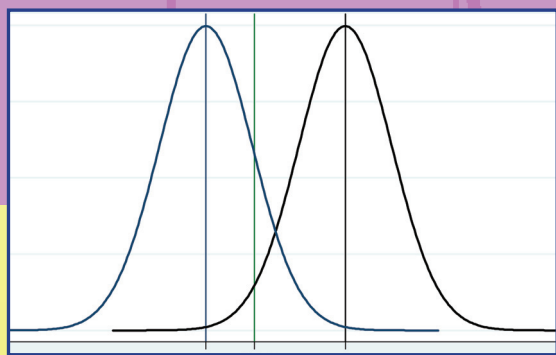
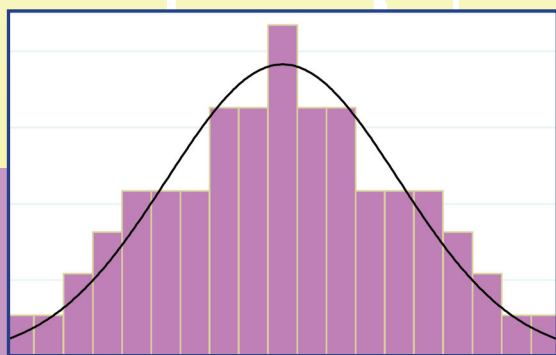


STATISTICS for ADVANCED PRACTICE NURSES and HEALTH PROFESSIONALS



MANFRED STOMMEL | KATHERINE J. DONTJE

Statistics for Advanced Practice Nurses and Health Professionals

Manfred Stommel, PhD, is a professor in the College of Nursing, Michigan State University. Dr. Stommel's recent research program has centered on the body mass index (BMI), its measurement in surveys, as well as antecedents and consequences of variations in the BMI. In addition, he has focused on the mortality and morbidity impacts of obesity and physical activity among different population groups. As the 2008/2009 Academy Health Senior Service Fellow at the CDC/National Center for Health Statistics, he specialized in the analysis of federal health surveys, such as the National Health Interview Survey (NHIS) and the National Health and Nutrition Examination Survey (NHANES) linked to the National Death Index. Dr. Stommel has published over 100 peer-reviewed papers and a book on clinical research methods. He is a regular reviewer for numerous medical and nursing research journals.

Katherine J. Dontje, PhD, FNP-BC, is director of the graduate clinical programs and assistant professor/nurse practitioner in the College of Nursing, Michigan State University. Dr. Dontje teaches in the doctorate of nursing practice and masters programs. She has an active clinical practice encompassing the full scope of primary care. Dr. Dontje has been principal investigator (PI) or co-PI on 12 research projects, including a W.K. Kellogg Foundation-funded project on university-based nurse-managed care and a Robert Wood Johnson Foundation-funded project on the quality of outcomes from three models of nurse-led primary care. She is presently a member of an interdisciplinary research team focusing on patient and provider aspects of the electronic health record. She has published numerous peer-reviewed articles on informatics and evidence-based practice for patients in primary care. Dr. Dontje has presented at regional, national, and international conferences. She has received numerous honors and awards, is a member of American Nurses Association (ANA), American Association of Nurse Practitioners (AANP), Midwest Nursing Research Society (MNRS), Sigma Theta Tau International (STTI), and the Michigan Council of Nurse Practitioners, and has served in leadership capacity for many of them. She serves as a reviewer for *Computers, Informatics, Nursing* and *The Journal for Nurse Practitioners*.

Statistics for Advanced Practice Nurses and Health Professionals

Manfred Stommel, PhD
Katherine J. Dontje, PhD, FNP-BC

 **SPRINGER PUBLISHING COMPANY**
NEW YORK

Copyright © 2014 Springer Publishing Company, LLC

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 Springer Publishing Company, LLC, or authorization through payment of the appropriate fees to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400, fax 978-646-8600, info@copyright.com or on the Web at www.copyright.com.

Springer Publishing Company, LLC
11 West 42nd Street
New York, NY 10036
www.springerpub.com

Acquisitions Editor: Margaret Zuccarini
Composition: Exeter Premedia Services Private Ltd.

ISBN: 978-0-8261-9824-2
e-book ISBN: 978-0-8261-9825-9

Data sets for SPSS, SAS, STATA, and Excel are available at www.springerpub.com/stommel.supplements

14 15 16 17 / 5 4 3 2 1

The author and the publisher of this Work have made every effort to use sources believed to be reliable to provide information that is accurate and compatible with the standards generally accepted at the time of publication. Because medical science is continually advancing, our knowledge base continues to expand. Therefore, as new information becomes available, changes in procedures become necessary. We recommend that the reader always consult current research and specific institutional policies before performing any clinical procedure. The author and publisher shall not be liable for any special, consequential, or exemplary damages resulting, in whole or in part, from the readers' use of, or reliance on, the information contained in this book. The publisher has no responsibility for the persistence or accuracy of URLs for external or third-party Internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Library of Congress Cataloging-in-Publication Data

Stommel, Manfred, author.

Statistics for advanced practice nurses and health professionals / Manfred Stommel, Katherine J. Dontje.

p. ; cm.

Includes bibliographical references.

ISBN 978-0-8261-9824-2—ISBN 978-0-8261-9825-9 (e-book)

I. Dontje, Katherine J., author. II. Title.

[DNLM: 1. Statistics as Topic—Nurses' Instruction. 2. Clinical Nursing Research—methods.

3. Nurse Practitioners. WA 950]

RT68

610.73072'7—dc23

2014002661

Special discounts on bulk quantities of our books are available to corporations, professional associations, pharmaceutical companies, health care organizations, and other qualifying groups. If you are interested in a custom book, including chapters from more than one of our titles, we can provide that service as well.

For details, please contact:

Special Sales Department, Springer Publishing Company, LLC
11 West 42nd Street, 15th Floor, New York, NY 10036-8002
Phone: 877-687-7476 or 212-431-4370; Fax: 212-941-7842
E-mail: sales@springerpub.com

Printed in the United States of America by McNaughton & Gunn.

Contents

Preface vii

Acknowledgments ix

Share Statistics for Advanced Practice Nurses and Health Professionals

PART I: FOUNDATIONS FOR STATISTICAL THINKING

1. Introduction: The Role of Statistics in Research and Clinical Practice 1
2. Properties of Variables: Levels of Measurement 7
3. Descriptive Univariate Statistics 13
4. Probabilities 29
5. Logic of Statistical Inference: The Sampling Distribution and Significance Tests 39
6. Standard Errors, Confidence Intervals, and the Power of Statistical Tests 53
7. Research Designs and Statistical Analysis 63

PART II: MODELS FOR CONTINUOUS/INTERVAL-LEVEL OUTCOME MEASURES

8. *t*-Test 71
9. One-Way Analysis of Variance 89
10. Linear Regression and Pearson's *r* Correlation 101
11. Factorial Analysis of Variance and Analysis of Covariance 119
12. Multiple Linear Regression 133
13. Repeated-Measures Analysis of Variance 149
14. Introduction to Mixed-Effects Regression Models 161

PART III: MODELS FOR CATEGORICAL OUTCOME MEASURES

15. Nonparametric/Ordinal Statistics 169
16. Frequency Cross-Tabulations: 2×2 Tables 185

17. Logistic Regression With One Independent Variable 199

18. Logistic Regression Models With Multiple Predictors 215

PART IV: MODELS FOR TIME-TO-EVENT DATA/SURVIVAL ANALYSIS

19. Incidence Rates, Life Tables, and Survival Function 231

20. Comparing Survival Functions in Different Groups and Hazard Regression 241

PART V: MEASUREMENT MODELS

21. Reliability Coefficients and Medical Test Evaluation 253

22. Factor Analysis 277

PART VI: ISSUES IN DATA MANAGEMENT

23. Data Management and Privacy Concerns 297

APPENDICES

A. Estimating Population Variance From Sample Variance 303

B. One-Sided Probabilities for z -Scores of the Standard Normal Distribution 305

C. Table of Critical t -Values for Several Significance Levels of t -Distributions With Different Degrees of Freedom (df) 307

D. Normalizing a Nonnormal Distribution 309

E. Table of Critical f -Values at the Significance Level of $\alpha = 0.05$ of f -Distributions With Different Degrees of Freedom (df) 311

F. Proof That $\Phi = \text{Pearson's } r$ 315

G. Table of Critical Chi-Square Values for Several Significance Levels of Chi-Square Distributions With Different Degrees of Freedom (df) 317

H. Refresher on Exponential and Logarithmic Transformations 319

I. Standardization of Interval-Level Variables 323

J. Answers to Selected Exercises 325

Index 333

Preface

This book has been written with a view of what the future practice of graduates from Doctor of Nursing Practice (DNP) programs and other advanced health care providers looks like. In particular, graduates of DNP programs and other health care providers need to be advocates of using the best research evidence available to facilitate both practice and system changes. National organizations of many health professionals have emphasized the importance of using the best evidence to improve health care of individuals, decrease the cost of care, and prevent errors across health care systems. All of this requires a solid grasp of statistical reasoning, which underlies much of the empirical information presented in research journals. In this book, our aim is to provide the reader with more than an introductory level comprehension of statistics. In particular, our emphasis is on *understanding* the most commonly encountered statistical models in the research journals. We believe that the “cookbook approach” to statistics, consisting of the memorization of formulas and rules, is not really helpful, as the correct interpretation of statistical tests and models requires an understanding of the underlying logic of the models employed.

The information provided in this book is divided into five parts, covering basic statistical reasoning and four different classes of statistical models. Part I covers the principles of statistical inference in clinical trials and observational studies, reasons for why we use statistical testing, and how we use it in the context of different research designs, as well as an overview of the basic descriptive statistics. Part II discusses statistical models used with continuous and interval-level outcome variables, which include *t*-tests, linear regression, analysis of variance, and some extensions of these models. Part III addresses statistical tests and models appropriate for categorical outcome variables. Part IV explores the use of time-to-event or survival analysis, which are often used in clinical research. Part V provides an overview of measurement models with an emphasis on reliability and validity of self-report and medical test data. In all chapters, we used examples relevant to clinical practice to provide information on how to use and interpret each of the statistical analysis models introduced. Exercise questions at the end of each chapter, and selected answers at the end of the book, serve the purpose of deepening the understanding. The book can be used as a stand-alone text for those readers primarily interested in understanding the models, but we offer some data sets in SPSS, STATA, SAS, and Excel formats on an accompanying website (www.springerpub.com/stommel.supplements) for those who want to engage in applied analysis themselves. The website also contains additional exercise questions and solutions.

The health care provider of the future needs to understand how to read statistical research and evaluate the quality of the research. Given the ever-increasing sophistication of statistical analyses in health care journals, it is important that clinicians acquire a level of understanding that enables them to interpret the results of research studies correctly and translate this information into practice. Readers of this book will be able to accomplish these tasks as well as to choose appropriate statistical methods for their own translational research. At least, readers should be able to recognize when more sophisticated analyses are necessary, and should have sufficient understanding of statistical reasoning to engage and converse with a statistical expert when needed. One important aspect of the translation model is to apply the best evidence available about clinical conditions in real-life clinical settings. The statistical methods described in this book will help health care providers to evaluate outcomes of quality improvement projects and system changes to determine the effectiveness of the evidence within their own clinical population.

Finally, we sincerely hope that readers find the information in this book useful and actually grow excited about the contributions that statistics can make to health care. Statistical reasoning provides a different way of looking at the world, which is particularly helpful when thinking about the health of populations or the evaluation of health care systems. In short, we believe that the study of statistics does not only consist of the acquisition of techniques and tools, however necessary, but opens up new ways of thinking about health-related problems. If we succeed in conveying both the ideas and our enthusiasm for statistical analysis, we shall have accomplished our goals.

Manfred Stommel, PhD
Katherine J. Dontje, PhD, FNP-BC

Acknowledgments

Textbooks do not get written by the authors alone, they also require support from many other persons. First, we would like to acknowledge Margaret Zuccarini at Springer Publishing Company, LLC, whose mix of flexibility and encouragement was just the right medicine to keep the authors on track. Dr. Stommel would also like to express his thanks to Dean Mary Mundt for her support and patience during the long period of textbook writing. Last, but not least, Dr. Stommel expresses his deep gratitude to his wife, Dr. Petra von Heimburg, for enduring many lost evenings, when he was “holed up” in the office. Dr. Dontje would like to acknowledge the encouragement of Dr. Barbara Given over the years and of Dr. Teresa Wehrwein throughout this project.

VISIT THIS BOOK'S WEB PAGE [↗](#)

BUY NOW [↗](#)

REQUEST AN EXAM OR REVIEW COPY [↗](#)

Share

**Statistics for Advanced Practice
Nurses and Health Professionals**



PART I. FOUNDATIONS FOR STATISTICAL THINKING

CHAPTER

1

Introduction: The Role of Statistics in Research and Clinical Practice

Many students in clinically oriented degree programs wonder why they are required to take statistics courses. There are several ways to answer this question, but it is probably best to start with a clinical example.

Suppose that you have a patient who has coronary artery disease and occasional episodes of angina, but only when exercising heavily. The patient comes to you and asks whether he should have a cardiac catheterization and possible stents put in with the procedure. How would you go about evaluating this? There are, of course, numerous different opinions on the subject, and the patient has friends who are encouraging him to get the procedure, as these friends had four to five stents implanted and reportedly are doing well. You search the literature and find that, for stable coronary artery disease, there are three main choices of treatment (medical therapy, angioplasty, or bypass surgery); however, the statistics indicate that, among stable angina cases, none of these have any significantly greater long-term benefit in terms of length of life or recurrence of heart attacks over the other (Boden et al., 2007; Hueb et al., 2004). As you review the studies, you are confronted with a variety of statistics, but how do you explain these convincingly to your patient, when “common sense” appears to suggest that opening up an artery should “save his life”?

Probably the first thing to emphasize would be that it is not enough to cite a few individual cases for whom a particular clinical intervention appears to have been successful. One reason we cannot rely on results from individual cases is that what works in one case may or may not work in another case. Human beings almost always show a range of responses to a given nursing or medical intervention. Thus the question arises: How do we then decide which treatment or intervention is better or worse? It turns out that statistics alone cannot answer this question either. The inference that an intervention is causally effective is also based on the quality of the research design of the intervention study. But statistical considerations are an essential aspect of how to design an effective intervention study that can answer the desired question.

Given the almost infinite variability of human responses to clinical treatments and interventions, we need a method by which we can separate “accidental” individual variability from *systematic*, treatment-related effects. As we will see, statistical models can be used to estimate *average* effects of interventions as well as provide information on the amount of uncertainty or relative certainty that must be attached to these estimates.

We use statistics not only in the evaluation of clinical interventions or treatments but also to generalize from the evidence obtained in a study sample to the target populations of interest. For example, data from the National Health Interview Survey (NHIS) conducted by the National Center for Health Statistics (NCHS) linked to the National Death Index have been employed to estimate the effects of adherence to the 2008 Physical Activity Guidelines (Centers for Disease Control and Prevention [CDC]) on the subsequent mortality risk of adult U.S. residents (Schoenborn & Stommel, 2011). In such studies, we use sample data to draw inferences about health conditions in larger target populations. While the aim is to describe patterns of mortality, morbidity, and health behaviors in the target population, we use surveys that employ random/probability sampling designs that, in some sense, “represent” or “reflect” the characteristics of the target population with sufficiently high accuracy.

Statistics are, in fact, all-pervasive in today's health care systems. We use it to evaluate the performance of medical interventions, the effectiveness of screening programs, to gauge quality improvement projects in health care delivery, to assess the performance of nursing students on the NCLEX or certification exams, to establish critical test scores that should trigger nursing or medical interventions, and so forth. Statistical evidence plays a major role in judging not only the quality of care delivered, but also its cost effectiveness. Statistics are also an important tool for providers and public health officials, as they engage in assessments of how well they are doing compared to other health delivery organizations, or compared to benchmarks derived from nationwide or statewide studies. Last, but not least, the evidence to support and evaluate clinical practice guidelines or guidelines for healthy behaviors (e.g., the CDC 2008 Physical Activity Guidelines) is grounded in statistical information.

By contrast, consider personal experience. In a way, all of us are “reckless generalizers” in our personal lives. We all believe we have an idea of what “human nature” is like, even though we get to know well only a few dozen individuals in a lifetime (and they are decidedly not a representative cross-section of the human race!). Similarly, from the very limited experience we have as patients with our primary care providers, dentists, or nurses, we draw inferences about their quality as providers. Suppose a patient with a diphtheria infection is misdiagnosed by her provider as having mononucleosis (easy to do in the initial phases of these diseases) and receives the wrong treatment, should she conclude that the provider is “incompetent?” Suppose you have evidence that the provider in question made such a diagnostic mistake only once in 25 years of practice, while another provider accumulated a long list of complaints for misdiagnoses. As this example shows, we cannot make credible inferences based on a single event; we need large amounts of data to discover a *pattern* of behavior. Hence, we need statistics. It allows us to distinguish among isolated events (“outliers”), systematic patterns of events (average differences or “effects”), and events whose occurrence cannot be predicted (“random errors”). All of this information is needed to evaluate outcomes of interest to health care providers.

Let us look at this a bit further. Not only are no two individuals exactly alike in terms of their biological characteristics and life experiences; as living organisms they are also subject to continual change over time. For example, as every nurse knows, people's “true” diastolic blood pressure (DBP) fluctuates, even during short time periods, before and after a meal, and so forth. On top of that, there are measurement errors associated with any clinical measure you can think of: For instance, blood pressure (BP) measures vary depending on whether the cuff is applied to the right or left arm, whether the cuff is more or less pressurized, whether the patient has more or less muscle tissue, and so forth. Similarly, any body temperature measure varies based on where the thermometer is applied (under the armpit, the tongue, etc.), and

any nurse, who has ever tried to establish the height of an infant, knows that it is impossible to get the “true” value. In short, uncertainty is part of everything we do, whether to estimate the likely survival of a patient with a recent Stage 4 lung cancer diagnosis or the recovery time after a triple bypass operation. Thus, we need to have realistic estimates of the uncertainty attached to our predictions, so that we can make rational decisions about which clinical interventions are better or worse. Statistical methods do just that. You might say statistics is the branch of mathematics that puts uncertainty (and probability) at the center of its models. It allows us to estimate the risks we engage in when we make informed decisions. It is for that reason that statistics has become a central part of clinical reality.

Finally, there is also a very pragmatic reason why providers have to become savvy in the evaluation of statistical information. On a daily basis, many of our patients follow mass media reports on health-related topics or search the Internet to get information about a disease or illness they might have or a medical treatment that they believe they might need. Such reports very often cite statistics from clinical trials or epidemiological studies. Certainly, advertisements for pharmaceutical drugs or claims on highway posters that this or that hospital is in the “top 100” for knee surgery, and so forth, all tend to cite statistics that may or may not be relevant to the claims involved.

For users of statistics, the most important issue has become how to evaluate all this statistical information and how to make intelligent choices based on it. For current and future clinicians, an additional problem is that the statistical information in medical and nursing research journals is becoming ever more sophisticated: just witness a special issue of *Nursing Research* (Volume 61(3), 2012) entirely devoted to newer statistical models used by nursing researchers. Yet knowing more about statistics is essential for clinicians to understand and interpret clinically relevant evidence. While today's clinicians do need a solid conceptual grounding in major statistical concepts, they do not need to know the particular mathematical structures of the major estimation techniques, for example, least squares, maximum likelihood, or partial likelihood estimation that underlie many statistical models (all of which require some knowledge of calculus). Instead, clinicians do need to understand the basic *logic* of statistical estimation, basic probability concepts, and how inferential statistical decisions are made. To draw correct inferences from statistical information also includes awareness that statistics are an integral part of the research design of a study. To take just one example: how one interprets the outcome of a *t*-test differs, depending on whether or not the data come from an experimental study with random assignment, a survey based on random sampling, or whether the data are cross-sectional or longitudinal. Thus, statistical evidence cannot be interpreted without knowing something about the study design context, the sampling design, as well as how measurement error can affect the results.

When an advertisement claims that a certain drug reduces bleeding by 35%, one should immediately ask how that figure was established. Without context, such a number is virtually impossible to interpret. What is the reference group compared to which the 35% reduction was observed? How large a sample of men/women was studied? Was the sample representative of the U.S. population at large or only of certain segments, for example, only women, only Whites, only persons younger than 40 or older than 65, or persons with a particular disease, and so forth? Were there measurement problems? Could the study show a causal connection between the drug or treatment and the outcome? Are there alternative treatments that are even better? Does the treatment work only under certain conditions, and so forth? Many of these questions involve statistical reasoning and statistical methods; so it is no accident that statistics has become a major component of the education and training of future clinicians.

This book is intended as a textbook for graduate nursing students and others who are preparing for advanced clinical practice roles. As it is addressed to clinicians, we present the major statistical models most often encountered in nursing, medical, and epidemiological research. Past statistics textbooks for nursing students have almost exclusively focused on statistical models derived from psychometrics and educational research, but have neglected models that should be of primary interest to advanced practice nurses, such as life table and survival analysis as well as the evaluation of diagnostic and screening tests. This book is intended to close this gap without sacrificing the more traditional topics from psychometrics (e.g., reliability and factor analysis), which continue to be relevant, particularly for behaviorally oriented nursing research. The overall goal of this book is to provide the learner with a larger range of statistical models and concepts, which are used in clinical research and as a basis for evidence-based clinical practice. This includes skills in “mining” data sets that are needed for the clinical management of patient populations.

Despite our goal of providing the reader with an introduction to more complicated statistical models than the t -test, analysis of variance (ANOVA), and linear regression, the presentation of the material does not require any calculus or matrix algebra, but relies on some knowledge of college algebra. Because students may have forgotten how to read and use exponential and logarithmic functions, a brief refresher has been provided in Appendix H to facilitate the discussion of logistic and survival regression models. On the whole, the emphasis in this book is on verbal explanations, and the use of worked out examples to explain the more complicated ideas of statistical inference. However, statistics is an inherently mathematical subject and it cannot be learned by completely shying away from mathematics. Neither can the science of nursing research be understood without mathematics (Henly, 2012). Some statistical formulas (e.g., standard deviations, covariance/correlations, odds ratios) are essential in understanding the material. Thus, they are not only introduced, but also accompanied by detailed verbal explanations. In addition, exercise questions at the end of each chapter (selected answers are provided in Appendix J) will provide opportunities to become familiar and comfortable with using such formulas and interpreting them correctly.

Our experience in teaching statistics for nurses at the graduate level shows that some students are apprehensive about taking a statistics course. We believe that part of the reason for this is that too many students were taught introductory statistics with an emphasis on memorizing formulas, but never really understood what the subject was about and how it relates to clinical practice. The information you will find in this book is designed to give you a solid understanding of the statistical methods we introduce. We believe that memorizing formulas alone does not really give you an understanding of the value of statistical analysis, and besides, formulas will be forgotten as soon as the course is over. In this book you will have the opportunity to learn about statistical reasoning and how it can provide the clinician with the contextual information necessary to make clinical decisions in particular instances. However, that requires understanding the conceptual basis for statistical methods, which, incidentally, also provides a much better aid to memory.

In sum, we hope we have made a convincing case for why statistics should be part of any clinician's tool kit, but more than that: We think of the subject of statistics as an exciting field that can transform the way you look at health care and clinical practice. If we succeed in changing students' outlook on statistics, it is our experience that it changes their outlook on clinical practice forever.

REFERENCES

- Boden, W. E., O'Rourke, R. A., Teo, K. K., Hartigan, P. M., Maron, D. J., Kostuk, W. J., . . . Weintraub, W. S.; COURAGE Trial Research Group. (2007). Optimal medical therapy with or without PCI for stable coronary disease. *The New England Journal of Medicine*, 356(15), 1503–1516.
- Centers for Disease Control and Prevention. (2008). *Physical Activity Guidelines Advisory Committee Report, 2008*. Washington, DC: USDHHS.
- Henly, S. J. (2012). Strength in numbers: Mathematics and nursing research. *Nursing Research*, 61(4), 241.
- Hueb, W., Soares, P. R., Gersh, B. J., César, L. A., Luz, P. L., Puig, L. B., . . . Ramires, J. A. (2004). The Medicine, Angioplasty, or Surgery Study (MASS-II): A randomized, controlled clinical trial of three therapeutic strategies for multivessel coronary artery disease: One-year results. *Journal of the American College of Cardiology*, 43(10), 1743–1751.
- Schoenborn, C. A., & Stommel, M. (2011). Adherence to the 2008 adult physical activity guidelines and mortality risk. *American Journal of Preventive Medicine*, 40(5), 514–521.