Gerardo Chowell
James M. Hyman
Luís M. A. Bettencourt
Carlos Castillo-Chavez
Editors

Mathematical and Statistical Estimation Approaches in Epidemiology

Springer

Mathematical and Statistical Estimation Approaches in Epidemiology

Gerardo Chowell · James M. Hyman ·
Luís M. A. Bettencourt · Carlos Castillo-Chavez
Editors

Mathematical and Statistical Estimation Approaches in Epidemiology

Springer

Editors
Dr. Gerardo Chowell
Arizona State University
School of Human Evolution & Social Change
Tempe AZ 85287-2402
USA
gchowell@asu.edu

Dr. Luís M. A. Bettencourt
Los Alamos National Laboratory
Los Alamos NM 87545
Mail Stop B284
USA
lmbett@lanl.gov

Dr. James M. Hyman
Los Alamos National Laboratory
Los Alamos NM 87545
Mail Stop B284
USA
mac@t7.lanl.gov

Dr. Carlos Castillo-Chavez
Arizona State University
Dept. Mathematics & Statistics
P.O.Box 871804
Tempe AZ 85287
USA
ccchavez@asu.edu

ISBN 978-90-481-2312-4 e-ISBN 978-90-481-2313-1
DOI 10.1007/978-90-481-2313-1
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2009926096

Preface

Mathematical and Statistical Estimation Approaches in Epidemiology compiles theoretical and practical contributions of experts in the analysis of infectious disease epidemics in a single volume. Recent collections have focused in the analyses and simulation of deterministic and stochastic models whose aim is to identify and rank epidemiological and social mechanisms responsible for disease transmission. The contributions in this volume focus on the connections between models and disease data with emphasis on the application of mathematical and statistical approaches that quantify model and data uncertainty.

The book is aimed at public health experts, applied mathematicians and scientists in the life and social sciences, particularly graduate or advanced undergraduate students, who are interested not only in building and connecting models to data but also in applying and developing methods that quantify uncertainty in the context of infectious diseases. Chowell and Brauer open this volume with an overview of the classical disease transmission models of Kermack-McKendrick including extensions that account for increased levels of epidemiological heterogeneity. Their theoretical tour is followed by the introduction of a simple methodology for the estimation of, the *basic reproduction number*, R_0. The use of this methodology is illustrated, using regional data for 1918–1919 and 1968 influenza pandemics. This chapter is followed by Greenwood and Gordillo's introduction to an analogous probabilistic framework. The emphasis is now on the computation of the distribution of the final epidemic size and the quantification of stochastically sustained oscillations. Next, the differences between *observable* and *unobservable* events in infectious disease epidemiology and their relationship to rigorous contact tracing and microbiological methodology are discussed in Chapter 3 by Nishiura et al. Furthermore, concepts like "dependent happening" and their role in identifying sources of infectious disease risk or in assessing vaccine efficacy are also discussed. In Chapter 4, Tennenbaum's engages us in a discussion of modeling perspectives and approaches through his discussion of the meaning of "contact". He challenges the reader to come up with novel approaches that bring together "ignored" biological and mechanistic aspects of the infection process.

Chapter 5 (Nishiura and Chowell) and Chapter 7 (Bettencourt) focus on real-time assessments of the reproduction number. The exposition is spiced with references to recent epidemic outbreaks. For example, Bettencourt uses his framework to estimate

disease epidemiological parameters and to assess the effects of interventions in real time using data from the 2005 outbreak of Marburg hemorrhagic fever in Angola. In Chapter 8, Burr and colleagues review the theoretical and practical challenges associated with biosurveillance including the detection of disease outbreaks using traditional diagnosed case rates or syndromic surveillance data. In Chapter 6, Lloyd notes that parameter estimates are subject to uncertainty that arise not only from errors (noise) in the data but also from the structure of the model used in the fitting process. In other words, he argues that uncertainty must be evaluated at multiple levels to account for our ignorance or for the balance that each modeler must reach between biological detail and model complexity and objectives. Parameter estimation, Lloyd argues, must include structural sensitivity analyses. The use of historical data in epidemiological research is highlighted in Chapter 9 by Acuña-Soto's contribution. As he notes epidemiologists are reluctant to consider systematically the possibility of working with historical data albeit, as we have seen in the first Chapter, it is possible to extract valuable information from such data on influenza outbreaks. In fact, we acquired the kind of quantitative knowledge that let us quantify some of the differences between seasonal and pandemic influenza. Acuña-Soto's work[1], for example, on the epidemic of 1576 that killed 45% of the entire population of Mexico, highlights but a myriad of new possibilities for which the quantitative methods and approaches highlighted in this book can be put to good use.

Banks and colleagues in Chapter 11 provide a succinct overview of the statistical and computational aspects associated with inverse or parameter estimation problems for deterministic dynamical systems. Their results illustrate the impact that the marriage between statistical theory and applied mathematics is having in the study of infectious diseases while Chapter 10 (Arriola and Hyman) provides a general and thorough introduction to the field of sensitivity and uncertainty analyses, a central piece of any scientific work that is based on modeling.

The challenges and opportunities generated by studies of disease outbreak or disease dynamics in specific contexts are highlighted in the final chapters. Shim and Castillo-Chavez (Chapter 12) evaluate the potential impact that ongoing age-dependent vaccination strategies (in the United States and Mexico) are likely to have in reducing the prevalence of severe rotavirus infections. Rios-Doria et al. (Chapter 13) analyze the spatial and temporal dynamics of rubella in Peru, 1997–2006 via a wavelet time series analysis and other methods. The study is carried out in the context of changing policies that include the introduction of a vaccine and/or increases in vaccination rates. Cintron-Arias and colleagues (Chapter 15) model drinking as a "communicable" disease and, in the process, they highlight a new set of opportunities and possibilities for the applications of the mathematical and statistical approaches used in this volume. The focus here is on the evaluation of the role of relapse (ineffective treatment) on drinking dynamics but as a function of social network heterogeneity.

---

[1] R Acuna-Soto, LC Romero, and JH Maguire; Large epidemics of hemorrhagic fevers in Mexico 1545–1815; Am. J. Trop. Med. Hyg, 62(6), 2000, pp. 733–739.

The Editors would like to thank Melania Ruiz our associated Editor at Springer for her encouragement and infinite patience. We also thank the Springer publication staff for their support in helping us put this volume together. We also thank Carlos A. Torre for his technical support and sense of humor. Finally, we would like to thank all the authors for their timely and innovative contributions and their patience during the full process.

Tempe, AZ — Gerardo Chowell
Los Alamos, NM — James M. Hyman
Los Alamos, NM — Luís M.A. Bettencourt
Tempe, AZ — Carlos Castillo-Chavez

Contents

The Basic Reproduction Number of Infectious Diseases: Computation and Estimation Using Compartmental Epidemic Models 1
Gerardo Chowell and Fred Brauer

Stochastic Epidemic Modeling 31
Priscilla E. Greenwood and Luis F. Gordillo

Two Critical Issues in Quantitative Modeling of Communicable Diseases: Inference of Unobservables and Dependent Happening 53
Hiroshi Nishiura, Masayuki Kakehashi, and Hisashi Inaba

The Chain of Infection, Contacts, and Model Parametrization 89
Stephen Tennenbaum

The Effective Reproduction Number as a Prelude to Statistical Estimation of Time-Dependent Epidemic Trends 103
Hiroshi Nishiura and Gerardo Chowell

Sensitivity of Model-Based Epidemiological Parameter Estimation to Model Assumptions 123
A.L. Lloyd

An Ensemble Trajectory Method for Real-Time Modeling and Prediction of Unfolding Epidemics: Analysis of the 2005 Marburg Fever Outbreak in Angola 143
Luís M. A. Bettencourt

Statistical Challenges in BioSurveillance 163
Tom Burr, Sarah Michalak, and Rick Picard

Death Records from Historical Archives: A Valuable Source of Epidemiological Information 189
Rodolfo Acuna-Soto

Sensitivity Analysis for Uncertainty Quantification in Mathematical Models 195
Leon Arriola and James M. Hyman

An Inverse Problem Statistical Methodology Summary 249
H. Thomas Banks, Marie Davidian, John R. Samuels, Jr., and Karyn L. Sutton

The Epidemiological Impact of Rotavirus Vaccination Programs in the United States and Mexico 303
Eunha Shim and Carlos Castillo-Chavez

Spatial and Temporal Dynamics of Rubella in Peru, 1997–2006: Geographic Patterns, Age at Infection and Estimation of Transmissibility 325
Daniel Rios-Doria, Gerardo Chowell, Cesar Munayco-Escate, Alvaro Witthembury, and Carlos Castillo-Chavez

The Role of Nonlinear Relapse on Contagion Amongst Drinking Communities 343
Ariel Cintrón-Arias, Fabio Sánchez, Xiaohong Wang, Carlos Castillo-Chavez, Dennis M. Gorman, and Paul J. Gruenewald

Index 361

Contributors

Rodolfo Acuna-Soto Departamento de Microbiología y Parasitología, Facultad de Medicina, Universidad Nacional Autónoma de México, Delegación Coyoacán, México D.F. 04510, México, yvonne@ibt.unam.mx

Leon Arriola Department of Mathematiccal and Computer Sciences, University of WisconsinWhitewater, Whitewater, WI 53190, USA, arriolal@uww.edu

Luís M. A. Bettencourt Theoretical Division, MS B284, Los Alamos National Laboratory, Los Alamos NM 87545, USA; Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA, lmbett@lanl.gov

Fred Brauer Department of Mathematics, The University of British Columbia, Vancouver, B.C., Canada V6T 1Z2, brauer@math.ubc.ca

Tom Burr Statistical Sciences Group, Los Alamos National Laboratory, Los Alamos NM 87545, USA, tburr@lanl.gov

Carlos Castillo-Chavez Mathematical, Computational, and Modeling Sciences Center, P.O. Box 871904, Arizona State University, Tempe, AZ 85287, USA; School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85287, USA; Department of Mathematics and Statistics, Arizona State University, Tempe, AZ, 85287, USA; Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA, chavez@math.asu.edu

Gerardo Chowell School of Human Evolution and Social Change, Arizona State University, Box 872402, Tempe, AZ 85287, USA; Mathematical, Computational, Modeling Sciences Center, Arizona State University, Tempe, AZ 85287, USA; Division of Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, MD, USA, gchowell@asu.edu

Ariel Cintron-Arias Center for Research in Scientific Computation, Box 8205, North Carolina State University, Raleigh, NC 27695-8212, USA, acintro@unity.ncsu.edu

Marie Davidian Center for Research in Scientific Computation and Center for Quantitative Sciences in Biomedicine, North Carolina State University, Raleigh, NC 27695-8212, USA, davidian@eos.ncsu.edu

Luis F. Gordillo Department of Mathematics, University of Puerto Rico-RUM, Mayaguez, PR 00681-9018, USA, gordillo@math.uprm.edu

Dennis M. Gorman Department of Epidemiology and Biostatistics, School of Rural Public Health, Texas A&M Health Science Center, P.O. Box 1266, College Station, TX 77843-1266, USA, dennis-m-gorman@tamu.edu

Priscilla E. Greenwood Department of Mathematics and Statistics, Arizona State University, Tempe, AZ 85287-1804, USA, pgreenw@math.la.asu.edu

Paul J. Gruenewald Prevention Research Center, 1995 University Avenue, Suite 450, Berkeley, CA 94704, USA, paul@prev.org

James M. Hyman Theoretical Division, MS B284, Los Alamos National Laboratory, Los Alamos, NM 87545, USA, mac@t7.lanl.gov; hyman@lanl.gov

Hisashi Inaba Graduate School of Mathematical Sciences, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan, inaba@ms.u-tokyo.ac.jp

Masayuki Kakehashi Graduate School of Health Sciences, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8551, Japan, kakehashi@hiroshima-u.ac.jp

Alun L. Lloyd Biomathematics Graduate Program and Department of Mathematics, North Carolina State University, Raleigh, NC 27695, USA, alun_lloyd@ncsu.edu

Sarah Michalak Statistical Sciences Group, Los Alamos National Lab, Los Alamos NM 87545, USA, michalak@lanl.gov

Cesar Munayco-Escate Dirección General de Epidemiología, Ministerio de Salud, Perú, Jr. Camilo-Carrillo 402, Jesus Maria-Lima 11, Peru, cmunayco@dge.gob.pe

Hiroshi Nishiura Theoretical Epidemiology, University of Utrecht, Yalelaan 7, 3584 CL, Utrecht, The Netherlands, h.nishiura@uu.nl

Rick Picard Statistical Sciences Group, Los Alamos National Lab, Los Alamos NM 87545, USA, picard@lanl.gov

Daniel Rios-Doria School of Human Evolution and Social Change, Arizona State University, Box 872402 Tempe, AZ 85287, USA; Mathematical, Computational, Modeling Sciences Center, Arizona State University, Tempe, AZ 85287, USA, daniel@mathpost.asu.edu

John R. Samuels, Jr. Center for Research in Scientific Computation and Center for Quantitative Sciences in Biomedicine, North Carolina State University, Raleigh, NC 27695-8205, USA, jrsamue2@ncsu.edu

Fabio Sanchez Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14853-7801, USA, fas9@cornell.edu

Eunha Shim Department of Epidemiology and Public Health, Yale School of Medicine, New Haven, CT, 06510, USA, eunha.shim@yale.edu

Karyn L. Sutton Center for Research in Scientific Computation and Center for Quantitative Sciences in Biomedicine, North Carolina State University, Raleigh, NC 27695-8212, USA, klsutton@unity.ncsu.edu

Stephen Tennenbaum Mathematical, Computational and Modeling Sciences Center, Arizona State University, Tempe, AZ 85287-1904, USA, set1@asu.edu

H. Thomas Banks Center for Research in Scientific Computation and Center for Quantitative Sciences in Biomedicine, North Carolina State University, Raleigh, NC 27695-8212; Department of Mathematics and Statistics, Arizona State University Tempe, AZ, 85287-1804, USA, htbanks@ncsu.edu

Xiaohong Wang Mathematical, Computational, and Modeling Sciences Center, P.O. Box 871904, Arizona State University, Tempe, AZ 85287, USA, Xiaohong.Wang@asu.edu

Alvaro Witthembury Dirección General de Epidemiología, Ministerio de Salud, Perú, Jr. Camilo-Carrillo 402, Jesus Maria-Lima 11, Perú, awhittembury@dge.gob.pe

The Basic Reproduction Number of Infectious Diseases: Computation and Estimation Using Compartmental Epidemic Models

Gerardo Chowell and Fred Brauer

Abstract The basic reproduction number (R_0) is a central quantity in epidemiology as it measures the transmission potential of infectious diseases. In this chapter we review the basic theory of the spread of infectious diseases using simple compartmental models based on ordinary differential equations including the simple Kermack-McKendrick epidemic model, SIR (susceptible-infectious-removed) models with demographics, the SIS (susceptible-infectious-susceptible) model, backward bifurcations, endemic equilibria, and the analytical derivation of R_0 using the next-generation approach. This theory is followed by simple methodology for the estimation of R_0 with its corresponding uncertainty from epidemic time series data. The 1918–1919 influenza pandemic in Winnipeg, Canada, and the 1968 influenza pandemic in US cities are used for illustration.

Keywords Influenza · Pandemic · Epidemiology · Basic reproduction number · Model

1 Thresholds in Disease Transmission Models

One of the fundamental results in mathematical epidemiology is that mathematical epidemic models, including those that include a high degree of heterogeneity exhibit a "threshold" behavior. In epidemiological terms, this can be stated as follows: *There is a difference in epidemic behavior when the average number of secondary infections caused by an average infective during his/her period of infectiousness, called the* basic reproduction number*, is less than one and when this quantity exceeds one*.

G. Chowell (✉)
School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85287, USA; Mathematical, Computational, Modeling Sciences Center, Arizona State University, Tempe, AZ 85287, USA; Division of Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, MD, USA
e-mail: gchowell@asu.edu

G. Chowell et al. (eds.), *Mathematical and Statistical Estimation Approaches in Epidemiology*, DOI 10.1007/978-90-481-2313-1_1,

There are two different situations. If the course of the disease outbreak is rapid enough that there are no significant demographic effects (births, natural deaths, recruitment) on the population being studied, then the disease will die out if the basic reproduction number is less than one, and if it exceeds one there will be an epidemic.

If, on the other hand, there is a flow into the population of individuals who may become infected, through births, recruitment, or recovery of infected individuals with no immunity against reinfection, then there is a different alternative. If the basic reproduction number is less than one, the disease dies out in the population. Mathematically this is expressed by the fact that there is a disease-free equilibrium approached by solutions of the model describing the situation. If the basic reproduction number exceeds one, the disease-free equilibrium is unstable and solutions flow away from it. There is also an *endemic* equilibrium, with a positive number of infective individuals, indicating that the disease remains in the population.

However, the situation may be more complicated. We shall see later that in certain circumstances it is possible to have an endemic equilibrium with a reproduction number less than one.

We begin by describing the threshold phenomenon and the basic reproduction number in epidemic models.

2 The Simple Kermack-McKendrick Epidemic Model

An epidemic, which acts on a short temporal scale, may be described as a sudden outbreak of a disease that infects a substantial portion of the population in a region before it disappears. Epidemics usually leave many members untouched. Often these attacks recur with intervals of several years between outbreaks, possibly diminishing in severity as populations develop some immunity.

One of the questions that first attracted the attention of scientists interested in the study of the spread of communicable diseases was why diseases would suddenly develop in a community and then disappear just as suddenly without infecting the entire community. One of the early triumphs of mathematical epidemiology [54] was the formulation of a simple model that predicted behavior very similar to that observed in countless epidemics. The Kermack-McKendrick model is a compartmental model based on relatively simple assumptions on the rates of flow between different classes of members of the population.

We formulate our descriptions as *compartmental models*, with the population under study being divided into compartments and with assumptions about the nature and time rate of transfer from one compartment to another. Diseases that confer immunity have a different compartmental structure from diseases without immunity. We will use the terminology SIR to describe a disease which confers immunity against re-infection, to indicate that the passage of individuals is from the susceptible class S to the infective class I to the removed class R. On the other hand, we will use the terminology SIS to describe a disease with no immunity against

re-infection, to indicate that the passage of individuals is from the susceptible class to the infective class and then back to the susceptible class. Other possibilities include $SEIR$ and $SEIS$ models, with an exposed period between being infected and becoming infective, and $SIRS$ models, with temporary immunity on recovery from infection.

In order to model such an epidemic we divide the population being studied into three classes labeled S, I, and R. We let $S(t)$ denote the number of individuals who are susceptible to the disease, that is, who are not (yet) infected at time t. $I(t)$ denotes the number of infected individuals, assumed infectious and able to spread the disease by contact with susceptibles. $R(t)$ denotes the number of individuals who have been infected and then removed from the possibility of being infected again or of spreading infection. Removal is carried out either through isolation from the rest of the population or through immunization against infection or through recovery from the disease with full immunity against reinfection or through death caused by the disease. These characterizations of removed members are different from an epidemiological perspective but are often equivalent from a modeling point of view which takes into account only the state of an individual with respect to the disease.

In formulating models in terms of the derivatives of the sizes of each compartment we are assuming that the number of members in a compartment is a differentiable function of time. This may be a reasonable approximation if there are many members in a compartment, but it is certainly suspect otherwise.

The basic compartmental models to describe the transmission of communicable diseases are contained in a sequence of three papers by W.O. Kermack and A.G. McKendrick in 1927, 1932, and 1933 [54–56]. The first of these papers described epidemic models. What is often called the Kermack-McKendrick epidemic model is actually a special case of the general model introduced in this paper. The general model included dependence on age of infection, that is, the time since becoming infected.

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

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

It is based on the following assumptions:

(i) An average member of the population makes contact sufficient to transmit infection with βN others per unit time, where N represents total population size (mass action incidence).
(ii) Infectives leave the infective class at rate αI per unit time.
(iii) There is no entry into or departure from the population; in particular there are no deaths from the disease. Thus population size is a constant N_0.