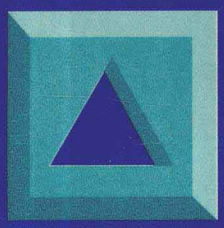
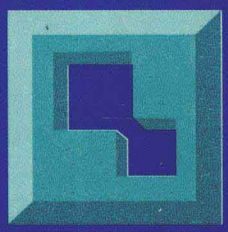
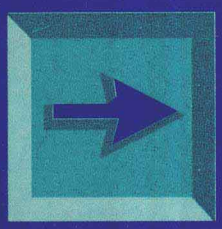
How To Report Statistics in Medicine

Annotated Guidelines for Authors, Editors, and Reviewers

医学统计数据报告指南







Thamas A.long
Michelle Secic

ACP SERIES MEDICAL WRITING AND COMMUNICATION

世界图书出版公司

How To Report Statistics in Medicine

ANNOTATED GUIDELINES FOR AUTHORS, EDITORS, AND REVIEWERS

Thomas A. Lang, MA

Manager, Medical Editing Services
Department of Scientific Publications
The Cleveland Clinic Foundation
Cleveland, Ohio

Michelle Secic, MS

Senior Biostatistician
Department of Biostatistics and Epidemiology
The Cleveland Clinic Foundation
Cleveland, Ohio

With a Foreword by
Edward J. Huth, MD, MACP
Editor Emeritus, Annals of Internal Medicine

American College of Physicians
Philadelphia, Pennsylvania

Acquisitions Editor: Mary K. Ruff
Manager, Book Publishing: David Myers
Administrator, Book Publishing: Diane M. McCabe
Production Supervisor: Allan S. Kleinberg
Production Editor: Victoria Hoenigke
Interior and Cover Design: DiDona Design

Printed in the United States of America.
Composition by Techsetters, Inc.
Printing/binding by Port City Press

American College of Physicians
Independence Mall West
Sixth Street at Race
Philadelphia, PA 19106-1572

Library of Congress Cataloging-in-Publication Data

Lang, Thomas A. (Thomas Allen)
How to report statistics in medicine: annotated guidelines for authors, editors, and reviewers / by Thomas A. Lang and Michelle Secic.
p. cm. – (Medical writing and communication)
Includes bibliographical references.
ISBN 0-943126-44-4
1. Medical statistics. 2. Medical writing. I. Secic, Michelle. 1967- . II. Series.
RA409.L357 1997
610'.72–dc20 96-30333
CIP

97 98 99 00 01 02 / 9 8 7 6 5 4 3 2

To Toni and Colin, for the real meaning of significance; and to Kathy F. Grupe, who long ago pointed a novice editor in the right direction.

TAL

To my close family and friends: John, for his constant love, support, and composure as my partner through every aspect of our lives; my father, Joseph, and my mother, Barbara, for their pride and encouragement; Joe, for his advice and leadership; Gina, for her kindness, sincerity, and concern; Dawn, for her loyalty and confidence; Rick, Jessi, and R.J., for the joy they have brought into my life; Mama and Tata, for always being there for me; Shari, for her guidance; and Dawn, Gina, Kim, and Roberta, for their unconditional friendship.

MS

A Note to the Reader

My books are water; those of the great geniuses are wine—everybody drinks water.

MARK TWAIN (1)

Both fine wines and biostatistics are characterized by complexities and subtleties that are truly appreciated only by the relatively few people who devote the time to master them. To these readers, we extend our apologies; this book was not written for you. Rather, it was written for a much larger group of readers: those who thirst for a basic understanding of statistics, but who do not aspire to appreciate the nuances. This is a book about reporting and interpreting statistical presentations, not about understanding theories of probabilities or mathematical concepts. This is a book for water drinkers.

It is exceedingly difficult to explain many statistical concepts in terms that are both technically accurate and easily understood by those with only a cursory knowledge of the topic. Thus, if our explanations do not include some of the finer points of a topic or if they have bypassed some distinction of meaning, it is because we believe that such fine points and distinctions would detract from an explanation otherwise adequate for most readers.

The medical examples in this book were created to illustrate statistical concepts. As such, the vast majority are hypothetical and should be accepted as teaching devices and not as medical fact.

1. Joseph M, editor. *Man is the Only Animal that Blushes . . . or Needs to. The Wisdom of Mark Twain*. New York: Random House; 1970.

FOREWORD

The need for quantitative evidence in medical judgments has been seen for at least two millennia. In the second century AD, Galen noted that

> [the empiricists] say that a thing seen but once cannot be accepted nor regarded as true, neither what was seen a few times only. They believe something can only be accepted and considered true, if it has been seen very many times, and in the same manner every time.

But for centuries this view was almost entirely ignored. Then, a bit more than 150 years ago, Pierre Charles-Alexandre Louis raised a closely related point:

> Let us bestow upon observation the care and time which it demands; let the facts be rigorously analyzed in order to a just appreciation of them; and it is impossible to attain this without classifying and counting them; and then therapeutics will advance not less than other branches of science.

Louis cannot be considered the father of medical statistics in its present state, but less than a century passed before the pioneers in modern biological and medical statistics—Karl Pearson and Ronald Fisher—began to set standards for quantitative evidence. The power of statistical methods became clearer and clearer during the following years. Adequate study design and statistical analysis began to lead to conclusions of great import for the health of all of us, as in the studies of Wynder and Graham and Doll and Hill on tobacco smoking and carcinoma of the lung. Today even the physician who knows nothing about statistical methods expects to find in reports of clinical trials of drugs statistical evidence for what they conclude.

Unfortunately, what passes before our eyes as statistical analysis and reporting does not always represent the proper use of statistical methods or the clear and adequate reporting of statistical findings. Editors of journals and their peer reviewers may catch statistical shortcomings in the papers they consider for publication, but the review system is not always infallible in judging statistical evidence and how it is presented. Hence, authors who know their duty to try to meet high standards of scientific

reporting must offer the strongest possible statistical evidence. However, this by itself is not enough; they must also strive to report this evidence clearly enough to convince even the most critical readers that the evidence is reliable and adequate. But up until now, authors have had available little published guidance in how to most effectively report their statistical data. Several biomedical style manuals have carried short sections on publication style for statistical data, but these have assumed that authors know enough about statistical reporting to do it clearly and convincingly. Now Lang and Secic, the authors of this manual, bring valuable specific and detailed help to authors who wish to make their papers as statistically convincing as possible. And authors are not the only persons who can profit from what they recommend. There are also the peer reviewers who may be uncertain of what standards for statistical reporting they should apply to papers they review. The same probably can be said of at least some editors who are ill informed on statistics and who are not fortunate enough to have statisticians on their staffs. The rest of the medical community—the readers of journals, other physicians, nurses—will eventually profit from Lang and Secic's advice when it is applied by authors and editors. And most important, there is the benefit that will trickle down in time to patients, who are the reason why our profession exists.

Edward J. Huth
Editor Emeritus, **Annals of Internal Medicine**

PREFACE

Standards governing the content and format of statistical aspects should be developed to guide authors in the preparation of manuscripts.

J.R. O'FALLON ET AL. (1)

Among the first physicians to have considered the implications of statistical probability in medical research is Donald Mainland, MD, of Dalhousie University, Halifax, Canada. He appears to have been the first to report statistics, in articles published in the *Canadian Medical Association Journal* and in the *British Medical Journal* in the 1930s (2,3). Since then, medical research has increasingly adopted the principles of experimental design and statistical analysis, with the result that biostatistics has emerged as a distinct field of study. Biostatistics has been essential in moving medical research from anecdotal case reports to experiments with control groups, and finally to the large-scale, randomized controlled trials that are now the preferred standard for scientific proof.

But there is a problem. Studies of the statistical quality of journal articles have consistently found high error rates in the application, reporting, and interpretation of statistical information, in even the most respected medical journals. Since the first such study—the earliest we found was published in 1959—error rates as high as 80% have been found, again, even in major medical journals (4–19). "These reviews [of statistical errors] reveal a remarkable and depressing consistency, with typically around fifty percent of reviewed papers being found to contain clear statistical errors" (20). Further, a large portion of these errors are so great as to cast doubts on the validity of the author's conclusions (6, 21).

At the same time, most of these errors are related to topics included in most introductory statistics books. It seems strange indeed that a problem seemingly so important, so widespread, and so long-standing should continue, despite apparently being so basic in nature.

Curiously, despite several calls to the contrary (1, 17, 20, 22–24), no comprehensive guidelines or reference books have been available to aid in statistical reporting. Several sets of general guidelines have been pub-

lished in biomedical journals (20, 25–30), but we believe that they are too general in nature, too limited in scope, and too specialized in vocabulary to be useful to most authors and editors. Obviously, statistical reporting errors will likely continue without the widespread adoption of suitable reporting guidelines.

Thus, our purpose in writing this book is to provide a set of detailed, comprehensive, and understandable guidelines for reporting statistical information in medicine. Further, we tried to make the guidelines more accessible to nonstatisticians through explanations and examples and by organizing the guidelines according to how they are used in a text rather than by the mathematical principles on which they are based.

As a result, this book is not a statistics book in the usual sense. We are not concerned with teaching research design, statistical theory or methods, or the calculations of statistical tests. We consider here only the presentation of statistical information in scientific publications and discuss some related concepts that should help put these presentations into perspective. We urge authors and researchers to collaborate with biostatisticians in all phases of research, but we also believe that one need not be a statistician to present or interpret elementary statistics correctly. One does need ready access to accurate, complete, and understandable information to do so, however. This book was written to help provide this access.

More than 60 years ago, the same Dr. Mainland who started it all expressed our hopes for the future of statistical reporting (2):

> . . . progress would be made if some fundamental ideas were more clearly understood, namely, that the principles underlying statistical methods are relatively simple, that the commonest methods are easy to learn, that the methods can be used as an instrument without a deep knowledge of their mathematical structure, that the methods do not impart a fictitious accuracy or an artificial quality to the results, and that these methods tend very often to show that conclusions are not so definite as the unaided observer would think they were. If these things were understood, the methods would be much more commonly used, and, more important still, workers would come to recognize when they should appeal to the statistician. This in turn would hasten the coming of the day when a consultant statistician will be considered necessary in every medical centre.

It is a truism of medical writing that in clarifying the meaning, we coincidentally reveal it. If our book will help clarify statistical analyses, it may also improve the way medical research is conducted and interpreted.

Tom Lang
Michelle Secic

REFERENCES

1. **O'Fallon JR, Duby SD, Salsburg DS, Edmonson JH, Soffer A, Colton T.** Should there be statistical guidelines for medical research papers? Biometrics, 1978;34(4):687-95.
2. **Mainland D.** Chance and the blood count. Can Med Assoc J. 1934;(June):656-8.
3. **Mainland D.** Problems of chance in clinical work. Br Med J. 1936;2:221-4.
4. **Hall JC, Hill D, Watts, JM.** Misuse of statistical methods in the Australasian surgical literature. Aust N Z J Surg. 1982;52:541-3.
5. **Schor S, Karten I.** Statistical evaluation of medical journal manuscripts. JAMA. 1966;195(13):1123-8.
6. **Glantz SA.** Biostatistics: how to detect, correct and prevent errors in the medical literature. Circulation. 1980;61(1):1-7.
7. **Lionel ND, Herxheimer A.** Assessing reports of therapeutic trials. BMJ. 1970;3(723):637-40.
8. **Altman DG.** Statistics in medical journals: developments in the 1980s. Stat Med. 1991;10(12):1897-913.
9. **White SJ.** Statistical errors in papers in the British Journal of Psychiatry. Br J Psychiatry. 1979;135:336-42.
10. **Gore SM, Jones IG, Rytter EC.** Misuse of statistical methods: critical assessment of articles in BMJ from January to March 1976. BMJ. 1977;1(6053):85-7.
11. **Freiman JA, Chalmers TC, Smith H Jr, Kuebler RR.** The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 negative trials. N Engl J Med. 1978;299(13):690-4.
12. **Reed JF, Slaichert W.** Statistical proof in inconclusive "negative" trials. Arch Intern Med. 1981;141(10):1307-10.
13. **Gardner MJ, Altman DG, Jones DR, Machin D.** Is the statistical assessment of papers submitted to the British Medical Journal effective? BMJ. 1983;286(6376):1485-8.
14. **MacArthur RD, Jackson GG.** An evaluation of the use of statistical methodology in the Journal of Infectious Diseases. J Infect Dis. 1984;149(3):349-54.
15. **Avram MJ, Shanks CA, Dykes MH, Ronai AK, Stiers WM.** Statistical methods in anesthesia articles: an evaluation of two American journals during two six-month periods. Anesth Analg. 1985;64(6):607-11.
16. **Godfrey K.** Comparing the means of several groups. N Engl J Med. 1985;313(23):1450-6.
17. **Pocock SJ, Hughes MD, Lee RJ.** Statistical problems in the reporting of clinical trials. A survey of three medical journals. N Engl J Med. 1987;317(7):426-32.
18. **Smith DG, Clemens J, Crede W, Harvey M, Gracely EJ.** Impact of multiple comparisons in randomized clinical trials. Am J Med. 1987;83(3):545-50.
19. **Gotzsche PC.** Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. Control Clin Trials. 1989;50(9):356.
20. **Murray GD.** Statistical aspects of research methodology. Br J Surg. 1991;78(7):777-81.
21. **Yancy JM.** Ten rules for reading clinical research reports [Editorial]. Am J Surg. 1990;159(6):553-9.
22. **Shott S.** Statistics in veterinary research. J Am Vet Med Assoc. 1985;187(2):138-41.
23. **Hayden GF.** Biostatistical trends in Pediatrics: implications for the future. Pediatrics. 1983;72(1):84-7.
24. **Altman DG, Bland JM.** Improving doctors' understanding of statistics. J R Statis Soc A. 1991;154:223-67.
25. **Altman DG, Gore SM, Gardner MJ, Pocock SJ.** Statistical guidelines for contributors to medical journals. BMJ. 1983; 286(6376):1489-93.
26. **International Committee of Medical Journal Editors.** Uniform requirements for manuscripts submitted to biomedical journals. N Engl J Med. 1991;324:424-8.
27. **Elenbaas RM, Elenbaas JK, Cuddy PG.** Evaluating the medical literature. Part II: Statistical analysis. Ann Emerg Med. 1983;

12(10):610-20.
28. **Murray GD.** Statistical guidelines for The British Journal of Surgery. Br J Surg. 1991; 78(7):782-4.
29. **Sumner D.** Lies, damned lies—or statistics? J Hypertens. 1992;10(1):3-8.
30. **Journal of Hypertension.** Statistical guidelines for the Journal of Hypertension. J Hypertens. 1992;10(1):6-8.

ACKNOWLEDGMENTS

We especially thank the following friends and colleagues for reviewing earlier drafts of our manuscript:

Daniel R. Browning, MA, MS
Computer-Assisted Reporting Specialist
Pioneer Press
Saint Paul, Minnesota

Michael Kutner, PhD
Chairman, Department of Biostatistics and Epidemiology
The Cleveland Clinic Foundation
Cleveland, Ohio

Ken Murray, MD, MS, MT (ASCP)
Medical Director, Independent Practice Association
Physician in Private Practice
Studio City, California
Clinical Assistant Professor of Family Medicine
University of Southern California
Los Angeles, California

Flauren Ricketts, MS
Instructor, Department of Mathematics
Normandale Community College
Bloomington, Minnesota

We owe a special debt of gratitude to Ed Huth, MD, MACP, who suggested that we submit our manuscript to the American College of Physicians. His encouragement of this project —which included agreeing to write the Foreword—was most welcome. We also thank Frank Davidoff, MD, FACP; R. Brian Haynes, MD, PhD, FRCP; Patricia Houston,

MD, MPH; Stephen Lock, MD; and Margaret Winker, MD, who graciously examined earlier drafts of the manuscript. Their comments, both verbal and written, were most appreciated.

Special thanks also to our colleagues in the Department of Biostatistics and Epidemiology, especially Dave P. Miller, MS; Nancy Obuchowski, PhD; Edward Mascha, MS; and Ming Tan, PhD, and in the Department of Scientific Publications, especially Cassandra Talerico, MA, and Jackie Werner. Finally, a number of people at the American College of Physicians deserve our sincere thanks for transforming our manuscript into a book: Diane M. McCabe, David Myers, Vicki Hoenigke, Allan Kleinberg, and Mary K. Ruff.

All of the people named above contributed substantially to the production of this book. Any errors that remain are, unfortunately, ours.

INTRODUCTION

The presentation of statistical guidelines should not be confused with statistical education.

S. L. GEORGE (1)

This book is not a statistics book in the usual sense. Rather, it is a guide to understanding and presenting statistical information. An earlier draft was titled *A Field Guide to the Statistical Flora and Fauna of the Biomedical Research Paper*. That is, we have written a book on wildflower identification and appreciation and not a book on plant biology, so to speak.

As indicated by its title, this book was written for authors, editors, and reviewers who prepare or evaluate biomedical research articles for publication. It should also be of value to students who are learning about biostatistics and medical research. It is divided into four parts:

Part 1, Annotated Guidelines for Reporting Statistical Information, is organized into 15 chapters that correspond to 15 general applications of biostatistics. These chapters were created to help nonstatisticians find the appropriate guidelines more easily. The guidelines themselves were derived from an extensive review of the clinical literature (see the Bibliography). The guidelines are usually accompanied by explanations, hypothetical examples, and cautions that aid in understanding, evaluating, and applying them correctly. Many guidelines are duplicated because we believe that they should be more readily available to readers than cross-referencing allows them to be. Authors and editors should refer to these guidelines when preparing manuscripts for publication.

We stress that these guidelines are just that: guidelines, not requirements.

Part 1 contains guidelines, subguidelines, cautions, methods of checking, and cross-references to related information. For example:


Subguidelines are given for special cases of the main guideline.


Potential problems identify possible reporting or interpretation problems associated with the guideline.


Methods of checking describe ways to verify or to question statistical presentations.


Related information cross-reference supplemental chapters, guidelines, tables, and figures within this text.

Part 2, Guide to Statistical Terms and Tests, is designed to help readers of scientific articles understand statistical information. Entries are descriptions of what the terms or concepts mean in the context of medical research; they are not intended to be mathematically or theoretically pure definitions. All are written to be understood by readers with only a basic knowledge of statistics. Readers who wish more detailed information should consult statistical texts.

Part 3, An Unannotated, Referenced List of Guidelines, is supplied for readers already familiar with statistical concepts, who will use the book primarily as a reference tool in writing, editing, and reviewing scientific articles.

Four appendices in Part 4 are included to assist in reporting: Appendix 1: Checklists for Reporting Clinical Trials; Appendix 2: Mathematical Symbols and Notation; Appendix 3: Rules for Presenting Numbers in Text; and Appendix 4: Spelling of Statistical Terms and Tests.

The bibliography contains the articles and books on which the guidelines are based.

The index uses a wide selection of terms and many cross-references to help readers find information.

REFERENCE

George SL. Statistics in medical journals: a survey of current policies and proposals for editors. Med Pediatr Oncol. 1985;13(2):109-12.

DIFFERENCES BETWEEN CLINICAL AND STATISTICAL SIGNIFICANCE

It has been said that a fellow with one leg frozen in ice and the other leg in boiling water is comfortable—on average.

J.M. YANCY (1)

One of the most common errors in reporting and interpreting medical research is the failure to distinguish between clinical and statistical significance. (Because in medical writing "significant" is reserved for its statistical meaning, we have used the phrase "clinical importance" throughout this book to refer to "clinical significance.") In general, a **clinically important finding** is a conclusion that has implications for patient care. A **statistically significant finding**, on the other hand, is a conclusion that there is evidence against the null hypothesis—that there is a low probability of getting a result as extreme or more extreme than the one observed in the data, given that the null hypothesis is true.

By itself, a statistically significant finding can have little to do with the practice of medicine. Similarly, a clinically important finding in a single case probably does not establish a biological relationship. A finding that is both clinically important and statistically significant is valuable because we are more likely to believe that the finding is the result of a biological process shared by a group of patients and that it is perhaps amenable to measurement, explanation, prediction, and control.

We call to your attention several aspects of the distinction between statistical significance and clinical importance:

1. Statistical significance essentially reflects the influence of chance on the outcome; clinical importance reflects the biological value of the outcome.

In general, *small differences between large groups* can be statistically sig-

nificant but clinically meaningless. A difference of 0.02 kg in the weights of two groups of adults is not likely to have any clinical importance even if such a difference would have occurred by chance less than 1 time in 100 ($P < 0.01$) or even less than 1 time in 100 000 ($P < 0.00001$).

It is also true that *large differences between small groups* can be clinically important but not statistically significant. In a study of 20 patients in which even 1 patient dies, the death is clinically important, whether or not it is statistically significant. The important question is whether the sample is large enough to detect a clinically important difference if, in fact, such a difference existed. This question is one of statistical power.

2. Statistics are derived from groups of individuals; medicine is practiced on specific individuals.

Because statistics is based on probability, not on biology, it is concerned with populations and not individual patients. Physicians who treat individual patients on the basis of medical research are in a real sense "playing the odds": They are hoping that what has been true for a group of similar patients will be true for one particular patient.

3. Statistical conclusions require adequate amounts of data to be valid; medical decisions must often be made with insufficient data.

Statistical comparisons involving small samples often have low statistical power. That is, researchers often do not collect enough information to be reasonably confident to conclude whether, say, a new treatment is as good as or better than the standard treatment. A study reporting a negative or statistically nonsignificant result for which the statistical power is low is actually not negative at all: it is inconclusive. For the same reason, when no statistically significant differences are found between the baseline values of small treatment and control groups, it is inappropriate to conclude that the groups are equivalent: absence of proof is not proof of absence.

4. Statistical answers are probabilistic; medical treatment requires committed decisions.

Statistics incorporates the notion of probability. When a result would be expected to occur by chance less than, say, 1 time in 1000 (that is, $P < 0.001$), the result nevertheless could be the result of chance; it is simply not probable that chance is the explanation. The result obtained in a sample is also an estimate of what might be expected to occur in the larger population. Although a 95% confidence interval provides a measure of precision for this estimated result, it, too, is a probabilistic statement and not a sure thing.

5. Statistical analysis always requires measurement; medicine sometimes requires intuition.

Science is measurement. Unfortunately, not everything in medical science is easily measured: depression, pain, quality of life; even the more