Single-case and Small-*n* Experimental Designs

A Practical Guide to Randomization Tests

Second Edition

Pat Dugard, Portia File,
and Jonathan Todman

Single-case and Small-*n* Designs

A Practical Guide to Randomization Tests

Second Edition

http://www.psypress.com/singlecase-and-smalln-experimental-designs-9780415886932

Single-case and Small-*n* Designs

A Practical Guide to Randomization Tests

Second Edition

Pat Dugard
Dundee University, UK

Portia File
University of Abertay Dundee, UK

Jonathan Todman
NHS, Scotland

Routledge
Taylor & Francis Group
New York London
http://www.psypress.com/singlecase-and-smalln-experimental-designs-9780415886932

Routledge
Taylor & Francis Group
711 Third Avenue
New York, NY 10017

Routledge
Taylor & Francis Group
27 Church Road
Hove, East Sussex BN3 2FA

Routledge is an imprint of Taylor & Francis Group, an Informa business

Printed in the United States of America on acid-free paper
Version Date: 20110805

International Standard Book Number: 978-0-415-88622-2 (Hardback) 978-0-415-88693-2 (Paperback)

Library of Congress Cataloging-in-Publication Data

Dugard, Pat.
Single-case and small-n experimental designs : a practical guide to randomization tests, second edition / Pat Dugard, Portia File, Jonathan Todman. -- 2nd ed.
p. cm.
Rev. ed. of: Single-case and small-n experimental designs / John B. Todman, Pat Dugard. 2001.
Summary: "Randomization tests are not a new idea, but they only became really useful after the advent of fast computing. Making randomization tests accessible to many more potential users by providing the means to use them within familiar statistical software, this book serves as an introduction and provides macros to perform in the familiar environments of SPSS and Excel. Though we expect that the book will still appeal to researchers, we believe the changes in the new edition will make the book an essential aid for graduate and senior undergraduate courses in statistics, data analysis, and/or research methods, taught in departments of psychology (especially clinical or counseling psychology), medicine, nursing, and other health and social sciences"-- Provided by publisher.
Includes bibliographical references and index.
ISBN 978-0-415-88693-2 (pbk.) -- ISBN 978-0-415-88622-2 ()
1. Statistical hypothesis testing. 2. Experimental design. I. File, Portia. II. Todman, John B. III. Todman, John B. Single-case and small-n experimental designs. IV. Title.

QA277.T63 2011
519.5'6--dc23 2011028087

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the Psychology Press Web site at
http://www.psypress.com

Dedication

To the memory of John Todman

Contents

Preface

This book is intended as a practical guide for students and researchers who are interested in the statistical analysis of data from single-case or very small-*n* experiments. We said this in the preface to the first edition, and it is true of this new edition as well. Randomization tests are not a new idea, but they became really useful only after the advent of fast computing. John Todman, lead author of the first edition of this book, wanted to make randomization tests accessible to many more potential users by providing the means to use them within familiar statistical software. As such, this book serves as an introduction to randomization tests and provides macros to perform some of them in the familiar environments of SPSS and Excel, which are already being used for data analysis by thousands of researchers. Minitab was dropped from this edition, because we believe fewer people are using it now, but the macros (plus a new one) for use with SPSS and Excel are still a central feature of the book. As well as the macros, we provide all the information you need to apply them to your own data.

Because most researchers are not familiar with randomization designs and tests, even if they are quite confident with data analysis and statistical inference, we have described the basic ideas of randomization tests using several extended examples. We also use examples to show how to choose a suitable design and, having made a choice, how to implement it and analyze the results. The emphasis throughout is on practical application and understanding when and how to use randomization tests. There is a book Web site (http://www.researchmethodsarena.com/9780415886932) from which you can take the macros and example data. There is a separate site for instructors only that contains the solutions to the exercises that are not included in the text.

If you are among that majority of researchers and students who are unfamiliar with randomization tests, studying this book will enable you to understand how randomization tests work, why they can be used in

many investigations that have too few participants for the familiar *t* tests or ANOVAs, and why their use had to await readily available fast computing. You will also learn how to choose, implement, and analyze randomization designs. We provide the tools, instructions, and demonstration examples for analysis in the familiar environment of SPSS or Excel. You will be able to apply these techniques to your own investigations and data.

New to this edition

In the years since the original publication, many readers have contacted us with queries, and we realized we had not gone far enough in explaining randomization tests to those who need to use them. So for this edition, we have reordered and revised most of the text and introduced many new examples from education and clinical research in psychology and medicine.

We have added exercises at the end of Chapters 2 to 7. Half of the solutions are provided at the back of the book for self-study, and the other half are on the Instructor's Teaching Resource Web site (http://www.researchmethodsarena.com/9780415886932).

We have added an appendix to help with skills such as choosing a random number, arranging objects in random order, listing and counting possible intervention points for phase designs, and checking for serial correlation. Many readers may not have used these skills much but are more familiar with taking random samples or random assignments to different conditions.

We have also added a glossary. The first time a technical term is used, it appears in italics, and a brief definition can be found in the glossary (but we also occasionally use italics for emphasis).

The first edition provided the macros on a CD, but for this edition they can be taken from the book Web site along with example datasets. Readers of the first edition will see that we have rationalized the names of the macros, and the example data file names match them. We use screenshots to show you how to use the macros and understand the results. We expect that the more extended explanations, wider range of examples, exercises, skills appendix, glossary, screenshots, and improved labeling of macros and datasets will make this new edition a much better tool for students and researchers who need randomization tests.

Though we expect that the book will still appeal to researchers, we believe the changes in the new edition will make the book an essential aid for graduate and senior undergraduate courses in statistics, data analysis, and/or research methods, taught in departments of psychology (especially

clinical or counseling psychology), medicine, nursing, and other health and social sciences. We hope that clinicians, who are among those most likely to need single-case or small-*n* designs, will be empowered to use these techniques in their own work. For more ambitious readers who may want to adapt our macros or write their own, we have extended and improved the explanations for the macros, and (in Chapter 8) we have demonstrated how to tackle such an adaptation or extension. The macro listings with the new explanations are now in Appendices 2 and 3, one each for SPSS and Excel. We have rationalized the variable names in the macros, so the explanations are easier to follow. There is a new macro for a small-*n* repeated measures design with replicates, which we wrote in response to a reader query. Macro listings for you to use, without the explanations, are on the book Web site along with the example datasets so you don't have to type them out yourself.

We have done some updating for this edition, but the main difference you will notice is that we have made much more effort to help you to see how you might use the randomization tests we demonstrate and understand how they work. We assume some knowledge of ordinary statistical work and a little knowledge of nonparametric tests would be useful. A basic introduction to SPSS or Excel is also assumed so that you can use the macros for your own data, but we provide step-by-step instruction in using the macros in Chapter 4 and the Appendices.

Contents

Chapter 1 puts randomization tests into the context of investigations with human participants to show how they can fit into the research process. Of course, they can be used in other investigations, but it is when we have human participants that we are most likely to have a single case or just a few.

Chapter 2 demonstrates the main ideas used in the rest of the book with some examples of randomization tests. For many readers, these ideas will be new. Instead of basing inferences on random sampling, we use random assignment in the design and random reassignments for comparison with the experimental results.

In Chapter 3, we consider the ways in which our choice of design may be constrained, how we address internal and external validity in randomization designs, and how different methods of randomization may be used. We show how to choose a suitable design and list the key features of the designs covered.

Chapter 4 shows how to implement the chosen design. This includes how to do the randomization, and where necessary draws attention to potentially inappropriate randomization methods.

Chapter 5 demonstrates the analysis using the macros. There are instructions on using them for those who may not have used macros before, and screenshots to show the results using example data. We also show how the results could be reported.

Chapter 6 shows how randomization tests fit into the body of statistical inference, and also why using a randomization design may be worth the effort even if you intend to use only a graphical analysis.

Chapter 7 discusses power. This can be problematical for phase designs, and they receive special consideration.

Chapter 8 is for those who want to take these ideas further and perhaps even write their own macros. We use the process of writing our new macro (for a small-*n* repeated measures design with replicates) to demonstrate how the macro listings and explanations in the appendices can enable readers to extend the range of designs in the book to include special problems of their own.

Appendix 1 gives useful tips for anyone not familiar with using random numbers and also offers help with other relevant skills that you may not have needed for ordinary statistical work. Appendices 2 and 3 list the macros with explanations for anyone contemplating more ambitious projects.

Acknowledgments

We want to thank the University of Dundee for support for this project, especially Trevor Harley, head of the Department of Psychology; Mark Bennett, who supported the appointment of one of us as honorary lecturer; and John Morris, whose technical support has always gone beyond the call of duty. We are also grateful to our friend and colleague Harry Staines who read some chapters, made useful comments, and gave us an example used in Chapter 7. Many readers have contacted us over the years with queries, and improvements in clarity are largely because you made us see the need for it! Among readers, we must mention Inna Tsirlin of York University in Canada, who prompted us to write a new macro and kindly allowed us to illustrate it with her example (which we have simplified a bit). Routledge/Taylor Francis reviewers Marie S. Hammond (Tennessee State University), Michele Ennis Soreth (Rowan University), Kimberly J. Vannest (Texas A & M University), and one anonymous reviewer made

helpful suggestions, but any deficiencies remaining are our own. Andrea Zekus at Routledge/Taylor Francis helped us by always answering our questions about the production and so smoothed the process. Most of all we, are grateful to John Todman for many years of inspiration. This edition owes a great deal to him, not only for his work on the first edition but also for his vision and enthusiasm. Before his death in 2009, he said how much he wished he could work on a new edition that would be accessible to a wider audience.

Pat Dugard

Portia File

Jonathan Todman

chapter three

Obtaining the data
Choosing the design

Introduction

In this chapter we consider how to go about choosing a design for the problem we want to investigate. There are advantages and disadvantages to any design, and whatever we are investigating, we want to find a design that will address the question as directly as possible and that will have few disadvantages. We have to consider any factors that limit our choice of design, and if several types of design could be suitable, we need to be sure we know the advantages and disadvantages of each. In this introductory section, we briefly describe the classes of designs. Later sections will give the conditions in which each design may be used.

We are assuming in this book that the first limitation on our choice of design is that we don't have large numbers of participants, so we are unable to use the well-known formats and huge knowledge base available for large-sample designs. This limitation is implicit in the whole chapter, and indeed throughout the book, but in this introduction we look at some other possible limitations that may determine which designs are available to us.

Although we have said that having only one or a small group of participants is a limitation on our choice of design, we should point out that sometimes there is particular value in finding out what happens to a few individual cases. Large-sample designs tell us about average effects, and there will always be variation in individual response. So we can take a positive view of single-case and small-*n* designs: They teach us about individuals and have an important place in research in the human sciences, as we explained in Chapter 1.

Many years of development have given us large-sample designs that ensure our results will not be tainted by observer bias or group selection bias, or be swamped in uncontrolled random variation. In the less well-trodden ways of small-*n* and single-case designs, we need to make sure we have attended to these issues. Because of this, we give a brief account of *internal* and *external validity* here in the introduction, but we also mention

validity when describing those particular designs for which validity may be problematic. We revisit this topic in Chapter 6.

Types of experimental design

We consider three classes of experimental design, as we explained in Chapter 2. The first class comprises those designs that are analogs of ANOVA. These are the most powerful designs and will be the best choice if we can find one that is suitable. Unfortunately, as the next section explains, there are plenty of situations where none of them is appropriate.

Phase designs are our next group, and as we shall see, one of them may be the design of last resort. It is often hard to ensure internal validity, as we explain in the case of the AB design below. These designs usually also have low power (in other words, they may have a high probability of missing an effect that is real), an issue we discuss in Chapter 7. Nevertheless, one of these designs could be the way to move an investigation forward, and in that case it is important to make sure the design is as good as we can make it.

We also consider order effects, though designs to examine these can sometimes be criticized as not being proper randomization designs. We show the kind of problem to which they are applied in this chapter, and we discuss their place in the scheme of randomization designs in Chapter 6.

Examples that can be used with either a small group or a single case are available for each of our classes of randomization designs: analogs of ANOVA, phase designs, and order effects. We now look at some features that may restrict our choice of design before concluding this introduction with a discussion of validity.

Limits to our choice of design

Our choice of design may often be limited by what kind of randomization is possible with the conditions we want to compare. Can the conditions we want to compare be randomly assigned to participants (if we have a small group) or to observation occasions (if we have a single case)? An example where the conditions could be randomly assigned would be the three methods of pain control in the single-case one-way design in Chapter 2. It does not matter if the conditions precede or follow one of the other methods of pain control because the pain control methods can be effective only while they are being used. They could therefore be assigned at random to the available observation periods for a single case, or if we had a group of eligible participants, the methods could be randomly assigned to them. A design analogous to ANOVA can usually be used if this kind of random assignment is possible.

There are some conditions or treatments that cannot be randomly assigned to observation periods. An example is an intervention such as surgery or a training course that has a long-lasting or permanent effect. If this is the subject of investigation, then the only option may be a before-and-after study, a phase design.

For some conditions or treatments, we may be in some doubt about whether random assignment to observation occasions is possible or appropriate. Our example for a phase design with a reversal phase in Chapter 2 is such a case. Here the conditions were normal food or a diet free from sugar and additives, and the observation occasions were school days. It would be possible to randomly assign the diet or normal food to each observation day, but there are two reasons why we may prefer not to do that if it can be avoided. First, it may be easier for our participant to stick to the diet for a block of consecutive days than for days assigned at random among normal food days. We do need his full compliance, so we should avoid making it harder than necessary for him. But also, it is possible that there is a short carry over effect from either a diet day or a normal food day into the next day. If there is, random assignment would allow this effect to obscure our results. The phase design we adopted would minimize any damage done to our results by a short carry over effect. In fact, if we believed such a thing might prejudice our results, we could decide, as part of the design, that the data from the first diet day and the first withdrawal day would be ignored.

As these examples illustrate, the first thing we have to consider when choosing a design is what limitations are imposed on the way we use randomization by the conditions or treatments we want to investigate. Other limitations may be imposed by our participants' ability or willingness to attend observation sessions, or record measurements or observations themselves. Some designs, especially phase designs, may require an unacceptably long series of observations. Cost can also impose limits on the number or type of observations that are possible.

Ethical considerations can also limit our choice of design, and the same constraints apply to small-*n* and single-case designs as to large-sample ones. Of particular relevance to us is the likelihood of being unable to use a phase design with reversal because if the participant becomes accustomed to, or benefits from, the intervention, then withdrawing it may be impossible.

Internal and external validity

When we set up an experiment to compare the effects of two or more conditions or treatments, we want to be able to say that any observed difference must be due to the different conditions or treatments, and cannot be

explained by such things as observer bias or preexisting differences among participants. This is the requirement that we attend to *internal validity*, and many threats to it have been uncovered over the years, including the placebo effect, unconscious or conscious observer bias, and unconsidered differences among participants. Even though randomization tests require fewer assumptions than parametric tests, careful consideration of possible threats to internal validity is just as important a part of choosing a design and setting up an experiment as it is for a large-sample experiment.

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 subjects of 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. This applies equally to large-sample experiments and the small-*n* and single-case experiments that are the subject of this book.

In the interests of internal validity, we often use "blinding" of participants or assessors to the condition being applied. Here is how it is used in a large randomized controlled trial, which has become the gold standard for comparing the efficacy of treatments. Eligible participants are randomly assigned to one of the treatments or else to the control group (which may receive an already established treatment). Normally, neither the participants nor their assessors know to which group they have been assigned, because many subtle effects have been found that are mediated by either the experimenters' or participants' beliefs about the treatments. So only the administrator, who never sees the participants, holds the key to who received which treatment. Sometimes considerable efforts must be made to make pills look alike or otherwise blind participants and experimenters to the treatment being received. This type of design generally removes any threat to internal validity.

With small groups or single cases, we usually have to modify the large-sample approach to achieving internal validity. In several of the descriptions below, we draw attention to how internal validity may be compromised and the kind of efforts we can make to improve it.

Choosing the randomization method

Before we consider example designs, there is one more general point to emphasize: When we do the randomization to obtain the reference set, we must use the same method as for setting up the experiment. The ESP example from Chapter 2 will show what we mean. We decided to assign four each

of conditions V and N randomly to the eight test periods. The randomization can easily be achieved by arranging four Vs and four Ns in random order as shown in Appendix 1. We could also toss a coin as shown in Chapter 2.

We chose to have equal numbers in the two conditions, but you may have wondered whether an equally good method would be to perform the coin toss for every test period and use the number of Vs and Ns thus obtained. This might make our ESP claimant's task harder, because she will no longer know that there are four of each, so perhaps this variant of the design would be an improvement. Remember that to set up the randomization test, we must calculate the test statistic for all the unused random assignments, the rest of the reference set. What would they be, and how many are there if we do it this way? Each of the eight test periods can be one of two conditions, so the first two have 2*2 possible assignments (VV, VN, NV, and NN). The first three have 2*2*2 possible assignments (VVV, VVN, VNV, NVV, VNN, NVN, NNV, and NNN). The eight test periods have $2^8 = 256$ possible assignments. Two of these would be all V or all N, so we may wish to reject them, leaving 254 possible assignments. We can obtain the reference set by listing them and counting the number correct for each assignment (compared with the one actually used), or else we can sample with replacement as described in Chapter 2. If we use the sampling method, generating each assignment by the computer equivalent of coin tossing, we need to make sure we reject the cases with all Vs or all Ns from the reference set if we exclude them from the design. This example illustrates the importance of deciding just how the randomization is to be done when choosing the design, and also how the reference set must be obtained using the same method of randomization.

In most circumstances, we shall want to collect equal or nearly equal numbers of observations in each condition, because the more nearly equal the numbers, the more powerful the test will be. However, for several designs the numbers do not *need* to be equal and may depend on constraints beyond our control. If one of the conditions is expensive or difficult to arrange, there may be a limit to the number of observations we can obtain for it, but it may be worth collecting a few more for the other conditions. In our example on pain control in Chapter 2, for instance, if the acupuncturist has a very full appointment book, we may be able to arrange only a small number of appointments within the time scale of the experiment. We now consider the uses, advantages, and disadvantages of the individual designs, grouped into their classes.

Designs analogous to ANOVA

If the treatments or conditions under investigation are suitable for random assignment to participants or observation occasions, then a design analogous to one of those used with ANOVA for large-*n* designs will give the

best opportunity to demonstrate any effect. This flexible class of designs enables us to compare two or more treatments or conditions. We can also incorporate repeated measures as well as consider two factors in a single experiment. This is as far as we go in this book, but randomization designs analogous to more complex ANOVAs could be devised using the five steps described in Chapter 2.

A single-case one-way design

This design is appropriate when the following conditions apply:

1. We have only one participant.
2. There are several available observation occasions that are all equivalent.
3. We have two or more conditions to compare.
4. It is possible to assign conditions to observation occasions at random.

Condition 2 would not be satisfied if, for instance, we are obliged to use both mornings and afternoons to obtain enough observation occasions, but we suspect that our participant will respond differently in the morning and afternoon. Likewise, it would not apply if we have to use two observers, and we suspect there may be slight differences in the way they make their assessments. If condition 2 does not apply, we need to consider a single-case randomized blocks design.

Condition 4 would not be satisfied if any of the conditions has a lasting or permanent effect, so that the next observation occasion would be affected by a condition applied previously. So this design might enable us to compare painkilling drugs, but not different surgical interventions. For investigating a long-lasting or permanent intervention, a phase design should be considered.

If there are only two conditions to compare, we can use a special case of this design that makes a one-tailed test possible. The macro is the same as the one for a small-*n* one-way design with two conditions and is described in Chapter 4.

As always with a single-case design, external validity will be achieved only after similar results are found with other participants in other contexts.

A small-n one-way design

This design is appropriate when the following conditions apply:

1. We have at least two participants.
2. Each participant will be measured once only.
3. We have two or more conditions to compare.
4. It is possible to assign conditions to participants at random.

Condition 2 will not apply if our investigation requires a series of two or more measurements on each participant, perhaps to record a process. In this case, a small-*n* repeated measures design should be considered. However, condition 2 could apply if the single measurement recorded for each participant is actually an average from several raters or observers.

Condition 4 of course implies that all participants are suitable and willing to be assigned to any of the conditions. Because each participant is observed only once, this time we may be less concerned if any of the conditions has a long-lasting effect.

If there are only two conditions to compare, we can use a special case of this design that makes a one-tailed test possible. This special case is described in Chapter 4 as a small-*n* one-way design with two conditions.

A single-case randomized blocks design

This design is appropriate when the following conditions apply:

1. We have only one participant.
2. The available observation occasions fall into two or more groups (days, for instance).
3. We have two or more conditions to compare.
4. It is possible to assign conditions to observation occasions at random.

We use a randomized blocks design if we are collecting data at several times, perhaps on different days, to reduce the chance of any treatment effect being obscured by variation among observation occasions. So, in case our participant has good and bad days, we can apply each treatment in random order within each day (block), so differences among treatments will be found within each block and then averaged over blocks. This is just the same idea that motivates randomized block designs for use with ANOVA.

As in the single-case one-way design, condition 4 will not apply if any of the conditions has a long-lasting or permanent effect.

If there are only two conditions to compare, we can use a special case of this design that makes a one-tailed test possible. This special case is described in Chapter 4 as a single-case randomized blocks design with two conditions.

As always with a single-case design, external validity will be achieved only after similar results are found with other participants in other contexts.

A small-n repeated measures design

This design is appropriate when the following conditions apply:

1. We have at least two participants.
2. We have two or more conditions to compare.

3. Each participant will receive each of the conditions.
4. It is possible to assign conditions in random order to each participant.

Because human participants are usually very variable, we may improve our ability to detect differences among the conditions by finding the differences within each participant and averaging the results. In this way, each participant acts as their own control. The same idea is used in repeated measures designs for large groups.

Condition 4 will not apply if any of the conditions has a long-lasting effect, just as in the single-case one-way and randomized blocks designs. This is again because a long-lasting effect would influence the next observation.

If there are only two conditions to compare, we can use a special case of this design that makes a one-tailed test possible. The macro is the same as the one used for the single-case randomized blocks design with two conditions, described in Chapter 4.

A small-n repeated measures design with replicates

This design is appropriate when the following conditions apply:

1. We have at least two participants.
2. We have two conditions to compare.
3. Each participant will receive each of the conditions on at least two occasions.
4. It is possible to assign conditions in random order to each participant.

As in the previous design, here each participant acts as his or her own control. As long as we have only two conditions to compare, we can use this design to improve our chance of finding a significant effect. Our macro assumes that we have equal numbers of observations in each condition and the same number of observations for each participant. The conditions will be assigned in random order to each participant.

Condition 4 will not apply if any of the conditions has a long-lasting effect, just as in the single-case one-way and randomized blocks designs. This is again because a long-lasting effect would influence the next observation.

It is possible to use the macro for this design for a randomized blocks design with replicates, and this use is outlined in Chapter 4.

A two-way factorial single-case design

Factorial designs allow us to investigate two or more variables in a single experiment. Here we consider only the two-way factorial, which means we consider two variables. An example would be an experiment to investigate the effectiveness of different types of painkilling drug and also

different relaxation methods (using music or exercise, perhaps). The two-way design described here allows just two levels for each variable (two drugs and two relaxation methods in the example just outlined).

This two-way factorial design is appropriate when the following conditions apply:

1. We have a single participant.
2. A series of observation occasions is available.
3. We have two factors to investigate (for instance, drug type and relaxation method).
4. There are two levels of each factor (for instance, drugs A and B, and relaxation methods X and Y).
5. It is possible to assign conditions in random order to observation occasions.

Notice that there are four conditions here (AX, BX, AY, and BY). We shall need equal numbers of observations for the four conditions, so will use 4, 8, 12, or some other multiple of four observation occasions. Condition 5 will not apply if any of the treatments has a long-lasting or permanent effect.

By observing a single participant, we remove the variability among participants, but as always with a single-case design, external validity will accumulate only as the experiment is repeated with different people in different situations.

A two-way factorial small-n design

This design is appropriate when the following conditions apply:

1. We have a multiple of four participants.
2. We have two factors to investigate (for instance, drug type and relaxation method).
3. There are two levels of each factor (for instance, drugs A and B, and relaxation methods X and Y).
4. It is possible to assign conditions at random to participants.

Each participant will receive one of the four conditions (AX, AY, BX, or BY). Because each is measured only once, condition 4 may still apply even if a condition has a long-lasting effect. Of course, all participants must be suitable and willing to receive any of the conditions.

Phase designs

Phase designs have their own particular problems with internal validity, because inevitably time passes during the study, and many things change with time. A phase design with *reversal* may help with this problem:

It is one where, after we have sufficient measurements in the intervention phase, we withdraw the intervention and continue taking measurements for the *withdrawal phase*. This is known as an ABA design. Of course, this is not always possible: There may be ethical reasons why we cannot withdraw a drug or therapy once offered, especially if the participant seems to be benefiting. Also our intervention may be irreversible, as would be the case if it took the form of surgery or a training session. However, evidence that any observed effect was due to the intervention may be strengthened if a reversal phase is possible.

Another problem with phase designs is that successive observations may be correlated, especially if the time between observations is short. Imagine recording blood pressure daily. Now imagine recording it every 10 minutes. If you record every 10 minutes, successive measurements are likely to be similar. If you record it daily, then successive measurements are less likely to be similar. If successive observations are similar, the effect is called *serial correlation*, and we show how to check for this in Appendix 1. Every effort should be made to ensure that the serial correlation is low, and this will mean that sufficient time must be allowed between observations. Another opportunity for reducing serial correlation concerns the way in which observations are made: When the same individual makes repeated measurements, there may be an increased chance of serial correlation. This might be avoided by assigning different raters at random to observation periods, but then unless interrater reliability is high we might introduce unnecessary extra variability into the data.

Phase designs qualify as randomization designs only if intervention and withdrawal points are chosen at random from those available. This is not the traditional way to implement a phase design, but doing the randomization not only allows us to use a randomization test but also removes a threat to internal validity. We expand on this point, and provide some evidence that even quite experienced researchers can be misled, in Chapter 6.

The single-case AB design

This design is appropriate when the following conditions apply:

1. We have a single participant.
2. We want to test the effectiveness of an intervention.
3. A series of observations is possible both before and after the intervention.
4. It is possible to choose the intervention point at random.

Conditions 3 and 4 will not apply if the participant urgently needs the intervention and cannot wait for a randomly assigned time to receive it.

This is the simplest of the phase designs, and the choice to use this design is often determined by the irreversible or long-lasting nature of the intervention being investigated: In fact, it may be the design of last resort. Nevertheless, it may move our investigation forward, and we need to make sure we do it as well as possible. Randomizing the intervention point is the first step.

The single-case ABA design

This design is appropriate when the following conditions apply:

1. We have a single participant.
2. We want to test the effectiveness of an intervention.
3. After sufficient observations are obtained, it will be possible to withdraw the intervention.
4. A series of observations is possible before the intervention, while the intervention continues, and after it is withdrawn.
5. It is possible to choose the intervention and withdrawal points at random.

Conditions 4 and 5 will not apply if the participant urgently needs the intervention, just as in the AB design. Condition 3 will not apply if the intervention is irreversible, such as surgery or training. It will also not apply if there are ethical objections to withdrawal.

If withdrawal is possible, the extra randomization will reduce the total number of observations needed. This effect can be seen in the examples used in Chapter 2 but will be shown in more detail in Chapter 4, where we implement the designs.

In addition to reducing the number of observations needed, the addition of a withdrawal phase may increase our confidence in the effectiveness of the intervention if the effect disappears when the intervention is withdrawn. The additional persuasive power of the withdrawal phase may be a strong reason to use this design if it is possible.

If a reversal phase is possible, then it may be useful to add a further intervention phase, giving us an ABAB design, and even more phases can be added. We can also have phase designs with two different interventions with a withdrawal phase in between, giving an ABAC design. You can think of more variants yourself, but multiple phases are more likely to lead to ethical problems, and implementation and analysis may also be more difficult. We discuss some of these possibilities further in Chapter 8.

The multiple baseline AB design

This design is appropriate when the following conditions apply:

1. We have at least two participants (a variant for one participant is given below).
2. We want to test the effectiveness of an intervention.
3. For each participant, a series of observations is possible both before and after the intervention.
4. It is possible to choose the intervention point at random for each participant.

Using the AB design with two or more participants will substantially reduce the number of observations needed on each participant, which may be a considerable advantage. However, it is important that the intervention point is chosen at random individually for each participant. Condition 4 will not apply if the intervention has to start at the same time for all participants.

If we find a significant effect, we shall be able to say only that the intervention produced a significant effect for at least one of the participants. We shall have to rely on visual inspection to suggest which ones showed the effect.

Conditions 3 and 4 will not apply if any participant needs the intervention urgently, just as in the single-case AB design.

This design may also be used with a single participant if we want to test the effectiveness of several interventions during a single series of observation occasions. An example of this use is given in Chapter 4.

The multiple baseline ABA design

This design is appropriate when the following conditions apply:

1. We have at least two participants (but see the note at the end for one participant).
2. We want to test the effectiveness of an intervention.
3. After sufficient observations are obtained, it will be possible to withdraw the intervention.
4. For each participant, a series of observations is possible before the intervention, while the intervention continues, and after it is withdrawn.
5. It is possible to choose the intervention and withdrawal points at random.

Using the ABA design with two or more participants will substantially reduce the number of observations needed on each participant, which may be an advantage. However, it is important that the intervention and withdrawal points are chosen at random individually for each participant. Condition 5 will not apply if intervention and withdrawal have to start at the same time for all participants.

If we find a significant effect, we shall only be able to say that the intervention produced a significant effect for at least one of the participants. We shall have to rely on visual inspection to suggest which.

Condition 3 will not apply if the intervention is irreversible or if there are ethical problems with withdrawal. Conditions 4 and 5 will not apply if any participant needs the intervention urgently.

As with the multiple baseline AB design, this design may be used with a single participant if we can take more than one type of measurement on each observation occasion, if each measurement has an associated intervention to be tested, and if the interventions can be made at individually chosen random points.

Advantages of phase designs

Phase designs allow us to investigate treatments or therapies with which we hope to help people who have rare or very individual disorders or disabilities. We may never have even a small group of people with a particular set of problems, much less the numbers needed for a conventional clinical trial. But still, we want to make as good and as objective an assessment as possible of any proposed intervention. A phase design may be the only option, and if so, using randomization of the intervention point gives the best hope of avoiding bias and demonstrating effectiveness using statistical significance.

Disadvantages of phase designs

Double-blind randomized controlled trials were developed to avoid the pitfalls that bedevil attempts to assess the effectiveness of treatments, including observer bias and the placebo effect. Phase designs usually are open to some of these problems. If the intervention is a drug, then the placebo effect may be countered if the participant can be given a dated strip of pills that all look alike, so the participant doesn't know when the intervention starts. Even then, the occurrence of side effects from the drug may let the participant know when the intervention starts. For other kinds of intervention such as training or surgery, where the participant inevitably knows when it occurs, we may not be able to distinguish an effect

from the passing effect of optimism or novelty. Considering this possibility may influence our choice of the minimum acceptable number of observations in the intervention phase, and it might help if we exclude from the analysis the first few observations immediately after the intervention. We may be able to reduce opportunities for observer bias by having someone unaware of when the intervention occurs taking the measurements.

The need for a long series of observations may be a problem, either because it is tedious for the participant or because the measurement may be affected by boredom or practice. If the necessity of avoiding highly correlated neighboring observations, or other constraints of the study, means that observations are quite widely spaced (e.g., perhaps once a week), then we have to consider whether the mere passage of time, over more than half a year perhaps, may bring about changes that will obscure the effect of our intervention.

Finally, if we want to use statistical tests on these designs, we have to be aware that they can be quite insensitive to small effects. In other words, we might not find enough evidence to support effects that are really there. We discuss this issue in Chapter 7 on size and power.

A design to investigate order effects

This design is appropriate when the following conditions apply:

1. We have a single participant and several observation occasions, or else several participants.
2. We want to test whether two variables are correlated.
3. Both variables are at least ordered.
4. It is possible to assign values of one of the variables at random to participants or observation occasions.

Condition 3 will not apply if one of the variables is something like POLITICAL AFFILIATION, with categories such as LABOUR, CONSERVATIVE, GREEN, LIBERAL, and OTHER, because these categories are not ordered. A variable such as FREQUENCY, with categories DAILY, WEEKLY, MONTHLY, and LESS THAN ONCE A MONTH, is ordered and so could be used here. Of course, any measurement is ordered and can also be used.

Condition 4 applies to something like number of training sessions, which could be randomly assigned to participants, but not to something like age, which is an unalterable property of a participant and so cannot be assigned at random.

Here is the kind of problem addressed by this design: One variable can be assigned at random (number of training sessions perhaps), and the other variable, which may be correlated, will be measured on each

participant (performance score on a task, for instance). Then, we can test whether or not the two variables are correlated or not.

It is possible to use this design even if there are ties in the ordering of one of the variables. For instance, instead of using the number of training hours, suppose we have just three training methods that are ordered by intensity. Assuming we have more than three participants, there will inevitably be ties in the ordering of the training method variable. If the methods can be randomly assigned to a small group of participants, then we can still investigate the possible association of training method with performance on a task. Our example in Chapter 2 was of this type. It is also possible to have ties on a variable because participants get equal scores or ratings on whatever is being observed.

Exercises

1. Why is it difficult to ensure internal validity for phase designs?
2. If we use a single-case or small-*n* design with no random sampling, how can external validity be achieved?
3. If we want to assess the effectiveness of a surgical intervention using a single participant, which designs may be suitable?
4. We need a design to compare four conditions. We shall be able to make only four observations per day. Why might we consider a randomized block rather than a one-way design?
5. If we have a small group of suitable participants who are willing to be randomly assigned to one of two conditions, either a surgical intervention or a period of observation with no treatment, which designs should we consider? Suggest a way in which we might encourage patients to participate, and also a way to improve internal validity.
6. If we are investigating the effectiveness of a possible treatment for a single participant, what considerations will influence our choice of design?
7. Are there conditions in which a multiple baseline phase design would be preferable to a small-*n* one-way or repeated measures design?