Interpreting Epidemiologic Evidence

Strategies for Study Design and Analysis

DAVID A. SAVITZ

Interpreting Epidemiologic Evidence

Strategies for Study Design and Analysis

DAVID A. SAVITZ

OXFORD
UNIVERSITY PRESS
2003

OXFORD
UNIVERSITY PRESS

Oxford New York
Auckland Bangkok Buenos Aires Cape Town Chennai
Dar es Salaam Delhi Hong Kong Istanbul Karachi Kolkata
Kuala Lumpur Madrid Melbourne Mexico City Mumbai
Nairobi São Paulo Shanghai Taipei Tokyo Toronto

Copyright © 2003 by Oxford University Press, Inc.

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York, 10016
http://www.oup-usa.org

Oxford is a registered trademark of Oxford University Press

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
electronic, mechanical, photocopying, recording, or otherwise,
without the prior permission of Oxford University Press.

Library of Congress Cataloging-in-Publication Data
Savitz, David A.
Interpreting epidemiologic evidence :
strategies for study design and analysis /
David A. Savitz.
p. ; cm. Includes bibliographical references and index.
ISBN 0-19-510840-X
1. Epidemiology—Technique. I. Title.
[DNLM: 1. Epidemiologic Research Design.
2. Bias (Epidemiology)
WA 950 S267i 2003]
RA652.4 .S286 2003
614.4—dc21 2002192672

9 8 7 6 5 4 3 2
Printed in the United States of America
on acid-free paper

Interpreting Epidemiologic Evidence

PREFACE

There was no shortage of epidemiology books when I started writing this book in the summer of 1996, and the intervening years have brought many new and very useful ones. Just as it is uninspiring to do a study that ends up as one more dot on a graph or one more line in a table, it was certainly not my goal to merely add one more book to an overcrowded shelf. But there was a particular need that did not seem to be very well addressed in books or journal articles—to bring together concepts and methods with real research findings in order to make informed judgments about the results.

One of the most difficult tasks facing new students and even experienced practitioners of epidemiology is to assess how much confidence one can have in a given set of findings. As I discussed such issues with graduate students and colleagues, and contemplated my own data, it was difficult to find a reasoned, balanced approach. It's easy but uninformative to simply acknowledge epidemiology's many pitfalls, and not much more difficult to mount a generic defense of the credibility of one's findings. What was difficult was to find empirical tools for assessing the study's susceptibility to specific sources of error in ways that could actually change preconceptions and go beyond intuition. This kind of examination is most fruitful when it can give good or bad news about one's study that was unexpected.

I knew that such approaches were out there because I found candidate tools

此为试读,需要完整PDF请访问: www.ertongbook.com

in the work of others. Sometimes our discussions would lead us to a technique applied elsewhere that was applicable to the current situation. Two elements were lacking, however, and I have tried to address them in this book. One is the link between methodological principles and the tools themselves, which involves taking stock of why the strategy for addressing the potential bias may or may not actually be informative, and how it could be misleading. The other is a full listing of the candidates to consider in addressing a potential problem, in the hope of improving our ability to draw upon one tool or another in the appropriate situation. My aspiration was to have a place to turn to when trying to interpret a study that deals with a challenging exposure measure, for example, where you could find a repertoire of tactics to assess how well it was measured as well as a reminder of what the consequences of error would most likely be. Ideally, at the point of planning the study, one would anticipate the challenge and collect the data needed to address the extent of the error upon completion of the study. For every topic in the book, there are chapters in other texts, many of which are excellent, and there are many journal articles on novel aspects of epidemiologic methods, but I could not find a source book of practical guidance on linking methodological principles with research practice. That is what this book aspires to provide.

My original goal in developing the book was to provide a reference that could be used when planning a study or needing to evaluate one. I believe that the book could also be useful in an intermediate or advanced epidemiologic methods course, supplementing a more rigorous methods text. Each chapter is intended to be freestanding, and could be referred to without having read the ones that precede it, but organizing the book along the lines of how research is conceptualized, conducted, and analyzed offers some benefits if the sequence of chapters is followed. Instructors might want to use this book along with evaluation of specific published papers for individual evaluation or group discussion. Using the above example, after considering the underlying algebra of exposure misclassification and its consequences in a methods text, the corresponding chapter in this book and a review of one or more real examples from the literature could help tie the theory to the practice, always a challenge for new (and old) epidemiologists. Making the connection between methodological principles and specific, substantive research applications is the most intellectually stimulating and challenging aspect of the field, in my view.

I thank Abigail Ukwuani for her excellent work in composing the manuscript and Jeff House for his patience and encouragement. I also express my appreciation to a number of distinguished epidemiologists who were kind enough to review various chapters and provide general guidance and encouragement as well: Marilie Gammon, Dana Loomis, Bob Millikan, Andy Olshan, Charlie Poole, Ken Rothman, and Noel Weiss. Joanne Promislow was extremely helpful in going through the manuscript in detail to capture the spirit of the book, and in com-

pleting the challenging task of identifying suitable illustrations from the published literature. Nonetheless, these colleagues and others who may read the book will undoubtedly find statements with which they disagree, so their help and acknowledgment should not be seen as a blanket endorsement of the book's contents. Following the spirit espoused in this book, critical evaluation is always needed and I welcome readers' comments on any errors in logic or omissions of potentially valuable strategies that they may find.

CONTENTS

1

INTRODUCTION

This book was written for both producers and consumers of epidemiologic research, though a basic understanding of epidemiologic principles will be necessary at the outset. Little of the technical material will be new to experienced epidemiologists, but I hope that my perspective on the application of those principles to interpreting research results will be distinctive and useful. For the large and growing group of consumers of epidemiology, which includes attorneys and judges, risk assessors, policy experts, clinicians, and laboratory scientists, the book is intended to go beyond the principles learned in an introductory course or textbook of epidemiology by applying them to concrete issues and findings. Although it is unlikely that ambiguous evidence can be made conclusive or that controversies will be resolved directly by applying these principles, a careful consideration of the underlying reasons for the ambiguity or opposing judgments in a controversy can represent progress. By pinpointing where the evidence falls short of certainty, we can give questions a sharper focus, leading to a clearer description of the state of knowledge at any point in time and thus helping identify the research that could contribute to a resolution of the uncertainty that remains.

Those who are called upon to assess the meaning and persuasiveness of epidemiologic evidence have a variety of approaches to consider. There are formal guidelines for judging causality for an observed association, which is defined as

此为试读,需要完整PDF请访问: www.ertongbook.com

a statistical dependence of two or more events (Last, 2001). A statistical association by itself does not indicate whether one causes the other, which is the issue of ultimate interest. The most widely cited framework for assessing the causal implications of associations is that of Hill (1965), which has been challenged by others (Rothman & Greenland, 1998) yet continues to be used widely by those who evaluate research findings in epidemiology. Hill's criteria serve as a means of reaching conclusions about reported positive associations that can help guide regulatory or policy decisions. The criteria, however, focus only on the interpretation of *positive* associations, neglecting the need to evaluate the validity of whatever association was measured or to consider the credibility of an observed absence of association.

Statistical methods have been used both to evaluate whether observed associations are statistically significant—for example, unlikely to have been observed if no association is present—and to quantify the strength of an association (Selvin, 1991; Clayton & Hills, 1993). A large methodological literature on confounding (in which the effect from an exposure of interest is entangled with other contributors), selection bias (in which the individuals who are included in the study results in erroneous measures of association), and misclassification (error in measurement) contributes importantly to making judgments about the validity of epidemiologic observations (Rothman & Greenland, 1998).

In parallel with the evolution of these more technical methodological considerations, interest in understanding and using epidemiology has grown considerably, reflected in media attention, courtroom applications, and interactions with scientists in other disciplines. The views of epidemiology within these groups are not always favorable. Many outside epidemiology have one of two extreme reactions to the evidence we generate: They may be so impressed with our findings on human beings exposed to the agents that may cause disease that observed associations are taken as direct reflections of causal effects with little need for scrutiny or caution. Or they may be more impressed with the lengthy list of potential sources of error, the ubiquitous potential confounders, and the seemingly unending controversy and flow of criticism among epidemiologists, and may come to believe that all our observations are hopelessly flawed and cannot be trusted as indicators of causal relations. Students often start with a naive, optimistic view of the power of the epidemiologic approach, become dismayed with the many sources of potential error, and then (hopefully) emerge with a sophistication that intelligently balances the promise and the pitfalls. More experienced epidemiologists appreciate that the truth lies somewhere between the extremes. Even for those who are familiar with the tools needed to evaluate evidence, however, the integration of that evidence into a global assessment is often a highly subjective process, and can be a contentious one.

This book is not a step-by-step manual for interpreting epidemiologic data that guarantees drawing the correct conclusion. It is simply not possible to reduce the

evaluation of evidence to an algorithm for drawing valid inferences. And because the truth is unknown, we could not tell whether any such algorithm worked. A more modest goal is to elucidate the underlying issues involved in the interpretation of evidence so that unbiased, knowledgeable epidemiologists can reach agreement or identify precisely where and why they disagree. In this book, I have tried to develop in some detail the array of considerations that should be taken into account in characterizing the epidemiologic evidence on a given topic, suggest how to identify the key considerations, and most importantly, offer a variety of strategies to determine whether a potential methodologic problem is likely to be influential and if so what magnitude and direction of influence it may have. The methodologic literature, particularly the recent synthesis by Rothman and Greenland (1998), provides the starting point for that evaluation. This book applies some methodological principles in specific and practical ways to the assessment of research findings in an effort to help reach sound judgments. In some cases traditional approaches to evaluating evidence are examined and found to be deficient. Because they are commonly used, however, they warrant careful examination here.

For instance, confounding is rather well-defined in theoretical terms (Greenland & Robins, 1986; Weinberg, 1993). According to Rothman and Greenland (1998), it is "a confusion of effects. Specifically, the apparent effect of the exposure of interest is distorted because the effect of an extraneous factor is mistaken for or mixed with the actual exposure effect" (p. 120). Statistical methods for controlling confounding have been clearly described (Kleinbaum et al., 1982; Rothman & Greenland, 1998), so that if the source of confounding, i.e., the extraneous factor of interest, can be measured accurately, then statistical tools can be used to minimize or eliminate its impact. The potential for confounding is inherent in observational studies where the exposure of interest is often only one of many correlated exposures. In contrast to studies where exposure is assigned randomly and thus isolated from other exposures, the potential for confounding cannot be readily quantified (Greenland, 1990). Nonetheless, in evaluating a measure of association from an observational study, we must judge how likely it is that the association has been distorted by confounding. What are the implications of a lack of knowledge of major risk factors for the disease (and thus candidate confounders) versus having measured and controlled for known strong risk factors? How likely is it that poor measurement of potential confounders has left substantial distortion in the association of interest? How effective is adjustment for indirect markers of potential confounders, such as educational attainment as a proxy for health behaviors? What is the likely direction of confounding, if present? What magnitude of association with disease or exposure would be required to have introduced a given amount of bias? How can we use other studies of the same association to help judge whether confounding is likely to be present?

The product of a careful evaluation of the study itself, drawing on the relevant methodological and substantive literature, is an informed judgment about the plausibility, direction, and strength of confounding, as well as specifying further research that would narrow uncertainty about the impact of confounding. Even when agreement among evaluators cannot be attained, the areas of disagreement should move from global questions about study validity to successively narrower questions that are amenable to empirical evaluation. To move from general disagreement about the credibility of a study's findings to asking such focused, answerable questions as whether a specific potential confounder has a sufficiently large association with disease to have markedly distorted the study results represents real progress. The methodologic principles are needed to refine the questions that give rise to uncertainty and controversy, which must then be integrated with substantive knowledge about the phenomenon of interest to make informed judgments. Much of this book focuses on providing that linkage between methodological principles and substantive knowledge in order to evaluate findings more accurately.

The challenges in interpretation relate to sets of studies of a given topic as well as individual results. For example, consistency across studies is often interpreted as a simple dichotomy: consistency in findings is supportive of a causal association and inconsistency is counter to it. But epidemiologic studies are rarely, if ever, pure replications of one another, and thus differ for reasons other than random error. When studies that have different methodologic features yield similar results, it can be tentatively assumed that the potential biases associated with those aspects of the study that differ have not introduced bias, and a causal inference is thus strengthened. Consistency across studies with features that should, under a causal hypothesis, yield different results suggests that some bias may be operating. There may also be meaningful differences in results from similarly designed studies conducted in different populations, suggesting that some important cofactors, necessary for the exposure to exert its impact, are present in some populations but not others. Clearly, when methodologically stronger studies produce different results than weaker ones, lack of consistency in results does not argue against causality.

The book has been organized to the extent possible in the order that issues arise. Chapter 2 sets the stage for evaluating epidemiologic evidence by clarifying the expected product of epidemiologic research, defining the goals. Next, I propose an overall strategy and philosophy for considering the quality of epidemiologic research findings (Chapter 3). The following chapters systematically cover the design, conduct, and analysis issues that bear on study interpretation.

Beginning with Chapter 4 and continuing through Chapter 9, sources of systematic error in epidemiologic studies are examined. The rationale for dividing the topics as the table of contents does warrants a brief explanation. Selection bias refers to "error due to systematic differences in characteristics between those who take part in a study and those who do not" (Last, 2001). It is the constitu-

tion of the study groups that is the potential source of such error. The construction of study groups is different in practice (though not in theory) in cohort and case–control studies. In cohort studies, groups with differing exposure status are identified and monitored for the occurrence of disease in order to compare disease incidence among them. In case–control studies, the sampling is based on health outcome; those who have experienced the disease of interest are compared to a sample from the population that gave rise to those cases of disease. Because the groups are constituted in different ways for different purposes in the two designs, the potential for selection bias is considered in separate chapters (Chapters 4 and 5). Furthermore, given that one particular source of selection bias, that due to non-participation, is so ubiquitous and often so large, Chapter 6 addresses this problem in detail.

Confounding, in which there is a mixing of effects from multiple exposures, is similar in many respects to selection bias, but its origins are natural as opposed to arising from the way the study groups were constituted. Evaluating the presence, magnitude, and direction of confounding is the subject of Chapter 7.

The consideration of measurement error, though algebraically similar regardless of what is being measured, is conceptually different and has different implications depending on whether the exposure, broadly defined as the potential causal factor of interest, the disease, again broadly defined as the health outcome of interest, is measured with error. The processes by which error arises (e.g., memory errors producing exposure misclassification, diagnostic errors producing disease misclassification) and their implications for bias in measures of association make it necessary to separate the discussions of exposure (Chapter 8) and disease (Chapter 9) misclassification.

The complex topic of random error, how it arises, affects study results, and should be characterized is addressed in Chapter 10. The sequence is intentional. Random error is discussed after the other factors to help counter the long-held view that it is the first or automatically the more important issue to consider in evaluating epidemiologic evidence.

There are several increasingly popular approaches to the integration of information from multiple studies, such as meta-analysis [defined as "a statistical synthesis of the data from separate by similar studies" (Last, 2001)] and pooling (combining data from multiple studies for reanalysis). These methods are discussed in Chapter 11. Chapter 12 deals with the integration and summary of information gained from the approaches covered in the previous chapters.

REFERENCES

Clayton D, Hills M. Statistical models in epidemiology. Oxford, England: Oxford University Press, 1993.

Greenland S. Randomization, statistics, and causal inference. Epidemiology 1990;1: 421–429.

Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. Int J Epidemiol 1986;15:413–419.

Hill AB. The environment and disease: association or causation? Proc R Soc Med 1965;58:295–300.

Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research. Principles and quantitative methods. Belmont, CA: Lifetime Learning Publications, 1982:312–319.

Last JM. A dictionary of epidemiology, Fourth Edition. New York: Oxford University Press, 2001.

Rothman KJ, Greenland S. Modern epidemiology, Second Edition. Philadelphia: Lippincott—Raven, 1998.

Selvin S. Statistical analysis of epidemiologic data. New York: Oxford University Press, 1991.

Weinberg CR. Toward a clearer definition of confounding. Am J Epidemiol 1993;137:1–8.

2

THE NATURE OF EPIDEMIOLOGIC EVIDENCE

GOALS OF EPIDEMIOLOGIC RESEARCH

To evaluate the quality or strength of epidemiologic evidence, we first need to clarify what information we can expect epidemiology to provide (Table 2.1). The effectiveness of epidemiologic research must be defined in relation to attainable, specific benchmarks in order to make judgments about how close the evidence comes to reaching the desired state of perfection. When we examine a study or set of studies, we typically make a statement as to whether those studies have individually or collectively fulfilled their expectations. Broad assessments such as "persuasive" or "inconclusive" are commonly applied to evidence, typically without much consideration of the standard against which it has been judged. It would be unwise to set an unrealistically high goal for epidemiologic research to yield absolute truth that directly benefits society, and equally unproductive to set a low (but readily attained) standard of merely providing clues or suggestions that will be resolved by other scientific approaches. Epidemiology functions somewhere between those extremes.

The highest expectation for epidemiologic research is to generate knowledge that contributes directly to improvements in the health of human populations. Such research would yield new knowledge, and that new knowledge would have beneficial applications to advancing public health. However appropriate this is

此为试读,需要完整PDF请访问: www.ertongbook.com