Secondary Data Sources for Public Health

A Practical Guide

Secondary data play an increasingly important role in epidemiology and public health research and practice; examples of secondary data sources include national surveys such as the BRFSS and NHIS, claims data for the Medicare and Medicaid systems, and public vital statistics records. Although a wealth of secondary data is available, it is not always easy to locate and access appropriate data to address a research or policy question.

This practical guide circumvents these difficulties by providing an introduction to secondary data and issues specific to its management and analysis, followed by an enumeration of major sources of secondary data in the United States. Entries for each data source include the principal focus of the data, years for which it is available, history and methodology of the data collection process, and information about how to access the data and supporting materials, including relevant details about file structure and format.

Sarah Boslaugh received her PhD from the City University of New York and her MPH from Saint Louis University. She is currently a Performance Research Analyst for BJC Healthcare in Saint Louis and an Adjunct Professor at the Washington University School of Medicine. She previously worked as a biostatistician and methodologist at Montefiore Medical Center in New York City, Saint Louis University School of Public Health, and the Washington University School of Medicine in Saint Louis. She has also written *An Intermediate Guide to SPSS Programming: Using Syntax for Data Management* (2004) and is Editor-in-Chief of the *Encyclopedia of Epidemiology* (2007).

Practical Guides to Biostatistics and Epidemiology

This is a series of short and practical but authoritative books for biomedical researchers, clinical investigators, public health researchers, epidemiologists, and nonacademic and consulting biostatisticians who work with data from biomedical and epidemiological and genetic studies. Some books are explorations of a modern statistical method and its application; others focus on a particular disease or condition and the statistical techniques most commonly used in studying it.

This series is for people who use statistics to answer specific research questions. The books explain the application of techniques, specifically the use of computational tools, and emphasize the interpretation of results, not the underlying mathematical and statistical theory.

Published in the series
Applied Multilevel Analysis, by **Jos W. R. Twisk**

Secondary Data Sources for Public Health

A Practical Guide

Sarah Boslaugh
BJC Healthcare

CAMBRIDGE UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment, a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9780521690232

First published 2007

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging-in-Publication data
Boslaugh, Sarah.
Secondary data sources for public health : a practical guide / Sarah Boslaugh.
p. ; cm. – (Practical guides to biostatistics and epidemiology)
Includes bibliographical references and index.
ISBN-13: 978-0-521-87001-6 (hardback)
ISBN-10: 0-521-87001-1 (hardback)
ISBN-13: 978-0-521-69023-2 (pbk.)
ISBN-10: 0-521-69023-4 (pbk.)
1. Public health – Research – Statistical methods. 2. Epidemiology – Research – Statistical methods. I. Title. II. Series.
[DNLM: 1. Data Collection – United States. 2. Epidemiology – United States. 3. Public Health – United States.WA 950 B743s 2007]
RA409.B66 2007
362.1072′7–dc22 2006034301

ISBN 978-0-521-87001-6 Hardback
ISBN 978-0-521-69023-2 Paperback

Contents

Preface

Secondary data analysis – meaning, in the broadest sense, analysis of data collected by someone else – plays a vital role in modern epidemiology and public health research and practice. This is partly because of the emphasis on population-based studies that is common to both fields. For instance, few individual researchers could hope to collect data sufficient to evaluate changes in the health status or health behaviors on a national scale. Fortunately, a wealth of data on health and related subjects, collected on a broad scale and over many years, is available for public use. However, locating secondary data appropriate to address a particular research question is not always easy, partly because an abundance of data is available and also because those data were collected by many different entities and are stored in many different locations. My primary purpose in writing *Secondary Data Sources for Public Health* is to facilitate use of those data sets in epidemiologic and public health research.

Chapter 1 introduces the topic of secondary data analysis, discusses some of its advantages and disadvantages, describes a general process for locating appropriate data to address a research question, and suggests some types of information that the researcher should try to acquire about any secondary data set being considered for analysis. Chapters 2 through 7 discuss the major secondary data sets and data archives available for studying health issues in the United States. These chapters are organized thematically, so Chapter 2 discusses health service utilization data; Chapter 3, health behaviors and risk factors data; Chapter 4, data sets dealing with multiple health topics; Chapter 5, fertility and mortality data; Chapter 6, Medicare and Medicaid data; and Chapter 7, other sources of data. The bibliography is organized by chapter and lists a

number of works, primarily theoretical and methodologic, relating to secondary data analysis and the data sets discussed. Appendix I lists the acronyms used in this volume, with the full name of the entity referred to and, if applicable, places the acronym in context. For instance, a term may be used primarily in conjunction with a particular data set, or a data set may be part of a larger project. Appendix II summarizes the data sets discussed in this volume, including the years for which data are available. Appendix III discusses data import and transfer.

Acknowledgments

This book would not have been written without the assistance and support of many individuals. In particular, I thank Elena Andresen for introducing me to secondary data analysis when I was a student at the Saint Louis University School of Public Health and for her steadfast belief in my abilities; Rand Ross at Washington University for helping me preserve my sanity; Neil Salkind, my agent at Studio B, for his unflagging support; Lauren Cowles, my editor at Cambridge University Press, for her patience and encouragement; and my husband, Dan Peck, for being there through it all.