

SINGLE-CASE  
AND SMALL-*n*  
EXPERIMENTAL  
DESIGNS

A Practical Guide to Randomization Tests

John B. Todman  
Pat Dugard

# **Single-Case and Small- $n$ Experimental Designs**

*A Practical Guide to Randomization Tests*



# Single-Case and Small-*n* Experimental Designs

*A Practical Guide to Randomization Tests*

**John B. Todman**

*Department of Psychology, University of Dundee*

**Pat Dugard**

*Independent Consultant*

 **Lawrence Erlbaum Associates**  
Taylor & Francis Group

---

New York London

This edition published in the Taylor & Francis e-Library, 2009.

To purchase your own copy of this or any of  
Taylor & Francis or Routledge's collection of thousands of eBooks  
please go to [www.eBookstore.tandf.co.uk](http://www.eBookstore.tandf.co.uk).

Copyright © 2001 by Lawrence Erlbaum Associates

All rights reserved. No part of this book may be reproduced in  
any form, by photostat, microfilm, retrieval system, or any other  
means, without prior written permission of the publisher.

Cover design by Kathryn Houghtaling Lacey

**Library of Congress Cataloging-in-Publication Data**

Todman, John.

Single-case and small-n experimental designs: a practical guide  
to randomization tests/John Todman, Pat Dugard.

p. cm.

Includes bibliographical references and index.

ISBN 0-8058-3554-7 (cloth: alk. paper)

1. Statistical hypothesis testing. 2. Experimental design.

I. Dugard, Pat. II. Title.

QA277.T63 2001

519.5'6-dc21 00-062294

ISBN 1-4106-0094-7 Master e-book ISBN

## **Acknowledgments**

*This book grew out of a paper published in *Augmentative and Alternative Communication* (Todman & Dugard, 1999) and we would like to express our thanks to the Associate Editor, Jan Bedrosian, who handled the review process. The existence of the book probably owes much to Jan's enthusiastic support. We are also grateful to the various reviewers, from whose comments we have benefitted greatly.*

*We are indebted to Eugene Edgington in two respects. Without his body of work on randomization tests, there would not have been a paper, let alone a book. Additionally, he was kind enough to read an early version of the paper, and his helpful and encouraging comments have been much appreciated.*

## **Dedication**

*To generations of students who probably taught us more than we taught them.*



# Contents

<b>Preface</b>	<b>xix</b>
<b>1 DATA ANALYSIS IN SINGLE-CASE AND SMALL-<i>n</i> EXPERIMENTS</b>	<b>1</b>
Varieties of Clinical Design	1
Random Allocation of Treatments to Participants or Test Occasions	3
Random Sampling Versus Random Allocation	3
Participants Versus Exposures	4
Testing Hypotheses in Single-Case and Small- <i>n</i> Designs	4
Time-Series Analysis Using ARIMA	4
Adaptations of Classical ANOVA and Regression Procedures	5
Nonparametric Tests	6
Randomization Tests and the Validity of Statistical Conclusions	6
The Versatility of Randomization Tests	7
Guidelines for Statistical Hypothesis Testing	7
<b>2 STATISTICAL AND VISUAL ANALYSIS</b>	<b>10</b>
Arguments Against Statistical Analysis of Single-Case Designs	10
The Operant Viewpoint	10
Control	11
Response-Guided Experimentation	13
<i>An Illustration of Response-Guidance Bias</i>	14
<i>An Experimental Test of Response-Guided Bias</i>	17
<i>Statistical Consequences of Response-Guided Experimentation</i>	18
Replication	18
Clinical and Statistical Significance	19
When to Use Statistical Analysis	21
<b>3 APPROACHES TO STATISTICAL INFERENCE</b>	<b>23</b>
Randomization Tests	23
The Classical Theory	24
Bayesian Inference	25

Randomization Tests Revisited	25
Assumptions of Randomization Tests	26
Distributional Assumptions	26
Autocorrelation	28
The Current Suitability of Randomization Tests	30
<b>4 PRINCIPLES OF RANDOMIZATION TESTS</b>	<b>32</b>
The Lady Tasting Tea Experiment	32
Examples With Scores as Data	33
Alternating Treatments (3 Observation Times for Each of 2 Treatments)	33
Alternating Treatments (4 Observation Times for Each of 2 Treatments)	36
A Phase (Baseline-Treatment) Design	36
Generalized Application of Principles of Randomization Tests	38
A Randomization Procedure	38
Selection of a Test Statistic	38
Computation of the Test Statistic for Arrangements of the Data	39
<i>Systematic Arrangements of the Data</i>	39
<i>Random Samples of Arrangements of the Data</i>	39
Reporting the Statistical Decision	40
Summary of Randomization Test Requirements	41
<b>5 RANDOMIZATION DESIGNS FOR SINGLE-CASE AND SMALL-<i>n</i> STUDIES</b>	<b>42</b>
Design 1 (AB)	43
Example for Design 1	44
Design 2 (ABA Reversal)	46
Example for Design 2	48
Design 3 (AB Multiple Baseline)	49
Example for Design 3	50
Design 4 (ABA Multiple Baseline)	52
Example for Design 4	52
Design 5	54
Design 5 (One-Way Small Groups)	54
<i>Example for Design 5 (Small Groups)</i>	55

Design 5 (Single-Case Randomized Treatment)	55
<i>Example for Design 5 (Single-Case)</i>	56
Design 5a (Small Groups or Single-Case—2 Randomized Treatments)	56
<i>Example for Design 5a</i>	56
Design 6	57
Design 6 (One-Way Small Group Repeated Measures)	57
<i>Example for Design 6 (Small Group)</i>	57
Design 6 (Single-Case Randomized Blocks)	58
<i>Example for Design 6 (Single-Case)</i>	59
Design 6a (2 Repeated Measures or 2 Randomized Blocks)	59
<i>Example for Design 6a</i>	59
Design 7 (Two-Way Factorial Single-Case)	60
Example for Design 7	61
Design 8 (Ordinal Predictions)	63
Example for Design 8	63
<b>6 RANDOMIZATION TESTS USING MINITAB</b>	<b>66</b>
Design 1 (AB)	67
Specifications for Design 1 Example	67
Commented Macro for Design 1 (Macro File Name: des1.txt)	68
Location of Sample Size (2000 and 2001 Entries) in Macro	69
Location of Design 1 Test Results	69
Randomization Test Results for Design 1 Example	69
Statistical Conclusion for Design 1 Example (Assuming a Directional Prediction)	69
Design 2 (ABA Reversal)	69
Specifications for Design 2 Example	69
Commented Macro for Design 2 (Macro File Name: des2.txt)	70
Location of Sample Size (2000 and 2001 Entries) in Macro	71
Location of Design 2 Test Results	71
Randomization Test Results for Design 2 Example	71
Statistical Conclusion for Design 2 Example (Assuming a Directional Prediction)	72
Design 3 (AB Multiple Baseline)	72

Specifications for Design 3 Example	72
Commented Macro for Design 3 (Macro File Name: des3.txt)	72
Location of Sample Size (2000 and 2001 Entries) in Macro	74
Location of Design 3 Test Results	74
Randomization Test Results for Design 3 Example	74
Statistical Conclusion for Design 3 Example (Assuming a Directional Prediction)	74
Design 4 (ABA Multiple Baseline)	75
Specifications for Design 4 Example	75
Commented Macro for Design 4 (Macro File Name: des4.txt)	75
Location of Sample Size (2000 and 2001 Entries) in Macro	77
Location of Design 4 Test Results	77
Randomization Test Results for Design 4 Example	78
Statistical Conclusion for Design 4 Example (Assuming a Directional Prediction)	78
Design 5 (One-Way Small Groups and Single-Case Randomized Treatment)	78
Specifications for Design 5 Example	78
Commented Macro for Design 5 (Macro File Name: des5.txt)	78
Location of Sample Size (2000 and 2001 Entries) in Macro	79
Location of Design 5 Test Results	79
Randomization Test Results for Design 5 Example	79
Statistical Conclusion for Design 5 (One-Way Small Groups) Example	79
Design 5a (Small Groups or Single-Case—Two Randomized Treatments)	80
Specifications for Design 5a Example	80
Commented Macro for Design 5a (Macro File Name: des5a.txt)	80
Location of Sample Size (2000 and 2001 Entries) in Macro	81
Location of Design 5a Test Results	81
Randomization Test Results for Design 5a Example	81
Statistical Conclusion for Design 5a (One-Tailed Single-Case) Example	81
Design 6 (One-Way Small Group Repeated Measures and Single-Case Randomized Blocks)	82

Specifications for Design 6 Example	82
Commented Macro for Design 6 (Macro File Name: des6.txt)	82
Location of Sample Size (2000 and 2001 Entries) in Macro	83
Location of Design 6 Test Results	83
Randomization Test Results for Design 6 Example	83
Statistical Conclusion for Design 6 (One-Way Small Groups) Example	83
Design 6a (2 Repeated Measures on Small Group or Single-Case Blocks)	83
Specifications for Design 6a Example	83
Commented Macro for Design 6a (Macro File Name: des6a.txt)	84
Location of Sample Size (2000 and 2001 Entries) in Macro	85
Location of Design 6a Test Results	85
Randomization Test Results for Design 6a Example	85
Statistical Conclusion for Design 6a (One-Tailed Single-Case) Example	85
Design 7 (Two-Way Factorial Single Case)	85
Specifications for Design 7 Example	85
Commented Macro for Design 7 (Macro File Name: des7.txt)	86
Location of Sample Size (2000 and 2001 Entries) in Macro	88
Location of Design 7 Test Results	88
Randomization Test Results for Design 7 Example	88
Statistical Conclusions for Design 7 (One-Tailed Main Effects) Example	88
Testing Simple Effects in the Factorial Design	89
Randomization Test Results for Design 7 (Simple Effect of Display Mode With Touch-Screen Interface) Example	89
Statistical Conclusion for One-Tailed Test of a Simple Effect	89
Design 8 (Ordinal Predictions Within Nonexperimental Designs)	90
Specifications for Design 8 Example	90
Commented Macro for Design 8 (Macro File Name: des8.txt)	90
Location of Sample Size (2000 and 2001 Entries) in Macro	91
Location of Design 8 Test Results	91
Randomization Test Results for Design 8 Example	91

Statistical Conclusion for Design 8 (Unique Order Prediction of Small Group Data) Example	91
Statistical Conclusion for Design 8 (Partial Order Prediction of Single-Case Data) Example	91
<b>7 RANDOMIZATION TESTS USING EXCEL</b>	<b>92</b>
Design 1 (AB)	94
Specifications for Design 1 Example	94
Commented Macro for Design 1 (Macro File Name: design 1.xls)	94
Location of Design 1 Test Results	97
Randomization Test Results for Design 1 Example	97
Statistical Conclusion for Design 1 Example (Assuming a Directional Prediction)	97
Design 2 (ABA Reversal)	97
Specifications for Design 2 Example	97
Commented Macro for Design 2 (Macro File Name: design2.xls)	98
Location of Design 2 Test Results	101
Randomization Test Results for Design 2 Example	101
Statistical Conclusion for Design 2 Example (Assuming a Directional Prediction)	102
Design 3 (AB Multiple Baseline)	102
Specifications for Design 3 Example	102
Commented Macro for Design 3 (Macro File Name: design3.xls)	102
Location of Design 3 Test Results	106
Randomization Test Results for Design 3 Example	106
Statistical Conclusion for Design 3 (Assuming a Directional Prediction)	106
Design 4 (ABA Multiple Baseline Design)	107
Specifications for Design 4 Example	107
Commented Macro for Design 4 (Macro File Name: design4.xls)	107
Location of Design 4 Test Results	112
Randomization Test Results for Design 4 Example	112
Statistical Conclusion for Design 4 Example (Assuming a Directional Prediction)	112
Design 5 (One-Way Small Groups and Single-Case Randomized Treatment)	112

Specifications for Design 5 Example	112
Commented Macro for Design 5 (Macro File Name: design5.xls)	113
Location of Design 5 Test Results	114
Randomization Test Results for Design 5 Example	114
Statistical Conclusion for Design 5 (One-Way Small Groups) Example	115
Design 5a (Small Groups or Single-Case—Two Randomized Treatments)	115
Specifications for Design 5a Example	115
Commented Macro for Design 5a (Macro File Name: design5a.xls)	115
Location of Design 5a Test Results	117
Randomization Test Results for Design 5a Example	117
Statistical Conclusion for Design 5a (One-Tailed Single-Case) Example	117
Design 6 (One-Way Small Group Repeated Measures and Single-Case Randomized Blocks)	118
Specifications for Design 6 Example	118
Commented Macro for Design 6 (Macro File Name: design6.xls)	118
Location of Design 6 Test Results	120
Randomization Test Results for Design 6 Example	120
Statistical Conclusion for Design 6 (One-Way Small Groups) Example	120
Design 6a (Two Repeated Measures on Small Group or Single-Case Blocks)	120
Specifications for Design 6a Example	120
Commented Macro for Design 6a (Macro File Name: design6a.xls)	121
Location of Design 6a Test Results	123
Randomization Test Results for Design 6a Example	123
Statistical Conclusion for Design 6a (One-Tailed Single-Case) Example	123
Design 7 (Two-Way Factorial Single Case)	123
Specifications for Design 7 Example	123
Commented Macro for Design 7 (Macro File Name: design7.xls)	124
Location of Design 7 Test Results	127
Randomization Test Results for Design 7 Example	127

Statistical Conclusions for Design 7 (One-Tailed Main Effects) Example	128
Testing Simple Effects in the Factorial Design	128
Randomization Test Results for Design 7 (Simple Effect of Display Mode With Touch-Screen Interface) Example	129
Statistical Conclusion for One-Tailed Test of a Simple Effect	129
Design 8 (Ordinal Predictions Within Nonexperimental Designs)	129
Specifications for Design 8 Example	129
Commented Macro for Design 8 (Macro File Name: design8.xls)	130
Location of Design 8 Test Results	131
Randomization Test Results for Design 8 Example	131
Statistical Conclusion for Design 8 (Unique Order Prediction of Small Group Data) Example	131
Statistical Conclusion for Design 8 (Partial Order Prediction of Single-Case Data) Example	132
<b>8 RANDOMIZATION TESTS USING SPSS</b>	<b>133</b>
Design 1 (AB)	135
Specifications for Design 1 Example	135
Commented Program for Design 1 (Program File Name: design 1.sps)	135
Randomization Test Results for Design 1 Example	137
Statistical Conclusion for Design 1 Example (Assuming a Directional Prediction)	137
Design 2 (ABA Reversal)	137
Specifications for Design 2 Example	137
Commented Program for Design 2 (Program File Name: design2. sps)	138
Randomization Test Results for Design 2 Example	140
Statistical Conclusion for Design 2 Example (Assuming a Directional Prediction)	140
Design 3 (AB Multiple Baseline)	141
Specifications for Design 3 Example	141
Commented Program for Design 3 (Program File Name: design3. sps)	141
Randomization Test Results for Design 3 Example	144

Statistical Conclusion for Design 3 Example (Assuming a Directional Prediction)	144
Design 4 (ABA Multiple Baseline)	144
Specifications for Design 4 Example	144
Commented Program for Design 4 (Program File Name: design4.sps)	144
Randomization Test Results for Design 4 Example	147
Statistical Conclusion for Design 4 Example (Assuming a Directional Prediction)	147
Design 5 (One-Way Small Groups and Single-Case Randomized Treatment)	148
Specifications for Design 5 Example	148
Commented Program for Design 5 (Program File Name: design5.sps)	148
Randomization Test Results for Design 5 Example	150
Statistical Conclusion for Design 5 (One-Way Small Groups) Example	150
Design 5a (Small Groups or Single-Case—Two Randomized Treatments)	150
Specifications for Design 5a Example	150
Commented Program for Design 5a (Program File Name: design5a.sps)	151
Randomization Test Results for Design 5a Example	153
Statistical Conclusion for Design 5a (One-Tailed Single-Case) Example	153
Design 6 (One-Way Small Group Repeated Measures and Single-Case Randomized Blocks)	153
Specifications for Design 6 Example	153
Commented Program for Design 6 (Program File Name: design6.sps)	154
Randomization Test Results for Design 6 Example	156
Statistical Conclusion for Design 6 (One-Way Small Groups) Example	156
Design 6a (Two Repeated Measures on Small Group or Single-Case Blocks)	157
Specifications for Design 6a Example	157

Commented Program for Design 6a (Program File Name: design 6a.sps)	157
Randomization Test Results for Design 6a Example	159
Statistical Conclusion for Design 6a (One-Tailed Single-Case) Example	159
Design 7 (Two-Way Factorial Single Case)	159
Specifications for Design 7 Example	159
Commented Program for Design 7 (Program File Name: design7. sps)	160
Randomization Test Results for Design 7 Example	163
Statistical Conclusions for Design 7 (One-Tailed Main Effects) Example	163
Testing Simple Effects in the Factorial Design	164
Randomization Test Results for Design 7 (Simple Effect of Display Mode With Touch-Screen Interface) Example	164
Statistical Conclusion for One-Tailed Test of a Simple Effect	164
Design 8 (Ordinal Predictions Within Nonexperimental Designs)	165
Specifications for Design 8 Example	165
Commented Program for Design 8 (Program File Name: design8. sps)	165
Randomization Test Results for Design 8 Example	167
Statistical Conclusion for Design 8 (Unique Order Prediction of Small Group Data) Example	167
Statistical Conclusion for Design 8 (Partial Order Prediction of Single-Case Data) Example	167
<b>9 OTHER SOURCES OF RANDOMIZATION TESTS</b>	<b>168</b>
Books and Journal Articles	168
Statistical Packages	169
RANDIBM	169
SCRT	169
StatXact	170
SPSS for Windows	171
SAS	171
<b>10 THE USE OF RANDOMIZATION TESTS WITH NONRANDOMIZED DESIGNS</b>	<b>172</b>
Nonrandomized Designs	173

Nonrandomized Classification Variables	173
Nonrandomized Phase Designs With Specific Predictions	174
<b>11 THE POWER OF RANDOMIZATION TESTS</b>	<b>176</b>
The Concept of Power	176
The Determinants of Power	177
The Probability of Type 1 Error (a Level)	177
Effect Size	178
Sample Size	179
Other Factors Influencing Power	179
<i>Control of Random Nuisance Variables</i>	179
<i>Increased Reliability of Measuring Instruments</i>	181
<i>Maximizing Effect Size</i>	181
<i>Choice of Statistic</i>	182
<i>Increased Precision of Prediction</i>	182
Estimating Power and Required Sample Size	182
Power in Single-Case Designs	183
Power for a Single-Case Randomized Treatment Design	184
Power for an AB Design With a Randomized Intervention Point	184
Power for a Phase Design With Random Assignment to Phases	186
Conclusions	187
<b>12 CREATING YOUR OWN RANDOMIZATION TESTS</b>	<b>188</b>
Steps in Creating a Macro for a Randomization Test	189
Writing Your Own Macros	190
Tinkering: Changing Values Within a Macro	191
Design Modifications Without Macro Changes	191
Data Modification Prior to Analysis	191
Downward Slope During Baseline	192
Upward Slope During Baseline	194
Sources of Design Variants and Associated Randomization Tests	197
<b>References</b>	<b>198</b>
<b>Author Index</b>	<b>201</b>
<b>Subject Index</b>	<b>203</b>



## *Preface*

We are, respectively, an academic psychologist with research interests in assistive communication and a statistician with a particular interest in using statistics to solve problems thrown up by researchers. Each has a long-standing involvement in teaching statistics from the different perspectives and needs of psychology students and others whose main interest is in applications rather than the methodology of statistics. The first author “discovered” randomization tests while searching for valid ways to analyze data from his single-case experiments. The second author already knew about the randomization test approach to making statistical inferences in principle but had not had occasion to use it. She developed an interest in his research problems and we began to work together on designs and associated randomization tests to tackle the issue of making sound causal inferences from single-case data, and the collaboration grew from there. It is the experience of both of us that communication between research psychologists and statisticians is often less than perfect, and we feel fortunate to have had such an enjoyable and (we hope) productive collaboration. We hope we have brought something of the effectiveness of our communication across the abstract statistics-messy research divide to this book. The way we approached the project was, broadly, this: The psychologist author produced a first draft of a chapter, the statistician author added bits and rendered it technically correct, and we then worked jointly at converting the technically correct version into a form likely to be comprehensible to clinical researchers with no more than average statistical skills. Where possible, we have tried to avoid using technically abstruse language and, where we have felt it necessary to use concepts that may be unfamiliar to some, we have tried to explain them in straightforward language. No doubt, the end product will seem unnecessarily technical to some and tediously “stating the obvious” to others. We hope, at least, that most readers will not find our treatment too extreme in either direction.

Our first publication together in the area of single-case research was a paper in *Augmentative and Alternative Communication*. That paper described a number of designs pertinent to assistive communication research, together with Minitab programs (macros) to carry out randomization tests on data from the designs. This book grew out of that paper. In general, the first author is responsible for the designs and examples and the second author is responsible for the programs, although this is undoubtedly an oversimplification of our respective roles. Although Minitab is quite widely used, there are plenty of researchers who do not have access to it. By implementing programs in two other widely used packages (Excel and SPSS) we hope to bring randomization tests into the familiar territory of many more researchers. Initially, we were uncertain about the feasibility of these implementations, and it is true that Excel in particular is relatively inefficient in terms of computing time. On the other hand, almost everyone using Windows has access to Excel and many people who would shy away from an overtly statistical package will have used Excel without fear and trepidation.

First and foremost, this book is supposed to be a practical guide for researchers who are interested in the statistical analysis of data from single-case or very small- $n$  experiments. That is not to say it is devoid of theoretical content, but that is secondary. To some extent, we have included theoretical content in an attempt to persuade skeptics that valid statistical tests are available and that there are good reasons why they should not ignore them. In

general, we have not set out to provide a comprehensive review of theoretical issues and we have tried to keep references to a minimum. We hope we have provided enough key references to put readers in touch with a more extensive literature relating to the various issues raised.

Our motivation for the book is our conviction that randomization tests are underused, even though in many cases they provide the most satisfactory means of making statistical inferences about treatment effects in small-scale clinical research. We identify two things holding clinical researchers back from the use of randomization tests. One is the need to modify familiar designs to use the tests, and the other is that tests are not readily available in familiar statistical packages.

In chapter 1, we consider the options for statistical analysis of single-case and small- $n$  studies and identify circumstances in which we believe that randomization tests should be considered.

In chapter 2, we consider when statistical analysis of data from single-case and small- $n$  studies is appropriate. We also consider the strong tradition of response-guided experimentation that is a feature of visual analysis of single-case data. We note that randomization tests require random assignment procedures to be built into experimental designs and that such randomization procedures are often incompatible with response-guided experimental procedures. We identify this need to change basic designs used in single-case research as one of the obstacles to the acceptance of randomization tests. We argue that, whether or not there is an intention to use statistical analysis, random assignment procedures make for safer causal inferences about the efficacy of treatments than do response-guided procedures. This is an obstacle that needs to be removed for reasons involving principles of good design.

In chapter 3, we take a look at how randomization tests relate to other approaches to statistical inference, particularly with respect to the assumptions that are required. In the light of this discussion we offer some conclusions concerning the suitability of randomization tests now that the necessary computational power is available on computers.

In chapter 4, we use examples to explain in some detail how randomization tests work.

In chapter 5, we describe a range of single-case and small- $n$  designs that are appropriate for statistical analysis using randomization tests. Readers who are familiar with the principles underlying randomization tests and already know that they want to use designs that permit the application of such tests may want to go directly to this chapter. For each design, we provide a realistic example of a research question for which the design might be appropriate. Using the example, we show how data are entered in a worksheet within the user's chosen statistical package (Minitab, Excel, or SPSS). We also explain how the data will be analyzed and a statistical inference arrived at. Almost all of the information in this chapter is common to the three statistical packages, with some minor variations that are fully described.

Chapters 6, 7, and 8 present the programs (macros) for running the randomization tests within Minitab, Excel, and SPSS, respectively. The programs, together with the sample data for the examples (from chapter 5), are also provided on CD-ROM. Our second obstacle to the use of randomization tests was availability of the tests in familiar statistical packages. We have attempted to reduce this obstacle by making tests for a range of designs available within three widely used packages. More efficient programs are undoubtedly possible

using lower level languages, but we think this misses the point. For researchers who are not computing wizards, we believe that it is not the time costs that prevent them from using randomization tests. Even for the slowest of our package implementations (Excel), the time costs are trivial in relation to the costs of running an experiment. We argue that providing programs in a familiar computing environment will be more likely to encourage researchers to “have a go” than providing very fast programs in an unfamiliar environment.

In planning our programs, we decided to go for consistency of computational approach, which led us to base all of the randomization tests on random sampling from all possible rearrangements of the data, rather than generating an exhaustive set of arrangements where that would have been feasible. This has the added advantage that users can decide how many samples to specify, permitting a trade-off between sensitivity and time efficiency. Another advantage is that programs based on random sampling can be made more general and therefore require less customization than programs based on an exhaustive generation of data arrangements.

These chapters also contain details of sample runs for all of the examples and explanations of how to set up and run analyses on users’ own data. For each example, information is provided on (a) the design specification, (b) the function of each section of the macro, (c) how to change the number of samples of data rearrangements on which the statistical conclusion is based, (d) where to find the results of the analysis, (e) results of three runs on the sample data, including average run times, and (f) an example of how to report the statistical conclusion.

In chapter 9, we give readers some help in finding sources of information about randomization tests and in locating other packages that provide some facilities for running randomization tests. We have not attempted to be comprehensive in our listing of alternative sources, but have confined ourselves to the few with which we have direct experience. Also, we do not provide detailed reviews of packages, but attempt to provide enough information to enable readers to decide whether a package is likely to be of interest to them.

In chapter 10, we consider the issue of whether, and in what circumstances, it may be acceptable to relax the requirement for a random assignment procedure as a condition for carrying out a randomization test. Our conclusions may be controversial, but we believe it is important that a debate on the issue takes place. As in classical statistics, it is not enough just to assert the ideal and ignore the messy reality.

In chapter 11, we take up another thorny issue—power—that, from available simulations, looks a bit like the Achilles’ heel of randomization tests, at least when applied to some designs. We are less pessimistic and look forward to further power simulations leading to both a welcome sorting of the wheat from the chaff and pressure for the development of more satisfactory designs.

In chapter 12, we attempt to help researchers who want to use variants of the designs we have presented. One of the attractions of the randomization test approach is that a test can be developed to capitalize on any random assignment procedure, no matter how unusual it may be. The provision of commented examples seems to us a good way to help researchers acquire the skills needed to achieve tailored programs. In this chapter we have tried to provide guidance to those interested in writing programs similar to ours to fit their own particular requirements.

Like others before us, we started out on our exploration of randomization tests full of enthusiasm and a conviction that a vastly superior “newstats” based on computational power was just around the corner. We remain enthusiastic, but our conviction has taken a beating. We now recognize that randomization tests are not going to be a panacea—there are formidable problems remaining, particularly those relating to power and serial dependency. We do, however, remain convinced that randomization tests are a useful addition to the clinical researcher’s armory and that, with the computing power now available, they are an approach whose time has come.



# Chapter 1

## *Data Analysis in Single-Case and Small-n Experiments*

### VARIETIES OF CLINICAL DESIGN

Research involving a clinical intervention normally is aimed at testing the efficacy of the treatment effect on a dependent variable that is assumed to be a relevant indicator of health or quality of life status. Broadly, such research can be divided into relatively large-scale clinical trials and single-case studies, with small-group studies in a rather vaguely specified intermediate position.

There are two problems that clinical designs are intended to solve, often referred to as internal and external validity. *Internal validity* refers to competing explanations of any observed effects: A good design will remove all threats to internal validity by eliminating explanations other than the different treatments being investigated. For example, if a well-designed experiment shows improved life expectancy with a new cancer treatment, there will be no possible alternative explanation, such as “The patients on the new treatment were all treated by an especially dedicated team that gave emotional support as well, whereas the others attended the usual clinic where treatment was more impersonal.” *External validity* refers to the general application of any result found: Most results are of little interest unless they can be applied to a wider population than those taking part in the experiment. External validity may be claimed if those taking part were a random sample from the population to which we want to generalize the results. In practice, when people are the participants in experiments, as well as in many other situations, random sampling is an unrealistic ideal, and external validity is achieved by repeating the experiment in other contexts.

A *true experiment* is one in which it is possible to remove all threats to internal validity. In the simplest clinical trials a new treatment is compared with a control condition, which may be an alternative treatment or a placebo. The treatment or control conditions are randomly allocated to a large group of participants, and appropriate measurements are taken before and after the treatment or placebo is applied. This is known as a *pretest-posttest control group design*. It can be extended to include more than one new treatment or more than one control condition. It is one of rather few designs in which it is possible to eliminate all threats to internal validity. The random allocation of participants to conditions is critical: It is the defining characteristic of a true experiment. Here we are using terminology introduced by Campbell and Stanley (1966) in their influential survey of experimental and quasi-experimental designs. They characterized only three designs that are true experiments in the sense that all threats to internal validity can be eliminated, and the pretest-posttest control group design is one of them.

There are well-established statistical procedures for evaluating the efficacy of treatments in true experiments, where treatments (or placebos) can be randomly allocated to a relatively large number of participants who are representative of a well-defined population. Parametric tests, such as *t*-tests, analyses of variance (ANOVAS) and analyses

## 2 *Single-Case and Small-n Experimental Designs*

of covariance (ANCOVAS) generally provide valid analyses of such designs, provided that some technical assumptions about the data populations are approximately met. For example, the appropriate statistical procedure for testing the treatment effect in a pretest-posttest control group design is the ANCOVA (Dugard & Todman, 1995).

In much clinical research, however, the availability of individuals within well-defined categories is limited, making large-*n* clinical trials impractical. It is no solution to increase the size of a clinical population by defining the category broadly. When this is done, the large functional differences between individuals within the broad category are likely to reduce drastically the power of the usual large-*n* tests to reveal real effects. For example, within the research area of aided communication, people who are unable to speak for a particular reason (e.g., cerebral palsy, stroke, etc.) generally differ widely in terms of associated physical and cognitive impairments. For these and other reasons, such as the requirements of exploratory research or a fine-grained focus on the process of change over time, single-case and small-*n* studies are frequently the methodologies of choice in clinical areas (Franklin, Allison, & Gorman, 1996; Remington, 1990).

For single-case designs, valid inferences about treatment effects generally cannot be made using the parametric statistical procedures typically used for the analysis of clinical trials and other large-*n* designs. Furthermore, although there is no sharp dividing line between small-*n* and large-*n* studies, the smaller the sample size, the more difficult it is to be confident that the parametric assumptions are met (Siegel & Castellan, 1988). Consequently, nonparametric alternatives usually are recommended for the analysis of studies with a relatively small number of participants. Bryman and Cramer (1990) suggested that the critical group size below which nonparametric tests are desirable is about 15. The familiar nonparametric tests based on rankings of scores, such as the Mann-Whitney U and Wilcoxon T alternatives to independent *t* and related *t* parametric tests, are not, however, a complete answer. These ranking tests lack sensitivity to real treatment effects in studies with very small numbers of participants. As with the large-*n* versus small-*n* distinction, there is no clear demarcation between designs with small and very small numbers of participants but, as a rough guide, we have in mind a total number of observations per treatment condition in single figures when we refer to very small-*n* studies.

For some very small-*n* designs, procedures known as *randomization tests* provide valid alternatives with greater sensitivity because they do not discard information in the data by reducing them to ranks. Importantly, randomization tests can also deliver valid statistical analysis of data from a wide range of single-case designs. It is true that randomization tests can be applied to large-*n* designs, but the parametric tests currently used to analyze such designs are reasonably satisfactory and the pressure to adopt new procedures, even if they are superior, is slight. It is with very small-*n* and single-case designs, where valid and sensitive alternatives are hard to come by, that randomization tests really come into their own.

We aim, first, to persuade clinical researchers who use or plan to use single-case or small-*n* designs that randomization tests can be a useful adjunct to graphical analysis techniques. Our second aim is to make a range of randomization tests readily accessible to researchers, particularly those who do not claim any great statistical sophistication. Randomization tests are based on randomization procedures that are built into the design of a study, and we turn now to a consideration of the central role of randomization in hypothesis testing.

## **RANDOM ALLOCATION OF TREATMENTS TO PARTICIPANTS OR TEST OCCASIONS**

As we noted earlier, randomization is a necessary condition for a true experimental design, but we need to be a little more specific about our use of the concept. There are two points to be made. First, random allocation is not the same thing as random sampling and, second, random allocation does not apply exclusively to the allocation of participants to treatment conditions. Each of these points is important for the rationale underlying randomization tests.

### **Random Sampling Versus Random Allocation**

Random sampling from a large, well-defined population or universe is a formal requirement for the usual interpretation of parametric statistics such as the *t* test and ANOVA. It is also often the justification for a claim of generalizability or external validity. However, usually it is difficult or prohibitively expensive even to define and list the population of interest, a prerequisite of random sampling. As an example of the difficulty of definition, consider the population of households. Does a landlord and the student boarder doing his or her own shopping and cooking constitute two households or one? What if the landlord provides the student with an evening meal? How many households are there in student apartments where they all have their own rooms and share some common areas? How can the households of interest be listed? Only if we have a list can we take a random sample, and even then it may be difficult. All the full-time students at a university will be on a list and even have a registration number, but the numbers will not usually form a continuous series, so random sampling will require the allocation of new numbers. This kind of exercise is usually prohibitively costly in time and money, so it is not surprising that samples used in experiments are rarely representative of any wider population.

Edgington (1995) and Manly (1991), among others, made the same point: It is virtually unheard of for experiments with people to meet the random sampling assumption underlying the significance tables that are used to draw inferences about populations following a conventional parametric analysis. It is difficult to conclude other than that random sampling in human experimental research is little more than a convenient fiction. In reality, generalization almost invariably depends on replication and nonstatistical reasoning.

The importance of randomization in human experimentation lies in its contribution to internal validity (control of competing explanations), rather than external validity (generalization). The appropriate model has been termed *urn randomization*, as opposed to sampling from a universe. Each of the conditions is regarded as an urn or container and each participant is placed into one of the urns chosen at random. A test based on the urn randomization approach requires that conditions be randomly assigned to participants. Provided this form of randomization has been incorporated in the design, an appropriate test would require that we take the obtained treatment and control group scores and assign them repeatedly to two urns. Then, an empirical distribution of mean differences arising exclusively from random assignment of this particular set of scores can be used as the criterion against which the obtained mean difference is judged. The empirical distribution is simply a description of the frequency with which repeated random assignments of the treatment

#### 4 *Single-Case and Small-n Experimental Designs*

and control group scores to the two urns produce differences of various sizes between the means of scores in the two urns. If urn differences as big as the actual difference between treatment and control group means occur infrequently, we infer that the difference between conditions was likely due to the effect of the treatment. This, in essence, is the randomization test, which is explained in detail with concrete examples in chapter 4.

### **Participants Versus Exposures**

The usual example of random allocation given is the random assignment of treatments to participants (or participants to treatments), but we really are talking about the random assignment of treatments to *exposure opportunities*, the points at which an observation will be made. This applies equally to the “to whom” of exposure (i.e., to different participants) and to the “when” of exposure (i.e., the order in which treatments are applied, whether to different participants or to the same participant). Provided only that some random procedure has been used to assign treatments to participants or to available times, a valid randomization test will be possible. This is the case whether we are dealing with large-*n*, small-*n*, or single-case designs, but the option of using a randomization test has far greater practical implications for very small-*n* and single-case designs.

### **TESTING HYPOTHESES IN SINGLE-CASE AND SMALL-*n* DESIGNS**

The reasons randomization tests are likely to have more impact on hypothesis testing in single-case designs (and, to a lesser extent, small-*n* designs) than in large-*n* designs are that the parametric methods applied in the latter are widely available, are easy to use, and, by and large, lead to valid inferences. There is, consequently, little pressure to change the methods currently used. The same cannot be said for methods generally applied to single-case designs.

### **Time-Series Analysis Using ARIMA**

Various approaches have been proposed for the analysis of single-case designs, and Gorman and Allison (1996) provided a very useful discussion of these. Among them, time-series analyses, such as the Box and Jenkins (1976) autoregressive integrated moving average (ARIMA) model, provide a valid set of procedures that can be applied to a range of single-case designs in which observations are made in successive phases (e.g., AB and ABA designs). They do, however, have some serious limitations. For example, they require a large number of observations, far more than are available generally in single-case phase designs (Busk & Marascuilo, 1992; Gorman & Allison, 1996; Kazdin, 1976). Furthermore, the procedure is far from straightforward and unlikely to appeal to researchers who do not want to grapple with statistical complexities. Having said that, for researchers who are prepared to invest a fair amount of effort in unraveling the complexities of ARIMA

modeling, provided they are using time-series designs with large numbers of observations (probably at least 50), this approach has much to recommend it.

### **Adaptations of Classical ANOVA and Regression Procedures**

It seems on the face of it that classical ANOVA and least-squares regression approaches might be applicable to single-case designs. It is well known that parametric statistics require assumptions of normality and homogeneity of variance of population distributions. It is also known that these statistical procedures are robust to violations of the assumptions (Howell, 1997). That means that even quite large departures from the assumptions may result in a low incidence of statistical decision errors; that is, finding a significant effect when the null hypothesis is true or failing to find an effect when the null hypothesis is false. This conclusion has to be modified in some circumstances, for example, when group sizes are relatively small and unequal. However, the general tolerance of violations of assumptions in large-*n* designs led some researchers (e.g., Gentile, Roden, & Klein, 1972) to suggest that parametric approaches can safely be applied to single-case designs.

This conclusion was mistaken. The problem is that there is an additional assumption necessary for the use of parametric statistics, which is often overlooked because usually it is not an issue in large-*n* designs. This is the requirement that errors (or residuals) are uncorrelated. This means that the deviation of one observation from a treatment mean, for example, must not be influenced by deviations of preceding observations. In a large-*n* design with random allocation of treatments, generally there is no reason to expect residuals of participants within a group to be affected by testing order. The situation is very different for a single-case phase design in which phases are treated as analogous to groups and observations are treated as analogous to participants. In this case, where all observations derive from the same participant, there is every reason to anticipate that the residuals of observations that are close together in time will be more similar than those that are more distant. This serial dependency problem, known as *autocorrelation*, was explained very clearly by Levin, Marascuilo, and Hubert (1978), and was discussed in some detail by Gorman and Allison (1996). They came to the same conclusion, that positive autocorrelation will result in many more significant findings than are justified and that, given the high probability of autocorrelation in single-case phase designs, the onus should be on researchers to demonstrate that autocorrelation does not exist in their data before using classical parametric analyses. The legendary robustness of parametric statistics does not extend, it seems, to violations of the assumption of uncorrelated errors in single-case designs.

The underlying problem with Gentile et al.'s (1972) approach is that there is a mismatch between the randomization procedure (if any) used in the experimental design (e.g., random ordering of phases) and the randomization assumed by the statistical test (e.g., random ordering of observations). This results in a gross overestimate of the degrees of freedom for the error term, which leads in turn to a too-small error variance and an inflated statistic. In pointing this out, Levin et al. (1978) also observed that Gentile et al.'s (1972) approach to the analysis of single-case phase designs is analogous to an incorrect analysis of group studies highlighted by Campbell and Stanley (1966). When treatments are randomized between intact groups (e.g., classrooms), such that all members of a group receive the same treatment, the appropriate unit of analysis is the group (classroom) mean, not the individual

score. Similarly, in a single-case randomized phase design, the appropriate unit of analysis is the phase mean rather than the individual observation. As we shall see, the match between the randomization procedures used in an experimental design and the form of randomization assumed in a statistical test procedure lies at the heart of the randomization test approach.

## Nonparametric Tests

Although we accept that there will be occasions when the use of time-series or classical analysis procedures will be appropriate for the analysis of single-case designs, we believe that their practical usefulness is likely to be limited to researchers with a fair degree of statistical experience. Fortunately, there is an alternative, simple to use, nonparametric approach available for dealing with a wide range of designs. There are well-known nonparametric “rank” alternatives to parametric statistics, such as Wilcoxon T, and Mann-Whitney U, for use in small-*n* designs, and in large-*n* designs when parametric assumptions may be seriously violated. These tests can also be applied to single-case designs that are analogous to group designs. These are generally designs in which the number of administrations of each treatment is fixed, in the same way that participant sample sizes are fixed for group designs (Edgington, 1996). Examples given by Edgington are use of the Mann-Whitney U and Kruskal-Wallis tests for single-case alternating treatment designs and the Wilcoxon T and Friedman’s ANOVA for single-case randomized block designs. Illustrative examples of these and other designs are provided in chapter 5.

These standard nonparametric tests are known as *rank tests* because scores, or differences between paired scores, are reduced to rank orders before any further manipulations are carried out. Whenever such tests are used, they could be replaced by a randomization test. In fact, the standard rank tests provide approximations for the exact probabilities obtained with the appropriate randomization tests. They are only approximations because information is discarded when scores are reduced to ranks. Furthermore, the statistical tables for rank tests are based on rearrangements (often referred to as *permutations*) of ranks, with no tied ranks, so they are only approximately valid when there are tied ranks in the data (Edgington, 1995). It is worth noting that standard nonparametric tests are equivalent, when there are no tied ranks, to the appropriate randomization tests carried out on ranks instead of raw data.

## RANDOMIZATION TESTS AND THE VALIDITY OF STATISTICAL CONCLUSIONS

As we shall see, randomization tests are based on rearrangements of raw scores. As such, within a particular experiment, they provide a “gold standard” against which the validity of statistical conclusions arrived at using other statistical tests is judged (e.g., Bradley, 1968). This is true for nonparametric and parametric tests alike, even when applied to large-*n* group experiments. For example, consider how the robustness of a parametric test is established when assumptions underlying the test are violated. This is achieved by demonstrating that the test does not lead to too many incorrect decisions, compared with a

randomization test of simulated data in which the null hypothesis is true and assumptions have been violated in a specified way. Notice that we qualified our statement about the gold standard status of randomization tests by restricting it to within a particular experiment. This is necessary because, in the absence of any random sampling requirement for the use of randomization tests, there can be no question of claiming any validity beyond the particular experiment. No inference about any larger population is permissible. In other words, we are talking exclusively about internal as opposed to external validity. This does not seem much of a sacrifice, however, in view of the unreality of the assumption of representative sampling underlying inferences based on classical parametric statistical tests applied to experiments with people.

### **The Versatility of Randomization Tests**

An attraction of the randomization test approach to hypothesis testing is its versatility. Randomization tests can be carried out for many single-case designs for which no standard rank test exists. Indeed, when the number of treatment times for each condition is not fixed, as is the case in most phase designs, there is no analogous multiparticipant design for which a rank test could have been developed (Edgington, 1996). Randomization tests are also versatile in the sense that whenever some form of randomization procedure has been used in the conduct of an experiment, an appropriate randomization test can be devised.

This versatility does have a downside, however. It is too much to expect most researchers, whose only interest in statistics is to get the job done in an approved way with the least possible effort, to construct their own randomization tests from first principles. There are numerous ways in which randomization can be introduced into experimental designs and for each innovation there is a randomization test inviting specification. Our aim is to provide a reasonably wide range of the most common examples of randomization procedures, each with “ready to run” programs to do the appropriate test within each of several widely available packages. In fact, the principles underlying randomization tests are, it seems to us, much more readily understandable than the classical approach to making statistical inferences. In our final chapter, therefore, we attempt to provide some guidance for more statistically adventurous readers who are interested in developing randomization tests to match their own designs.

## **GUIDELINES FOR STATISTICAL HYPOTHESIS TESTING**

To summarize, we suggest the following guidelines for selecting a statistical approach to testing hypotheses:

1. For large- $n$  designs in which assumptions are reasonably met, the classical parametric tests provide good asymptotic approximations to the exact probabilities that would be obtained with randomization tests. There seems little reason to abandon the familiar parametric tests.
2. For large- $n$  designs in which there is doubt about assumptions being reasonably met, standard nonparametric rank tests provide good asymptotic approximations to the exact

## 8 *Single-Case and Small-n Experimental Designs*

probabilities that would be obtained with randomization tests. It is satisfactory to use these rank tests.

3. For small-*n* designs (less than 15 observations per treatment condition) for which standard nonparametric rank tests are available, it is acceptable to use these tests, although their validity is brought more into question as the number of tied ranks increases. There is no disadvantage to using a randomization test and there may well be a gain in terms of precision or validity (Edgington, 1992).
4. For very small-*n* designs (less than 10 observations per treatment condition), the use of randomization tests is strongly recommended. Our admittedly nonsystematic exploration of data sets suggests that the smaller the number of observations per treatment condition and the larger the number of different treatment conditions, the stronger this recommendation should be.
5. For single-case designs with multiple-participant analogues, classical parametric statistics may be acceptable provided that the absence of autocorrelation can be demonstrated, although this is unlikely to be possible for designs with few observations (Busk & Marascuilo, 1992). If the use of parametric tests cannot be justified, the advice is as for Points 3 and 4, depending on the number of observations per treatment condition.
6. For single-case designs with a large number of observations (e.g., a minimum of 50), but without multiple-participant analogues (e.g., phase designs and multiple baseline designs), ARIMA-type time-series analyses may be worth considering if the researcher is statistically experienced or highly motivated to master the complexities of this approach. The advantage gained is that this approach deals effectively with autocorrelation.
7. For single-case designs without a large number of observations and without multiple-participant analogues, randomization tests are the only valid option. Researchers should, however, at least be aware of uncertainties regarding power and effects of autocorrelation with respect to these designs (see chaps. 3 and 11).

These guidelines are further summarized in Fig. 1.1 in the form of a decision tree, which shows rather clearly that statistical procedures other than randomization tests are only recommended in quite restricted conditions. Before going on (in chap. 3) to consider a range of approaches to statistical inference and (in chap. 4) to provide a detailed rationale for randomization tests and examples of how they work in practice, we take up the issue of the relation between statistical and graphical analysis of single-case designs in chapter 2.

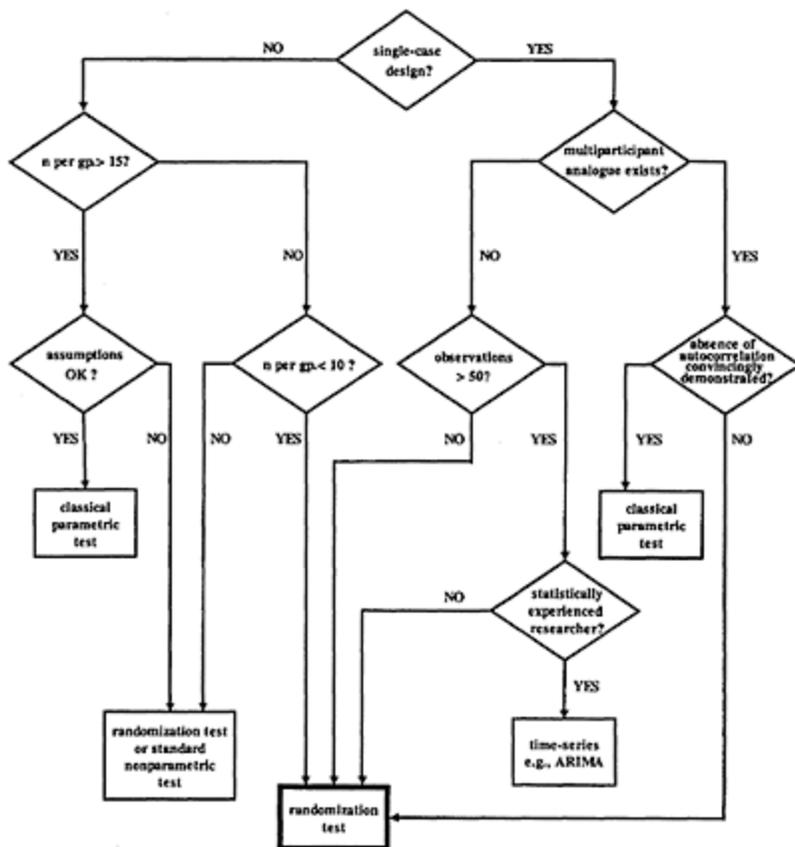


FIG. 1.1. A decision tree for selecting a statistical approach to testing hypotheses.

## References

- Allison, D.B., Silverstein, J.M., & Gorman, B.S. (1996). Power, sample size estimation, and early stopping rules. In R.D.Franklin, D.B.Allison, & B.S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 335–371). Mahwah, NJ: Lawrence Erlbaum Associates.
- Baker, R.D. (1995). Modern permutation test software. In E.S.Edgington, *Randomization tests* (3rd ed., pp. 391–401). New York: Dekker.
- Barlow, M.D.H., & Hersen, M. (1984). *Single case experimental designs: Strategies for studying behavior change* (2nd ed., pp. 285–324). New York: Pergamon.
- Borenstein, M., Rothstein, H., & Cohen, J. (1997). SamplePower™ 1.0. Chicago: SPSS Inc.
- Box, G.E.P., & Jenkins, G.M. (1976). *Time series analysis, forecasting and control*. San Francisco: Holden-Day.
- Bradley, J.V. (1968). *Distribution-free statistical tests*. Englewood Cliffs, NJ: Prentice Hall.
- Bryman, A., & Cramer, D. (1990). *Quantitative data analysis for social scientists*. London: Routledge.
- Busk, P.L., & Marascuilo, L.A. (1992). Statistical analysis in single-case research: Issues, procedures, and recommendations, with applications to multiple behaviors. In T.R.Kratochwill & J.R.Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 159–185). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Campbell, D.T., & Stanley, J.C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Chen, R.S., & Dunlap, W.P. (1993). SAS procedures for approximate randomization tests. *Behavior Research Methods, Instruments, & Computers*, 25, 406–409.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dugard, P., & Todman, J. (1995). Analysis of pretest-posttest control group designs. *Educational Psychology*, 15, 181–198.
- Edgington, E.S. (1969). Approximate randomization tests. *The Journal of Psychology*, 72, 143–149.
- Edgington, E.S. (1975). Randomization tests for one-subject operant experiments. *The Journal of Psychology*, 90, 57–68.
- Edgington, E.S. (1984). Statistics and single case analysis. In M.Hersen, R.M. Eisler, & P.M.Miller (Eds.), *Progress in behavior modification* (Vol. 16, pp. 83–119) Orlando, FL: Academic.
- Edgington, E.S. (1987). *Randomization tests* (2nd ed.). New York: Dekker.
- Edgington, E.S. (1992). Nonparametric tests for single-case experiments. In T. R.Kratochwill & J.R.Levin (Eds.), *Single-case research design and analysis* (pp. 133–157). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Edgington, E.S. (1995). *Randomization tests* (3rd ed.). New York: Dekker.
- Edgington, E.S. (1996). Randomized single-subject experimental designs. *Behaviour Research and Therapy*, 34, 567–574.
- Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. Philadelphia: Society for Industrial and Applied Mathematics.
- Ferron, J., & Onghena, P. (1996). The power of randomization tests for single-case phase designs. *The Journal of Experimental Education*, 64, 231–239.
- Ferron, J., & Ware, W. (1994). Using randomization tests with responsive singlecase designs. *Behaviour Research and Therapy*, 32, 787–791.
- Ferron, J., & Ware, W. (1995). Analyzing single-case data: The power of randomization tests. *The Journal of Experimental Education*, 63, 167–178.
- Fisher, R.A. (1935). *The design of experiments*. Edinburgh, Scotland: Oliver & Boyd.

- Franklin, R.D., Allison, D.B., & Gorman, B.S. (1996). Introduction. In R.D. Franklin, D.B. Allison, & B.S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 1–11). Mahwah, NJ: Lawrence Erlbaum Associates.
- Franklin, R.D., Gorman, B.S., Beasley, T.M., & Allison, D.B. (1996). Graphical display and visual analysis. In R.D. Franklin, D.B. Allison, & B.S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 119–158). Mahwah, NJ: Lawrence Erlbaum Associates.
- Gentile, J.R., Roden, A.H., & Klein, R.D. (1972). An analysis of variance model for the intrasubject replication design. *Journal of Applied Behavior Analysis*, 5, 193–198.
- Good, P. (1994). *Permutation tests: A practical guide to resampling methods for testing hypotheses*. New York: Springer-Verlag.
- Gorman, B.S., & Allison, D.B. (1996). Statistical alternatives for single-case designs. In R.D. Franklin, D.B. Allison, & B.S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 159–214). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hayes, A.F. (1998). SPSS procedures for approximate randomization tests. *Behavior Research Methods, Instruments, & Computers*, 30, 536–543.
- Howell, D.C. (1997). *Statistical methods for psychology* (4th ed.). Belmont, CA: Duxbury.
- Jeffreys, H. (1939). *Theory of probability*. Oxford, UK: Clarendon.
- Kazdin, A.E. (1973). Methodological and assessment considerations in evaluating reinforcement programs in applied settings. *Journal of Applied Behavior Analysis*, 6, 517–531.
- Kazdin, A.E. (1975). *Behavior modification in applied settings*. Homewood, IL: Dorsey.
- Kazdin, A.E. (1976). Statistical analyses for single-case experimental designs. In M. Hersen & D.H. Barlow (Eds.), *Single case experimental designs: Strategies for studying behavior change* (pp. 265–316). New York: Pergamon.
- Kazdin, A.E. (1980). Obstacles in using randomization tests in single-case experimentation. *Journal of Educational Statistics*, 5, 253–260.
- Kazdin, A.E. (1982). *Single-case research designs: Methods for clinical and applied settings*. London: Oxford University Press.
- Kazdin, A.E. (1984). Statistical analyses for single-case experimental designs. In D.H. Barlow & M. Hersen (Eds.), *Single case experimental designs: Strategies for studying behavior change* (2nd ed., pp. 285–324). New York: Pergamon.
- Levin, J.R., Marascuilo, L.A., & Hubert, L.J. (1978). N=nonparametric randomization tests. In T.S. Kratochwill (Ed.), *Single subject research: Strategies for evaluating change* (pp. 167–196). New York: Academic.
- Lindley, D.V. (1965). *Introduction to probability and statistics from a Bayesian viewpoint*. Cambridge, UK: Cambridge University Press.
- Manly, B.F.J. (1991). *Randomization and Monte Carlo methods in biology*. London: Chapman & Hall.
- Marascuilo, L.A., & Busk, P.L. (1988). Combining statistics for multiple-baseline AB and replicated ABAB designs across subjects. *Behavioral Assessment*, 10, 1–28.
- Matyas, T.A., & Greenwood, K.M. (1996). Serial dependency in single-case time series. In R.D. Franklin, D.B. Allison, & B.S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 215–243). Mahwah, NJ: Lawrence Erlbaum Associates.
- May, R.B., Masson, M.E.J., & Hunter, M.A. (1990). *Application of statistics in behavioral research*. New York: Harper & Row.
- Minium, E.W., King, B.M., & Bear, G. (1993). *Statistical reasoning in psychology and education* (3rd ed.). New York: Wiley.
- Neyman, J., & Pearson, E.S. (1967). *Collected joint statistical papers*. Cambridge, UK: Cambridge University Press.
- Noreen, E.W. (1989). *Computer-intensive methods for testing hypotheses: An introduction*. New York: Wiley.

- Ongheña, P. (1992). Randomization tests for extensions and variations of ABAB single-case experimental designs: A rejoinder. *Behavioral Assessment, 14*, 153–171.
- Ongheña, P. (1994). *The power of randomization tests for single-case designs*. Unpublished doctoral dissertation, Katholieke Universiteit Leuven, Belgium.
- Ongheña, P., & Edgington, E.S. (1994). Randomized tests for restricted alternating treatments designs. *Behaviour Research and Therapy, 32*, 783–786.
- Ongheña, P., & May, R.B. (1995). Pitfalls in computing and interpreting randomization  $p$  values: A commentary on Chen and Dunlap. *Behavior Research Methods, Instruments, & Computers, 27*, 408–411.
- Ongheña, P., & Van Damme, G. (1994). SCRT1.1: Single case randomization tests. *Behavior Research Methods, Instruments, & Computers, 26*, 369.
- Ongheña, P., & Van Den Noortgate, W. (1997). Statistical software for microcomputers: Review of StatXact-3 for Windows and LogXact (version 2) for Windows. *British Journal of Mathematical and Statistical Psychology, 50*, 370–373.
- Parsonson, B.S., & Baer, D.M. (1978). The analysis and presentation of graphic data. In T.R.Kratochwill (Ed.), *Single subject research: Strategies for evaluating change* (pp. 101–165). New York: Academic.
- Pitman, E.J.G. (1937). Significance tests which maybe applied to samples from any populations. *Journal of the Royal Statistical Society: Section B, 4*, 119–130.
- Primavera, L.H., Allison, D.B., & Alfonso, V.C. (1996). Measurement of dependent variables. In R.D.Franklin, D.B.Allison, & B.S.Gorman (Eds.), *Design and analysis of single-case research* (pp. 41–91). Mahwah, NJ: Lawrence Erlbaum Associates.
- Remington, R. (1990). Methodological challenges in applying single case designs to problems in AAC. In J.Brodin & E.Bjorck-Akesson (Eds.), *Proceedings from the first ISAAC Research Symposium in Augmentative and Alternative Communication* (pp. 74–78). Stockholm: Swedish Handicap Institute.
- Shaughnessy, J.J., & Zechmeister, E.B. (1994). *Research methods in psychology* (3rd ed.). New York: McGraw-Hill.
- Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychological research*. New York: Basic Books.
- Siegel, S., & Castellan, N.J., Jr. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). Singapore: McGraw-Hill.
- Skinner, B.F. (1956). A case history in scientific method. *American Psychologist, 11*, 221–233.
- Skinner, B.F. (1966). Operant behavior. In W.K.Honig (Ed.), *Operant behavior: Areas of research and application* (pp. 12–32). New York: Appleton-Century-Crofts.
- Todman, J., & Dugard, P. (1999). Accessible randomization tests for single-case and small- $n$  experimental designs in AAC research. *Augmentative and Alternative Communication, 15*, 69–82.
- Wampold, B.E., & Furlong, M.J. (1981). Randomization tests in single-subject designs: Illustrative examples. *Journal of Behavioral Assessment, 3*, 329–341.
- Wilcox, R.R. (1987). *New statistical procedures for the social sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.