This is a chapter excerpt from Guilford Publications.
Quasi-Experimentation: A Guide to Design and Analysis.
Charles S. Reichardt. Copyright © 2019.
Purchase this book now: www.guilford.com/p/reichardt

1

Introduction

For as Hume pointed out, causation is never more than an inference; and any inference involves at some point the leap from what we see to what we can't see. Very well. It's the purpose of my Inquiry to shorten as much as humanly possible the distance over which I must leap. . . .

—Barth (1967, p. 214)

Although purely descriptive research has an important role to play, we believe that the most interesting research in social science is about questions of cause and effect.

—Angrist and Pischke (2009, p. 3)

The theory of quasi-experimentation . . . is one of the twentieth century's most influential methodological developments in the social sciences.

—Shadish (2000, p. 13)

Copyright © 2019 The Guilford Press

Overview

Researchers often assess the effects of treatments using either randomized experiments or quasi-experiments. The difference between the two types of designs lies in how treatment conditions are assigned to people (or other study units). If treatment conditions are assigned at random, the design is a randomized experiment; if treatment conditions are not assigned at random, the design is a quasi-experiment. Each design type has its place in the research enterprise, but quasi-experiments are particularly well suited to the demands of field settings. This volume explicates the logic underlying the design and analysis of quasi-experimentation, especially when they are implemented in field settings.

1.1 INTRODUCTION

We frequently ask questions about the effects of different treatments. Will attending college produce more income in the long run than attending a trade school? Which form of psychotherapy most effectively reduces depression? Which smoking cessation

program leads to the greatest reduction in smoking? Will this innovative reading program help close the gap in reading abilities between preschool children from high- and low-socioeconomic strata? How much does this criminal justice reform reduce recidivism among juveniles? And so on.

As these examples suggest, estimating the effects of treatments or interventions is of broad interest. Indeed, the task of estimating effects is of interest across the entire range of the social and behavioral sciences, including the fields of criminology, economics, education, medicine, political science, public health, public policy, psychology, social work, and sociology. In all of these fields, estimating effects is a mainstay in both basic research and applied research.

One of the primary tasks of **basic research** is testing theories—where a theory is tested by identifying its implications and making empirical observations to see if the implications hold true. And some of the most significant implications of theories involve predictions about the effects of treatments. So testing theories often involves estimating the effects of treatments. For example, consider tests of the theory of cognitive dissonance, a theory that specifies that when beliefs and behaviors are incongruent, people will change their beliefs to bring the beliefs into accord with the behaviors (Festinger, 1957). In a classic test of this theory, Aronson and Mills (1959) offered study participants membership in a discussion group if they would perform a disagreeable task (i.e., reading obscene material out loud). As predicted by the theory of cognitive dissonance, Aronson and Mills found that performing the disagreeable task increased the participants' liking for the discussion group. The idea was that cognitive dissonance would be aroused by performing the disagreeable task unless membership in the discussion group was viewed as desirable. So Aronson and Mills's research tested the theory of cognitive dissonance by estimating the effects of a treatment (performing a disagreeable task) on an outcome (liking for the group). In ways such as this, estimating the effects of treatments often plays a central role in theory testing.

Applied research is also concerned with estimating effects, though not always in the service of testing theories. Instead, applied research in the social sciences most often focuses on assessing the effects of programs intended to ameliorate social and behavioral problems. For example, in the service of finding treatments that can improve people's lives, economists have estimated the effects of job training programs on employment; educators have estimated the effects of class size on academic performance; and psychologists have assessed the effectiveness of innovative treatments for substance abuse. Applied researchers want to know what works better, for whom, under what conditions, and for how long—and this involves estimating effects (Boruch, Weisburd, Turner, Karpyn, & Littell, 2009).

The present volume is concerned with the task of estimating the effects of treatments whether for testing theories or ameliorating social problems. Of course, assessing treatment effects is not the only task in the social and behavioral sciences. Other central research tasks include discovering intriguing phenomena and devising theories to explain them. The fact that other tasks are important in no way diminishes the

Copyright © 2019 The Guilford Press

importance of the task of assessing the effects of treatments. Without knowing the effects of treatments, we cannot know if our theories of behavior are correct. Nor can we know how to intervene to improve the human condition. The point is that we need to understand the effects of treatments if we are to have a good understanding of nature and function in the world. This book will show you how to estimate the effects of treatments using quasi-experiments.

1.2 THE DEFINITION OF QUASI-EXPERIMENT

As will be explained in greater detail in Chapter 2, estimating the effects of treatments requires drawing comparisons between different treatment conditions. Often, a comparison is drawn between a treatment condition and a no-treatment condition; but the comparison could instead be drawn between two alternative treatments. For example, a comparison could be drawn between an innovative treatment and the usual standard-of-care treatment.

Comparisons used to estimate the effects of treatments can be partitioned into two types: **randomized experiments** and **quasi-experiments** (Shadish, Cook, & Campbell, 2002). The difference has to do with how people (or other observational units such as classrooms, schools, or communities) are assigned to treatment conditions. In randomized experiments, study units are assigned to treatment conditions at random. Assigning treatment conditions at random means assigning treatments based on a coin flip, the roll of a die, the numbers in a computer-generated table of random numbers, or some equivalently random process. In quasi-experiments, units are assigned to treatment conditions in a nonrandom fashion, such as by administrative decision, self-selection, legislative mandate, or some other nonrandom process. For example, administrators might assign people to different treatment conditions based on their expectations of which treatment would be most effective for people with different characteristics. Alternatively, people might self-select treatments based on which treatment appears most desirable or most convenient.

I use the label **experiment** to refer to any randomized or quasi-experiment that estimates the effects of treatments. The label of experiment is sometimes used either more broadly or more narrowly. On the one hand, experiments sometimes are defined more broadly to include studies where no attempt is made to estimate effects. For example, a demonstration to show that an innovation can be successfully implemented might be called an experiment even if there is no attempt to assess the effects of the innovation. On the other hand, experiments are sometimes defined more narrowly, such as when the term is restricted to studies in which an experimenter actively and purposefully intervenes to implement a treatment (whether at random or not). Such usage would exclude what are called "natural" experiments which can arise, for example, when nature imposes a hurricane or earthquake or when ongoing social conventions, such as lotteries, serve to introduce interventions. And experiment is sometimes narrowly used

Copyright © 2019 The Guilford Press

to be synonymous with randomized experiments. My usage is neither so broad nor so narrow. To me, an experiment is any attempt to estimate the effect of a treatment or an intervention using an empirical comparison. An experiment could involve an experimenter actively implementing a treatment, but the rubric of experiment also includes natural experiments. In any case, the exact usage of experiment is not as important as the distinction between randomized and quasi-experiments, both of which are experiments according to my nomenclature.

As another aside, randomized experiments are sometimes called "true" experiments. I avoid the label "true," for it suggests that alternatives to randomized experiments (i.e., quasi-experiments) are pejoratively false. Quasi-experiments are not the same as randomized experiments, but they are not false in any meaningful sense of the word. In addition, although they are often used interchangeably, I will not use the label randomized clinical trials (RCTs) instead of randomized experiment simply because RCT suggests to some readers an unnecessary restriction to medical or other health uses. For all intents and purposes, however, RCTs and randomized experiments mean the same thing.

Regardless of names and labels, the focus of this book is on quasi-experiments and especially on the logic and practice of quasi-experiments in field settings. However, this volume also considers randomized experiments. A researcher cannot fully appreciate quasi-experiments without reference to randomized experiments, so differences between the two are cited throughout the book. One of the first chapters is devoted to randomized experiments to provide a baseline with which to draw distinctions and lay the groundwork for the presentation of quasi-experiments.

Copyright © 2019 The Guilford Press

1.3 WHY STUDY QUASI-EXPERIMENTS?

Randomized experiments are generally considered the gold standard for estimating effects, with quasi-experiments being relegated to second-class status (Boruch, 1997; Campbell & Stanley, 1966). So why do we need quasi-experiments when randomized experiments hold such an exalted position? One answer is that randomized experiments cannot always be implemented, much less implemented with integrity (West, Cham, & Liu, 2014). Implementing a randomized experiment would often be unethical. For example, it would be unethical to assess the effects of HIV by assigning people at random to be infected with the virus. Nor would it be ethical for researchers to randomly assign children to be physically abused or for couples to divorce. Even if we were to randomly assign children to be physically abused, we might not be sufficiently patient to wait a decade or two to assess the effects when the children become adults. Similarly, it is impractical, if not impossible, to assess the effects of such massive social interventions as recessions or wars by implementing them at random. Even when random assignment would be both ethical and physically possible, it can be difficult to convince both administrators and prospective participants in a study that randomized

experiments are desirable. Or funding agencies might require that investigators serve all of those most in need of a presumed ameliorative intervention, so that none of those most in need could be relegated to a presumed less effective comparison condition. People sometimes perceive random assignment to be unfair and therefore are unwilling to condone randomized experiments. Sometimes, too, data analyses are conducted after the fact when a randomized experiment has not been implemented or when a randomized experiment was implemented but became degraded into a quasi-experiment.

Even though randomized experiments can be superior to quasi-experiments in theory, they are not always superior in practice. Strict controls that can often be imposed in laboratory settings can enable randomized experiments to be implemented with high fidelity. But field settings often do not permit the same degree of control as in the laboratory; as a result, randomized experiments can become degraded when implemented in the field. For example, randomized experiments can be degraded when participants fail to comply with the assigned treatment conditions (which is called **noncompliance**) or drop out of the study differentially across treatment conditions (which is called differential **attrition**). Under some conditions, quasi-experiments can be superior to such corrupted randomized experiments. That is, it can sometimes be better to implement a planned quasi-experiment than to salvage a randomized experiment that has been corrupted in unplanned and uncontrolled ways.

Finally, even though randomized experiments are often superior to quasi-experiments, they are not perfect. Even well-implemented randomized experiments have weaknesses as well as strengths. Their strengths and weaknesses often complement the strengths and weaknesses of quasi-experiments. Science benefits from an accumulation of results across a variety of studies. Results accumulate best when the methods used to create knowledge are varied and complementary in their strengths and weaknesses (Cook, 1985; Mark & Reichardt, 2004; Rosenbaum, 2015b; Shadish, Cook, & Houts, 1986). Hence, the results of randomized experiments combined with quasi-experiments can be more credible than the results of randomized experiments alone (Boruch, 1975; Denis, 1990). Bloom (2005a, p. 15) expressed the idea in applied terms: "combining experimental and nonexperimental statistical methods is the most promising way to accumulate knowledge that can inform efforts to improve social interventions." Because randomized experiments are more often degraded in field settings than in the laboratory, quasi-experiments often best complement randomized experiments in field settings. Indeed, the strengths of the one particularly well offset the weaknesses of the other in field settings.

Randomized experiments are generally preferred to quasi-experiments in the laboratory and they are relatively more difficult to implement in the field than are quasi-experiments. For these reasons, quasi-experiments are more common in the field than in the laboratory (Cook & Shadish, 1994; Shadish & Cook, 2009), and the examples in this book are drawn mostly from the use of quasi-experiments in field settings. Nonetheless, the theory of quasi-experimentation is the same in both field and laboratory settings. What follows applies to both equally.

Copyright © 2019 The Guilford Press

1.4 OVERVIEW OF THE VOLUME

This volume explains the logic of the design of quasi-experiments and the analysis of the data they produce to provide the most credible estimates of treatment effects that can be obtained under the many demanding constraints of research practice. The volume brings together the insights of others that are widely scattered throughout the literature. In addition, a few of my own insights are added that come from various locations in the literature. In this way, this work provides a compendium of material that would take substantial effort to assemble otherwise. In many cases, this volume makes this material easier to understand than if you read the original literature on your own. The purpose of this book is to provide accessible and helpful guidance to practitioners and methodologists alike.

Although estimating the size of effects can be highly quantitative and statistical, the presentation is directed to those who have no more than a basic understanding of statistical inference and the statistical procedure of multiple regression (and have taken an undergraduate course in research methods). Even then, elementary statistical and methodological topics are reviewed when it would be helpful. All told, the presentation relies on common sense and intuition far more than on mathematical machinations. I emphasize the conceptual foundation of quasi-experimentation so that readers will be well equipped to explore the more technical literature for themselves. When I have a choice to cite either a more or a less technical reference, I favor the less technical reference for its ease of understanding.

When I describe statistical procedures, please keep in mind that almost always multiple analysis strategies can be applied to data from any given quasi-experimental design (and new approaches are seemingly being developed every day). It would be impossible to present all the current possibilities, much less anticipate future developments. My purpose is to present the most common and basic analyses that will provide readers the framework they will need to understand both alternative variations and more sophisticated techniques—as well as future advances.

Chapter 2 defines a treatment effect and provides the background and conventions that undergird the ensuing presentation. A **treatment effect** is defined as a counterfactual difference between potential outcomes. Such a comparison is impossible to obtain in practice, so the definition of a treatment effect establishes the hurdles over which one must leap to draw credible causal inferences.

Chapter 3 explains that the effect of a treatment is a function of five **size-of-effect factors:** the treatment or cause; the participants in the study; the times at which the treatments are implemented and effects are assessed; the settings in which the treatments are implemented and outcomes assessed; and the outcome measures used to estimate the effects of the treatment. These five size-of-effect factors play a prominent role in distinguishing among types of quasi-experiments and in supplementing quasi-experimental designs to better withstand **threats to validity,** especially threats to internal validity (which are alternative explanations for obtained results). The five

Copyright © 2019 The Guilford Press

size-of-effect factors are both the fundamental elements of an effect size and the fundamental components in the design of comparisons to estimate effects. Chapter 3 also introduces four types of validity and the notion of threats to validity. **Construct validity** is concerned with correctly labeling the five size-of-effect factors in a causal relationship. **Internal validity** is a special case of construct validity and is concerned with influences that are confounded with treatment assignment. **Statistical conclusion validity** addresses two questions: (1) Is the degree of uncertainty that exists in the estimate of a treatment effect correctly represented? and (2) Is that degree of uncertainty sufficiently small (i.e., is the estimate of the treatment effect sufficiently precise, and is a test of statistical significance sufficiently powerful?)? **External validity** concerns the generalizability of the study results.

Chapter 4 introduces randomized experiments because they serve as a benchmark with which to compare quasi-experiments and because randomized experiments can become degraded into quasi-experiments. As already noted, noncompliance to treatment assignment and differential attrition from treatment conditions can degrade randomized experiments. The means of coping with both types of degradation are presented (which can be considered part of the theory of quasi-experimentation). **Selection differences** are also addressed as a threat to internal validity.

Chapter 5 switches gears. While Chapter 4 concerns what is often said to be the gold standard of causal research design (i.e., the randomized experiment), Chapter 5 begins the discussion of alternatives to randomized experiments by starting with a design that is not even experimental because it does not entail an explicit comparison of treatment conditions. It is important to consider such a **pre-experimental design** because, despite its weaknesses, it is still used even though it is usually not considered acceptable research practice.

Chapters 6–9 present four prototypical quasi-experiments. In general, the presentation proceeds from the least to the most credible designs. Chapter 6 introduces the **pretest–posttest design** which is susceptible to what is often a debilitating range of biases, although the design can be used to good effect under limited circumstances. Chapter 7 introduces the **nonequivalent group design,** which is one of the most widely used quasi-experiments. Chapters 8 and 9 introduce the **regression discontinuity design** and the **interrupted time-series design**, respectively, which are the quasi-experiments that tend to produce the most credible results, though they are often the most demanding to implement. The threats to internal validity that most commonly arise in each design are described, and methods for coping with these threats (including statistical analyses) are explicated.

Although the designs presented in Chapters 6–9 are prototypical quasi-experimental designs, they do not cover the entire terrain of quasi-experiments. Because researchers need to be able to creatively craft designs to best fit their specific research settings, they need to know the full range of design options. Toward this end, Chapter 10 presents a typology of designs for estimating treatment effects that goes beyond the prototypical designs described in Chapters 6–9. The typology of designs distinguishes between

Copyright © 2019 The Guilford Press

randomized experiments and quasi-experiments, as well as between two types of quasi-experiments: those where treatment assignment is determined according to a quantitative variable (e.g., the regression discontinuity and interrupted time-series designs) and those where treatment assignment is not so controlled (e.g., the pretest–posttest design and the nonequivalent group design). Cross-cutting the types of assignment to treatment conditions are four types of units that can be assigned to treatment conditions. The different units that can be assigned to treatments are participants, times, settings, and outcome measures. Chapter 10 shows where the four prototypical quasi-experimental designs fall within the more general typology and notes how the logic of the design and the analysis of the four prototypical designs generalize to the other designs in the typology.

Chapter 11 expands upon the typology presented in Chapter 10, showing how each fundamental design type in the typology can be elaborated by adding one or more supplementary comparisons. The additional comparisons help rule out specific threats to internal validity. Such elaborated designs employ comparisons that differ in one of four ways: they involve different participants, times, settings, or outcome measures. The design typology presented in Chapter 10, together with the elaborations, provide the tools by which researchers can tailor designs to fit the circumstances of their research settings. All too often the literature implies that researchers are to take a prototypical quasi-experimental design off the shelf, so to speak. That is, researchers are too frequently led to believe that the four prototypical quasi-experiments are the only choices available. The typology of designs in Chapter 10 and the elaborations in Chapter 11 provide a broader range of design options than is generally recognized. The task in designing a study is not to choose from among just four prototypes but to craft one's own design by selecting components from among the complete range of design options (Rosenbaum, 2015b, 2017; Shadish & Cook, 1999).

Chapter 12 distinguishes between focused and unfocused design elaborations. In **focused design elaborations,** separate estimates are used to address a shared threat to validity. Such focused design elaborations are explicated in Chapter 11. In **unfocused design elaborations**, separate estimates of the treatment effect are subject to different threats to validity. Chapter 12 provides examples of unfocused design elaboration and explains how unfocused design elaboration addresses the multiple threats to validity that are present. Chapter 12 also conceptualizes the process of estimating treatment effects as a task of **pattern matching**. To estimate a treatment effect, the researcher must collect data wherein the treatment is predicted to result in certain patterns of outcomes (should a treatment effect be present), while threats to validity are predicted to result in alternative patterns of outcomes (should they be present). The researcher then compares the predicted patterns to the data that are obtained. To the extent that the pattern predicted by the treatment fits the data better than the pattern predicted by threats to validity, the treatment is declared the winner and a treatment effect is plausible. Often, the best patterns for distinguishing treatment effects from the effects of threats to validity are complex. In quasi-experimentation, complex patterns are obtained by both

Copyright © 2019 The Guilford Press

focused and unfocused design elaboration. Chapter 12 illustrates the benefits of complex patterns and accompanying complex designs in the context of unfocused design elaborations. Complex patterns can often be created by combining treatment comparisons. Therefore, a treatment comparison that is relatively weak if implemented on its own might nonetheless add substantially to the credibility of results when combined with other comparisons.

Chapter 13 concludes the presentation by describing underlying principles of good design and analysis. These principles include the following. Threats to internal validity are best ruled out using design features rather than statistical analyses. When in doubt about which underlying assumptions are correct, use multiple statistical analyses based on a range of plausible assumptions. **Treatment effect interactions**, and not just average treatment effects, should be assessed. Knowledge is best accumulated by critically combining results from a variety of perspectives and studies.

This volume also includes a glossary containing definitions of technical terms. Terms included in the glossary are bold-faced the first time they appear in the body of the text.

1.5 CONCLUSIONS

Both randomized and quasi-experiments estimate the effects of treatments. Some commentators have opined that only randomized experiments can satisfactorily fulfill that purpose. Such commentators fail to recognize both the frequent limitations of randomized experiments and the potential benefits of quasi-experiments. While randomized experiments are to be preferred to quasi-experiments in many instances, randomized experiments cannot always be implemented, especially in field settings—in which case quasi-experiments are the only option. Even when randomized experiments can be implemented, there can be benefits to quasi-experiments. Hence, researchers need to understand how to conduct quasi-experiments if they are to be able to estimate treatment effects when confronted with the diverse array of research needs and circumstances.

Copyright © 2019 The Guilford Press

1.6 SUGGESTED READING

Campbell, D. T. (1969b). Reforms as experiments. *American Psychologist, 24,* 409–429.
—A classic call to action for using experimental methods to determine which social programs best ameliorate social problems.

Shadish, W. R., & Cook, T. D. (2009). The renaissance of field experimentation in evaluating interventions. *Annual Review of Psychology, 60,* 607–629.
—Presents a history of field experiments as used to ameliorate social problems.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin.
—A classic, must-read text on quasi-experimentation. The present volume provides more up-to-date coverage of quasi-experimentation than Shadish, Cook, and Campbell (2002). But if you want to know more about quasi-experimentation (especially experimental design) after reading the present volume, read Shadish et al.

West, S. G., Cham, H., & Liu, Y. (2014). Causal inference and generalizations in field settings: Experimental and quasi-experimental designs. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 49–80). New York: Cambridge University Press.
—Also provides an insightful overview of randomized and quasi-experiments as implemented in the field.

Copyright © 2019 The Guilford Press

Copyright © 2019 The Guilford Press.
No part of this text may be reproduced, translated, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise, without written permission from the publisher.
Purchase this book now: www.guilford.com/p/reichardt

Guilford Publications
370 Seventh Avenue
New York, NY 10001
212-431-9800
800-365-7006
www.guilford.com