Cambridge University Press
978-0-521-19134-0 - Introduction to Statistical Methods for Biosurveillance: With an Emphasis on Syndromic Surveillance
Ronald D. Fricker
Frontmatter
More information

Introduction to Statistical Methods for Biosurveillance

Bioterrorism is not a new threat but, in an increasingly interconnected world, the potential for catastrophic outcomes is greater today than ever. The medical and public health communities are establishing biosurveillance systems designed to proactively monitor populations for possible disease outbreaks as a first line of defense.

The ideal biosurveillance system should identify trends not visible to individual physicians and clinicians in near-real time. Many of these systems use statistical algorithms to look for anomalies and to trigger epidemiologic investigation, quantification, localization, and outbreak management.

This book is focused on the design and evaluation of statistical methods for effective biosurveillance. Weaving public health and statistics together, it presents both basic and more advanced methods, all with a focus on empirically demonstrating added value. Although the emphasis is on epidemiologic surveillance and syndromic surveillance, the statistical methods can also be applied to a broad class of public health surveillance problems.

Ronald D. Fricker, Jr. is an Associate Professor of Operations Research at the Naval Postgraduate School (NPS). He holds a Ph.D. in statistics from Yale University. Prior to joining NPS, Dr. Fricker was a Senior Statistician at the RAND Corporation and the Associate Director of the National Security Research Division. Published widely in leading professional journals, he is a Fellow of the American Statistical Association, an Elected Member of the International Statistical Institute, and a former chair of the ASA Section on Statistics in Defense and National Security. He is a contributing editor to *Interfaces* and is on the editorial boards of *Statistics, Politics, and Policy* and the *International Journal of Quality Engineering and Technology*. Fricker's current research is focused on studying the performance of various statistical methods for use in biosurveillance, particularly syndromic surveillance, and statistical process control methodologies more generally.

Introduction to Statistical Methods for Biosurveillance

With an Emphasis on Syndromic Surveillance

RONALD D. FRICKER, JR.
Naval Postgraduate School

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Mexico City

Cambridge University Press
32 Avenue of the Americas, New York, NY 10013-2473, USA

www.cambridge.org
Information on this title: www.cambridge.org/9780521191340

First published 2013

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication data

Fricker, Ronald D., 1960–
Introduction to statistical methods for biosurveillance : with an emphasis on syndromic surveillance /
Ronald D. Fricker, Jr.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-0-521-19134-0 (hardback)
I. Title.
[DNLM: 1. Biosurveillance – methods. 2. Bioterrorism – prevention & control.
3. Communicable Disease Control – methods. 4. Disease Outbreaks – prevention & control.
5. Models, Statistical. WA 950]
363.325′3–dc23 2012035196

ISBN 978-0-521-19134-0 Hardback

Dedicated to all who are
working to protect the world
from disease and terrorism.

Cambridge University Press
978-0-521-19134-0 - Introduction to Statistical Methods for Biosurveillance: With an Emphasis on Syndromic Surveillance
Ronald D. Fricker
Frontmatter
More information

Contents

Part V Appendices

Preface

This book is about basic statistical methods useful for biosurveillance. The focus on basic methods has a twofold motivation. First, there is a need for a text that starts from the fundamentals, both of public health surveillance and statistics, and weaves them together into a foundation for biosurveillance. Only from a solid foundation can an enduring edifice be built.

Second, while there is a large and growing literature about biosurveillance that includes the application of some very complicated and sophisticated statistical methods, it has been my experience that more complicated methods and models do not always result in better performance. And even when they do, there is often an inherent trade-off made in terms of transparency and interpretability.

Indeed, a real challenge in today's data-rich environment is deciding when enough complication is enough. More is not always better, whether we're talking about eating dessert or building a model or developing a detection algorithm. There is a rich history that speaks to this point:

> Occam's razor: "All other things being equal, a simpler explanation is better than a more complex one."
>
> Blaise Pascal (1623–1662): "Je n'ai fait cette lettre – ci plus longue que parce que je n'ai pas eu le loisir de la faire plus courte." (I have made this letter longer than usual, only because I have not had time to make it shorter.)
>
> Albert Einstein (1879–1955): "Make everything as simple as possible, but not simpler," and "Any intelligent fool can make things bigger, more complex. . . . It takes a touch of genius . . . to move in the opposite direction."

Note the theme in these quotes is not one of just simplicity but also that it takes effort and insight to *appropriately* simplify. Hence, I do not claim that the methods in this book are necessarily the best or most correct ones for biosurveillance. Most of the research necessary to reach such a determination is yet to be done. However, the philosophy on which this book is predicated is that biosurveillance should start with basic methods such as those described herein and, only after *empirically demonstrating the added value of more complicated methods*, extend from there.

This text presumes a familiarity with basic probability and statistics at the level of an advanced undergraduate or beginning graduate-level course. For readers requiring a probability refresher, Appendix A provides a brief review of many of the basic concepts used throughout the text. However, the text also uses some statistical methods that are often not taught in introductory courses, such as ROC (receiver operating characteristic) curves, imputation, and time series modeling. In presenting these and other methods, the goal has been to make the exposition as accessible and as relevant to the widest audience possible. However, this inevitably means that some of the concepts and methods will be insufficiently explained for some readers, while others may have preferred a more advanced treatment. In an attempt to accommodate all levels of interest, the end of each chapter contains an "additional reading" section with pointers to other resources, some providing more background and introductory material and others providing a more advanced treatment of the material.

That said, this book is largely focused on univariate temporal data. More complicated data, whether multivariate or spatio-temporal, will by definition require more complicated statistical methods. In this book, I touch on these types of data, but they require a treatment more in depth than a text of this length will allow.

As a statistician with a background in industrial quality control, I approach the problem of biosurveillance early event detection from the perspective of statistical process control (SPC). This is, of course, only one way to approach the problem, and different disciplines have different viewpoints.

SPC methods were first developed to monitor industrial processes, which are generally more controlled and for which the data are often easier to distributionally characterize than biosurveillance data. Nonetheless, I am of the opinion that, appropriately applied to biosurveillance data, these methods have much to offer in terms of (1) their performance and (2) a rich, quantitatively rigorous literature that both develops the methods and describes their performance characteristics. Thus, returning to a previous point, my motivation for starting from an SPC perspective is that it provides biosurveillance with a solid methodological foundation on which to build.

It is also important to note that I tend to look at biosurveillance as a tool for guarding against bioterrorism. Of course, a system designed to detect a bioterrorism attack is also useful for detecting natural disease outbreaks, but it's not necessarily true that a biosurveillance system designed for natural disease detection will be optimal for bioterrorism applications. Just as the person who tries to please everyone ends up pleasing no one, so it is with biosurveillance. Thus, while these systems do have dual-use possibilities, I am of the opinion that first and foremost they should be designed for thwarting bioterrorism.

Additional material related to this book, including errata, can be found at http://facultly.nps.edu/rdfricke/biosurveillance_book/. Please feel free to e-mail me at rdfricker@nps.edu with any comments, thoughts, or material that might be relevant and useful in the next revision.

In conclusion, I hope this book contributes to the effective design and implementation of biosurveillance systems. Given the increasingly dangerous threats that face humankind, some of natural origin and some not, and all magnified by our increasingly interconnected world, biosurveillance systems are truly a first line of defense.

Monterey, California | R. D. Fricker, Jr.
September 2012 | Associate Professor

Cambridge University Press
978-0-521-19134-0 - Introduction to Statistical Methods for Biosurveillance: With an Emphasis on Syndromic Surveillance
Ronald D. Fricker
Frontmatter
More information

Acknowledgments

This book has benefited from discussions and interactions with, and the assistance of, many people. The following is surely an incomplete list.

In the academic community: Bill Woodall, Dan Jeske, Howard Burkom, David Buckeridge, Doug Montgomery, Ken Kleinman, Galit Shmueli, Lance Waller, Karen Kafadar, Mike Stoto, Kwok Tsui, Yajun Mei, Al Ozonoff, and Abel Rodriguez.

In the public health community: Henry Rolka, Taha Kass-Hout, Lori Hutwagner, Jerry Tokars, Myron Katzoff, Wendy Wattigney, and Kathy O'Connor of the Centers for Disease Control and Prevention and Krista Hanni, Suzie Barnes, Kristy Michie, and Bryan Rees of the Monterey County Health Department (MCHD).

I would particularly like to thank Krista Hanni and the MCHD for sharing data. One of the major impediments to improving biosurveillance is the lack of access to real data. Krista and the MCHD were uncommonly forward leaning; rather than finding reasons why something could not be done, they constantly looked for ways to make things work. The rest of the public health community would do well to follow their lead.

A large portion of the research for this book was conducted while I was on sabbatical at the University of California, Riverside (UCR). My thanks to Dan Jeske, chair of the UCR Department of Statistics, for hosting my sabbatical. Thanks also to the Naval Postgraduate School (NPS) for sponsoring the sabbatical.

While on sabbatical, I taught a course using an early version of this book. I am very appreciative of the UCR students who took the class: Tatevik Ambartsoumian, Fei He, Quan Tuong Truong Le, Rebecca Phuonganh Le, Xin Zhang, Joyce Yingzhuo, Anne Hansen, and Judy Li. Their involvement and comments helped significantly improve the text.

At NPS, I also used a draft for a reading course, and the material was again improved with the feedback of those who participated in that course: Krista Hanni, Suzie Barnes, Manny Ganuza, Katie Hagen, and Randi Korman. And much of the material in Chapters 9 and 10 is the result of joint research with NPS students, including Katie Hagen, Ben Hegler, Matt Knitt, Andy Dunfee, and Cecilia Hu.

I am indebted to many people who, over the years, have supported and nurtured my research and academic career, including Randy Spoeri, who first introduced me to statistics and sparked my interest in the field; Joe Chang, advisor and researcher extraordinaire, without whose patience and support I would not have survived the dissertation process; and Nancy Spruill, a great friend and mentor, whose leadership, management, and organizational abilities I admire and to which I can only aspire.

Many thanks also to Lauren Cowles, my editor at Cambridge University Press. Lauren's encouragement and calm patience are major reasons this book actually made it to completion.

For many reasons I will always be beholden to and grateful for my spouse, Christine Arruda. In terms of this book, she good-naturedly endured, and often encouraged, my research and writing efforts. The time invested in this book often came at the expense of time we would otherwise have spent together.

Finally, I would be terribly remiss if I did not acknowledge the broader community of researchers and practitioners from whom I've benefited over the years and on whose work this book is based. Of course, any errors or omissions – and all opinions – are my own.