

Cambridge University Press
978-0-521-53407-9 - Study Design and Statistical Analysis: A Practical Guide for Clinicians
Mitchell H. Katz
Frontmatter
[More information](#)

Study Design and Statistical Analysis

A Practical Guide for Clinicians

This book takes the reader through the entire research process: choosing a question, designing a study, collecting the data, using univariate, bivariate and multivariable analysis, and publishing the results. It does so by using plain language rather than complex derivations and mathematical formulae. It focuses on the nuts and bolts of performing research by asking and answering the most basic questions about doing research studies. It has numerous tables, graphs and tips to help demystify the process. It is filled with up-to-date examples from the clinical literature on how to use statistical analyses to answer important questions.

Cambridge University Press
978-0-521-53407-9 - Study Design and Statistical Analysis: A Practical Guide for Clinicians
Mitchell H. Katz
Frontmatter
[More information](#)

Cambridge University Press

978-0-521-53407-9 - Study Design and Statistical Analysis: A Practical Guide for Clinicians

Mitchell H. Katz

Frontmatter

[More information](#)

Study Design and Statistical Analysis

A Practical Guide for Clinicians

Mitchell H. Katz



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Cambridge University Press is part of the University of Cambridge.
It furthers the University’s mission by disseminating knowledge in the pursuit of
education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9780521534079

©M.H. Katz 2006

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 2006
7th printing 2013

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

ISBN 978-0-521-53407-9 Paperback

Cambridge University Press 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.

.....

Every effort has been made in preparing this book to provide accurate and
up-to-date information which is in accord with accepted standards and practice
at the time of publication. Although case histories are drawn from actual cases,
every effort has been made to disguise the identities of the individuals involved.
Nevertheless, the authors, editors and publishers can make no warranties that the
information contained herein is totally free from error, not least because clinical
standards are constantly changing through research and regulation. The authors,
editors and publishers therefore disclaim all liability for direct or consequential
damages resulting from the use of material contained in this book. Readers
are strongly advised to pay careful attention to information provided by the
manufacturer of any drugs or equipment that they plan to use.

Cambridge University Press
978-0-521-53407-9 - Study Design and Statistical Analysis: A Practical Guide for Clinicians
Mitchell H. Katz
Frontmatter
[More information](#)

To best friends: Perri Klass and Adam Lowe

Cambridge University Press
978-0-521-53407-9 - Study Design and Statistical Analysis: A Practical Guide for Clinicians
Mitchell H. Katz
Frontmatter
[More information](#)

Contents

	Preface	<i>page xi</i>
1	Introduction	1
	1.1 Why is statistical analysis so important for clinical research?	1
2	Designing a study	8
	2.1 How do I choose a research question?	8
	2.2 How do I choose a study design?	11
	2.3 What are the differences between randomized and observational studies?	11
	2.4 What are the different types of randomized controlled trials?	18
	2.5 What are the different methods of allocating subjects within a randomized design?	20
	2.6 What are the different types of observational studies?	23
	2.7 Do I need to specify a particular hypothesis for my study?	32
	2.8 Can I specify an alternative hypothesis with a specific direction?	33
	2.9 Can my study have more than one question?	34
	2.10 What kind of measures should I use?	35
	2.11 How many subjects will I need for my study?	36
	2.12 How do I obtain an institutional review board approval to perform a research study?	37
3	Data management	38
	3.1 How do I manage my data?	38
	3.2 What procedures should I follow in collecting data?	38
	3.3 How do I create data collection instruments?	39
	3.4 How do I enter my data?	43
	3.5 How do I clean my data?	45

viii	Contents	
	3.6 How do I recode a variable?	45
	3.7 How do I transform a variable?	50
	3.8 When will I need to derive variables?	50
	3.9 When should I export my data to a statistical program?	50
4	Univariate statistics	52
	4.1 How should I describe my data?	52
	4.2 How should I describe my interval and ordinal variables?	52
	4.3 How should I describe my dichotomous variables?	57
	4.4 How should I describe my nominal variables?	59
	4.5 How should I describe my ordinal variables?	60
	4.6 How should I describe events that occur over time?	60
5	Bivariate statistics	66
	5.1 How do I assess an association between two variables?	66
	5.2 How do I assess an association between two dichotomous variables (comparison of proportions)?	66
	5.3 How do I test an association between a nominal variable and a dichotomous variable or between two nominal variables?	77
	5.4 How do I test an association involving an interval variable? (When do I use parametric statistics versus non-parametric statistics?)	79
	5.5 How do I test an association of a dichotomous variable with an interval variable?	84
	5.6 How do I test an association of a nominal variable with an interval variable?	88
	5.7 How do I test an association between two interval variables? (How do I determine if an association is linear?)	92
	5.8 How do I test an association of two variables when one or both of the variables are ordinal?	100
	5.9 How do I compare outcomes that occur over time?	102
	5.10 How do I analyse repeated observations of the same subject?	107
	5.11 How do I test bivariate associations with matched data?	116
6	Multivariable statistics	120
	6.1 What is multivariable analysis? Why is it necessary?	120
	6.2 How do I choose what type of multivariable analysis to use?	123
	6.3 What should I do if my outcome variable is ordinal or nominal?	123

6.4	How do I assess the impact of an individual variable on an outcome in a multivariable analysis?	124
6.5	What assumptions underlie multivariable models?	125

7

Sample size calculations 127

7.1	How do I determine the number of subjects needed for my study?	127
7.2	How do I determine the sample size needed for univariate statistics?	129
7.3	How do I determine the sample size needed for a univariate analysis of a dichotomous variable (proportion)?	130
7.4	How do I determine the sample size needed for a univariate analysis of an interval variable (mean)?	131
7.5	How do I determine the sample size needed for bivariate analysis?	131
7.6	How do I determine the sample size needed for comparison of two proportions (two dichotomous variables)?	133
7.7	How do I determine the sample size needed for comparison of two means (association of a dichotomous variable with a normally distributed interval variable)?	134
7.8	How do I determine the sample size needed for comparison of two normally distributed interval variables (Pearson's correlation coefficient)?	135
7.9	How do I determine the sample size needed for comparison of two survival times (log-rank statistic)?	135
7.10	How do I determine the sample size needed for multivariable analyses?	136
7.11	How do I determine the sample size needed to prove that two treatments are equal?	137
7.12	What if the sample size needed exceeds the sample size I can obtain?	138

8

Studies of diagnostic and prognostic tests (predictive studies) 141

8.1	How do predictive studies differ from explanatory studies?	141
8.2	What are sensitivity and specificity, and how are they related to one another?	143
8.3	What are the positive and negative predictive values of a test?	144
8.4	How do I determine the accuracy of a test?	145
8.5	How do I calculate the characteristics of a test with an interval scale?	146

x	Contents	
	8.6 What is Bayes' theorem?	148
	8.7 How do I choose the best standard for predictive studies?	153
	8.8 What population should I use for determining the predictive ability of a test?	154
	8.9 How is validity determined for predictive studies?	154
9	Statistics and causality	155
	9.1 When can statistical association establish causality?	155
	9.2 Can the results be statistically significant and clinically unimportant?	161
	9.3 Can the results be statistically insignificant and clinically important?	163
10	Special topics	165
	10.1 What is the difference between the relative risk and the absolute risk?	165
	10.2 What other effect measures are available in addition to relative risk and absolute risk?	165
	10.3 Do I need to use statistical analysis if I have population data?	170
	10.4 How do I choose what statistical program to use for analyzing data?	171
11	Publishing research	172
	11.1 How do I write my study up for publication?	172
	11.2 How do I determine authorship for the paper?	174
	11.3 How do I resolve disagreements about authorship?	175
	11.4 How do I decide what journal to send the paper to?	176
	11.5 What if my paper is rejected but I am asked to revise and resubmit it?	179
	11.6 What if my paper is rejected?	180
	11.7 How should I deal with the media?	181
12	Conclusion	183
	12.1 Would you review the steps for designing and analyzing data from a clinical study?	183
	Index	185

Preface

I decided to write this book based on the many favorable responses I received about my first book: *Multivariable Analysis: A Practical Guide for Clinicians*. Readers who found the conceptual, non-mathematical approach to multivariable analysis helpful, asked me to write a basic statistics book using the same format. My hope is that the two books together will enable clinical researchers to design rigorous studies and analyse the data using both basic and advanced statistical techniques. Although oriented for researchers performing their own studies, the book will also enable readers of clinical research to understand how statistics are used – and misused – in the published literature.

My experience teaching statistics has led me to believe that most statistics textbooks present the material backwards. Typically the formulas and derivations are presented first; only after you have slogged your way through the mathematics are you rewarded with the fun part – analyzing data to answer important questions. The problem with this approach is that many readers will be bored or overwhelmed during the mathematical approach, and will have lost interest in the subject before they get to the fun part.

I have tried to do the opposite by putting the fun part first. I have included clinical examples at the beginning and throughout the text so that you can experience the intellectual pleasure of identifying a question and using statistical analyses to answer it. To ensure that the book would not be intimidating I have excluded derivations, minimized the use of algebraic expressions, and, where possible, used words rather than mathematical symbols to express the underlying statistical concepts. As readily available statistical programs, such as Stata or SAS or Epi Info, will correctly perform the mathematics for you, I think that what is most important is to understand the concepts.

Once hooked on clinical research I hope you will want to learn more. An excellent book that includes derivations and a more thorough review of many of the concepts discussed in this book is: S. Glantz's *Primer of Biostatistics* (5th edition, McGraw-Hill, 2001). For a more comprehensive approach, I recommend B. Rosner's *Fundamentals of Biostatistics* (5th edition, Duxbury, 2000).

I have organized the book to fit the chronologic order of how clinical research is performed: identification of a question, study design, data collection, univariate, bivariate, and multivariable analysis, manuscript writing and publication of the results. This organization should allow you to read each chapter as you are working on that part of the study.

One exception to the chronologic order of this book is that I have placed the sample size section after the section on statistics. Even though you will need to determine the needed sample size prior to collecting and analyzing your data, you can't calculate a sample size without knowing what type of statistical analysis you will be performing.

As much as possible I have included practical advice on the nuts and bolts of performing clinical research, such as how to recode and transform variables. This information is rarely included in statistics books but if done incorrectly will lead you to the wrong answer.

I have minimized overlap between this book and my multivariable book, just released in a new 2nd edition (Cambridge University Press, 2005). If you want to know more about multivariable analysis than contained in Chapter 6, I hope you will read it.

In writing this book I am indebted to my teachers, students, and colleagues. I include among my teachers several epidemiologists and biostatisticians I have never met but whose books I have benefited from. Rather than name them all here I have cited them liberally in the footnotes. One reference I found particularly helpful at several points was B.S. Everett's *Medical Statistics from A to Z* (Cambridge University Press, 2003). My colleagues at the Department of Public Health and the University of San Francisco, California have taught me much about identifying and answering important clinical questions. Several years of students in the University of California, San Francisco, Training in Clinical Research Program have sharpened my teaching skills by letting me try out different methods of presenting the material. Warren Browner, Susan Buchbinder, Jeffrey Martin, and Rani Marx reviewed the manuscript and made many helpful suggestions. If any errors crept in despite their review, I alone am to blame.

In writing this book, I appreciate the support of my editor Peter Silver and the staff at Cambridge University Press.

If you have questions or suggestions for future editions e-mail me at mhkatz59@yahoo.com