www.google.com/covid19/mobility> (visited on 04/21/2020).
- Gostic, Katelyn M., Lauren McGough, Edward B. Baskerville, Sam Abbott, Keya Joshi, Christine Tedijanto, Rebecca Kahn, Rene Niehus, James A. Hay, Pablo M. De Salazar, Joel Hellewell, Sophie Meakin, James D. Munday, Nikos I. Bosse, Katharine Sherratt, Robin N. Thompson, Laura F. White, Jana S. Huisman, Jérémie Scire, Sebastian Bonhoeffer, Tanja Stadler, Jacco Wallinga, Sebastian Funk, Marc Lipsitch, and Sarah Cobey. "Practical Considerations for Measuring the Effective Reproductive Number,  $R_t$ ". In: *PLOS Computational Biology* 16.12 (Dec. 10, 2020), e1008409.
- Greenhalgh, David. "Some Results on Optimal Control Applied to Epidemics". In: *Mathematical Biosciences* 88.2 (Apr. 1, 1988), pp. 125–158.
- Hashizume, Masahiro, Ben Armstrong, Shakoor Hajat, Yukiko Wagatsuma, Abu S. G. Faruque, Taiichi Hayashi, and David A. Sack. "The Effect of Rainfall on the Incidence of Cholera in Bangladesh". In: *Epidemiology* 19.1 (Jan. 2008), pp. 103–110.
- Hashizume, Masahiro, Luis Fernando Chaves, A. S. G. Faruque, Md Yunus, Kim Streftfield, and Kazuhiko Moji. "A Differential Effect of Indian Ocean Dipole and El Niño on Cholera Dynamics in Bangladesh". In: *PLOS ONE* 8.3 (Mar. 29, 2013), e60001.
- He, Xi, Eric H. Y. Lau, Peng Wu, Xilong Deng, Jian Wang, Xinxin Hao, Yiu Chung Lau, Jessica Y. Wong, Yujuan Guan, Xinghua Tan, Xiaoneng Mo, Yanqing Chen, Baolin Liao, Weilie Chen, Fengyu Hu, Qing Zhang, Mingqiu Zhong, Yanrong Wu, Lingzhai Zhao, Fuchun Zhang, Benjamin J. Cowling, Fang Li, and Gabriel M. Leung. "Temporal Dynamics in Viral Shedding and Transmissibility of COVID-19". In: *Nature Medicine* (Apr. 15, 2020), pp. 1–4.
- Heesterbeek, Hans, Roy M. Anderson, Viggo Andreasen, Shweta Bansal, Daniela De Angelis, Chris Dye, Ken T. D. Eames, W. John Edmunds, Simon D. W. Frost, Sebastian Funk, T. Deirdre Hollingsworth, Thomas House, Valerie Isham, Petra Klepac, Justin Lessler, James O. Lloyd-Smith, C. Jessica E. Metcalf, Denis Mollison, Lorenzo Pellis, Juliet R. C. Pulliam, Mick G. Roberts, Cecile Viboud, and Isaac Newton Institute IDD Collaboration. "Modeling Infectious Disease Dynamics in the Complex Landscape of Global Health". In: *Science* 347.6227 (Mar. 13, 2015).
- Hill, Vincent R., Nicole Cohen, Amy M. Kahler, Jessica L. Jones, Cheryl A. Bopp, Nina Marano, Cheryl L. Tarr, Nancy M. Garrett, Jacques Boncy, Ariel Henry, Gerardo A. Gómez, Michael Wellman, Maurice Curtis, Molly M. Freeman, Maryann Turnsek, Ronald A. Benner, Georges Dahourou, David Espey, Angelo DePaola, Jordan W. Tappero, Tom Handzel, and Robert V. Tauxe. "Toxigenic Vibrio Cholerae O1 in Water and Seafood, Haiti". In: *Emerging Infectious Diseases* 17.11 (Nov. 2011), pp. 2147–2150.
- Hinch, Robert, William J. M. Probert, Anel Nurtay, Michelle Kendall, Chris Wymant, Matthew Hall, Katrina Lythgoe, Ana Bulas Cruz, Lele Zhao, Andrea Stewart, Luca Ferretti, Daniel Montero, James Warren, Nicole Mather, Matthew Abueg, Neo Wu, Anthony Finkelstein, David G. Bonsall, Lucie Abeler-Dörner, and Christophe Fraser. "OpenABM-Covid19 - an Agent-Based Model for Non-Pharmaceutical Interventions against COVID-19 Including Contact Tracing". In: *medRxiv* (Sept. 22, 2020), p. 2020.09.16.20195925.
- HIT COVID Team. *Health Interventions Tracking for COVID-19 (HIT-COVID)*. Zenodo, Apr. 24, 2020.

- HSL. A Collection of Fortran Codes for Large Scale Scientific Computation.*
- Huang, Chaolin, Yeming Wang, Xingwang Li, Lili Ren, Jianping Zhao, Yi Hu, Li Zhang, Guohui Fan, Jiuyang Xu, Xiaoying Gu, Zhenshun Cheng, Ting Yu, Jiaan Xia, Yuan Wei, Wenjuan Wu, Xuelei Xie, Wen Yin, Hui Li, Min Liu, Yan Xiao, Hong Gao, Li Guo, Jungang Xie, Guangfa Wang, Rongmeng Jiang, Zhancheng Gao, Qi Jin, Jianwei Wang, and Bin Cao. "Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan, China". In: *The Lancet* 395.10223 (Feb. 15, 2020), pp. 497–506.
- Huffman, George J., David T. Bolvin, Eric J. Nelkin, David B. Wolff, Robert F. Adler, Guojun Gu, Yang Hong, Kenneth P. Bowman, and Erich F. Stocker. "The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales". In: *Journal of Hydrometeorology* 8.1 (Feb. 1, 2007), pp. 38–55.
- Huisman, Jana S., Jérémie Scire, Daniel C. Angst, Jinzhou Li, Richard A. Neher, Marloes H. Maathuis, Sebastian Bonhoeffer, and Tanja Stadler. "Estimation and Worldwide Monitoring of the Effective Reproductive Number of SARS-CoV-2". In: *medRxiv* (June 24, 2021), p. 2020.11.26.20239368.
- IHME, COVID-19 health service utilization forecasting team and Christopher JL Murray. "Forecasting the Impact of the First Wave of the COVID-19 Pandemic on Hospital Demand and Deaths for the USA and European Economic Area Countries". In: *medRxiv* (Apr. 26, 2020), p. 2020.04.21.20074732.
- Ionides, Edward L., Dao Nguyen, Yves Atchadé, Stilian Stoev, and Aaron A. King. "Inference for Dynamic and Latent Variable Models via Iterated, Perturbed Bayes Maps". In: *Proceedings of the National Academy of Sciences* 112.3 (Jan. 20, 2015), pp. 719–724.
- IRI/LDEO. "Climate Data Library" (2016).
- Jutla, Antarpreet, Ali Shafqat Akanda, Anwar Huq, Abu Syed Golam Faruque, Rita Colwell, and Shafiqul Islam. "A Water Marker Monitored by Satellites to Predict Seasonal Endemic Cholera". In: *Remote Sensing Letters* 4.8 (Aug. 1, 2013), pp. 822–831.
- Kadire, Siri R., Robert M. Wachter, and Nicole Lurie. "Delayed Second Dose versus Standard Regimen for Covid-19 Vaccination". In: *New England Journal of Medicine* o.o (Feb. 17, 2021), e28.
- Kaper, J. B., J. G. Morris, and M. M. Levine. "Cholera". In: *Clinical Microbiology Reviews* 8.1 (Jan. 1, 1995), pp. 48–86.
- Kermack, William Ogilvy and A. G. McKendrick. "A Contribution to the Mathematical Theory of Epidemics". In: *Proceedings of the Royal Society A* 115.772 (Aug. 1, 1927), pp. 700–721.
- Kerr, Cliff C., Robyn M. Stuart, Dina Mistry, Romesh G. Abeysuriya, Katherine Rosenfeld, Gregory R. Hart, Rafael C. Núñez, Jamie A. Cohen, Prashanth Selvaraj, Brittany Hagedorn, Lauren George, Michał Jastrzębski, Amanda Izzo, Greer Fowler, Anna Palmer, Dominic Delport, Nick Scott, Sherrie Kelly, Caroline S. Bennette, Bradley Wagner, Stewart Chang, Assaf P. Oron, Edward Wenger, Jasmina Panovska-Griffiths, Michael Famulare, and Daniel J. Klein. "Covasim: An Agent-Based Model of COVID-19 Dynamics and Interventions". In: *medRxiv* (Apr. 1, 2021), p. 2020.05.10.20097469.
- Khamsi, Roxanne. "If a Coronavirus Vaccine Arrives, Can the World Make Enough?" In: *Nature* 580.7805 (7805 Apr. 9, 2020), pp. 578–580.
- Khan, Rakib, Rifat Anwar, Shafqat Akanda, Michael D. McDonald, Anwar Huq, Antarpreet Jutla, and Rita Colwell. "Assessment of Risk of Cholera in Haiti Following Hurricane Matthew". In: *The American Journal of Tropical Medicine and Hygiene* 97.3 (July 24, 2017), pp. 896–903.
- King, Aaron A., Edward L. Ionides, Mercedes Pascual, and Menno J. Bouma. "Inapparent Infections and Cholera Dynamics". In: *Nature* 454.7206 (Aug. 2008), p. 877.
- King, Aaron A., Dao Nguyen, and Edward L. Ionides. *Statistical Inference for Partially Observed Markov Processes via the R Package Pomp*. Sept. 1, 2015. URL: <http://arxiv.org/abs/1509.00503> (visited on 05/26/2019).
- Kirpich, Alexander, Thomas A. Weppelmann, Yang Yang, Afsar Ali, J. Glenn Morris Jr, and Ira M. Longini. "Cholera Transmission in Ouest Department of Haiti: Dynamic Modeling and the Future of the Epidemic". In: *PLOS Neglected Tropical Diseases* 9.10 (Oct. 21, 2015), e0004153.
- Kirpich, Alexander, Thomas A. Weppelmann, Yang Yang, John Glenn Morris Jr, and Ira M. Longini Jr. "Controlling Cholera in the Ouest Department of Haiti Using Oral Vaccines". In: *PLOS Neglected Tropical Diseases* 11.4 (Apr. 14, 2017), e0005482.
- Kissler, Stephen M., Christine Tedijanto, Edward Goldstein, Yonatan H. Grad, and Marc Lipsitch. "Projecting the Transmission Dynamics of SARS-CoV-2 through the Postpandemic Period". In: *Science* (Apr. 14, 2020).

- Koelle, Katia, Mercedes Pascual, and Associate Editor: Benjamin M. Bolker. "Disentangling Extrinsic from Intrinsic Factors in Disease Dynamics: A Nonlinear Time Series Approach with an Application to Cholera". In: *The American Naturalist* 163.6 (2004), pp. 901–913.
- Kucharski, Adam J., Petra Klepac, Andrew J K Conlan, Stephen M Kissler, Maria L Tang, Hannah Fry, Julia R Gog, W John Edmunds, Jon C Emery, Graham Medley, James D Munday, Timothy W Russell, Quentin J Leclerc, Charlie Diamond, Simon R Procter, Amy Gimma, Fiona Yueqian Sun, Hamish P Gibbs, Alicia Rosello, Kevin van Zandvoort, Stéphane Hué, Sophie R Meakin, Arminder K Deol, Gwen Knight, Thibaut Jombart, Anna M Foss, Nikos I Bosse, Katherine E Atkins, Billy J Quilty, Rachel Lowe, Kiesha Prem, Stefan Flasche, Carl A B Pearson, Rein M G J Houben, Emily S Nightingale, Akira Endo, Damien C Tully, Yang Liu, Julian Villabona-Arenas, Kathleen O'Reilly, Sebastian Funk, Rosalind M Eggo, Mark Jit, Eleanor M Rees, Joel Hellewell, Samuel Clifford, Christopher I Jarvis, Sam Abbott, Megan Auzenberg, Nicholas G Davies, and David Simons. "Effectiveness of Isolation, Testing, Contact Tracing, and Physical Distancing on Reducing Transmission of SARS-CoV-2 in Different Settings: A Mathematical Modelling Study". In: *The Lancet Infectious Diseases* 20.10 (Oct. 1, 2020), pp. 1151–1160.
- Kucharski, Adam J., Timothy W. Russell, Charlie Diamond, Yang Liu, John Edmunds, Sebastian Funk, Rosalind M. Eggo, Fiona Sun, Mark Jit, James D. Munday, Nicholas Davies, Amy Gimma, Kevin van Zandvoort, Hamish Gibbs, Joel Hellewell, Christopher I. Jarvis, Sam Clifford, Billy J. Quilty, Nikos I. Bosse, Sam Abbott, Petra Klepac, and Stefan Flasche. "Early Dynamics of Transmission and Control of COVID-19: A Mathematical Modelling Study". In: *The Lancet Infectious Diseases* 0.0 (Mar. 11, 2020).
- Kühn, Juliane, Flavio Finger, Enrico Bertuzzo, Sandrine Borgeaud, Marino Gatto, Andrea Rinaldo, and Melanie Blokesch. "Glucose- but Not Rice-Based Oral Rehydration Therapy Enhances the Production of Virulence Determinants in the Human Pathogen *Vibrio Cholerae*". In: *PLOS Neglected Tropical Diseases* 8.12 (Dec. 4, 2014), e3347.
- Laboratory, Los Alamos National. *COVID-19 Cases and Deaths Forecasts*. URL: <https://covid-19.bsvgateway.org/> (visited on 07/23/2021).
- Lai, Shengjie, Nick W. Ruktanonchai, Liangcai Zhou, Olivia Prosper, Wei Luo, Jessica R. Floyd, Amy Wesolowski, Mauricio Santillana, Chi Zhang, Xiangjun Du, Hongjie Yu, and Andrew J. Tatem. "Effect of Non-Pharmaceutical Interventions to Contain COVID-19 in China". In: *Nature* 585.7825 (May 2020), pp. 410–413.
- Lai, Ying-Si, Patricia Biedermann, Uwem F. Ekpo, Amadou Garba, Els Mathieu, Nicholas Midzi, Pauline Mwinzi, Eliézer K. N'Goran, Giovanna Raso, Rufin K. Assaré, Moussa Sacko, Nadine Schur, Idrissa Talla, Louis-Albert Tchuente, Seydou Touré, Mirko S. Winkler, Jürg Utzinger, and Penelope Vounatsou. "Spatial Distribution of Schistosomiasis and Treatment Needs in Sub-Saharan Africa: A Systematic Review and Geostatistical Analysis". In: *The Lancet. Infectious Diseases* 15.8 (Aug. 2015), pp. 927–940.
- Lam, Siu Kwan, Antoine Pitrou, and Stanley Seibert. "Numba: A LLVM-Based Python JIT Compiler". *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC. LLVM '15*. New York, NY, USA: Association for Computing Machinery, Nov. 15, 2015, pp. 1–6.
- Lauer, Stephen A., Kyra H. Grantz, Qifang Bi, Forrest K. Jones, Qulu Zheng, Hannah R. Meredith, Andrew S. Azman, Nicholas G. Reich, and Justin Lessler. "The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application". In: *Annals of Internal Medicine* (Mar. 10, 2020).
- Lauer, Stephen A., Shaun A. Truelove, and Kyra Grantz. *HopkinsIDD/covidSeverity: Initial Release of covidSeverity R Package*. Zenodo, May 23, 2020.
- Lee, Elizabeth C, Dennis L Chao, Joseph C Lemaitre, Laura Matrajt, Damiano Pasetto, Javier Perez-Saez, Flavio Finger, Andrea Rinaldo, Jonathan D Sugimoto, M Elizabeth Halloran, Ira M Longini, Ralph Ternier, Kenia Vissieres, Andrew S Azman, Justin Lessler, and Louise C Ivers. "Achieving Coordinated National Immunity and Cholera Elimination in Haiti through Vaccination: A Modelling Study". In: *The Lancet Global Health* 8.8 (Aug. 1, 2020), e1081–e1089.
- Lemaitre, Joseph, Damiano Pasetto, Javier Perez-Saez, Carla Sciarra, Joseph Francis Wamala, and Andrea Rinaldo. "Rainfall as a Driver of Epidemic Cholera: Comparative Model Assessments of the Effect of Intra-Seasonal Precipitation Events". In: *Acta Tropica* 190 (Feb. 1, 2019), pp. 235–243.
- Lemaitre, Joseph C., Kyra H. Grantz, Joshua Kaminsky, Hannah R. Meredith, Shaun A. Truelove, Stephen A. Lauer, Lindsay T. Keegan, Sam Shah, Josh Wills, Kathryn Kaminsky, Javier Perez-Saez, Justin Lessler, and Elizabeth C. Lee. "A Scenario Modeling Pipeline for COVID-19 Emergency Planning". In: *Scientific Reports* 11.1 (1 Apr. 6, 2021), p. 7534.

- Lemaitre, Joseph C., Damiano Pasetto, Mario Zanon, Enrico Bertuzzo, Lorenzo Mari, Stefano Miccoli, Renato Casagrandi, Marino Gatto, and Andrea Rinaldo. "Optimizing the Spatio-Temporal Allocation of COVID-19 Vaccines: Italy as a Case Study". In: *medRxiv* (May 13, 2021), p. 2021.05.06.21256732.
- Lemaitre, Joseph C., Javier Perez-Saez, Andrew S. Azman, Andrea Rinaldo, and Jacques Fellay. "Assessing the Impact of Non-Pharmaceutical Interventions on SARS-CoV-2 Transmission in Switzerland". In: *Swiss Medical Weekly* 150.2122 (May 30, 2020).
- Lessler, Justin, Sean M Moore, Francisco J Luquero, Heather S McKay, Rebecca Grais, Myriam Henkens, Martin Mengel, Jessica Dunoyer, Maurice M'bangombe, Elizabeth C Lee, Mamoudou Harouna Djingarey, Bertrand Sudre, Didier Bom-pangue, Robert S M Fraser, Abdinasir Abubakar, William Perea, Dominique Legros, and Andrew S Azman. "Mapping the Burden of Cholera in Sub-Saharan Africa and Implications for Control: An Analysis of Data across Geographical Scales". In: *The Lancet* (Mar. 1, 2018).
- Lessler, Justin, Henrik Salje, M. Kate Grabowski, and Derek A. T. Cummings. "Measuring Spatial Dependence for Infectious Disease Epidemiology". In: *PLoS One* 11.5 (2016), e0155249.
- Leung, Tiffany and Laura Matrajt. "Protection Afforded by Previous Vibrio Cholerae Infection against Subsequent Disease and Infection: A Review". In: *PLOS Neglected Tropical Diseases* 15.5 (May 20, 2021), e0009383.
- Levine, M. M., R. E. Black, M. L. Clements, L. Cisneros, D. R. Nalin, and C. R. Young. "Duration of Infection-Derived Immunity to Cholera". In: *The Journal of Infectious Diseases* 143.6 (June 1, 1981), pp. 818–820.
- Lindeløv, Jonas Kristoffer. *Mcp: An R Package for Regression With Multiple Change Points*. preprint. Open Science Framework, Jan. 5, 2020.
- Lipp, Erin K., Anwar Huq, and Rita R. Colwell. "Effects of Global Climate on Infectious Disease: The Cholera Model". In: *Clinical Microbiology Reviews* 15.4 (Oct. 2002), pp. 757–770.
- Lipsitch, Marc and Natalie E. Dean. "Understanding COVID-19 Vaccine Efficacy". In: *Science* 370.6518 (Nov. 13, 2020), pp. 763–765.
- Lipsitch, Marc, Keya D. Joshi, and Sarah E. Cobey. "Comment on Pan A, Liu L, Wang C, et al. Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China." (Apr. 23, 2020).
- Lipsitch, Marc and Cécile Viboud. "Influenza Seasonality: Lifting the Fog". In: *Proceedings of the National Academy of Sciences* 106.10 (Mar. 10, 2009), pp. 3645–3646.
- Liu, Yang, Centre for Mathematical Modelling of Infectious Diseases nCoV Working Group, Sebastian Funk, and Stefan Flasche. "The Contribution of Pre-Symptomatic Infection to the Transmission Dynamics of COVID-2019" (2020), 5:58.
- Luquero, Francisco J., Lise Grout, Iza Ciglenecki, Keita Sakoba, Bala Traore, Melat Heile, Alpha Amadou Diallo, Christian Itama, Anne-Laure Page, Marie-Laure Quilici, Martin A. Mengel, Jose Maria Eiros, Micaela Serafini, Dominique Legros, and Rebecca F. Grais. "Use of Vibrio Cholerae Vaccine in an Outbreak in Guinea". In: *New England Journal of Medicine* 370.22 (May 29, 2014), pp. 2111–2120.
- Luquero, Francisco J., Lise Grout, Iza Ciglenecki, Keita Sakoba, Bala Traore, Melat Heile, Alpha Amadou Diallo, Christian Itama, Micaela Serafini, Dominique Legros, and Rebecca F. Grais. "First Outbreak Response Using an Oral Cholera Vaccine in Africa: Vaccine Coverage, Acceptability and Surveillance of Adverse Events, Guinea, 2012". In: *PLOS Neglected Tropical Diseases* 7.10 (Oct. 17, 2013), e2465.
- Luquero, Francisco J., Marc Rondy, Jacques Boncy, André Munger, Helmi Mekaoui, Ellen Rymshaw, Anne-Laure Page, Brahma Toure, Marie Amelie Degail, Sarala Nicolas, Francesco Grandesso, Maud Ginsbourger, Jonathan Polonsky, Kathryn P. Alberti, Mego Terzian, David Olson, Klaudia Porten, and Iza Ciglenecki. "Mortality Rates during Cholera Epidemic, Haiti, 2010–2011". In: *Emerging Infectious Diseases* 22.3 (Mar. 2016), pp. 410–416.
- Magny, Guillaume Constantin de, Raghu Murtugudde, Mathew R. P. Sapiano, Azhar Nizam, Christopher W. Brown, Antonio J. Busalacchi, Mohammad Yunus, G. Balakrish Nair, Ana I. Gil, Claudio F. Lanata, John Calkins, Byomkesh Manna, Krishnan Rajendran, Mihir Kumar Bhattacharya, Anwar Huq, R. Bradley Sack, and Rita R. Colwell. "Environmental Signatures Associated with Cholera Epidemics". In: *Proceedings of the National Academy of Sciences of the United States of America* 105.46 (Nov. 18, 2008), pp. 17676–17681.
- Magny, Guillaume Constantin de, Wassila Thiaw, Vadlamani Kumar, Noël M. Manga, Bernard M. Diop, Lamine Gueye, Mamina Kamara, Benjamin Roche, Raghu Murtugudde, and Rita R. Colwell. "Cholera Outbreak in Senegal in 2005: Was Climate a Factor?" In: *PLOS ONE* 7.8 (Aug. 31, 2012), e44577.

- Mari, L., E. Bertuzzo, L. Righetto, R. Casagrandi, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo. "Modelling Cholera Epidemics: The Role of Waterways, Human Mobility and Sanitation". In: *Journal of The Royal Society Interface* 9.67 (Feb. 7, 2012), pp. 376–388.
- Mari, Lorenzo, Enrico Bertuzzo, Flavio Finger, Renato Casagrandi, Marino Gatto, and Andrea Rinaldo. "On the Predictive Ability of Mechanistic Models for the Haitian Cholera Epidemic". In: *Journal of The Royal Society Interface* 12.104 (Mar. 6, 2015), p. 20140840.
- Marziano, Valentina, Giorgio Guzzetta, Bruna Maria Rondinone, Fabio Boccuni, Flavia Riccardo, Antonino Bella, Piero Poletti, Filippo Trentini, Patrizio Pezzotti, Silvio Brusaferro, Giovanni Rezza, Sergio Iavicoli, Marco Ajelli, and Stefano Merler. "Retrospective Analysis of the Italian Exit Strategy from COVID-19 Lockdown". In: *Proceedings of the National Academy of Sciences* 118.4 (Jan. 26, 2021).
- Matrajt, Laura, Julia Eaton, Tiffany Leung, and Elizabeth R. Brown. "Vaccine Optimization for COVID-19: Who to Vaccinate First?" In: *Science Advances* 7.6 (Feb. 1, 2020), eabfr374.
- Mavian, Carla, Taylor K. Paisie, Meer T. Alam, Cameron Browne, Valery Madsen Beau De Rochars, Stefano Nembrini, Melanie N. Cash, Eric J. Nelson, Taj Azarian, Afsar Ali, J. Glenn Morris, and Marco Salemi. "Toxigenic Vibrio Cholerae Evolution and Establishment of Reservoirs in Aquatic Ecosystems". In: *Proceedings of the National Academy of Sciences* 117.14 (Apr. 7, 2020), pp. 7897–7904.
- Mccormack, William M., Md Shafiqul Islam, and Md Fahimuddin. "A Community Study of Inapparent Cholera Infections". In: *American Journal of Epidemiology* 89.6 (June 1, 1969), pp. 658–664.
- McCreesh, Nicky and Mark Booth. "Challenges in Predicting the Effects of Climate Change on Schistosoma Mansoni and Schistosoma Haematobium Transmission Potential." In: *Trends in parasitology*. 29.11 (11 Nov. 1, 2013), pp. 548–555.
- McCreesh, Nicky, Grigory Nikulin, and Mark Booth. "Predicting the Effects of Climate Change on Schistosoma Mansoni Transmission in Eastern Africa". In: *Parasites & Vectors* 8.1 (Jan. 6, 2015), p. 4.
- Medlock, Jan and Alison P. Galvani. "Optimizing Influenza Vaccine Distribution". In: *Science* 325.5948 (Sept. 25, 2009), pp. 1705–1708.
- Meibom, Karin L., Melanie Blokesch, Nadia A. Dolganov, Cheng-Yen Wu, and Gary K. Schoolnik. "Chitin Induces Natural Competence in Vibrio Cholerae". In: *Science* 310.5755 (Dec. 16, 2005), pp. 1824–1827.
- Miller Neilan, Rachael and Suzanne Lenhart. "Optimal Vaccine Distribution in a Spatiotemporal Epidemic Model with an Application to Rabies and Raccoons". In: *Journal of Mathematical Analysis and Applications* 378.2 (June 15, 2011), pp. 603–619.
- Mitchell, Cristina and https://www.facebook.com/pahowho. *PAHO/WHO | Haiti Reaches One-Year Free of Cholera*. Pan American Health Organization / World Health Organization. Jan. 23, 2020. URL: [https://www3.paho.org/hq/index.php?option=com\\_content&view=article&id=15684:haiti-reaches-one-year-free-of-cholera&Itemid=1926&lang=en](https://www3.paho.org/hq/index.php?option=com_content&view=article&id=15684:haiti-reaches-one-year-free-of-cholera&Itemid=1926&lang=en) (visited on 07/27/2021).
- Moghadas, Seyed M., Affan Shoukat, Meagan C. Fitzpatrick, Chad R. Wells, Pratha Sah, Abhishek Pandey, Jeffrey D. Sachs, Zheng Wang, Lauren A. Meyers, Burton H. Singer, and Alison P. Galvani. "Projecting Hospital Utilization during the COVID-19 Outbreaks in the United States". In: *Proceedings of the National Academy of Sciences* 117.16 (Apr. 21, 2020), pp. 9122–9126.
- Moore, Sam, Edward M. Hill, Mike J. Tildesley, Louise Dyson, and Matt J. Keeling. "Vaccination and Non-Pharmaceutical Interventions: When Can the UK Relax about COVID-19?" In: *medRxiv* (Jan. 26, 2021), p. 2020.12.27.20248896.
- Morton, R. and K. H. Wickwire. "On the Optimal Control of a Deterministic Epidemic". In: *Advances in Applied Probability* 6.4 (Dec. 1974), pp. 622–635.
- Mutreja, Ankur, Dong Wook Kim, Nicholas R. Thomson, Thomas R. Connor, Je Hee Lee, Samuel Kariuki, Nicholas J. Croucher, Seon Young Choi, Simon R. Harris, Michael Lebents, Swapan Kumar Niyogi, Eun Jin Kim, T. Ramamurthy, Jongsik Chun, James L. N. Wood, John D. Clemens, Cecil Czerkinsky, G. Balakrish Nair, Jan Holmgren, Julian Parkhill, and Gordon Dougan. "Evidence for Several Waves of Global Transmission in the Seventh Cholera Pandemic". In: *Nature* 477.7365 (Sept. 2011), pp. 462–465.
- Nair, G. Balakrish, Firdausi Qadri, Jan Holmgren, Ann-Mari Svenserholm, Ashrafus Safa, Nurul A. Bhuiyan, Q. Shafi Ahmad, Shah M. Faruque, A. S. G. Faruque, Yoshifumi Takeda, and David A. Sack. "Cholera Due to Altered El Tor Strains of Vibrio Cholerae O1 in Bangladesh". In: *Journal of Clinical Microbiology* 44.11 (Nov. 1, 2006), pp. 4211–4213.

- National Academies of Sciences Engineering, and Medicine. *Framework for Equitable Allocation of COVID-19 Vaccine*. Ed. by Benjamin Kahn, Lisa Brown, William Foege, and Helene Gayle. The National Academies Collection: Reports Funded by National Institutes of Health. Washington (DC): National Academies Press (US), 2020.
- Neher, Richard A., Robert Dyrda, Valentin Druelle, Emma B. Hodcroft, and Jan Albert. "Potential Impact of Seasonal Forcing on a SARS-CoV-2 Pandemic". In: *Swiss Medical Weekly* 150.1112 (Mar. 16, 2020).
- Nelson, Eric J., Jason B. Harris, J. Glenn Morris, Stephen B. Calderwood, and Andrew Camilli. "Cholera Transmission: The Host, Pathogen and Bacteriophage Dynamic". In: *Nature reviews. Microbiology* 7.10 (Oct. 2009).
- O'Dea, S. *Smartphone Users in Switzerland 2018-2024*. Statista. Apr. 20, 2020. URL: <https://www.statista.com/statistics/494640/smartphone-users-in-switzerland/> (visited on 05/02/2020).
- OAG. *Flight Data in OAG*. 2020.
- Obadia, Thomas, Romana Haneef, and Pierre-Yves Boëlle. "The Ro Package: A Toolbox to Estimate Reproduction Numbers for Epidemic Outbreaks". In: *BMC Medical Informatics and Decision Making* 12.1 (Dec. 18, 2012), p. 147.
- OFSP. *Nouvelles Règles d'hygiène et de Conduite Pour Se Protéger Contre Le Nouveau Coronavirus*. 2020. URL: <https://www.admin.ch/gov/fr/accueil/documentation/communiques.msg-id-78304.html> (visited on 04/28/2020).
- OFSP, Office fédéral de la santé publique. *Rapport sur la situation épidémiologique en Suisse et dans la Principauté de Liechtenstein*. 2020. URL: <https://www.bag.admin.ch/bag/fr/home/krankheiten/ausbrueche-epidemien-aktuelle-ausbrueche-epidemien/novel-cov/situation-schweiz-und-international.html> (visited on 04/26/2020).
- openZH. *openZH/Covid\_19*. Specialist Unit for Open Government Data Canton of Zurich, Apr. 25, 2020.
- Pacini, Filippo. *Osservazioni microscopiche e deduzioni patologiche sul Cholera asiatico*. Firenze, 1854.
- Park, Joonha and Edward L. Ionides. *A Guided Intermediate Resampling Particle Filter for Inference on High Dimensional Systems*. Aug. 28, 2017. URL: <http://arxiv.org/abs/1708.08543> (visited on 07/19/2019).
- "Inference on High-Dimensional Implicit Dynamic Models Using a Guided Intermediate Resampling Filter". In: *Statistics and Computing* 30.5 (Sept. 1, 2020), pp. 1497–1522.
- Park, Shin Young, Young-Man Kim, Seonju Yi, Sangeun Lee, Baeg-Ju Na, Chang Bo Kim, Jung-il Kim, Hea Sook Kim, Young Bok Kim, Yoojin Park, In Sil Huh, Hye Kyung Kim, Hyung Jun Yoon, Hanaram Jang, Kyungnam Kim, Yeonhwa Chang, Inhye Kim, Hyeyoung Lee, Jin Gwack, Seong Sun Kim, Miyoung Kim, Sanghui Kweon, Young June Choe, Ok Park, Young Joon Park, and Eun Kyong Jeong. "Early Release - Coronavirus Disease Outbreak in Call Center, South Korea - Volume 26, Number 8—August 2020 - Emerging Infectious Diseases Journal - CDC" (2020).
- Parker, Lucy A, John Rumunu, Christine Jamet, Yona Kenyi, Richard Laku Lino, Joseph F Wamala, Allan M Mpairwe, Iza Ciglenecki, Francisco J Luquero, Andrew S Azman, and Jean-Clement Cabrol. "Adapting to the Global Shortage of Cholera Vaccines: Targeted Single Dose Cholera Vaccine in Response to an Outbreak in South Sudan". In: *The Lancet Infectious Diseases* 17.4 (Apr. 1, 2017), e123–e127.
- "Adapting to the Global Shortage of Cholera Vaccines: Targeted Single Dose Cholera Vaccine in Response to an Outbreak in South Sudan". In: *The Lancet. Infectious Diseases* 17.4 (Apr. 2017), e123–e127.
- Pascual, Mercedes, Xavier Rodó, Stephen P. Ellner, Rita Colwell, and Menno J. Bouma. "Cholera Dynamics and El Niño-Southern Oscillation". In: *Science* 289.5485 (Sept. 8, 2000), pp. 1766–1769.
- Pasetto, D., J.C. Lemaitre, E. Bertuzzo, M. Gatto, and A. Rinaldo. "Range of Reproduction Number Estimates for COVID-19 Spread". In: *Biochemical and Biophysical Research Communications* (2020).
- Pasetto, Damiano, Flavio Finger, Anton Camacho, Francesco Grandesso, Sandra Cohuet, Joseph C. Lemaitre, Andrew S. Azman, Francisco J. Luquero, Enrico Bertuzzo, and Andrea Rinaldo. "Near Real-Time Forecasting for Cholera Decision Making in Haiti after Hurricane Matthew". In: *PLOS Computational Biology* 14.5 (May 16, 2018), e1006127.
- Pasetto, Damiano, Flavio Finger, Andrea Rinaldo, and Enrico Bertuzzo. "Real-Time Projections of Cholera Outbreaks through Data Assimilation and Rainfall Forecasting". In: *Advances in Water Resources* 108 (Oct. 1, 2017), pp. 345–356.
- Patel, Rajan, Ira M. Longini, and M. Elizabeth Halloran. "Finding Optimal Vaccination Strategies for Pandemic Influenza Using Genetic Algorithms". In: *Journal of Theoretical Biology* 234.2 (May 21, 2005), pp. 201–212.
- Perez-Saez, Javier, Aaron A. King, Andrea Rinaldo, Mohammad Yunus, Abu S. G. Faruque, and Mercedes Pascual. "Climate-Driven Endemic Cholera Is Modulated by Human Mobility in a Megacity". In: *Advances in Water Resources* 108 (Oct. 2017), pp. 367–376.

- Perez-Saez, Javier, Theophile Mande, Natalie Ceperley, Enrico Bertuzzo, Lorenzo Mari, Marino Gatto, and Andrea Rinaldo. "Hydrology and Density Feedbacks Control the Ecology of Intermediate Hosts of Schistosomiasis across Habitats in Seasonal Climates". In: *Proceedings of the National Academy of Sciences* 113.23 (June 7, 2016), pp. 6427–6432.
- Periago, Mirta Roses, Thomas R. Frieden, Jordan W. Tappero, Kevin M. De Cock, Bernt Aasen, and Jon K. Andrus. "Elimination of Cholera Transmission in Haiti and the Dominican Republic". In: *The Lancet* 379.9812 (Jan. 21, 2012), e12–e13.
- Pezzoli, Lorenzo. "Global Oral Cholera Vaccine Use, 2013–2018". In: *Vaccine* (Sept. 10, 2019).
- Piarroux, Renaud, Robert Barrais, Benoît Faucher, Rachel Haus, Martine Piarroux, Jean Gaudart, Roc Magloire, and Didier Raoult. "Understanding the Cholera Epidemic, Haiti". In: *Emerging Infectious Diseases* 17.7 (July 2011), pp. 1161–1168.
- Probst, Daniel. *Daenuprobst/Covid19-Cases-Switzerland*. May 2, 2020.
- Qadri, Firdausi, Mohammad Ali, Julia Lynch, Fahima Chowdhury, Ashraful Islam Khan, Thomas F. Wierzba, Jean-Louis Excler, Amit Saha, Md Taufiqul Islam, Yasmin A. Begum, Taufiqur R. Bhuiyan, Farhana Khanam, Mohiul I. Chowdhury, Iqbal Ansary Khan, Alamgir Kabir, Baizid Khoorshid Riaz, Afroza Akter, Arifuzzaman Khan, Muhammad Asaduzzaman, Deok Ryun Kim, Ashraf U. Siddik, Nirod C. Saha, Alejandro Cravioto, Ajit P. Singh, and John D. Clemens. "Efficacy of a Single-Dose Regimen of Inactivated Whole-Cell Oral Cholera Vaccine: Results from 2 Years of Follow-up of a Randomised Trial". In: *The Lancet Infectious Diseases* 18.0 (Mar. 14, 2018).
- Quirynen, Rien, Milan Vukov, and Moritz Diehl. "Multiple Shooting in a Microsecond". *Multiple Shooting and Time Domain Decomposition Methods*. Ed. by Thomas Carraro, Michael Geiger, Stefan Körkel, and Rolf Rannacher. Cham: Springer International Publishing, 2015, pp. 183–201.
- Raftery, Adrian E., Michael A. Newton, Jaya M. Satagopan, and Pavel N. Krivitsky. "Estimating the Integrated Likelihood via Posterior Simulation Using the Harmonic Mean Identity". *Bayesian Statistics*. 2007, pp. 1–45.
- Ramírez, Iván J. and Sue C. Grady. "El Niño, Climate, and Cholera Associations in Piura, Peru, 1991–2001: A Wavelet Analysis". In: *EcoHealth* 13.1 (Mar. 1, 2016), pp. 83–99.
- Rawlings, James B., David Q. Mayne, and Moritz Diehl. *Model Predictive Control: Theory, Computation, and Design*. 2nd ed. Nob Hill Publishing, 2017.
- Rebaudet, Stanislas, Gregory Bulit, Jean Gaudart, Edwige Michel, Pierre Gazin, Claudia Evers, Samuel Beaulieu, Aaron Aruna Abedi, Lindsay Osei, Robert Barrais, Katilla Pierre, Sandra Moore, Jacques Boncy, Paul Adrien, Edouard Beigbeder, Florence Duperval Guillaume, and Renaud Piarroux. "The National Alert-Response Strategy against Cholera in Haiti: A Four-Year Assessment of Its Implementation". In: *bioRxiv* (Feb. 5, 2018), p. 259366.
- Rebaudet, Stanislas, Gregory Bulit, Jean Gaudart, Edwige Michel, Pierre Gazin, Claudia Evers, Samuel Beaulieu, Aaron Aruna Abedi, Lindsay Osei, Robert Barrais, Katilla Pierre, Sandra Moore, Jacques Boncy, Paul Adrien, Florence Duperval Guillaume, Edouard Beigbeder, and Renaud Piarroux. "The Case-Area Targeted Rapid Response Strategy to Control Cholera in Haiti: A Four-Year Implementation Study". In: *PLOS Neglected Tropical Diseases* 13.4 (Apr. 16, 2019), e0007263.
- Rebaudet, Stanislas, Bertrand Sudre, Benoît Faucher, and Renaud Piarroux. "Cholera in Coastal Africa: A Systematic Review of Its Heterogeneous Environmental Determinants". In: *The Journal of Infectious Diseases* 208 Suppl 1 (Nov. 1, 2013), S98–106.
- "Environmental Determinants of Cholera Outbreaks in Inland Africa: A Systematic Review of Main Transmission Foci and Propagation Routes". In: *The Journal of Infectious Diseases* 208 Suppl 1 (Nov. 1, 2013), S46–54.
- Reidl, Joachim and Karl E. Klose. "Vibrio Cholerae and Cholera: Out of the Water and into the Host". In: *FEMS Microbiology Reviews* 26.2 (June 1, 2002), pp. 125–139.
- Reiner, Robert C., Aaron A. King, Michael Emch, Mohammad Yunus, A. S. G. Faruque, and Mercedes Pascual. "Highly Localized Sensitivity to Climate Forcing Drives Endemic Cholera in a Megacity". In: *Proceedings of the National Academy of Sciences* 109.6 (Feb. 7, 2012), pp. 2033–2036.
- Righetto, L., E. Bertuzzo, L. Mari, E. Schild, R. Casagrandi, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo. "Rainfall Mediations in the Spreading of Epidemic Cholera". In: *Advances in Water Resources* 60 (Oct. 1, 2013), pp. 34–46.
- Rinaldo, Andrea, Enrico Bertuzzo, Melanie Blokesch, Lorenzo Mari, and Marino Gatto. "Modeling Key Drivers of Cholera Transmission Dynamics Provides New Perspectives for Parasitology". In: *Trends in Parasitology* 33.8 (Aug. 1, 2017), pp. 587–599.
- Rinaldo, Andrea, Enrico Bertuzzo, Lorenzo Mari, Lorenzo Righetto, Melanie Blokesch, Marino Gatto, Renato Casagrandi, Megan Murray, Silvan M. Vesenbeckh, and Ignacio Rodriguez-Iturbe. "Reassessment of the 2010–2011 Haiti Cholera Out-

- break and Rainfall-Driven Multiseason Projections". In: *Proceedings of the National Academy of Sciences* 109.17 (Apr. 24, 2012), pp. 6602–6607.
- Rinaldo, Andrea, Marino Gatto, and Ignacio Rodriguez-Iturbe. *River Networks as Ecological Corridors: Species, Populations, Pathogens*. Cambridge University Press, Oct. 22, 2020. 457 pp.
- Riou, Julien and Christian L. Althaus. "Pattern of Early Human-to-Human Transmission of Wuhan 2019 Novel Coronavirus (2019-nCoV), December 2019 to January 2020". In: *Eurosurveillance* 25.4 (Jan. 30, 2020), p. 2000058.
- Rodó, Xavier, Mercedes Pascual, Francisco J. Doblas-Reyes, Alexander Gershunov, Dáithí A. Stone, Filippo Giorgi, Peter J. Hudson, James Kinter, Miquel-Àngel Rodríguez-Arias, Nils Ch. Stenseth, David Alonso, Javier García-Serrano, and Andrew P. Dobson. "Climate Change and Infectious Diseases: Can We Meet the Needs for Better Prediction?" In: *Climatic Change* 118.3 (June 1, 2013), pp. 625–640.
- Roth, Gregory A. et al. "Global, Regional, and National Age-Sex-Specific Mortality for 282 Causes of Death in 195 Countries and Territories, 1980–2017: A Systematic Analysis for the Global Burden of Disease Study 2017". In: *The Lancet* 392.10159 (Nov. 10, 2018), pp. 1736–1788.
- Ruiz-Moreno, Diego, Mercedes Pascual, Menno Bouma, Andrew Dobson, and Benjamin Cash. "Cholera Seasonality in Madras (1901–1940): Dual Role for Rainfall in Endemic and Epidemic Regions". In: *EcoHealth* 4.1 (Mar. 1, 2007), pp. 52–62.
- Russell, Timothy W., Joel Hellewell, Christopher I. Jarvis, Kevin van Zandvoort, Sam Abbott, Ruwan Ratnayake, CMMID COVID-19 working Group, Stefan Flasche, Rosalind M. Eggo, W. John Edmunds, and Adam J. Kucharski. "Estimating the Infection and Case Fatality Ratio for Coronavirus Disease (COVID-19) Using Age-Adjusted Data from the Outbreak on the Diamond Princess Cruise Ship, February 2020". In: *Eurosurveillance* 25.12 (Mar. 26, 2020), p. 2000256.
- Saad-Roy, Chadi M., Sinead E. Morris, C. Jessica E. Metcalf, Michael J. Mina, Rachel E. Baker, Jeremy Farrar, Edward C. Holmes, Oliver G. Pybus, Andrea L. Graham, Simon A. Levin, Bryan T. Grenfell, and Caroline E. Wagner. "Epidemiological and Evolutionary Considerations of SARS-CoV-2 Vaccine Dosing Regimes". In: *medRxiv* (Feb. 3, 2021), p. 2021.02.01.21250944.
- Sack, David A., R. Bradley Sack, G Balakrish Nair, and AK Siddique. "Cholera". In: *The Lancet* 363.9404 (Jan. 17, 2004), pp. 223–233.
- Sack, David A., R. Bradley Sack, and Claire-Lise Chaignat. "Getting Serious about Cholera". In: *New England Journal of Medicine* 355.7 (Aug. 17, 2006), pp. 649–651.
- Sah, Pratha, Jan Medlock, Meagan C. Fitzpatrick, Burton H. Singer, and Alison P. Galvani. "Optimizing the Impact of Low-Efficacy Influenza Vaccines". In: *Proceedings of the National Academy of Sciences* 115.20 (May 15, 2018), pp. 5151–5156.
- Salje, Henrik, Cécile Tran Kiem, Noémie Lefrancq, Noémie Courtejoie, Paolo Bosetti, Juliette Paireau, Alessio Andronico, Nathanaël Hozé, Jehanne Richet, Claire-Lise Dubost, Yann Le Strat, Justin Lessler, Daniel Levy-Bruhl, Arnaud Fontanet, Lulla Opatowski, Pierre-Yves Boelle, and Simon Cauchemez. "Estimating the Burden of SARS-CoV-2 in France". In: *Science* (May 13, 2020).
- Savorgnan, C., C. Romani, A. Kozma, and M. Diehl. "Multiple Shooting for Distributed Systems with Applications in Hydro Electricity Production". In: *Journal of Process Control* 21 (2011), pp. 738–745.
- Sciarrà, Carla, Andrea Rinaldo, Francesco Laio, and Damiano Pasetto. *Mathematical Modeling of Cholera Epidemics in South Sudan*. Jan. 8, 2018. URL: <http://arxiv.org/abs/1801.03125> (visited on 05/10/2021).
- Scire, Jeremie, Sarah Nadeau, Timothy Vaughan, Gavin Brupbacher, Simon Fuchs, Jürg Sommer, Katrin N. Koch, Reto Misteli, Lukas Mundorff, Thomas Götz, Tobias Eichenberger, Carlos Quinto, Miodrag Savic, Andrea Meienberg, Thilo Burkard, Michael Mayr, Christoph A. Meier, Andreas Widmer, Richard Kuehl, Adrian Egli, Hans H. Hirsch, Stefano Bassetti, Christian H. Nickel, Katharina S. Rentsch, Werner Kübler, Roland Bingisser, Manuel Battegay, Sarah Tschudin-Sutter, and Tanja Stadler. "Reproductive Number of the COVID-19 Epidemic in Switzerland with a Focus on the Cantons of Basel-Stadt and Basel-Landschaft". In: *Swiss Medical Weekly* 150.1920 (May 4, 2020).
- Seidlein, Lorenz von and Jacqueline L. Deen. "Preventing Cholera Outbreaks through Early Targeted Interventions". In: *PLOS Medicine* 15.2 (Feb. 27, 2018), e1002510.
- Sethi, Suresh P. and Preston W. Staats. "Optimal Control of Some Simple Deterministic Epidemic Models". In: *Journal of the Operational Research Society* 29.2 (Feb. 1, 1978), pp. 129–136.

- Shea, Katriona, Michael C. Runge, David Pannell, William J. M. Probert, Shou-Li Li, Michael Tildesley, and Matthew Ferrari. "Harnessing Multiple Models for Outbreak Management". In: *Science* 368.6491 (May 8, 2020), pp. 577–579.
- Silva, Petrônio C. L., Paulo V. C. Batista, Hélder S. Lima, Marcos A. Alves, Frederico G. Guimarães, and Rodrigo C. P. Silva. "COVID-ABS: An Agent-Based Model of COVID-19 Epidemic to Simulate Health and Economic Effects of Social Distancing Interventions". In: *Chaos, Solitons & Fractals* 139 (Oct. 1, 2020), p. 110088.
- Skellam, J. G. "The Frequency Distribution of the Difference between Two Poisson Variates Belonging to Different Populations". In: *Journal of the Royal Statistical Society. Series A (General)* 109 (Pt 3 1946), p. 296.
- Snow, John. *On the Mode of Communication of Cholera*. John Churchill, 1855. 216 pp.
- Spassiani, Ilaria, Lorenzo Gubian, Giorgio Palù, and Giovanni Sebastiani. "Vaccination Criteria Based on Factors Influencing COVID-19 Diffusion and Mortality". In: *Vaccines* 8.4 (4 Dec. 2020), p. 766.
- SSNBS. *Population Projections for South Sudan by County 2015 - 2020*. South Sudan National Bureau of Statistics, 2015.
- Stringhini, Silvia, Ania Wisniak, Giovanni Piumatti, Andrew S. Azman, Stephen A. Lauer, Helene Baysson, David De Ridder, Dusan Petrovic, Stephanie Schrempf, Kailing Marcus, Isabelle Arm-Vernez, Sabine Yerly, Olivia Keiser, Samia Hurst, Klara Posfay-Barbe, Didier Trono, Didier Pittet, Laurent Getaz, Francois Chappuis, Isabella Eckerle, Nicolas Vuilleumier, Benjamin Meyer, Antoine Flahault, Laurent Kaiser, and Idris Guessous. "Repeated Seroprevalence of Anti-SARS-CoV-2 IgG Antibodies in a Population-Based Sample from Geneva, Switzerland". In: *medRxiv* (May 6, 2020), p. 2020.05.02.20088898.
- Sugimoto, Jonathan D., Amanda A. Koepke, Eben E. Kenah, M. Elizabeth Halloran, Fahima Chowdhury, Ashraful I. Khan, Regina C. LaRocque, Yang Yang, Edward T. Ryan, Firdausi Qadri, Stephen B. Calderwood, Jason B. Harris, and Ira M. Longini Jr. "Household Transmission of Vibrio Cholerae in Bangladesh". In: *PLOS Neglected Tropical Diseases* 8.11 (Nov. 20, 2014), e3314.
- Thul, Lawrence and Warren Powell. *Stochastic Optimization for Vaccine and Testing Kit Allocation for the COVID-19 Pandemic*. Jan. 4, 2021. URL: <http://arxiv.org/abs/2101.01204> (visited on 02/04/2021).
- Tohme, Rania A., Jeannot François, Kathleen Wannemuehler, Preetha Iyengar, Amber Dismer, Paul Adrien, Terri B. Hyde, Barbara J. Marston, Kashmira Date, Eric Mintz, and Mark A. Katz. "Oral Cholera Vaccine Coverage, Barriers to Vaccination, and Adverse Events Following Vaccination, Haiti, 2013". In: *Emerging Infectious Diseases* 21.6 (June 2015), pp. 984–991.
- Truelove, Shaun, Stephen A. Lauer, Joseph Lemaitre, Kathryn Kaminsky, and jkamins7. *HopkinsIDD/covidImportation: Initial Release of covidImportation R Package*. Zenodo, May 22, 2020.
- Truelove, Shaun, Luis Mier-y-Teran-Romero, Paul Gastanaduy, Allison Taylor Walker, Andre Berro, Justin Lessler, and Michael A. Johansson. "Epidemics, Air Travel, and Elimination in a Globalized World: The Case of Measles". In: *medRxiv* (May 12, 2020), p. 2020.05.08.20095414.
- Tuite, A.R., L. Zhu, D.N. Fisman, and J.A. Salomon. "Alternative Dose Allocation Strategies to Increase Benefits from Constrained COVID-19 Vaccine Supply". In: *Annals of Internal Medicine* (2021), pp. 1–18.
- Tuite, Ashleigh R., Joseph Tien, Marisa Eisenberg, David J. D. Earn, Junling Ma, and David N. Fisman. "Cholera Epidemic in Haiti, 2010: Using a Transmission Model to Explain Spatial Spread of Disease and Identify Optimal Control Interventions". In: *Annals of Internal Medicine* 154.9 (May 3, 2011), pp. 593–601.
- USAfacts. *US COVID-19 Cases and Deaths by State*. 2021. URL: [usafacts.org](https://usafacts.org).
- Van de Linde, P. A. M. and G. I. Forbes. "Observations on the Spread of Cholera in Hong Kong, 1961-63". In: *Bulletin of the World Health Organization* 32.4 (1965), pp. 515–530.
- Vehtari, Aki, Andrew Gelman, and Jonah Gabry. "Practical Bayesian Model Evaluation Using Leave-One-out Cross-Validation and WAIC". In: *Statistics and Computing* 27.5 (5 Sept. 1, 2017), pp. 1413–1432.
- Verity, Robert, Lucy C Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, Gina Cuomo-Dannenburg, Hayley Thompson, Patrick G T Walker, Han Fu, Amy Dighe, Jamie T Griffin, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Anne Cori, Zulma Cucunubá, Rich FitzJohn, Katy Gaythorpe, Will Green, Arran Hamlet, Wes Hinsley, Daniel Laydon, Gemma Nedjati-Gilani, Steven Riley, Sabine van Elsland, Erik Volz, Haowei Wang, Yuanrong Wang, Xiaoyue Xi, Christl A Donnelly, Azra C Ghani, and Neil M Ferguson. "Estimates of the Severity of Coronavirus Disease 2019: A Model-Based Analysis". In: *The Lancet Infectious Diseases* (Mar. 30, 2020).
- Vezzulli, Luigi, Rita R. Colwell, and Carla Pruzzo. "Ocean Warming and Spread of Pathogenic Vibrios in the Aquatic Environment". In: *Microbial Ecology* 65.4 (May 2013), pp. 817–825.

- Vezzulli, Luigi, Chiara Grande, Philip C. Reid, Pierre Hélaouët, Martin Edwards, Manfred G. Höfle, Ingrid Brettar, Rita R. Colwell, and Carla Pruzzo. "Climate Influence on Vibrio and Associated Human Diseases during the Past Half-Century in the Coastal North Atlantic". In: *Proceedings of the National Academy of Sciences* 113.34 (Aug. 23, 2016), E5062–E5071.
- Vezzulli, Luigi, Elisabetta Pezzati, Ingrid Brettar, Manfred Höfle, and Carla Pruzzo. "Effects of Global Warming on Vibrio Ecology". In: *Microbiology Spectrum* 3.3 (June 2015).
- Vrugt, Jasper A. "Markov Chain Monte Carlo Simulation Using the DREAM Software Package: Theory, Concepts, and MATLAB Implementation". In: *Environmental Modelling & Software* 75 (Supplement C Jan. 1, 2016), pp. 273–316.
- Wächter, Andreas and Lorenz T. Biegler. "On the Implementation of an Interior-Point Filter Line-Search Algorithm for Large-Scale Nonlinear Programming". In: *Mathematical Programming* 106.1 (Mar. 1, 2006), pp. 25–57.
- Wallinga, Jacco and Peter Teunis. "Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures". In: *American Journal of Epidemiology* 160.6 (Sept. 15, 2004), pp. 509–516.
- Weissman, Gary E., Andrew Crane-Droesch, Corey Chivers, ThaiBinh Luong, Asaf Hanish, Michael Z. Levy, Jason Lubken, Michael Becker, Michael E. Draugelis, George L. Anesi, Patrick J. Brennan, Jason D. Christie, C. William Hanson, Mark E. Mikkelsen, and Scott D. Halpern. "Locally Informed Simulation to Predict Hospital Capacity Needs During the COVID-19 Pandemic". In: *Annals of Internal Medicine* 173.1 (July 7, 2020), pp. 21–28.
- WHO. *Background Paper on the Integration of Oral Cholera Vaccines into Global Cholera Control Programmes*. 2009.
- *Background Paper on Whole-Cell, Killed, Oral Cholera Vaccines*. 2017.
  - "Cholera Vaccines: WHO Position Paper" (2017).
  - "Cholera, 2019". In: *Weekly Epidemiological Record* 95.37 (Sept. 11, 2020), pp. 441–448.
  - *Ending Preventable Child Deaths from Pneumonia and Diarrhoea by 2035 - The Integrated Global Action Plan for Pneumonia and Diarrhoea (GAPPD)*. Executive Summary. Apr. 10, 2013.
  - "First Steps for Managing an Outbreak of Acute Diarrhoea". 2010.
  - *WHO Director-General's Opening Remarks at the Media Briefing on COVID-19*. Mar. 11, 2020. URL: <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020> (visited on 07/23/2021).
  - *WHO Situation Report 90*. 2020. URL: [https://www.who.int/docs/default-source/coronavirus/situation-reports/20200419-sitrep-90-covid-19.pdf?sfvrsn=551d47fd\\_4](https://www.who.int/docs/default-source/coronavirus/situation-reports/20200419-sitrep-90-covid-19.pdf?sfvrsn=551d47fd_4) (visited on 04/26/2020).
- Wilder, Bryan, Marie Charpignon, Jackson A. Killian, Han-Ching Ou, Aditya Mate, Shahin Jabbari, Andrew Perrault, Angel N. Desai, Milind Tambe, and Maimuna S. Majumder. "Modeling Between-Population Variation in COVID-19 Dynamics in Hubei, Lombardy, and New York City". In: *Proceedings of the National Academy of Sciences* 117.41 (Oct. 13, 2020), pp. 25904–25910.
- Wölfel, Roman, Victor M. Corman, Wolfgang Guggemos, Michael Seilmaier, Sabine Zange, Marcel A. Müller, Daniela Niemeyer, Terry C. Jones, Patrick Vollmar, Camilla Rothe, Michael Hoelscher, Tobias Bleicker, Sebastian Brünink, Julia Schneider, Rosina Ehmann, Katrin Zwirglmaier, Christian Drosten, and Clemens Wendtner. "Virological Assessment of Hospitalized Patients with COVID-2019". In: *Nature* (Apr. 1, 2020), pp. 1–5.
- Woodward, William E. "Cholera Reinfection in Man". In: *The Journal of Infectious Diseases* 123.1 (1971), pp. 61–66.
- Yan, Ping. "Separate Roles of the Latent and Infectious Periods in Shaping the Relation between the Basic Reproduction Number and the Intrinsic Growth Rate of Infectious Disease Outbreaks". In: *Journal of Theoretical Biology* 251.2 (Mar. 21, 2008), pp. 238–252.
- Yan, Ping and Gerardo Chowell. *Quantitative Methods for Investigating Infectious Disease Outbreaks*. Texts in Applied Mathematics. Springer International Publishing, 2019.
- Yang, Juan, Wen Zheng, Huilin Shi, Xuemei Yan, Kaige Dong, Qian You, Guangjie Zhong, Hui Gong, Zhiyuan Chen, Mark Jit, Cecile Viboud, Marco Ajelli, and Hongjie Yu. "Who Should Be Prioritized for COVID-19 Vaccination in China? A Descriptive Study". In: *BMC Medicine* 19.1 (Feb. 10, 2021), p. 45.
- Zakary, Omar, Mostafa Rachik, and Ilias Elmouki. "On the Analysis of a Multi-Regions Discrete SIR Epidemic Model: An Optimal Control Approach". In: *International Journal of Dynamics and Control* 5.3 (Sept. 1, 2017), pp. 917–930.
- Zhu, Na, Dingyu Zhang, Wenling Wang, Xingwang Li, Bo Yang, Jingdong Song, Xiang Zhao, Baoying Huang, Weifeng Shi, Roujian Lu, Peihua Niu, Faxian Zhan, Xuejun Ma, Dayan Wang, Wenbo Xu, Guizhen Wu, George F. Gao, and Wenjie Tan. "A Novel Coronavirus from Patients with Pneumonia in China, 2019". In: *New England Journal of Medicine* (Jan. 24, 2020).





# Joseph Lemaitre | PhD Student

Route du Jorat 68 – 1000 Lausanne – Switzerland  
+41 78 608 23 11 • jo.lemaitresamra@gmail.com • jcblemai  
in j-lemaitre • 0000-0002-2677-6574 • Joseph C. Lemaitre

My thesis involves infectious disease modeling towards the design of effective control policies, with practical applications on COVID-19 and Cholera. We use models to answer questions related to diseases dynamics, and communicate results to decision makers. These models combine novel data sources, scientific knowledge and state-of-the-art statistical methods to guide effective policies from noisy, incomplete and often misleading data.

## Education

- 2017 – 2021 **PhD, Laboratory of Ecohydrology (ECHO)**, Swiss Federal Institute of Technology (EPFL), Switzerland.  
Advisor: Prof. Andrea Rinaldo; Co-advisor: Prof. Damiano Pasetto.
- 2020 **PhD Mobility, Infectious Disease Dynamics Group**, Johns Hopkins Bloomberg School of Public Health, MD, USA.  
with Prof. Justin Lessler.
- 2013 – 2017 **Master of Science in Computational Science and Engineering**, EPFL.
- 2010 – 2013 **Bachelor in Microengineering**, EPFL.

## Research Projects

- 2017 – Now **PhD on infectious diseases modeling towards the design of effective control policies.**  
My objective was to guide public-health decision makers towards the design of informed policies, using state-of-the-art statistical models to forecast epidemics. During these 4.5 years my main projects were:
- **COVID-19:**
    - contributing the JHU COVID Scenario Pipeline, a flexible modeling framework that projects epidemic trajectories and healthcare impacts under different suites of interventions in order to aid in scenario planning. It produces reports aimed at various governmental entities, and the pipeline is actively developed to meet the ever-changing decision maker needs.
    - assessing the effectiveness of non-pharmaceutical interventions in Switzerland for the first COVID-19 wave. In collaboration with local hospitals, we benefitted from detailed healthcare data which allowed us to mitigate the underreporting due to the low test capacity.
  - **Cholera:**
    - providing a model-based evaluation of the potential of a mass vaccination campaign against cholera in Haiti, as one of four groups of a multi-model comparison study.
    - formal model comparison study to gauge the effect of intra-seasonal rainfall events on cholera transmission in Juba, South Sudan.
  - **Optimal control of infectious diseases:** using automatic differentiation and non-linear programming, we explored methods to construct optimal vaccination strategies in space and time around any spatial epidemiological model, such that the impact of each dose is maximal.
- 2017 **Master Thesis: Agent-based Modelling of Schistosomiasis Transmission Dynamics.**  
ECHO.
- 2014–2015 **Semester Projects.**
  - **Left-Atrium Segmentation**, LHTC, CHUV, Switzerland. Development of a 4D segmentation framework to reconstruct the left-atrium from MRI scans, with GUI for practitioners.
  - **Autonomous Vehicle Formations**, DISAL, EPFL.

---

## Professional Expertise

### Soft skills

- Soft I enjoy working in multi-cultural environments, and collaborating across disciplinary boundaries. Creative, self-motivated, I strive to tackle complex, fast-moving, and ambiguous problems.
- Comm. Public-speaking and presenting results to audiences with diverse, non-quantitative backgrounds.

### Quantitative methods

- Data Exploratory Data Analysis (EDA), data visualization, predictive modeling, mechanistic or statistical modeling. Statistical inference using MCMC, particle filtering, frequentist or bayesian methods...
- Optimization Optimal Control, Automatic Differentiation (AD), Non-Linear Programming (NLP).
- Epi. modeling Compartmental (SIR), agent-based (ABM), spatial, stochastic or ODE-based models.

### Computer skills

- Python Numpy, Pandas, Numba, Dask, Matplotlib/Seaborn, CasADI...
- Programming R (incl. tidyverse), C, C++ (incl. Pybind11), Matlab, a bit of Stan.
- Development Git, Mercurial, CI and TDD, Docker, Slurm, High-Performance Computing.
- Other Latex, Markdown, GNU/Linux, Wordpress, Microsoft Office Suite, Adobe Creative Suite.

---

## Publications (& Preprints)

- 2021 **Optimizing the spatio-temporal allocation of COVID-19 vaccines: Italy as a case study**, *J. Lemaître, D. Pasetto, M. Zanon, E. Bertuzzo, L. Mari, S. Miccoli, R. Casagrandi, M. Gatto, A. Rinaldo*.  
medRxiv.
- 2021 **Modeling of future COVID-19 cases, hospitalizations, and deaths by vaccination and nonpharmaceutical interventions scenarios — United States, April–September 2021**, *R. Borchering, C. Viboud, E. Howerton, C. Smith, S. Truelove, M. Runge, N. Reich, L. Contamin, J. Levander, J. Salerno, W. van Panhuis, M. Kinsey, K. Tallaksen, R. F. Obrecht, L. Asher, C. Costello, M. Kelbaugh, S. Wilson, L. Shin, M. Gallagher, L. Mullany, K. Rainwater-Lovett, J. Lemaître, J. Dent, K. Grantz, J. Kaminsky, S. Lauer, E. Lee, H. Meredith et al.*  
MMWR. Morbidity and mortality weekly report 70.19, 719-724.
- 2021 **Wastewater monitoring outperforms case numbers as a tool to track COVID-19 incidence dynamics when test positivity rates are high**, *X. Fernandez-Cassi, A. Scheidegger, C. Bänziger, F. Cariti, A. Tuñas Corzon, P. Ganesanandamoorthy, J. Lemaître, C. Ort, T. Julian, T. Kohn*.  
Water Research, 117252.
- 2021 **Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the US**, *E. Cramer, V. Lopez, J. Niemi, G. George, J. Cegan, I. Dettwiler, W. England, M. Farthing, R. Hunter, B. Lafferty, I. Linkov, M. Mayo, M. Parno, M. Rowland, B. Trump, L. Wang, L. Gao, Z. Gu, M. Kim, Y. Wang, J. Walker, R. Slayton, M., Matthew B. et al.*  
medRxiv.
- 2021 **A scenario modeling pipeline for COVID-19 emergency planning**, *J. Lemaître\*, K. Grantz\*, J. Kaminsky\*, H. Meredith\*, S. Truelove\*, S. Lauer, L. Keegan, S. Shah, J. Wills, K. Kaminsky, J. Perez-Saez, J. Lessler, E. Lee*.  
Nature Scientific Reports 11 (1), 1-13.

- 2021 **Effect of specific non-pharmaceutical intervention policies on SARS-CoV-2 transmission in the counties of the United States**, *B. Yang, A. Huang, B. Garcia-Carreras, W. Hart, A. Staid, M. Hitchings, E. Lee, C. Howe, K. Grantz, A. Wesolowski, J. Lemaitre, S. Rattigan, C. Moreno, B. Borgert, C. Dale, N. Quigley, A. Cummings, A. McLorg, K. LoMonaco, S. Schlossberg, D. Barron-Kraus, H. Shrock, UFCOVID Interventions Team, J. Lessler, C. Laird, D. Cummings.*  
*Nature Communications* 12 (1), 1-10.
- 2020 **Assessing the impact of non-pharmaceutical interventions on SARS-CoV-2 transmission in Switzerland**, *J. Lemaitre\*, J. Perez-Saez\*, A. Azman, A. Rinaldo, J. Fellay.*  
*Swiss Medical Weekly* 150, w20295.
- 2021 **Range of reproduction number estimates for COVID-19 spread**, *D. Pasetto, J. Lemaitre, E. Bertuzzo, M. Gatto, A. Rinaldo.*  
*Biochemical and Biophysical Research Communications* 538, 253-258.
- 2019 **Achieving coordinated national immunity and cholera elimination in Haiti through vaccination: a modelling study**, *E. Lee\*, D. Chao\*, J. Lemaitre\*, L. Matrajt\*, D. Pasetto, J. Perez-Saez, F. Finger, A. Rinaldo, J. Sugimoto, M. E. Halloran, I. Longini, R. Ternier, K. Vissieres, A. Azman, J. Lessler, L. Ivers.*  
*The Lancet Global Health* 8 (8), e1081-e1089.
- 2019 **Rainfall as a driver of epidemic cholera: comparative model assessments of the effect of intra-seasonal precipitation events**, *J. Lemaitre, D. Pasetto, J. Perez-Saez, C. Sciarra, JF. Wamala, A. Rinaldo.*  
*Acta tropica* 190, 235-243.
- 2018 **Near real-time forecasting for cholera decision making in Haiti after Hurricane Matthew**, *D. Pasetto, F. Finger, A. Camacho, F. Grandesso, S. Cohuet, J. Lemaitre, A. Azman, F. Luquero, E. Bertuzzo, A. Rinaldo.*  
*PLoS Computational Biology* 14 (5), e1006127.
- 2015 **Distributed graph-based convoy control for networked intelligent vehicles**, *A. Marjovi, M. Vasic, J. Lemaitre, A. Martinoli.*  
2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS).
- 2015 **Cloud Base Height Estimation Using High-Resolution Whole Sky Imagers**, *F. Savoy, J. Lemaitre, S. Dev, Y. Lee, S. Winkler.*  
2015 IEEE Intelligent Vehicles Symposium (IV).

---

## Other academic contributions

### Community work

- 2020–2021 **Peer-reviews.**  
for PNAS<sup>(×2)</sup>, Swiss Medical Weekly<sup>(×1)</sup>, Royal Society Open Science<sup>(×4)</sup>, Nature Scientific Reports<sup>(×1)</sup>, PLoS One<sup>(×2)</sup>, Lancet Regional Health – Europe<sup>(×1)</sup>.
- 2017–2021 **Mentoring of Master students**, EPFL.  
On various infectious disease dynamics topics. Students: Thomas Fry, Cristiano Trevisin (now PhD student), Eloi Benvenuti, Maëlle Romero Grass, Briac Virey, Cloé Keller.
- 2012–2020 **Teaching Assistant**, EPFL.  
for bachelor and master level courses: Analysis (2012, 2016), Programming (2012 – 2014), Logical Systems (2012 – 2014), Mechanical Design (2013), Physics (2016), Hydrology (2017 – 2019), Water Resources Engineering (2019).

## **Outreach**

Links Interviewed or quoted in RTS, Der Bund, Berner Zeitung, Tages-Anzeiger, 24 Heures, Tribune de Genève, Le Temps, Le Temps (2), Le Figaro, Radio Canada, Heise, EPFL news.

## **Awards**

2021 **Swiss National Fund, PhD mobility grant.**  
15'000.- CHF for a mobility at Johns Hopkins University, Baltimore, USA.

## **Invited Talks**

Dec. 2021 **Environmental Engineering Seminar Series, EPFL.**  
Epidemiological modelling during the COVID-19 pandemic: between Science and Public Health.

## **Secondary Experience**

2015 **Junior Scientist, Illinois at Singapore Pte Ltd, Singapore.**  
6 month internship on cloud segmentation using MODIS Satellites and Whole-Sky Imagers.

2009 – 2013 **Embedded Developer for Mobile Robots, Club Vaudois de Robotique Autonome, Switzerland.**  
Participation to the European robotic contest *Eurobot*. Worked on computer vision, embedded programming, control systems and electrical design.

Sum. 2012 **Machining Traineeship, École des métiers de Lausanne.**  
Physical and precision work in loud environments.

## **Languages**

French Mother tongue.

English Advanced, professional working proficiency.

## **Personal Informations**

Nationalities French, Swiss.

Civil status Married.

Age 28 year old.

Misc. B driving license, no military obligations.