Rolf Steyer
Ulf Kröhne

Probability and Causality

Version: July 7, 2016

University of Jena

Preface

What can we do to reduce global warming? How can we prevent another global financial crisis? How to fight AIDS? How can we reduce hunger in the world? These questions ask about causal effects of interventions. Obviously, interventions based on the wrong causal theories and hypotheses will cost the life of many and huge amounts of money that could be spent more appropriately. Even if our daily problems are less dramatic, they are of the same nature. Just think about your own actions that you have to chose in your responsibilities as a student, scientist, teacher, physician, psychologist, politician, or just as a parent! Whatever you do has effects, and these effects might be different if you take one action instead of another one. It is these kind of thoughts that make us believe that there is no other issue in the methodology of empirical sciences that deserves and needs more attention and effort than causality. And because the dependencies we are investigating are of a nondeterministic nature, we need a *probabilistic theory of causality*. In other words, we need to understand *probability* and *causality*.

What This Book is About

Empirical causal research involves several inferences and interpretations. Among these are:

(a) statistical inference, i. e., the inference from sample data to parameters characterizing the distributions of random variables,
(b) causal inference, i. e., the inference from parameters characterizing the distributions of random variables to causal effects and/or dependencies,
(c) interpretation of the putative cause,
(d) interpretation of the outcome variable,
(e) interpretation of the random experiment considered.

This book does not deal with all these points. We will neither discuss the mathematics of statistical inference nor the content issues of construct validity or external validity (Campbell & Stanley, 1963; Cook & Campbell, 1979; Shadish, Cook, & Campbell, 2002) involved in points (c) to (e). Instead we will focus on the second point: causal inference, i. e., the inference from parameters (such as the expectations of an outcome variable in two treatment conditions) to causal effects and/or causal dependencies. This is what the probabilistic theory of causality presented in this book is about. As will be shown in this book, causal effects are

also parameters that characterize the joint distributions of the random variables considered in a random experiment. However, their definitions are less obvious than 'ordinary' expectations and their differences.

Basic Idea

In order to get a first impression of what this means, let us briefly formulate the basic idea that can most easily be explained if the putative (or presumed) cause is a treatment variable. Suppose an individual, or in more general terms, an observational unit, could be treated by condition 1 or it could be treated by condition 0, *everything else being invariant*. If there is a difference in the outcome considered (some measure of success of the treatment), then this difference is due to the difference in the two treatment conditions. This conception goes back at least to Mill (1843/1865).

Multiple Determinacy

The problem with this first version of the basic idea is that most outcomes are *multiply determined*, i. e., they are not only influenced by the treatment variable, but by many other variables as well. In the field of agricultural research, e. g., the *yield* (*outcome*) of a *variety* not only depends on the variety (*treatment*) itself, but it also depends on the quality of the *plot* (*observational unit*), such as the average hours of sunshine on the plot per day, the amount of water reaching the plot, and the number of microbes in the plot, etc. Although Mill's idea sounds perfect, it is not immediately clear which implications it has for practice, because the number of other causes is often too large for keeping constant all of them. Furthermore Mill's idea fails to distinguish between covariates and intermediate variables. Holding constant all intermediate as well — and not only all covariates — would imply that there is no treatment effect any more, if we assume that all treatment effects have to be transmitted by some intermediate(s) (see section 4.2.4 for a more detailed discussion).

Because of the problem of multiple determinacy, Mills conception has been complemented by Sir Ronald A. Fisher (1925/1946) and by Jerzy S. Neyman (1923/1990) in the second and third decades of the last century. Simply speaking, introducing the randomized experiment, Fisher replaced the *ceteris paribus* clause ('everything else invariant') by the *ceteris paribus distributionibus* clause: *all other possible causes (the 'covariates') having the same distribution*. This is what random assignment of units to treatment conditions secures.

A Metaphor — The Invisible Man and his Shadow

Imagine an invisible man. Although we cannot see him, suppose we know that he is there, because we can see his shadow. Furthermore, suppose we would like to measure his size. Doing that, we have two problems, a theoretical and a practical one. The *theoretical problem* is to define *size*. We have to clarify that we do not

mean 'volume' or 'weight', but 'height' — without shoes, and without hat and hair. Unfortunately, actual height varies slightly in the course of a day. Hence, we define *size* to be the average of the actual heights at the different times of the day. This solves the theoretical problem; now we know what we want to measure.

However, because the man is invisible, we cannot measure his *size* directly — and this is not only because his size slightly varies over the day. The crucial problem is that we can only observe his shadow. And this is the *practical problem*: How to determine his size from his shadow? Sometimes, there is almost no shadow at all, sometimes it is huge. Some geometrical reflection yields a first simple solution: measuring the shadow when the sun has an angle of 45°. But what if it is winter and the sun does not reach this angle and if traveling to another point of the earth is too expensive? Now we need more geometrical knowledge, taking into account the actual angle of the sun and the observed length of the shadow. This will yield an exact measure of the *size* of the invisible man as well.

Determining a causal effect we face the same kind of problems. First, we have to define a *causal effect*, and second, we have to find out how to determine it from empirical estimable parameters such as true means, i. e., from expectations. The simple solution — corresponding to the 45° angle of the sun in the metaphor — is the perfect randomized experiment. The sample mean differences we see in a randomized experiment only randomly deviate from the causal effect (due to random sample variation). In contrast, in quasi-experiments and observational studies, solutions to the practical problem are more sophisticated. They are also more sophisticated than in the problem of the invisible man, because it is not only *one* other variable (the angle) that determines the length of the shadow; instead there often are *many* other variables systematically determining the sample means as well as the true means that are estimated by these sample means. This is again the problem of multiple determinacy.

This book presents a solution to the theoretical and the practical problems mentioned above. Unfortunately, both solutions are not as simple and obvious as in our metaphor. Furthermore, there is not only one single kind of causal effects. (In the paragraphs above we referred to total causal effects.) To our knowledge, the first pioneer tackling the theoretical *and* the practical problems was Jerzy S. Neyman (1923/1990).

Individual and Average Causal Effects

While Fisher introduced the design technique of randomization, Neyman introduced the concepts of individual and average causal effects, thus attempting a first solution to the theoretical problem mentioned above. (Note, however, that he used different terms for these concepts). He assumed that, for each individual plot, there is an intra-individual (plot-specific) distribution of the outcome variable, say Y, under each treatment. He then simply defined the *individual causal effect of treatment x compared to treatment x'* to be the difference between the intra-individual (plot-specific) expectation of Y (the "true yield") given treatment ("variety") x and the intra-individual (plot-specific) expectation of Y given

treatment ("variety") x'. Having defined the individual causal effect, the *average treatment effect* is simply the expectation of the corresponding individual (plot-specific) causal effects in the population of observational units (plots). Similarly, several kinds of *conditional effects* can be defined, conditioning, for instance, on covariates, i. e ., on other causes of Y that cannot be affected by X, such as measures of the *quality of the soil*, *average hours of sunshine*, *average hours of rain*, etc.

Total, Direct, and Indirect Effects

At about the same time as Neyman and Fisher developed their ideas, Sewall Wright (Wright, 1918, 1921, 1923, 1934, 1960a, 1960b) developed his ideas on path analysis and the concepts of total, direct, and indirect effects. While his *total effect* aims at the same idea as the average causal effect, his *direct* and *indirect effects* were new. Simply speaking, in the context of an experiment or quasi-experiment, a direct effect of the treatment is the effect that is not transmitted through an intermediate variable; it is the conditional effect of the treatment variable holding constant the intermediate variable on one of its values. In contrast, the *indirect effect* is the difference between the total effect and the direct effect.

Fundamental Problem of Causal Inference

Whereas the basic ideas outlined above are relatively simple and straightforward, trying to put them into practice — i. e., solving the practical problem mentioned above — is often difficult and needs considerable sophistication. The "fundamental problem of causal inference" (Holland, 1986) is that we cannot expose an observational unit to treatment 1 and, at the same time, to treatment 0. However, this is exactly what is necessary if we want to be sure that 'everything else is invariant', a clause that is also an implicit idea in the solution proposed by Neyman.

Pre-Post Designs

If we choose to first observe a unit under 'no treatment' and then observe it again after 'treatment', we may be tempted to interpret the pre-post differences as estimates of the individual causal effects of the treatment given in between. However, this interpretation might be wrong, because the unit may have developed (maturated, learned), may have suffered from critical life events, may have experienced historical change, etc. (see, e. g., Campbell & Stanley, 1963; Cook & Campbell, 1979; Shadish et al., 2002). Hence, in these *pre-post designs* or synonymously, *within-group designs*, we have to make assumptions on the nature of these possible alternative interpretations of the pre-post comparisons, e. g., that they do not hold in the application considered or that they have a certain structure that can be taken into account when making causal inferences based on pre-post comparisons.

Between-Group Designs

If, instead of making comparisons within a unit, we compare different units to each other in *between-group experiments*, we certainly lose the possibility of estimating the *individual* causal effects. However, what we can hope for is that we are still able to estimate the *average causal effect* and certain conditional causal effects. But how to estimate the average of the individual causal effects if the individual causal effects are not estimable? Both, between-group experiments and quasi-experiments, have a set of (observational) units, at least two experimental conditions ('treatment conditions', 'expositions', 'interventions', etc.), and at least one outcome variable ('response', 'criterion', 'dependent variable') Y. In the medical sciences, the units are usually patients. In psychology the observational units are often persons, but it could be persons-in-a-situation, or groups as well. In economics it could be subjects, companies, or countries, for instance. In educational sciences the units might be school classes, schools, communities, districts, or countries. In sociology and the political sciences, the units could be persons, but also communities, countries, etc.

Scope of the Theory

In order to delineate the scope of the theory, consider the following kind of *random experiment*: Draw an observational unit u (e. g., a person) out of a set of units, observe the value z of a (possibly multivariate qualitative or quantitative) covariate Z for this unit, assign the unit or observe its assignment to one of several experimental conditions, observe the value m of an intermediate variable M, and record the numerical value y of the outcome variable Y. We will use U to denote the random variable representing with its value u the unit drawn. Note that many observations can be made additional to observing U, Z, X, M, and Y. Although this simple single-unit trial is a prototype of the kind of empirical phenomena the theory is dealing with, there are other single-unit trials in which the theory can be applied as well (see ch. 2). In fact, the theory is applicable far beyond the true experiment and the quasi-experiment. This includes applications in which the putative causes are *not* manipulable and in which the putative cause is a continuous random variable. The theory has its limitations only if there is no clear ordering of the random variables considered as putative causes or outcomes.

True Experiments and Quasi-Experiments

The single-unit trial described above is a random experiment, but not necessarily a randomized experiment. A *randomized experiment* is a special random experiment in which the unit drawn is *randomly assigned* to one of the treatment conditions, e. g., depending on the outcome of a coin toss. (In empirical applications, the single-unit trials are repeated n times, where n denotes the sample size.) Referring to single-unit trials, we can distinguish the *true experiment* from

the *quasi-experiment* as follows: In the *true experiment*, there are at least two treatment conditions and the assignment to one of the treatment conditions is randomized, e. g., by flipping a coin. In a traditional *randomized experiment*, for instance, the treatment probabilities are chosen to be equal for all units. However, equal treatment probabilities for all units are neither essential for the definition of the true experiment nor for drawing valid causal inferences. We may as well have treatment probabilities depending on the units and/or on another covariate (see section 7.5), as long as these treatment probabilities are fixed or known by the researcher. Note, however, that in designs, in which different units have different treatment probabilities, standard data analysis techniques such as t-tests or analysis of variance do not test the correct hypotheses any more.

For between-group designs, the *quasi-experiment* may be defined such that there are at least two treatment conditions; however, in contrast to the true experiment, the treatment probabilities are unknown. Nevertheless, valid causal inferences can be drawn in quasi-experiments *provided that we can rely on certain assumptions*. In specific applications these assumptions might be wrong. If they are actually wrong, causal inferences can be completely wrong as well.

Beyond Experiments and Quasi-Experiments

As it turns out, formalizing the ideas outlined above in probabilistic terms results in a theory of probabilistic causality that is applicable far beyond experiments and quasi-experiments, thus bringing together the experimental tradition of Fisher and Neyman on one side and Wright's observational studies tradition on the other side. Furthermore, causal dependencies of manifest variables measuring latent variables as well as causal dependencies between latent variables can be treated in the framework presented in this book. Hence, the scope of the theory also includes what in the past has been addressed only within structural equation modeling (see, e. g., Bentler & Wu, 2002; Jöreskog & Sörbom, 1996/2001; Muthén & Muthén, 1998-2007) and/or graphical modeling (see, e. g., Pearl, 2009; Spirtes, Glymour, & Scheines, 2000). Furthermore, specific psychometric problems such as 'differential item functioning' and 'measurement invariance' turn out to be problems of causal modeling that can be treated within the same theoretical framework as the analysis of causal effects in experimental and quasi-experimental designs.

Who Should Study This Book?

The Methodologist

In the first place, we would like to address the *methodologist*, i. e., the expert in empirical research methodology, especially in the social, economic, behavioral, cognitive, medical, agricultural, and biological sciences. This book provides answers to some of the most important and fundamental questions of these empirical sciences: What do we mean by terms like 'X affects Y', 'X has an effect on Y',

'X influences Y', 'X leads to Y', etc. used in our informal theories and hypotheses? How can we translate these terms into a language that is compatible with the statistical analysis of empirical data? How to design a study and how to look at the resulting data if we want to probe our theories empirically and learn about the causal dependencies postulated in these theories and hypotheses? And last but not least: How to evaluate interventions, treatments, or expositions to (possibly detrimental) environments and learn about how effective they are for which kind of subjects or observational-units, and under which circumstances?

The Statistician

Many statisticians believe that causality is beyond their horizon. Causality might be a matter of empirical researchers and philosophers, they say, but not their own. They think that it cannot be treated mathematically and therefore a statistician cannot be helpful. As a consequence, they ignore the issue of causality. Reading this book will prove that all these beliefs should be abandoned. Probabilistic causality, as presented here, is a branch of probability theory, which itself, at least since Kolmogorov (1956), is a part of pure mathematics — although with an enormous potential for applications in many empirical sciences and even beyond. The main purpose of this book is to translate the informal concepts about causality shared by many methodologists and applied statisticians into the well-defined terms of mathematical probability theory. The principle is not to use any undefined term, and the result is a pure mathematical theory of probabilistic causality. Of course, this will make it harder for the methodologist and those not yet trained in probability theory. However, the reward is a much deeper understanding of what is essential and a much better grasp of the nature of our theories about the real world.

Of course, undefined terms are still used in this book, but only in the examples, in the interpretations, and in the motivations of the definitions. The theory itself is pure mathematics, just in the same way as Kolmogorov's probability theory presented in 1933, which explicated the mathematical, measure-theoretical structure of probabilistic concepts. Substantive meaning only results if we interpret the core components of the formal structure in a specific random experiment considered. And this is also true for the theory of probabilistic causality presented in this book.

The Empirical Scientist

The empirical scientist in the fields mentioned above has at least two good reasons to study this book. The *first* is that some crucial parts of his theories and hypotheses are explicated, at least when it comes to considering a concrete experiment or study. The ambiguity in causal language such as 'X affects Y', 'X has an effect on Y', 'X influences Y', 'X leads to Y' are not necessary any more. Reading this book will make it possible to replace these ambiguous terms by well-

understood and well-defined terms, improving quality of empirical research and theories.

The *second* motivation of the empirical scientist is that even if he knows his own theoretical concepts and hypotheses, he still has to know how to design experiments and studies that enable him to test them. Furthermore, the standard ways of analyzing data offered in the textbooks of applied statistics and in the available computer programs often do not estimate and test the correct causal effects and dependencies. And this is not only bad for the empirical scientist but also for all those relying on the validity of his inferences and his expertise. Just think about all the harmful consequences of wrong causal theories in various empirical research fields, if they are applied to solving concrete problems!

The Experimental Scientist

This book has two messages for those who do their research with experiments, a good one and a bad one. The good news is that, in perfect randomized experiments, the average causal total treatment effect is indeed estimated when comparing means between two different treatment conditions. The bad news is that *we can not rely on randomized assignment of units to treatment conditions* when it comes to estimating *direct* and *indirect* effects. More specifically, in such an analysis it is usually not sufficient to consider intermediates, treatment and outcome variables. Instead we also have to include in our analysis *pre-treatment variables* such as a pre-test of the intermediate and a pre-test of the outcome variable and apply adjustment methods, very much in the same way as we have to use these techniques in quasi-experiments — *even though we have randomized*! Hence if you want to look into the black box between the treatment and the outcome variables, you have to adopt the techniques of causal modeling that are far beyond traditional comparisons of means and analysis of variance.

The Philosopher of Science

Philosophers of science study and teach the methodology of empirical sciences. In that respect, their task is very similar to that of the methodologist, perhaps only more general and less specific for a certain discipline. Therefore, it is not surprising that probabilistic causality has also been tackled by philosophers of science (see, e. g., Cartwright, 1979; Spohn, 1980; Stegmüller, 1983; Suppes, 1970). Compared to these approaches, our emphasis is more on those parts of the theory that have implications for the *design* of empirical studies and the *analysis of data* resulting from such studies.

The Students in These Fields

We believe that probabilistic causality is the most rewarding topic in methodology. Although it is tough to get into it, you will get insights why all this methodology stuff was useful and what it was good for. At least this is what our students

say at the end of our curriculum, even if they did not have the choice whether or not to take our course on probabilistic causality.

Research Traditions in Stochastic Causality

Several research traditions have been contributing to the theory probabilistic causality in various ways. From the *Neyman-Rubin tradition*, we adopted the idea that it is important to define various causal effects such as individual, conditional, and average causal effects, even though we modified and extended these concepts in important aspects. Defining causal effects is important for proving that certain methods of data analysis yield estimates of these effects if certain assumptions can be made. Are there conditions under which the analysis of change scores (between pre- and post-tests) and repeated-measures analysis of variance yield causal effects? Under which conditions do we test causal effects in the analysis of covariance? Which are the assumptions under which propensity score methods yield estimates of causal effects? Which are the assumptions under which an instrumental variable analysis estimates a causal effect? All these questions and their answers presuppose that we have a clear definition of causal effects and/or of causal probabilistic dependencies.

From the *Campbellian tradition* (see, e. g., Campbell & Stanley, 1966; Cook & Campbell, 1979; Shadish et al., 2002) we learned that there are questions and problems beyond stochastic causality itself that are relevant in empirical causal research, such as: How to generalize beyond the study? What does the treatment variable mean? What is the meaning of the outcome variable? And, perhaps the most important question: Are there alternative explanations for the effect? The vast majority of social scientists (including ourselves) have been educated in this research tradition to some degree. Although this training is still very useful as a general methodology framework, it lacks precision and clarity in a number of issues — and causality is one of these.

From the *graphical modeling tradition* (see, e. g., Cox & Wermuth, 2004; Pearl, 2009; Spirtes et al., 2000), we learned that conditional independence plays an important role in causal modeling. This research tradition has also been developing techniques to estimate causal effects and to search for causal models if specific assumptions can be made. The fact that randomization in a true experiment in no way guarantees the validity of causal inferences on *direct* effects has been brought up by this research tradition.

Structural equation modeling and *psychometrics* have been teaching us how to use latent variables and structural equation modeling in testing causal hypotheses. Due to a number of statistical programs such as AMOS (Arbuckle, 2006), EQS (Bentler, 1995), lavaan (Rosseel, 2012), LISREL (Jöreskog & Sörbom, 1996/2001), Mplus (Muthén & Muthén, 1998-2007), OpenMx (OpenMx, 2009), RAMONA (Browne & Mels, 1998), structural equation modeling became extremely popular in the Social Sciences. Although many users of these programs hope to find causal answers, it should be clearly stated that structural equation modeling — and this is true for all kinds of statistical models (including analysis of vari-

ance) — does neither automatically estimate and test causal effects, nor does it provide a satisfactory *theory* of causal effects and dependencies. Nevertheless, this research tradition contributes — just like other areas of statistics — a number of statistical techniques that can be very useful in causal modeling.

In this book, we also aim at embedding — and, where necessary, extending — conventional statistical procedures such as analysis of covariance, nonorthogonal analysis of variance, and latent variable modeling, but also more recent techniques based on propensity scores, or on instrumental variables into a coherent theory of probabilistic causality.

How to Use This Book

This book is self-contained. It is written such that standard mathematical probability theory is sufficient for a complete understanding, provided one takes the time that these topics require. In many parts, this is not a book one can just *read*; instead it is a book to be *studied*. This includes working on the questions and exercises. We presume that the reader is familiar with — or learns while studying this book — the essentials of probability theory, including conditional expectations, as well as conditional independence and conditional distributions. These essentials of probability theory are dealt with in Steyer and Nagel (in press).

We devoted this book almost entirely to the *theory* of causal effects and probabilistic causality, although, in chapter 13, we outline the implications of the theory for *design* and for *data analysis in experiments and quasi-experiments*. We also developed the PC program *Causal Effects Explorer* (Nagengast, Kröhne, Bauer, & Steyer, 2007) that can be used for exploring prima facie effects, conditional and average total effects given certain parameters. We believe that this program is useful for teaching and learning the fundamentals of the theory. Furthermore, the program *EffectLiteR* (Mayer, Dietzfelbinger, Rosseel, & Steyer, in press), can be used to estimate total, direct, and indirect effects from empirical data in experiments and quasi-experiments. Both programs, which are available at *www.causal-effects.de*, may be used together with this book in a course on causal modeling. In fact, this is the content of our workshops on the analysis of total, direct, and indirect causal effects, which are available both as videos-on-demand on the internet and on DVDs, again at *www.causal-effects.de*.

Acknowledgements

This book has been written with the help of several colleagues and students. Werner Nagel (FSU Jena) helped whenever we felt lost in probability spaces. Stephen G. West (Arizona State University) and Felix Thömmes (University of Tübingen) made detailed suggestions for improving readability of the book. Safir Yousfi and Sonja Hahn contributed and/or suggested concrete ideas, Sonja being extremely helpful also in checking some of the mathematics. Our students Lisa Dietzfelbinger, Niclas Heider, Marc Heigener, Remo Kamm, Lawrence Lo, Axel Mayer, David Meder, Marita Menzel, Yuka Morikawa, Sebastian Nitsche, Michael

Temmerman, Sebastian Weirich, Anna Zimmermann, and other students critically commented on previous versions, helped minimizing errors, or organizing the references. Sven Hartenstein, Christoph Nachtigall, Marc Müller, Steffi Pohl, Norman Rose and Andreas Wolf together with the others mentioned above provided the intellectual climate in which this book could be written. We are also grateful to the students and colleagues participating at our courses on the analysis of causal effects asking questions and making important comments. Over the years, this helped a lot to improve this book.

Jena, April 15, 2016

Contents

List of Figures

List of Tables

Part I
Introduction

Chapter 1
Introductory Examples

For more than a century there have been examples in the statistical literature showing that comparing means or comparing probabilities (e. g., of success of a treatment) between a group exposed to a treatment and a comparison group (unexposed or exposed to a different treatment) does not necessarily answer our questions: 'Which treatment is better overall?' or 'Which treatment is better for which kind of person?' Differences between means and differences between probabilities (or any other comparison between probabilities such as odds ratios, log odds ratios, or relative risk) are usually not the treatment effects we are looking for (see, e. g., Pearson, Lee, & Bramley-Moore, 1899; Yule, 1903; Simpson, 1951). They are just *effects at first sight* or "prima facie effects" (Holland, 1986).

Just like the shadow in the metaphor of the invisible man (see the preface), prima facie effects reflect the effects of the treatment (the size of the invisible man), but also of other causes (the angle of the sun). The goal of analyzing *causal* effects is to estimate the effect of the treatment alone, isolating it from other potential influences, e. g., of sex, educational background, socio-economic status, etc. The general idea is to compute a treatment effect that is not biased by differences between treatment groups that would also exist *without treatment*.

Overview

We will illustrate systematic bias in determining *total* treatment effects in quasi-experiments by two examples. The first one deals with a dichotomous outcome variable, the second with a quantitative one. While the problems described in these two examples cannot occur in a randomized experiment, our third example will show that the randomized assignment of units to treatment conditions does not help to prevent systematic bias in determining *direct* treatment effects with respect to an intermediate variable that may transmit the effects of the treatment on the outcome variable.

1.1 Example 1 — Simpson's Paradox

In our first example, the prima facie effect reverses if we switch from comparing $P(Y{=}1\,|X{=}1)$ to $P(Y{=}1\,|X{=}0)$, the conditional probabilities of success between treatment and control, to comparing $P(Y{=}1\,|X{=}1, Z{=}z)$ to $P(Y{=}1\,|X{=}0, Z{=}z)$,

Table 1.1. Joint Probabilities of Treatment and Success

	Treatment		
Success	No (X=0)	Yes (X=1)	
No (Y=0)	.240	.232	.472
Yes (Y=1)	.360	.168	.528
	.600	.400	1.000

the corresponding probabilities additionally controlling for Z = *sex* with values m (males) and f (females). This kind of phenomenon, which is already known at least since Yule (1903), is called *Simpson's paradox* (Simpson, 1951), and it is still being debated (see, e. g., Hernán, Clayton, & Keiding, 2011).

1.1.1 Prima Facie Effect

Table 1.1 shows the joint distribution of treatment and success, i. e., the joint probabilities $P(X{=}x, Y{=}y)$ of treatment and success, as well as the marginal probabilities $P(X{=}x)$ and $P(Y{=}y)$ of treatment x and success y, respectively. Comparing the conditional probability of success (Y=1) given the *treatment condition* (X=1) to the conditional probability of success given the *control condition* (X=0) would lead us to the conclusion that the *treatment is harmful*. These two conditional probabilities can be computed by

$$P(Y{=}1\,|\,X{=}1) = \frac{P(Y{=}1, X{=}1)}{P(X{=}1)} = \frac{.168}{.168 + .232} = .42$$

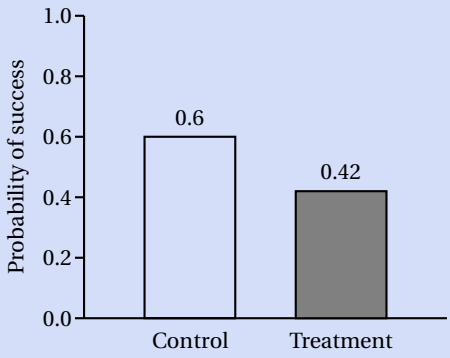


Figure 1.1. Probability of success given treatment

and

$$P(Y=1\,|\,X=0) \;=\; \frac{P(Y=1, X=0)}{P(X=0)} \;=\; \frac{.360}{.360+.240} \;=\; .60,$$

respectively (see, e. g., Steyer & Nagel, in press, section 4.2). Figure 1.1 displays both conditional probabilities in a histogram.

These two conditional probabilities can be compared to each other in different ways. The simplest one is looking at the *difference* $P(Y=1\,|\,X=1) - P(Y=1\,|\,X=0)$. This is a particular case of the difference $E(Y\,|\,X=1) - E(Y\,|\,X=0)$ between two conditional expectation values, in which the outcome variable Y is dichotomous with values 0 and 1. Following Holland (1986), we will call this difference the (unconditional) *prima facie effect* and use the notation PFE_{10}. Other possibilities of comparing the two conditional probabilities are to look at the odds ratio, or the logarithm of the odds ratio (see chapter 4 of Rothman, Greenland, & Lash, 2008, for a detailed discussion of these and other effect parameters).

1.1.2 Prima Facie Effects Controlling for Sex

The conclusion about the effect of the treatment is completely different if we look at the dependencies separately for males and females. Table 1.2 (p. 6) shows the joint distributions of treatment, success and $Z :=$ *sex* with values 0 (*male*) and 1 (*female*). The probabilities of the two values are $P(Z=0) = P(Z=1) = .50$. According to this table, the probability of success for the males in the treatment condition is

$$P(Y=1\,|\,X=1, Z=0) \;=\; \frac{.016}{.016+.004} \;=\; .80$$

(see Exercise 1-7), whereas the probability of success in the control condition is

$$P(Y=1\,|\,X=0, Z=0) \;=\; \frac{.336}{.336+.144} \;=\; .70.$$

Hence, the difference

$$P(Y=1\,|\,X=1, Z=0) \;-\; P(Y=1\,|\,X=0, Z=0) \tag{1.1}$$

is $.80-.70 = .10$, which may lead us to conclude that *the treatment is beneficial for males*. Again, because Y is dichotomous with values 0 and 1, this difference is a particular case of the difference $PFE_{10;\,Z=0} := E(Y\,|\,X=1, Z=0) - E(Y\,|\,X=0, Z=0)$, which we call the *conditional prima facie effect* given $Z=0$.

What about the treatment effects for females? Table 1.2 shows that the probability of success for the females in the treatment condition is $.152/(.152+.228) = .40$, whereas it is $.024/(.024+.096) = .20$ in the control condition. Figure 1.2 shows these conditional probabilities in a histogram. Considering the difference $.40 - .20 = .20$ may lead us to conclude that *the treatment is also beneficial for females*.

Hence, we can conclude that the treatment seems to be *beneficial for both, males and females*. This, however, seems to contradict our finding ignoring sex. Just considering the difference $E(Y\,|\,X=1) - E(Y\,|\,X=0)$, the *treatment seemed to be harmful*.

Table 1.2. Joint Probabilities of Treatment, Sex and Success

Males ($Z=0$); $P(Z=0)=0.50$

	Treatment		
Success	No ($X=0$)	Yes ($X=1$)	
No ($Y=0$)	.144	.004	.148
Yes ($Y=1$)	.336	.016	.352
	.480	.020	.500

Females ($Z=1$); $P(Z=1)=0.50$

	Treatment		
Success	No ($X=0$)	Yes ($X=1$)	
No ($Y=0$)	.096	.228	.324
Yes ($Y=1$)	.024	.152	.176
	.120	.380	.500

1.1.3 Prima Facie Effect vs. Average of the Prima Facie Effects

In contrast to our intuition, the *prima facie effect* $E(Y|X=1)-E(Y|X=0)$ is *neither* the *simple average nor any weighted average of the corresponding prima fa-*

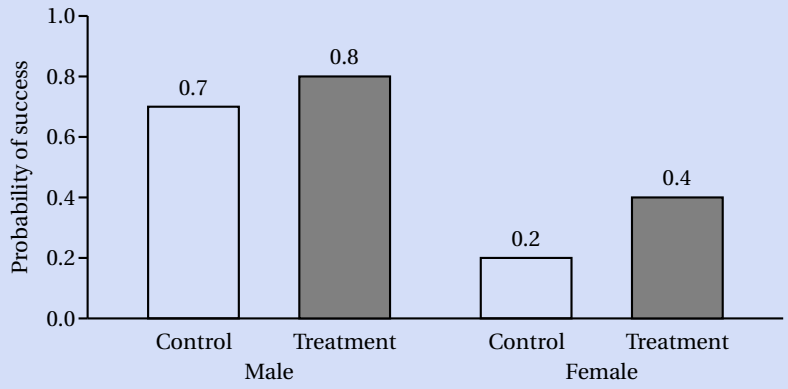


Figure 1.2. Probabilities of success given treatment (and sex)

cie effects $E(Y|X=1,Z=z)-E(Y|X=0,Z=z)$ controlling for Z = *sex*. This is now studied in more detail.

Prima Facie Effect

The probability $P(Y=1|X=0)$ of success in the control condition is the sum of the corresponding probabilities, $P(Y=1|X=0,Z=0)$ and $P(Y=1|X=0,Z=1)$, *weighted by the conditional probabilities* $P(Z=0|X=0)$ and $P(Z=1|X=0)$, respectively, i. e.,

$$\begin{aligned} P(Y=1\,|\,X=0) &= P(Y=1\,|\,X=0,Z=0)\cdot P(Z=0\,|\,X=0) + \\ &\quad P(Y=1\,|\,X=0,Z=1)\cdot P(Z=1\,|\,X=0) \\ &= .70\cdot\frac{.48}{.60} + .20\cdot\frac{.12}{.60} = .60 \end{aligned}$$

[see Box 9.2 (ii) of Steyer & Nagel, in press, and Exercise 1-8]. Because the difference between the conditional probabilities $P(Z=0\,|\,X=0)$ = .48/.60 and $P(Z=1\,|\,X=0)$ = .12/.60 is large, the probability of success in treatment 0 is much closer to .70 than to .20 (see the dots above $X=0$ in Fig. 1.3).

Similarly, the probability $P(Y=1\,|\,X=1)$ of success in the treatment condition ($X=1$) is the sum of the two corresponding probabilities, $P(Y=1\,|\,X=1,Z=0)$ and $P(Y=1\,|\,X=1,Z=1)$, *weighted by the conditional probabilities* $P(Z=0\,|\,X=1)$ and $P(Z=1\,|\,X=1)$, respectively, i. e.,

$$\begin{aligned} P(Y=1\,|\,X=1) &= P(Y=1\,|\,X=1,Z=0)\cdot P(Z=0\,|\,X=1) + \\ &\quad P(Y=1\,|\,X=1,Z=1)\cdot P(Z=1\,|\,X=1) \\ &= .80\cdot\frac{.02}{.40} + .40\cdot\frac{.38}{.40} = .42. \end{aligned}$$

Hence, the *prima facie effect* is $P(Y=1\,|\,X=1)-P(Y=1\,|\,X=0)$ = .42 − .60. = −.18. Because the two conditional probabilities $P(Z=0\,|\,X=1)$ = .02/.40 and $P(Z=1\,|\,X=1)$ = .38/.40 are very different, the probability of success in treatment 1 is much closer to .40 than to .80 (see the dots above $X=1$ in Fig. 1.3). (The size of the area of the dotted circles represent the joint probabilities $P(X=x,Z=z)$. For $X=1$ and $Z=0$, this probability is very small such that the circle is not visible. This kind of graphics has been adopted from Agresti, 2007).

Average of the Conditional Prima Facie Effects

In contrast to the prima facie effect, the *average of the conditional prima facie effects* is the expectation of the function $PFE_{10;Z}$, the values of which are the two prima facie effects $PFE_{10;Z=0}$ and $PFE_{10;Z=1}$ for males and females, i. e.,

$$E(PFE_{10;Z}) = \sum_z PFE_{10;Z=z}\cdot P(Z=z). \tag{1.2}$$

Because the conditional prima facie effect of the treatment is $PFE_{10;Z=0}$ = .10 for males and $PFE_{10;Z=1}$ = .20 for females, the average prima facie effect is simply:

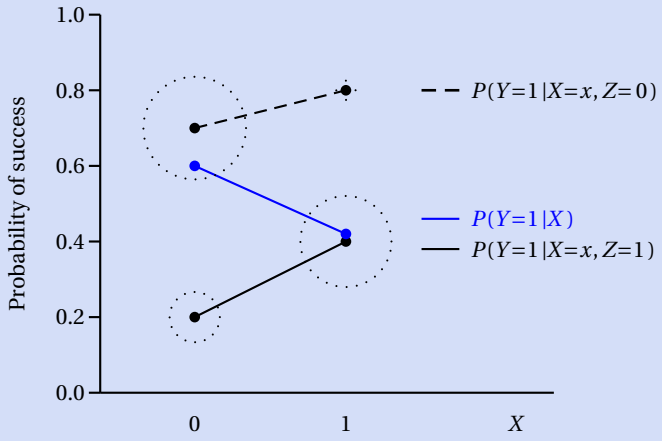


Figure 1.3. Probabilities of success given treatment and sex

$$E(PFE_{10;Z}) \;=\; .10\cdot P(Z=0) + .20\cdot P(Z=1) \;=\; .10\cdot\frac{1}{2} + .20\cdot\frac{1}{2} \;=\; .15\,.$$

Hence, whereas the *prima facie effect* $E(Y|X=1)-E(Y|X=0)$ is *negative*, namely $-.18$, the *average of the* $(Z=z)$*-conditional prima facie effects* $E(Y|X=1,Z=z)-E(Y|X=0,Z=z)$ is *positive*, namely .15.

1.1.4 How to Evaluate the Treatment?

Because the conclusions drawn from the differences $E(Y|X=1)-E(Y|X=0)$ and $E(Y|X=1,Z=z)-E(Y|X=0,Z=z)$ are contradictory, which of these comparisons should we trust? Is the treatment harmful — as $E(Y|X=1)-E(Y|X=0)$ suggests? Or is it beneficial as suggested by the differences $E(Y|X=1,Z=z)-E(Y|X=0,Z=z)$? Which of these comparisons are meaningful for evaluating the causal effect of the treatment? Before we come back to these questions, let us consider another example.

1.2 Example 2 — Nonorthogonal Two-Factorial Experiment

In this section, we treat an example with three treatment conditions, three values of a discrete covariate, and a quantitative outcome variable. In this example, we use a 3×3 factorial design with crossed, non-orthogonal factors. The analysis of such designs has been puzzling many statisticians (see, e. g., Aitkin, 1978; Appelbaum & Cramer, 1974; Carlson & Timm, 1974; Gosslee & Lucas, 1965; Jennings & Green, 1984; Keren & Lewis, 1976; Kramer, 1955; Overall & Spiegel, 1969, 1973b,

Table 1.3. Conditional Expectations Given Treatment

Treatment	Expectation of Y in treatment conditions $E(Y\|X{=}x)$	Treatment probabilities $P(X{=}x)$
$X{=}0$ (Control)	111.25	1/3
$X{=}1$ (Treatment 1)	100.00	1/3
$X{=}2$ (Treatment 2)	114.25	1/3
$E(Y)$	108.50	

1973a; Overall, Spiegel, & Cohen, 1975; Williams, 1972), and it continues to do so (see, e. g., Langsrud, 2003; Nelder & Lane, 1995).[1]

1.2.1 Prima Facie Effects

In the example presented in Table 1.3, there are *three treatment conditions* representing two treatments and a control. The outcome variable Y is now a quantitative measure of success. The expectations of the outcome variable Y in the three treatment conditions are displayed in Table 1.3. The ratios in the last column are the treatment probabilities $P(X{=}x)$ which are, in this example, the same for all three treatment conditions. However, although the probabilities $P(X{=}x)$ are the same for all three groups, this is *not* a randomized design as will become obvious if we look at the second factor and the 'cell probabilities' (see Table 1.4). Discussing the example at the level of conditional expectation values will again make clear that the contradictory inferences are not due to errors in *statistical inference* (from sample statistics to true parameters), but due to errors in *causal inference*, i. e., they are due to the misinterpretation of the differences between the expectations $E(Y|X{=}x)$ of the outcome variable Y in the three treatment conditions as causal effects.

If our evaluation of the treatment effects were based on these differences between the expectations of Y in the three treatment conditions, we would conclude that there are two treatment effects: a *negative effect* (namely, $100.00 - 111.25 = -11.25$) of treatment 1 compared to the control, and a *positive effect* (namely, $114.25 - 111.25 = 3.00$) of treatment 2 compared to the control.

[1] In fact, none of the statistical packages such as SAS, SysStat, or SPSS with their Type I, II, III or IV sums of squares provide correct estimates and tests of the average effects (or main effects) for such a design unless the covariate (the second factor) has a uniform distribution, with equal probabilities for all values of the covariate. In this case Type III analysis yields correct results, at least, if the second factor is assumed to be fixed. However, in most applications in the Social Sciences, the covariate (second factor) is not fixed but stochastic with varying sample means, etc. In chapter 13, we will outline a correct analysis including the average total effects.

Table 1.4. Conditional Expectations Given Treatment and Neediness

	Neediness						
Treatment	Low (Z=0)		Medium (Z=1)		High (Z=2)		
X=0	120	(20/120)	110	(17/120)	60	(3/120)	(40/120)
X=1	100	(7/120)	100	(26/120)	100	(7/120)	(40/120)
X=2	80	(3/120)	90	(17/120)	140	(20/120)	(40/120)
		(30/120)		(60/120)		(30/120)	

Note. Probabilities $P(X{=}x, Z{=}z)$, $P(Z{=}z)$, and $P(X{=}x)$ in parentheses.

1.2.2 Prima Facie Effects Controlling for Neediness

A second way to evaluate the 'effects' of the three *treatment conditions* is to look at the differences between the expectations of Y in the three treatment conditions *within each of the three classes of neediness* for the therapy: low, medium, and high. Table 1.4 displays the expectations of the outcome variable Y in the nine cells of the 3×3 design. The ratios in parentheses are the probabilities that the pairs (x, z) of values of X and Z are observed. Hence, this table contains the conditional expectation values (true cell means) of the outcome variable Y, and the probabilities $P(X{=}x, Z{=}z)$ determining the true joint distribution of X and Z.

In the *low neediness condition* ($Z{=}0$), there are large negative effects, both of treatment 1 and of treatment 2 compared to the control:

$$PFE_{10;Z=0} := E(Y|X{=}1, Z{=}0) - E(Y|X{=}0, Z{=}0) = 100 - 120 = -20$$

and

$$PFE_{20;Z=0} := E(Y|X{=}2, Z{=}0) - E(Y|X{=}0, Z{=}0) = 80 - 120 = -40.$$

In the *medium neediness condition* ($Z{=}1$), there are also negative effects of treatment 1 and of treatment 2 compared to the control:

$$PFE_{10;Z=1} := E(Y|X{=}1, Z{=}1) - E(Y|X{=}0, Z{=}1) = 100 - 110 = -10$$

and

$$PFE_{20;Z=1} := E(Y|X{=}2, Z{=}1) - E(Y|X{=}0, Z{=}1) = 90 - 110 = -20.$$

Finally, in the *high neediness condition* ($Z{=}2$), the effects of treatment 1 and treatment 2 are both positive:

$$PFE_{10;Z=2} := E(Y|X{=}1, Z{=}2) - E(Y|X{=}0, Z{=}2) = 100 - 60 = 40$$

and

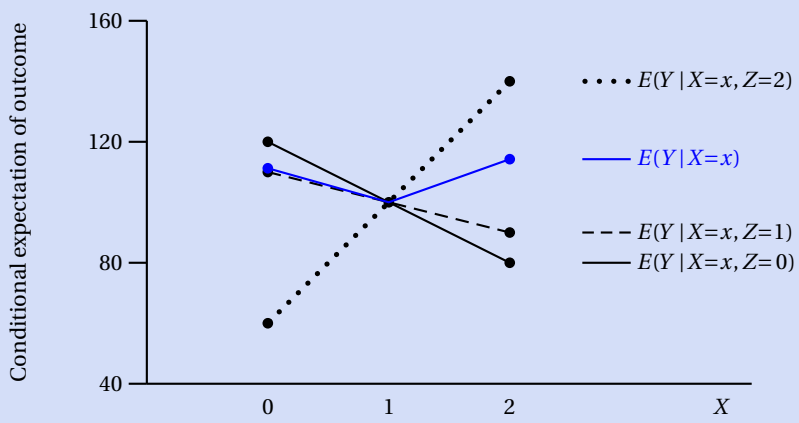


Figure 1.4. Conditional expectation values of Y given treatment and neediness

$$PFE_{20;Z=2} := E(Y|X=2,Z=2) - E(Y|X=0,Z=2) = 140 - 60 = 80.$$

Based on these comparisons, we can conclude that the 'effects' of the treatments depend on the neediness of the subjects: the differences between the expectations of Y are negative for subjects with low and medium neediness, and they are positive for the subjects with high neediness.

1.2.3 Prima Facie Effects vs. Average of the Prima Facie Effects

There is no doubt that the conditional effects given neediness, which are sometimes also called *simple effects*, are more informative than average treatment effects if we want to know which treatment is the best for which level of neediness. Nevertheless, we might ask: What are the 'treatment effects' on average? Or, in other words which are the 'main effects'? In fact, all major statistical programs compute 'main effects' (see Langsrud, 2003 for a list on which program suggests what solution to this problem). Note that we have two average effects in this example, because we can compare treatment 1 *and* treatment 2 to the control. Because we already looked at the corresponding conditional effects, we just have to compute their averages, i. e., the expectations of these conditional effects over the distribution of neediness:

$$E(PFE_{10;Z}) = \sum_z PFE_{10;Z=z} \cdot P(Z=z) = -20 \cdot \frac{1}{4} + (-10) \cdot \frac{1}{2} + 40 \cdot \frac{1}{4} = 0.$$

Hence, the average effect of treatment 1 compared to the control is zero.

Comparing treatment 2 to the control yields on average:

$$E(PFE_{20;Z}) = \sum_z PFE_{20;Z=z} \cdot P(Z{=}z) = -40 \cdot \frac{1}{4} + (-20) \cdot \frac{1}{2} + 80 \cdot \frac{1}{4} = 0.$$

According to this result, the average effect of treatment 2 compared to the control is zero as well.

1.2.4 How to Evaluate the Treatment?

To summarize, we discussed three ways that may, at first sight, be used to evaluate the treatment effects: *First*, we may compare the differences between the expectations $E(Y|X{=}x)$ of the outcome variable in the three treatment conditions $X{=}0$, $X{=}1$, and $X{=}2$. *Second*, we may consider the corresponding differences between the conditional expectation values $E(Y|X{=}x, Z{=}z)$ *within each of the three values* $Z{=}0$, $Z{=}1$, *and* $Z{=}2$ *of neediness. Third*, we may compare the averages of these differences between the conditional expectation values over the distribution of Z (see Box 1.1 for a summary of these effects).[2] All these comparisons yield different results. Which of them are meaningful for the evaluation of the treatment effects? All three of them, or only two, just one, or none at all?

1.3 Example 3 — Direct Effect in a Randomized Experiment

The problems described in the examples treated in the preceding sections occur because there are covariates (in the examples, *sex* and *neediness*) that are related to the treatment variable *and* the outcome variable. Hence, these problems can *not* occur in a randomized experiment, in which, by definition, all covariates and the treatment are (stochastically) independent. Hence, if in a randomized experiment, we are only interested in the *total effects* of the treatment on the outcome variable, the effects that are estimated by the differences between means in the treatment groups are the total effects of the treatment. However, often we are also interested in the mediation processes producing these total effects. A typical question in educational research is: 'Is there a direct effect of the treatment that is not transmitted through *motivation after treatment*?' In medical research we may ask: 'Is there a direct effect of the treatment that is not transmitted through the *amount of antibodies*?'

1.3.1 Conditional Expectation of Y Given Treatment and Intermediate Variables

Suppose that Table 1.5 displays the true means, variances, covariances, and correlations of a treatment variable X with values 0 and 1, an intermediate vari-

[2] In fact, there are even more than three ways. Types II and III of computing the sums of squares in nonorthogonal ANOVA are not yet considered in our discussion. In chapter 13, we show that all four types of computing sums of squares in such a design yield wrong results in our example (see also Exercise 1-14).

Table 1.5. Covariances, Correlations, and Expectations (Omitting Pre-Tests)

		X	M	Y
Treatment (yes = 1, no = 0)	X	0.25	*.727*	*.597*
Post-test motivation	M	5.00	189.00	*.893*
Post-test achievement	Y	5.00	205.70	280.45
Expectations		0.50	90.00	140.00

Note. Correlations (in italics) are rounded.

able M, and an outcome variable Y. (This example is adopted from Mayer, Thoemmes, Rose, Steyer, & West, 2014.)

First of all, let us consider the conditional expectation $E(Y \mid X, M)$, assuming that it can be written as a linear function of X and M (see Fig. 1.5). In fact, the covariance matrix presented in Table 1.5 has been constructed such that this linearity assumption holds. Using the covariances and expectations displayed in this table, we receive

$$E(Y \mid X, M) \approx 34.9924 - 3.7528 \cdot X + 1.1876 \cdot M \tag{1.3}$$

(see Exercise 1-12). According to textbook wisdom (see, e. g., MacKinnon, 2008, but also Baron & Kenny, 1986), the direct effect of X on Y, controlling for M, is approximately -3.75.

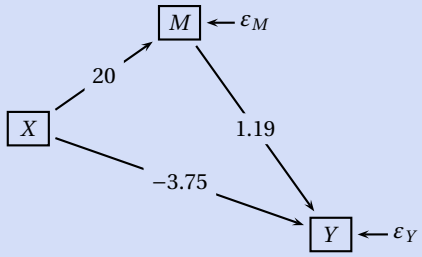


Figure 1.5. Path diagram of $E(M \mid X)$ and $E(Y \mid X, M)$

Table 1.6. Covariances, Correlations, and Expectations (Including Pre-Tests)

		W	Z	X	M	Y
Pre-test achievement	W	100.00	*.850*	*.000*	*.495*	*.740*
Pre-test motivation	Z	85.00	100.00	*.000*	*.582*	*.696*
Treatment (yes=1, no=0)	X	0.00	0.00	0.25	*.727*	*.597*
Post-test motivation	M	68.00	80.00	5.00	189.00	*.893*
Post-test achievement	Y	124.00	116.50	5.00	205.70	280.45
Expectations		100.00	100.00	0.50	90.00	140.00

Note. Correlations (in italics) are rounded.

1.3.2 Conditional Expectation of Y Given Treatment, Intermediate, and Pre-Test Variables

Suppose M represents *post-test motivation* in a randomized experiment designed to evaluate two teaching methods represented by $(X=0)$ and $(X=1)$, respectively. In this case, even if not observed, there will be a variable, say Z representing *pre-test motivation* with respect to which students will differ before treatment. Furthermore, there will be a variable, say W, representing *pre-test achievement* with respect to which students will differ prior to treatment as well. Furthermore, the two pre-test variables Z and W will be correlated. This is a plausible scenario for such a teaching experiment, and this is how the complete variance-covariance matrix and the expectations presented in Table 1.6 have been generated.

Hence, if instead of $E(Y|X,M)$, we consider the conditional expectation of Y given X, M, Z, and W, again assuming linearity — and this is how the parameters presented in Table 1.6 have been generated — we receive

$$E(Y|X,M,Z,W) \;=\; .00 + 10\cdot X + 0.50\cdot M + 0.00\cdot Z + .90\cdot W \tag{1.4}$$

(see Exercise 1-13). Now the coefficient 10 of X might be interpreted to be the direct treatment effect, 'direct' with respect to the intermediate variable M. It is the effect of X controlling for the intermediate variable M *and* for all covariates, in this example, the two pre-test variables Z and W.

How can we explain this seemingly paradoxical result? How can there be confounding in a perfect randomized experiment? The answer is that even though X and the bivariate random variable (W,Z) are independent, *conditional independence* of X and (W,Z) given M does *not* hold. Instead, conditioning on M induces conditional *dependence* of X and Z, if both Z and X are related to M. Intuitively speaking, because both Z and X affect M, a high value of *post-test motivation* M means that both, X and Z tend to be high, whereas a low value of M means that both, X and Z tend to be low (see Fig. 1.6). Hence, conditioning on M, the treatment variable X and the *pre-test motivation* Z will be dependent, even though X and Z are unconditionally independent, due to randomization (see also Pearl,

2009, ch. 1, p. 17, or Spirtes et al., 2000). This conditional dependence between X and Z given M is also reflected by a non-zero partial correlation $Corr(X, Z; M)$ (see section 11.6 of Steyer & Nagel, in press).

1.3.3 Conditional Expectation of Y Given Treatment Variable

Finally, let us consider the *average total treatment effect*. In this example, in which X and all covariates are independent, the average total treatment effect is the coefficient of X in the equation

$$E(Y|X) = 130 + 20 \cdot X, \tag{1.5}$$

where the intercept $\alpha_0 = 130$ is obtained by $\alpha_0 = E(Y) - \alpha_1 \cdot E(X) = 140 - 20 \cdot 0.50 = 130$ and the slope by $\alpha_1 = Cov(X, Y)/Var(X) = 5.00/0.25 = 20$ [see Steyer & Nagel, in press, Eqs. (12.58) and (12.59)]. Therefore, in this example, the *indirect treatment effect* is the difference $20 - 10 = 10$. In this model with no interaction, this indirect effect is also equal to the product $20 \cdot .50$ (see Fig. 1.6), which is in accordance with the rules of path analysis developed by Sewall Wright in the twenties of last century (see, e. g., Wright, 1918, 1921, 1923).

1.3.4 How to Analyze Direct Effects?

We discussed two different ways to analyze the direct effect of the treatment variable on the outcome variable. The first one is recommended in traditional textbooks such as MacKinnon (2008) and in one of the most frequently cited papers Baron and Kenny (1986). It yields the negative direct effect of -3.75. The second one also controls for the pre-tests of the intermediate variable and the outcome variables. This second analysis yields a direct treatment effect of 10. Hence, the effect is reversed as compared to the first analysis. Which is the correct direct effect? Or are both wrong?

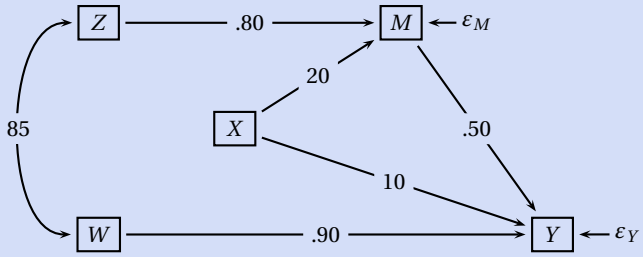


Figure 1.6. Path diagram of $E(M|X, Z, W)$ and $E(Y|X, M, Z, W)$

1.4 Summary and Conclusions

In this chapter, we treated three examples. In the first example, a dichotomous treatment variable X has a negative 'effect' $E(Y|X=1)-E(Y|X=0)$ on a *dichotomous outcome variable* Y ('success'), although the corresponding treatment 'effects' $E(Y|X=1,Z=z)-E(Y|X=0,Z=z)$ are positive if we condition on males ($Z=m$) and females ($Z=f$). Taking the expectation of these two conditional effects also yielded a positive 'effect'. In the second example, there are nonzero differences $E(Y|X=1) - E(Y|X=0)$ and $E(Y|X=2) - E(Y|X=0)$, where Y is a *quantitative outcome variable*, and nonzero conditional 'effects' $E(Y|X=1,Z=z)$ $- E(Y|X=0,Z=z)$ and $E(Y|X=2,Z=z) - E(Y|X=0,Z=z)$ for the different values of *neediness*. The expectations of these conditional 'effects' over the three neediness conditions, i. e., the average 'effects', are zero. In the third example, we discussed two different ways of analyzing the direct treatment effect. The first yields a negative 'direct effect' and the second a positive 'direct effect'.

The Problem

Because the conclusions drawn from these analyses are contradictory, which of these should we trust? In Simpson's paradox: Is the treatment harmful — as the difference $E(Y|X=1)-E(Y|X=0)$ suggests? Or is it beneficial as suggested by the differences $E(Y|X=1,Z=z)-E(Y|X=0,Z=z)$, controlling for sex? Which of these comparisons are meaningful for the evaluation of the causal effects of the treatment? Similarly, in the second example: are there treatment effects, overall? Or are the effects nil on average? And, are the conditional effects dependable, or could it be that they would also be reversed if we condition on an additional covariate, such as *age* or *educational status*? As demonstrated in Simpson's paradox, we can neither expect that the difference $E(Y|X=1)-E(Y|X=0)$ is the average of the corresponding differences $E(Y|X=1,Z=z)-E(Y|X=0,Z=z)$, nor can we expect that a difference $E(Y|X=1,Z=z)-E(Y|X=0,Z=z)$ is the average over the corresponding differences if we condition on an additional covariate such as age. Note, these questions are not related to *statistical* inference; they are not raised at the sample level, but on the level of true parameters!

Hence our examples show that the conditional expectation values and their differences, the prima facie effects, can be totally misleading in evaluating the effects of a treatment variable X on an outcome variable Y. This conclusion can also be extended to conditional probabilities, to correlations and to all other parameters describing relationships and dependencies between random variables. They all are like the shadow in the metaphor of the invisible man (see the preface).

If this is true, is the whole idea of *learning from experience* — the core of empirical sciences — wrong? Our answer is 'No'. However, we have to be more explicit in what we mean by terms like 'X affects Y', 'X has an effect on Y', 'X influences Y', 'X leads to Y', etc. used in our theories and hypotheses. How can these terms be translated into a language compatible with statistical analyses of empirical data?

Box 1.1 Glossary of New Concepts

$PFE_{xx'}$	*Prima facie effect* of treatment x compared to treatment x'. It is defined by $$PFE_{xx'} := E(Y \mid X{=}x) - E(Y \mid X{=}x').$$
$PFE_{xx';Z=z}$	$(Z{=}z)$-*Conditional prima facie effect* of treatment x compared to treatment x'. It is defined by $$PFE_{xx';Z=z} := E(Y \mid X{=}x, Z{=}z) - E(Y \mid X{=}x', Z{=}z).$$
$E(PFE_{xx';Z})$	*Expectation of the* $(Z{=}z)$-*conditional prima facie effects* of treatment x compared to treatment x'. It is defined by $$E(PFE_{xx';Z}) := \sum_z PFE_{xx';Z=z} \cdot P(Z{=}z).$$

How to design a study and how to look at the resulting data if we want to probe our theories empirically and learn about the causal dependencies postulated in these theories and hypotheses?

We know that a reversal of total effects does not occur in the randomized experiment, i. e., in an experiment in which observational units (in the social and behavioral sciences, usually the subjects or individuals) are randomly assigned to one of at least two treatment conditions. In the randomized experiment comparing expectation values *is* informative about total causal treatment effects. But why? What is so special in the randomized experiment? Which are the conditions allowing for causal inference in the randomized experiment? Can we create these conditions also in quasi-experimental studies? How can we estimate causal effects in quasi-experiments? And why does randomization not help if we analyze direct treatment effects? Obviously, conclusive answers to these questions can be hoped for only within a theory of causal effects.

Relevance of the Problem

These questions are of fundamental importance for the methodology of empirical sciences and for the empirical sciences themselves. The answers to these questions have consequences for the design and analysis of experiments, quasi-experiments, and other studies aiming at estimating the effects of *treatments*, *interventions*, or *expositions* on certain outcome variables. No *prevention study* can be meaningfully conducted without knowing the concepts of causal effects and how they can be estimated from empirical data, and the same is true for the *evaluation of institutions* such as schools, universities, or clinics with respect to their effects on the outcomes of their clients. Similarly, without a clear concept

of causal effects we are not able to learn from our data about the effects of a certain (possibly harmful) environment on our health, or about the effects of certain behaviors such as smoking or drug abuse. Again, this is similar to the problem of measuring the invisible man's size via the length of his shadow: only with a clear concept of *size*, some basic knowledge in geometry, and the additional information such as the angle of the sun at the time of measurement, are we able to determine his size from the length of his shadow.

Furthermore, without an explicit theory of causal effects we are not able to study direct and indirect effects, and *this is true even in a perfect randomized experiment*. For example, if we are interested in whether or not the effect of vaccination is completely transmitted through the amount of a certain type of antibodies, then this cannot be done relying only on the benefits of a perfect randomized trial. Instead we have to apply certain adjustment techniques. In terms of our metaphor, the 45° angle (the randomized experiment) does not help in determining the parameters we are looking for (the direct effects).

Research Traditions

Of course, raising these questions and attempting answers is not new. Immense knowledge and wisdom about experiments and quasi-experiments has been collected in the Campbellian tradition of experiments and quasi-experiments (see, e. g., Campbell & Stanley, 1963; Cook & Campbell, 1979; Shadish et al., 2002). In the last decades, a more formal approach has been developed supplementing the Campbellian theory and terminology in important aspects: the theory of causal effects in the Neyman-Rubin tradition (see, e. g., Splawa-Neyman, 1923/1990; Rubin, 1974, 2005). Many papers and books indicate the growing influence of this theory (see, e. g., Greenland, 2000, 2004; Höfler, 2005; Rosenbaum, 2002; Rubin, 2006; Winship & Morgan, 1999; Morgan & Winship, 2007) and formidable efforts have already been made to integrate it into the Campbellian framework (West, Biesanz, & Pitts, 2000). Furthermore, these questions have also been dealt with in the graphical modeling tradition (see, e. g., Pearl, 2009; Spirtes et al., 2000) as well as in biometrics, econometrics, psychometrics, and other fields dealing with the methodology of empirical research fields.

Outlook

In this book, we present the theory of total, direct, and indirect causal effects in terms of classical probability theory. We show that a number of questions that have been debated controversially and inconclusively can now be given a clear-cut answer. What kinds of causal effects can be meaningfully defined? Which design techniques guarantee unbiased estimation of causal effects? How to analyze nonorthogonal ANOVA designs (cf., e. g., Aitkin, 1978; Appelbaum & Cramer, 1974; Gosslee & Lucas, 1965; Maxwell & Delaney, 2004; Overall et al., 1975)? How to analyze non-equivalent control-group designs (cf., e. g., Reichardt, 1979)? Should we compare pre-post differences between treatment groups (cf., e. g.,

Lord, 1967; Senn, 2006; van Breukelen, 2006; Wainer, 1991)? Should we use analysis of covariance to adjust for differences in treatment and control that already existed prior to treatment (cf., e. g., Maxwell & Delaney, 2004; Cohen, Cohen, West, & Aiken, 2003)? Should we use new techniques such as propensity score methods instead of the more traditional procedures mentioned above (cf., e. g., Rosenbaum & Rubin, 1984)? How do we deal with non-compliance to treatment assignment (cf., e. g., Cheng & Small, 2006; Dunn et al., 2003; Jo, 2002a, 2002b, 2002c; Jo, Asparouhov, Muthén, Ialongo, & Brown, 2008; J. Robins & Rotnitzky, 2004; J. M. Robins, 1998)? How to analyze direct and indirect effects? We do not treat the statistical sampling models with their distributional assumptions, their implications for parameter estimation, and the evaluation (or tests) of hypotheses about these parameters. However, in chapter 13 we discuss the virtues and problems of general strategies of data analysis such as the analysis of difference scores, analysis of covariance, its generalizations, analysis based on propensity scores, and instrumental variables.

1.5 Exercises

▷ **Exercise 1-1** Why do we need the concept of a causal treatment effect?

▷ **Exercise 1-2** What is the relationship between the unconditional prima facie effect PFE_{10} and the expectations $E(Y|X=0)$ and $E(Y|X=1)$ of the outcome variable Y in the two treatment conditions?

▷ **Exercise 1-3** Verify that Table 1.1 (p. 4) is in fact obtained by collapsing the two corresponding tables for males and females (see Table 1.2, p. 6).

▷ **Exercise 1-4** Which are the three kinds of prima facie effects treated in this chapter?

▷ **Exercise 1-5** What is the difference between statistical inference and causal inference?

▷ **Exercise 1-6** Why are the conditional expectation values $E(Y|X=x)$ in treatment conditions x also probabilities for $Y=1$ in the first example treated in this chapter?

▷ **Exercise 1-7** Compute the conditional probability $P(Y=1 \mid X=1, Z=0)$ from Table 1.2 (p. 6).

▷ **Exercise 1-8** Compute the probability $P(Y=1|X=0)$ of success in the control condition.

▷ **Exercise 1-9** What are the unconditional prima facie effects of the treatments, i. e., the prima facie effects $E(Y|X=1)-E(Y|X=0)$ and $E(Y|X=2)-E(Y|X=0)$ in the second example of this chapter?

▷ **Exercise 1-10** What are the conditional prima facie effects of the treatments, i. e., the prima facie effects $E(Y|X=1,Z=z)-E(Y|X=0,Z=z)$ and $E(Y|X=2,Z=z)-E(Y|X=0,Z=z)$ in the second example of this chapter?

▷ **Exercise 1-11** What are the averages of the conditional prima facie effects

$$E(Y|X=1,Z=z) - E(Y|X=0,Z=z) \qquad \text{and} \qquad E(Y|X=2,Z=z) - E(Y|X=0,Z=z)$$

in the second example of this chapter?

▷ **Exercise 1-12** Compute the coefficients of the equation for the conditional expectation $E(Y|X,M)$ presented in Equation (1.3).

▷ **Exercise 1-13** Compute the coefficients of the equation for the conditional expectation $E(Y|X,M,Z,W)$ presented in Equation (1.4).

▷ **Exercise 1-14** Download *table.1.4.10000.sav* from *www.causal-effects.de*. This data set has been generated from Table 1.4 (p. 10) for a sample of size $N = 10.000$.

(a) Estimate the cell means and the relative frequencies of observations in each of the nine cells of the 3×3 table.

(b) Use each of the procedures offered by your statistical program package to analyze the data including a test of the main effects of the treatment factor (most programs offer Typ I, II and III sums of squares for such an analysis).

(c) Compare the results of these analyses to the parameters presented in Table 1.4 (p. 10).

▷ **Exercise 1-15** Download *table.1.6.10000.sav* from *www.causal-effects.de*. This data set has been generated from Table 1.6 (p. 14) for a sample of size $N = 10.000$.

(a) Estimate the conditional expectation of Y given X and M.

(b) Estimate the conditional expectation of Y given X, M, Z and W.

(c) Compare the estimated regression coefficients to the parameters presented in Equations (1.3) and (1.4), respectively.

Solutions

▷ **Solution 1-1** We need the concept of a causal treatment effect, because Simpson's paradox shows that differences between expectations are meaningless for the evaluation of the effects of a treatment, unless we can show how the differences between expectations are related to the causal effects. Without a definition of causal treatment effects, this would not be possible. Estimating causal treatment effects is crucial for answering questions such as 'Does the treatment help our patients with respect to the outcome variable considered?'

▷ **Solution 1-2** The unconditional prima facie effect PFE_{10} is defined as the difference between the two expectations $E(Y|X=1)$ and $E(Y|X=0)$.

▷ **Solution 1-3** This can easily be verified by adding the probabilities for the observations of the pairs (x,z) of X and Z over males and females. This yields $.144 + .096 = .240$, $.004 + .228 = .232$, $.336 + .024 = .360$ and $.016 + .152 = .168$.

▷ **Solution 1-4** The three kinds of prima facie effects treated in this chapter are: the *unconditional prima facie effect*, the *conditional prima facie effect* given the value z of a covariate Z, and the *average of the* $(Z=z)$*-conditional prima facie effects*. The unconditional prima facie effect of treatment 1 compared to treatment 0 is the difference $PFE_{10} := E(Y|X=1) - E(Y|X=0)$ between the expectations of an outcome variable Y in the two treatment conditions. The $(Z=z)$-conditional prima facie effect is the difference $PFE_{10;Z=z} := E(Y|X=1,Z=z) - E(Y|X=0,Z=z)$ between the $(Z=z)$-conditional expectation values of the outcome variable Y in the two treatment conditions. The average prima facie effect is the expectation of the conditional prima facie effects [see Eq. (1.2)].

▷ **Solution 1-5** In *statistical* inference we estimate and test hypotheses about parameters characterizing the distribution of a random variable from sample data. In *causal* inference we interpret some of these parameters as causal effects.

▷ **Solution 1-6** $E(Y|X{=}x) = P(Y{=}1|X{=}x)$, because, in this example, Y is dichotomous with values 0 and 1. In this case, $E(Y|X{=}x) := \sum_y y \cdot P(Y{=}y|X{=}x)$ [see Steyer & Nagel, in press, Eq. (9.19)] yields $E(Y|X{=}x) = 0 \cdot P(Y{=}0|X{=}x) + 1 \cdot P(Y{=}1|X{=}x) = P(Y{=}1|X{=}x)$.

▷ **Solution 1-7** According to Table 1.2 (p. 6) ,

$$P(Y{=}1|X{=}1,Z{=}0) = \frac{P(X{=}1,Y{=}1,Z{=}0)}{P(X{=}1,Z{=}0)} = \frac{.016}{.016+.004} = .80.$$

▷ **Solution 1-8** First of all, note that the theorem of total probability, can also be applied to conditional probabilities, in this exercise, the $(X{=}0)$-conditional probabilities. Hence, according to this theorem,

$$P(Y{=}1|X{=}0) = P(Y{=}1|X{=}0,Z{=}0)\cdot P(Z{=}0|X{=}0) + P(Y{=}1|X{=}0,Z{=}1)\cdot P(Z{=}1|X{=}0).$$

The probabilities $P(Y{=}1|X{=}0,Z{=}0) = .70$ and $P(Y{=}1|X{=}0,Z{=}1) = .20$ are computed analogously to Exercise 1-7 and the other two probabilities occurring in this formula are $P(Z{=}0|X{=}0) = .48/.60$ and $P(Z{=}1|X{=}0) = .12/.60$ (see Table 1.2, p. 6). Hence,

$$P(Y{=}1|X{=}0) = \frac{.70\cdot.48}{.60} + \frac{.20\cdot.12}{.60} = .60.$$

▷ **Solution 1-9** The prima facie effects $E(Y|X{=}1) - E(Y|X{=}0)$ and $E(Y|X{=}2) - E(Y|X{=}0)$ can be computed from Table 1.3 (p. 9). They are as follows:

$$PFE_{10} = E(Y|X{=}1) - E(Y|X{=}0) = 100.00 - 111.25 = -11.25$$

and

$$PFE_{20} = E(Y|X{=}2) - E(Y|X{=}0) = 114.25 - 111.25 = 3.00.$$

▷ **Solution 1-10** The conditional prima facie effects $E(Y|X{=}1,Z{=}z) - E(Y|X{=}0,Z{=}z)$ and $E(Y|X{=}2,Z{=}z) - E(Y|X{=}0,Z{=}z)$ can be computed from Table 1.4 (p. 10). For *low neediness* $(Z{=}0)$, they are:

$$PFE_{10;Z=0} = E(Y|X{=}1,Z{=}0) - E(Y|X{=}0,Z{=}0) = 100 - 120 = -20$$

$$PFE_{20;Z=0} = E(Y|X{=}2,Z{=}0) - E(Y|X{=}0,Z{=}0) = 80 - 120 = -40.$$

For *medium neediness* $(Z{=}1)$, they are:

$$PFE_{10;Z=1} = E(Y|X{=}1,Z{=}1) - E(Y|X{=}0,Z{=}1) = 100 - 110 = -10$$

$$PFE_{20;Z=1} = E(Y|X{=}2,Z{=}1) - E(Y|X{=}0,Z{=}1) = 90 - 110 = -20.$$

Finally, for *high neediness* $(Z{=}2)$, the conditional prima facie effects are:

$$PFE_{10;Z=2} = E(Y|X{=}1,Z{=}2) - E(Y|X{=}0,Z{=}2) = 100 - 60 = 40$$

$$PFE_{20;Z=2} = E(Y|X{=}2,Z{=}2) - E(Y|X{=}0,Z{=}2) = 140 - 60 = 80.$$

▷ **Solution 1-11** Using the results of the last exercise, the average of the $(Z{=}z)$-conditional prima facie effects can be computed from the conditional effects as follows:

$$\begin{aligned} E(PFE_{10;Z}) &= PFE_{10;Z=0}\cdot P(Z{=}0) + PFE_{10;Z=1}\cdot P(Z{=}1) + PFE_{10;Z=2}\cdot P(Z{=}2) \\ &= -20\cdot\frac{1}{4} - 10\cdot\frac{1}{2} + 40\cdot\frac{1}{4} = 0. \end{aligned}$$

$$\begin{aligned} E(PFE_{20;Z}) &= PFE_{20;Z=0}\cdot P(Z{=}0) + PFE_{20;Z=1}\cdot P(Z{=}1) + PFE_{20;Z=2}\cdot P(Z{=}2) \\ &= -40\cdot\frac{1}{4} - 20\cdot\frac{1}{2} + 80\cdot\frac{1}{4} = 0. \end{aligned}$$

▷ **Solution 1-12** The two coefficients $\beta_1 \approx -3.7528$ and $\beta_2 \approx 1.1876$ are obtained by

$$\begin{aligned} \boldsymbol{\beta} = \boldsymbol{\Sigma}_{VV}^{-1}\boldsymbol{\Sigma}_{Vy} &= \begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix} \approx \begin{pmatrix}0.25 & 5.00\\ 5.00 & 189\end{pmatrix}^{-1}\begin{pmatrix}5.00\\ 205.70\end{pmatrix} \\ &\approx \begin{pmatrix}8.4944 & -0.2247\\ -0.2247 & 0.0112\end{pmatrix}\begin{pmatrix}5.00\\ 205.70\end{pmatrix} \approx \begin{pmatrix}-3.7528\\ 1.1876\end{pmatrix} \end{aligned}$$

[see Steyer & Nagel, in press, Eq. (12.54)]. The appropriate statements in R are:

```
a=matrix(c(.25,5,5,189),byrow=T,nrow=2,ncol=2)
b=matrix(c(5,205.7),byrow=T,nrow=2,ncol=1)
round(solve(a,b),4)
```

In this equation, $\boldsymbol{\Sigma}_{VV}^{-1}$ denotes the inverse of the covariance matrix of $V := (X, M)$ and $\boldsymbol{\Sigma}_{Vy}$ the covariance vector of $V = (X, M)$ and Y. The intercept $\beta_0 \approx 34.989$ is obtained by

$$\begin{aligned} \beta_0 &\approx E(Y) - \beta_1\cdot E(X) + \beta_2\cdot E(M) \\ &\approx E(Y) + 3.7528\cdot E(X) - 1.1876\cdot E(M) \\ &\approx 140 + 3.7528\cdot 0.50 - 1.1876\cdot 90 \approx 34.9924 \end{aligned}$$

[see Steyer & Nagel, in press, Eq. (12.53)].

▷ **Solution 1-13** The coefficients γ_1 to γ_4 of

$$E(Y|X, M, Z, W) = \gamma_0 + \gamma_1\cdot X + \gamma_2\cdot M + \gamma_3\cdot Z + \gamma_4\cdot W$$

are obtained by

$$\begin{aligned} \boldsymbol{\gamma} = \boldsymbol{\Sigma}_{RR}^{-1}\boldsymbol{\Sigma}_{Ry} &= \begin{pmatrix}\gamma_1\\ \gamma_2\\ \gamma_3\\ \gamma_4\end{pmatrix} = \begin{pmatrix}0.25 & 5.00 & 0.00 & 0.00\\ 5.00 & 189 & 80 & 68\\ 0.00 & 80 & 100 & 85\\ 0.00 & 68 & 85 & 100\end{pmatrix}^{-1}\begin{pmatrix}5.00\\ 205.70\\ 116.50\\ 124.00\end{pmatrix} \\ &= \begin{pmatrix}20.0000 & -0.8000 & 0.6400 & 0.0000\\ -0.8000 & 0.0400 & -0.0320 & 0.0000\\ 0.6400 & -0.0320 & 0.0616 & -0.0306\\ 0.0000 & 0.0000 & -0.0306 & 0.0360\end{pmatrix}\begin{pmatrix}5.00\\ 205.70\\ 116.50\\ 124.00\end{pmatrix} = \begin{pmatrix}10.00\\ 0.50\\ 0.00\\ 0.90\end{pmatrix} \end{aligned}$$

[see again Steyer & Nagel, in press, Eq. (12.54)]. In this equation, $\boldsymbol{\Sigma}_{RR}^{-1}$ denotes the inverse of the covariance matrix of $R := (X, M, Z, W)$ and $\boldsymbol{\Sigma}_{Ry}$ the covariance vector of $R = (X, M, Z, W)$ and Y. The appropriate statements in R are:

```
a=matrix(c(.25,5,0,0,5,189,80,68,0,80,100,85,0,68,85,100),
         byrow=T,nrow=4,ncol=4)
b=matrix(c(5,205.7,116.5,124),byrow=T,nrow=4,ncol=1)
round(solve(a,b),4).
```

The intercept $\gamma_0 = 0.00$ is obtained by

$$\begin{aligned}\gamma_0 &= E(Y) - [\gamma_1 \cdot E(X) + \gamma_2 \cdot E(M) + \gamma_3 \cdot E(Z) + \gamma_4 \cdot E(W)] \\ &= E(Y) - 10 \cdot E(X) - 0.50 \cdot E(M) - 0.00 \cdot E(Z) - .90 \cdot E(W) \\ &= 140 - 10 \cdot 0.50 - 0.50 \cdot 90 - 0.00 \cdot 100 - .90 \cdot 100 \;=\; 0.00.\end{aligned}$$

[see again Steyer & Nagel, in press, Eq. (12.53)].

▷ **Solution 1-14** No solution provided. Just compare your results to the parameters presented in Table 1.4 (p. 10).

▷ **Solution 1-15** No solution provided. Just compare your estimated parameters to the true parameters presented in Equations (1.3) and (1.4).

Chapter 2
Some Typical Random Experiments

In chapter 1 we have shown that comparing conditional expectation values of an outcome variable between treatment groups can be completely misleading if used for the evaluation of treatment effects. We have also shown that regression coefficients and the conditional expectations they describe can be completely misleading even in the randomized experiment, if used to determine the direct treatment effect with respect to an intermediate variable M. In this chapter we will prepare the stage for the theory of causal effects, describing the kind of empirical phenomena it refers to: single-unit trials of experiments or quasi-experiments, but also single-unit trials of observational studies in which causal effects can be investigated.

A single-unit trial is a specific random experiment. Note the distinction between a *random experiment* and a *randomized experiment*. Stochastic dependencies between events and between random variables always refer to a random experiment, but not necessarily to a *randomized experiment* in which a subject is assigned to one of the treatment conditions by a randomization procedure. In the simplest case of such a randomization we assign the subject to treatment or control according to the outcome of flipping a coin. In contrast, a *random experiment* is the concrete empirical phenomenon to which stochastic dependencies between events and random variables described by conditional distributions, probabilities, correlations, and conditional expectations refer to.

The single-unit trial *is not the sample* dealt with in statistical models. In a sample, the single-unit trial is repeated many times in one way or another. This is necessary when it comes to estimating parameters and testing hypotheses about these parameters, some of which might be causal effects. The single-unit trial does *not allow* treating problems of parameter estimation or hypothesis testing. However, it is sufficient for defining causal effects and studying how to identify them, i. e., studying under which conditions and how they can be computed from empirically estimable parameters.

A single-unit trial is also what we refer to in hypotheses and theories of the empirical sciences. Furthermore, single-unit trials are what is of interest in practical work. How does the treatment of a patient affect the outcome of this patient if compared to another possible treatment? What is the treatment effect for a male, and what is its effect for a female? What is the direct treatment effect (e. g., of vaccination) on the outcome variable (e. g., influenza) that is *not* transmitted through a specific intermediate variable (e. g., a measure of certain antibodies)?

Which variables explain inter-individual differences in individual causal effects? All these questions are raised using concepts referring to single-unit trials.

Overview

We start with the single-unit trial of simple experiments and then treat increasingly more complex ones introducing additional design features. Specifically, we will introduce the single-unit trials of experiments and quasi-experiments with fallible covariates, a multifactorial design with more than one treatment, multilevel experiments and quasi-experiments, and experiments and quasi-experiments with intermediate variables and latent outcome variables.

We also discuss different kinds of random variables that will play a crucial role in the chapters to come. Among these random variables are the observational-unit variable, other manifest and latent covariates, treatment variables, intermediate variables, as well as manifest and latent outcome variables. In this chapter, we confine ourselves to an informal description of single-unit trials and the random variables involved, preparing the stage for their mathematical representations in the following chapters.

2.1 Simple Experiments

As a first class of random experiments, we consider the single-unit trials of *simple experiments and quasi-experiments*. Such single-unit trials are between-subjects experiments and quasi-experiments in which *no fallible covariates* are assessed.

Such a single-unit trial consists of:

(a) sampling an observational unit u (e. g., a person) from a set (sometimes called 'population') of units,
(b) assigning the unit or observing its assignment to one of several experimental conditions (represented by the value x of the treatment variable X),
(c) recording the numerical value y of the outcome variable Y.

Figure 2.1 displays a tree representation of the set of possible outcomes of this single-unit trial. Note that this is the kind of random experiment we (implicitly) referred to describing Simpson's paradox in chapter 1. The random variables X (treatment), Y (success), and Z (sex), the conditional expectation values $E(Y|X{=}x)$ and $E(Y|X{=}x,Z{=}z)$, as well as the probabilities $P(X{=}x)$, $P(Z{=}z)$, $P(X{=}x,Z{=}z)$ all referred to such a single-unit trial. Of course, all these conditional expectation values and probabilities are unknown in applications. Nevertheless, they are the parameters that stochastically determine the outcome of the single-unit trial, just in the same way as the probability of tossing *heads* stochastically determines the outcome of flipping a coin.

In order to illustrate this point, imagine flipping a deformed coin that has the shape of a Chinese wok and suppose that in this case the probability of flipping *heads* is .80 instead of .50. Although these probabilities do not deterministically

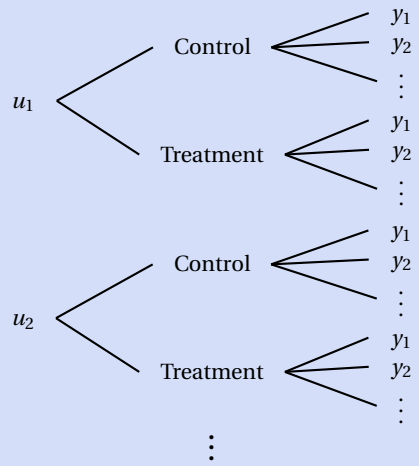


Figure 2.1. A simple experiment or quasi-experiment

determine the outcomes of flipping the coins, they stochastically determine the outcomes.

In fact, we may consider the single-unit trial of (a) sampling a coin u from a set of coins, (b) forming ($X=1$) or not forming ($X=0$) a wok out of it, and (c) observing whether ($Y=1$) or not ($Y=0$) we then toss *heads*. In this single-unit trial, the difference .80 − .50 = .30 would be the causal effect of the treatment variable X on the outcome variable Y. Let us emphasize that the probabilities .80 and .50 and their difference .30 refer to this single-unit trial, although these probabilities can only be estimated if we conduct many of these single-unit trials, i. e., if we draw a sample. However, if these probabilities were known, we could dispense with a sample and the data that would result from drawing it (see Exercise 2-1), and still have a perfect prediction for the outcome of such a single-unit trial.

2.1.1 Sampling a Unit

The first part of this single-unit trial consists of sampling an observational unit. In the social sciences, units often are persons, but they might be groups, school classes, schools and even countries. Usually such units change over time. Therefore, it should be emphasized that, in simple experiments and quasi-experiments, we are talking about the *units at the onset of treatment*. Later we will see that we have to distinguish between *units at the onset of treatment* and *units at the time of assessment of the outcome variable*, which might be months or even years later. In a single-unit trial of simple experiments and quasi-experiments, the units can

be represented by the observational-unit variable U, whose possible values u are the *units at the onset of treatment*.

Note that the unit at the onset of treatment also comprises his or her experiences a year and/or the day before treatment, as well as the psycho-bio-social situation in which he or she is at the onset of treatment. Both, the experiences and the situation, already happen *before* the onset of treatment. Therefore, they are attributes of the observational units u, and this is true although they once were just possible events that had some (unknown) probabilities to occur. Looking at them at the time of the onset of treatment, they are no events any more that may or may not occur. Instead, these prior experiences and situations are then *fixed attributes* of the units. They can be treated in the same way as other attributes such as sex and educational status.

2.1.2 Treatment Variable

In an experiment or quasi-experiment, there is always a treatment variable, which we usually denote by X. The unit drawn is either assigned — e. g., by the experimenter or by some other person (such as a physician, a psychologist, or a social worker) — to one of the possible treatments, or we observe self-selection to one of the treatment conditions. In the simplest case there are at least two treatment conditions, e. g., *treatment* and *control*. These treatment conditions are the possible values of the treatment variable X. For simplicity, we use the values $0, 1, \ldots, J$ to represent $J+1$ treatment conditions. Furthermore, unless stated otherwise, we presume that treatment *assignment* and actual *exposure* to treatment will be equivalent, i. e., unless stated otherwise, we assume that there is perfect compliance.

Selection of a unit into one of the treatment conditions x may happen with unknown probabilities, e. g., when there is self-selection or assignment by an unknown physician. This is often the case in quasi-experiments. However, assignment can also be done with known probabilities that are equal for different units (such as in the simple randomized experiment) or with known probabilities that may be unequal for different units (such as in the conditionally randomized experiment). In this case, these treatment probabilities may also depend on a covariate Z representing pre-treatment attributes of the units. *Conditional and unconditional randomized assignment, distinguish the true experiment from the quasi-experiment*, in which the assignment or selection probabilities are unknown. (See section 7.5 for more details on randomization and conditional randomization.)

2.1.3 Covariates

In simple experiments and quasi-experiments, the focus is usually on the treatment effects on an outcome variable. Hence, if we are interested in the treatment variable as a cause, then each attribute of the observational units is a covariate. Examples are *sex*, *race*, *educational status*, and *socio-economic status*. Once

the unit is drawn, its *sex, race, educational status,* and *socio-economic status* are fixed as well. This means that there is no additional sampling process associated with assessing these covariates. This is also the reason why they do not appear in points (a) to (c) describing the single-unit trial (see p. 26).

Because covariates represent attributes of the unit *at the onset of treatment* they can never be affected by the treatment. However, there can be (stochastic) dependencies between the treatment variable and covariates. In Simpson's paradox, for instance, there is a strong correlation between *sex* and the treatment variable.

Multidimensional Covariates

Covariates may be uni- or multi-dimensional, qualitative (such as Z_1 := *sex* and Z_2 := *educational background*) or quantitative (such as Z_3 := *height* and Z_4 := *body mass index*) or, if it is a multivariate variable made up of several uni-dimensional variables, it may consist of qualitative *and* quantitative covariates such as $Z_5 := (Z_1, Z_4)$.

Specific Covariates

Note that the U-conditional treatment probabilities $P(X{=}x|U)$ and the Z-conditional treatment probabilities $P(X{=}x\,|\,Z)$ are covariates as well, provided that Z is a covariate. (The mathematical background for this statement are chapters 2 and 10 of Steyer and Nagel (in press).) Furthermore, the *assignment* to treatment x with values 'yes' and 'no' is also covariate, if *assignment to treatment* and *exposure to treatment* (again with values 'yes' and 'no') are not identical. This distinction is useful in experiments with non-compliance (see, e. g., Jo, 2002a, 2002b, 2002c; Jo et al., 2008).

Unobserved Covariates

Even if we consider a multivariate covariate Z consisting of several univariate covariates, there are always unobserved variables that are prior or simultaneous to treatment. Such variables are called *unobserved covariates*. Sometimes they are also called *hidden confounders* (cf., e. g., Rosenbaum, 2002). Of course such an unobserved covariate may bias the conditional expectation values of the outcome variable just in the same way as an observed covariate. Whether or not the conditional expectation values of the outcome variable in the treatment conditions are unbiased such that their differences represent causal effects does not only depend on the relationship between the measured variables such as X, Y, and the observed (possible multivariate) covariate, say Z, but also on the relationship of these variables to the unobserved covariates. In other words, covariates exert their maleficent effects irrespective of whether or not we observe them.

2.1.4 Outcome Variable

Of course, the outcome variable Y refers to a time at which the treatment might have had its impact. Hence, treatment variables are always prior to the outcome variable. In principle, we may also observe several outcome variables, e. g., in order to study how effects of a treatment grow or decline over time or to study effects that are not limited to a single outcome variable. All random variables mentioned above refer to a concrete single-unit trial and they have a joint distribution. Each combination of unit, treatment condition, and score of the outcome variable may be an observed result of such a single-unit trial. This implies that the variables U, Z, X, and Y, as well as unobserved covariates, say W, have a joint distribution. (See, e. g., section 5.3 of Steyer and Nagel (in press).) Once we specified the random experiment to be studied, this joint distribution is fixed, even though it might be known only in parts or even be unknown completely.

2.1.5 Causal Effects and Causal Dependencies

There is already a plenitude of different kinds of causal effects and causal dependencies that can be considered in the single-unit trial of a simple experiment or quasi-experiment. For simplicity, suppose the treatment has just two values, say *treatment* and *control*. *First*, there is the *average total effect* of treatment (compared to control) on the outcome variable Y. *Second*, there are the *conditional total treatment effects* on Y, where we may condition on any function of the observational-unit variable U. If, e. g., $Z :=$ *sex* with values m for *male* and f for *female*, then we may consider the $(Z=m)$-conditional total treatment effect on Y, i. e., the average total treatment effect for males, and the $(Z=f)$-conditional total treatment effect on Y, i. e., the average total treatment effect for females. Similarly, if $Z :=$ *socio-economical status*, we may consider the conditional total treatment effects on Y for each status group, etc. *Third*, although difficult and often impossible to estimate, we may also consider the *individual total effect* of *treatment* compared to *control* on Y.

By definition, within a *simple* experiment and quasi-experiment we cannot consider any *direct* treatment effects with respect to a specified intermediate variable, i. e., the effects of the treatment on the outcome variable that *are not* transmitted through a specified intermediate variable M. However, the total treatment effects discussed above are, of course, transmitted through intermediate variables, irrespective of whether or not we observe (or are aware of) these intermediate variables. (See section 2.5 for experiments and quasi-experiments with observed intermediate variables).

Aside from the treatment effects discussed above we can also consider the causal effects of a covariate. Among the causal effects of such a covariate are its average total effect on the outcome variable Y, its conditional direct effects on Y given the different values of the treatment variable — which now takes the role of an intermediate variable — the average of these X-conditional direct effects and its indirect effect mediated by X. In a randomized experiment, e. g., the

causal effects of all covariates on X will be zero. In other words, the zero prima facie effects created by randomization will *not be biased* or *spurious*. In contrast, in quasi-experiments, causal effects of some covariates on X might be different from zero. In self-selection, e. g., *neediness* for a therapy might have strong average effects on the treatment variable. Furthermore, neediness often has strong $(X{=}x)$-conditional direct effects and a strong average direct effect on the outcome variable if it measures some aspects of health.

These effects of the covariates on the treatment variable and on the outcome variable are discussed in the literature on structural equation modeling (see, e. g., Bentler, 1995; Bollen, 2002; Kaplan, 2000; S.-Y. Lee, 2007; Little, Card, Bovaird, Preacher, & Crandall, 2007; MacCallum & Austin, 2000; Muthén & Muthén, 1998-2007) and graphical modeling (see, e. g., Cox & Wermuth, 2004; Greenland, Pearl, & Robins, 1999; Spirtes et al., 2000; Pearl, 1995, 1998, 2009), whereas they have been criticized in the Rubin tradition (see, e. g., Holland, 1986). What should be noted is that the causal effects of covariates have no 'individual causal interpretation' (see, e. g., Borsboom, Mellenbergh, & van Heerden, 2003). While the average total effect of a treatment variable can be interpreted as the effect of the treatment on an unknown, randomly drawn unit that can be exposed to treatment or control, the effects of a covariate such as *neediness* or *sex* on an outcome variable do *not* have such an individual causal interpretation. The unit (at the onset of treatment) *has* a certain degree of *neediness* and it *has* a sex, but it cannot be *exposed* to a neediness condition or a sex. Nevertheless, neediness effects and sex effects can be 'spurious' or 'biased', and we can define and aim at estimating 'unbiased' or 'causal' neediness and sex effects. As we discuss in more detail in chapter 5, we can consider both, a treatment variable or an attribute of the units, as causes. The conceptual framework provided in the chapters to come will cover both kinds of causal effects which seem — and with respect to manipulability at the individual level *are* — different from a content point of view.

2.2 Experiments With Fallible Covariates

Another class of random experiments are single-unit trials of experiments and quasi-experiments in which we assess a fallible covariate. In this case, there is at least one covariate that is *not* a (deterministic) attribute of the observational units. The single-unit trial of such an experiment of quasi-experiment consists of:

(a) sampling an observational unit u (e. g., a person) from a population of units,
(b) assessing the values $z_1, \ldots, z_k$ of the covariates $Z_1, \ldots, Z_k$, $k \geq 1$.
(c) assigning the unit or observing its selection to one of several experimental conditions (represented by the value x of the treatment variable X),
(d) recording the numerical value y of the outcome variable Y.

The crucial difference to a simple (quasi-) experiment is that there is variability of the manifest covariate *given the observational unit u* (see Fig. 2.2). In this case,

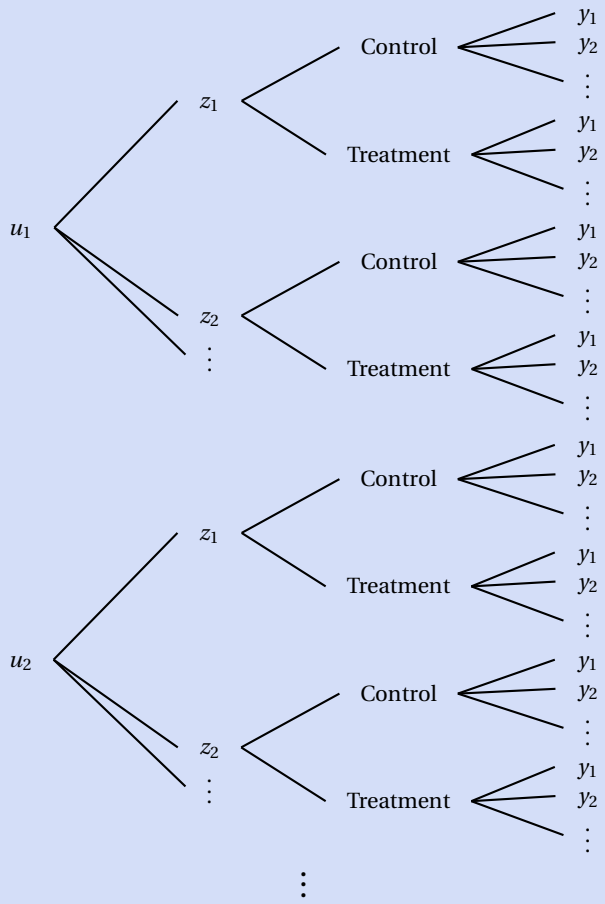


Figure 2.2. Experiment or quasi-experiment with a fallible covariate

we may distinguish between the *latent covariate*, say ξ, representing the attribute to be assessed and its *fallible measures*, the manifest variables $Z_1, \ldots, Z_k$ actually observed. This distinction does not only open up the possibilities to study the effects of the latent covariate on the treatment variable and on the outcome variable, but also for investigating whether or not the dependency of the fallible measures on the latent variable is causal.

Furthermore, this distinction also implies that the *unit* whose attributes are measured *at the time when the covariate is assessed* is not identical any more with the *unit at the onset of treatment* (see section 2.1). The covariate might be assessed some months before the treatment is given — enough time and plenty of possibilities for the unit to change in various ways, e. g., due to maturation,

learning, critical life events, and other experiences that are not fixed at the time of assessing the covariate. As a consequence, a variable, say W, representing such intermediate events and experiences may also affect the outcome variable Y over and above (a) the covariate Z, (b) the treatment variable X, and (c) the observational-unit variable U, which now represents the *observational units at the time of the assessment of the covariate* $Z := (Z_1, \ldots, Z_k)$. In other words, the outcome variable Y does not necessarily only depend on the units, covariates , and the treatment variable alone. Instead, it may also depend on an unobserved covariate W lurking in between the assessment of the observed covariate Z and the onset of treatment. This is one of the reasons why we need to define causal effects in a more general way than in the Neyman-Rubin tradition (see chs. 4 and 5).

Note that assessing a fallible covariate does not only change the interpretation of the observational-unit variable U, but it also changes the random experiment, and with it, the empirical phenomenon we are considering. Assessing, prior to treatment, a fallible covariate such as a pre-test of an ability, an attitude, or a personality trait, may change the observational units and their attributes, as well the effects of the treatment on a specified outcome variable, which usually is related to such a pre-test. This has already been discussed by Campbell and Stanley (1963), who also recommended designs for studying the effects of pre-treatment assessment.

Covariates

What are the covariates in such a single-unit trial? First of all, we have to choose the cause to be considered. If it is the treatment variable X, then each attribute of the unit at the time of the assessment of the observed covariates $Z_1, \ldots, Z_k$ is a covariate pertaining to X as well. This does not only include variables such as *sex*, *race*, and *educational status*, but also the latent covariate, say ξ, (which might be multi-dimensional). Furthermore, aside from the manifest covariates $Z_1, \ldots, Z_k$, each variable W representing an intermediate event or experience of the unit (occurring in between the assessment of the observed covariates and the onset of the treatment), as well as any attribute of the *unit at the onset of treatment* is a covariate as well, irrespective of whether or not these covariates are observed.

Note that a latent covariate ξ may be considered a cause of its fallible measures $Z_1, \ldots, Z_k$ but also of the outcome variable Y. This is not in conflict with the theory that the treatment variable X is a cause of Y as well. In this kind of single-unit trial, we have several causes and several outcome variables that are affected by these causes. Again it would be possible to consider the treatment variable X to be causally dependent on the manifest or latent covariates. In other words, we may also raise the question if the treatment probabilities $P(X{=}1 \mid Z_1, \ldots, Z_k)$ or $P(X{=}1 \mid \xi)$ describe causal dependencies. This makes clear that the term 'covariate' can only be defined with respect to a focused cause.

2.3 Two-Factorial Experiments

As a third class of random experiments we consider two-factorial experiments. The single-unit trial of such a two-factorial experiment or quasi-experiment consists of:

(a) sampling an observational unit u (e. g., a person) from a population of units,
(b) assigning the unit or observing its assignment to one of several experimental conditions that are defined by the pair (x, z) of levels of two treatment variables X and Z, respectively.
(c) recording the numerical value y of the outcome variable Y.

Sampling a Unit

Because we presume that no fallible covariates such as 'severity of symptoms', 'motivation for treatment', etc. are assessed before treatment, sampling an observational unit means that we are sampling a *unit at the onset of treatment*.

Treatment Variables

As a simple example, let us consider an experiment in which we study the effects — including the joint effects — of two treatment factors, say *individual therapy* represented by X (with values 'yes' and 'no') and *group therapy* represented by Z (with values 'yes' and 'no').

In such a two-factorial experiment, we consider *group therapy* as a covariate and *individual therapy* to be the treatment variable in order to ask for the conditional and average total effects of individual therapy given group therapy. In contrast, we may also consider *individual therapy* to be a covariate and *group therapy* to be the focused treatment variable. Finally, we may also consider the two-dimensional variable (X, Z) as the treatment variable. Which option is chosen depends on the causal effects we are interested in (see below).

Outcome Variable

Again, the outcome variable Y refers to a time at which the treatment might have had the impact to be estimated. Hence, both treatment variables are prior to the outcome variable considered. And again, we may also observe several outcome variables, e. g., in order to study how effects of a treatment grow or decline over time or to study effects that are not limited to a single outcome variable.

Causal Effects

There are several causal effects we might look at. If X and Z have only two values, then we may be interested in the following effects on the outcome variable Y:

(a_1) the conditional total effect of 'individual therapy' as compared to 'no individual therapy' given that the unit treated also receives 'group therapy',
(b_1) the corresponding conditional total effect given that the unit does *not* receive 'group therapy', and
(c_1) in the average of these conditional total effects of 'individual therapy' as compared to 'no individual therapy', averaging over the two values of Z.

Vice versa, we might also be interested in the following effects on the outcome variable Y:

(a_2) the conditional total effect of 'group therapy' as compared to 'no group therapy' given that the unit treated also receives 'individual therapy',
(b_2) the corresponding conditional total effect given that the unit does *not* receive 'individual therapy', and
(c_2) in the average of these conditional total effects of 'group therapy' as compared to 'no group therapy', averaging over the two values of X.

Furthermore, there are other causal effects on Y we might study, namely

(a_3) the total effect of receiving 'individual therapy' *and* 'group therapy' as compared to receiving none of the two treatments.
(b_3) the total effect of receiving 'individual therapy' *and* 'group therapy' as compared to receiving 'individual therapy' only.
(c_3) the total effect of receiving 'individual therapy' *and* 'group therapy' as compared to receiving 'group therapy' only.
(d_3) the total effect of receiving 'individual therapy' *and* 'no group therapy' as compared to receiving 'group therapy' *and* 'no individual therapy'.

All these effects may answer meaningful causal questions and in fact, there are even more causal effects than those listed above even if we do not count the various conditional total effects we might want to study if additional covariates such as *sex* or *educational status* are considered.

Covariates

If we focus the effect of X (individual therapy), then we consider Z (group therapy) as a covariate , whereas we treat X as a covariate if we study the effects of Z (group therapy). In both cases, each attribute of the unit at the onset of treatment (such as *sex* or *educational status*) could be considered as covariates as well. Assessing these covariates does not appear in points (a) to (c) of the random experiment, because these covariates are (deterministic) functions of the observational-unit variable. Therefore, there is no additional sampling process associated with their assessment.

This is also true for other covariates, e. g., variables characterizing the situation in which the unit is at the onset of treatment, the number of *hours slept* last night, or *day time* at which the unit receives its treatment. Even variables that characterize early experiences in the childhood of the unit such as a *broken home* or

mother's child care behavior are covariates in this single-unit trial. They are there and exert their effects even if they are not assessed.

Note, again that assessment of these covariates in a questionnaire filled in by the person constitutes a new random experiment that may differ in important ways from a random experiment in which the unit has no such task (see section 2.2). In psychology, an assessment often is a treatment of its own.

2.4 Multilevel Experiments

In multilevel experiments and quasi-experiments we also study the effect of a treatment on an outcome variable. However, in such a design, the observational units are nested within higher hierarchical units referred to as *clusters*. Examples include experiments, in which students are nested within classrooms, patients are nested within groups of treated patients, and inhabitants are nested in neighborhoods. Multilevel designs can be classified as designs with treatment assignment at the unit-level or at the cluster-level. Furthermore, multilevel designs differ with respect to the assignment of units to clusters. There are designs with pre-existing clusters and there are designs with assignment of units to clusters. All these designs involve different single-unit trials.

A single-unit trial with *pre-existing clusters* consists of:

(a) sampling a cluster c (e. g., a school class, a neighborhood or a hospital) from a set of clusters,
(b) sampling an observational unit u (e. g., a person) from a set of units within the cluster,
(c) assigning the unit or the cluster (depending on the design) or observing their assignment to one of several experimental conditions (represented by the value x of the treatment variable X),
(d) recording the numerical value y of the outcome variable Y.

In contrast, a *single-unit trial with assignment of units to clusters* consists of:

(a) sampling an observational unit u (e. g., a person) from a population of units,
(b) assigning the unit or observing its assignment to one of several clusters (represented by the value c of the cluster variable C),
(c) assigning the unit or the cluster (depending on the design) or observing their assignment to one of several experimental conditions (represented by the value x of the treatment variable X),
(d) recording the numerical value y of the outcome variable Y.

In the experiment with pre-existing clusters, each unit can only appear in one cluster, whereas in the experiment with assignment of units to clusters, each unit can appear in more than one cluster. Hence, in the latter designs the cluster variable can bias the dependency of the outcome variable on the treatment variable *on the level of the observational unit*. In this aspect this design resembles the multifactorial design described in section 2.3.

Covariates

What are the covariates in multilevel designs if the treatment variable X is considered as the cause? The answer depends on the type of design considered: In designs with treatment assignment at the unit-level, attributes of the observational unit (such as *sex*, *race* or *educational status*) are covariates, but also attributes of the cluster (such as *school type*, *hospital ownership* or cluster-specific expectations of covariates at the unit-level, such as *school-level of socio-economic status* or *school-level intelligence*). In these designs, clusters may not only be considered as covariates, but also as treatments, because some of the effects observed later on may depend on the composition of the group to which a particular unit, say Joe, is assigned. Receiving group therapy together with beautiful Ann in the same group might make a great difference as compared to getting it together with awful Jim. In designs in which clusters as a whole are assigned to treatment conditions, only attributes of the cluster can influence the assignment. Hence, in data analysis we would focus on controlling for the covariates on the cluster level (see, e. g., Nagengast, 2009 for more details).

2.5 Experiments With Intermediate Variables

Another class of random experiments are experiments with *intermediate variables*. The basic goal of such a experiment is to investigate if and to which degree the effect of a cause X (such as a treatment variable) on an outcome variable Y may be *mediated* or *transmitted* by another variable, say M. A first example is mediation of the effect of *vaccination* (with values *yes* or *no*) on *the severity of influenza symptoms* by the *amount of antibodies*. Another example is mediation of the effect of *teachers encouragement* (with values *yes* or *no*) on the *achievement* by the *amount of time spent on learning* (see, e. g., Holland, 1988; Sobel, 2008, or Rubin, 2004).

In the simplest case with a single manifest intermediate variable we consider the following single-unit trial that consists of:

(a) sampling a person u out of a set of persons (the population of persons),
(b) assigning the unit or observing its selection to one of several experimental conditions (represented by the value x of the treatment variable X),
(c) assessing the value m of an intermediate variable M, and
(d) recording the numerical value y of the outcome variable Y.

In this single-unit trial, the values u of the observational-unit variable U again represent the observational *unit at the onset of treatment*, while the intermediate variable M represents some attribute of the *unit at the time point at which the intermediate variable is assessed*. This time point is *in between* the onset of treatment and the assessment of the outcome variable Y (see Fig. 2.3). If M is fallible, then we distinguish between M and the latent variable to be measured by M. In this case we would need an additional layer in the tree representation for the

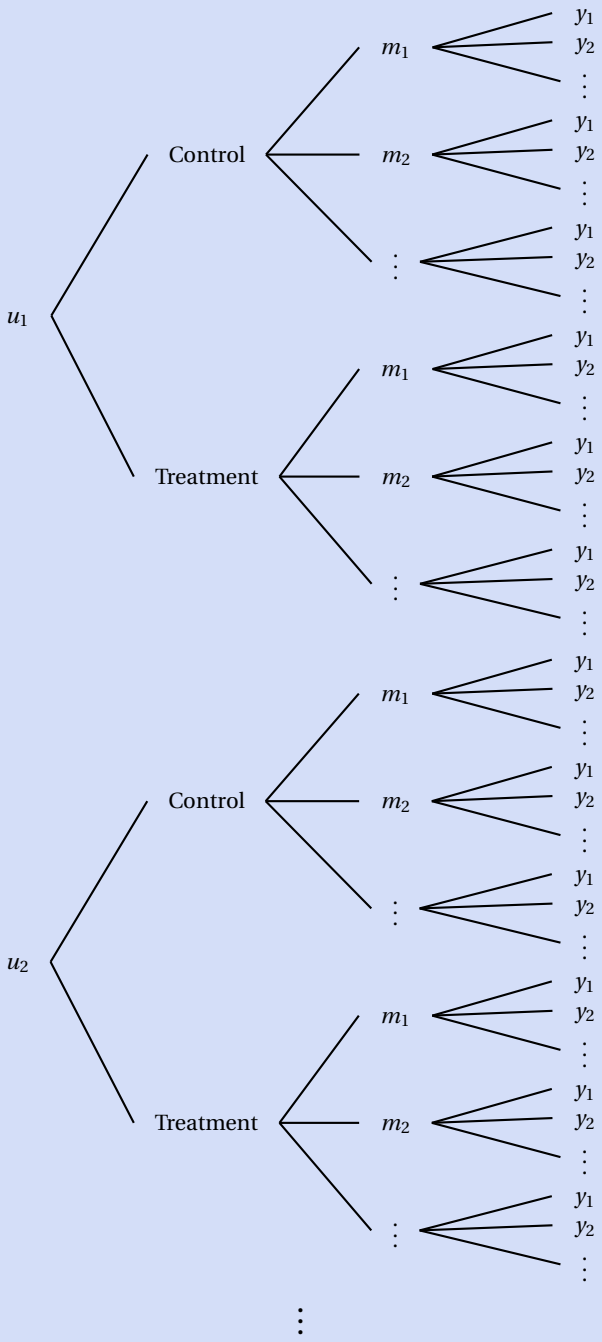


Figure 2.3. Experiment or quasi-experiment with an intermediate variable

latent intermediate variable. Furthermore, instead of a single manifest intermediate variable, we would need several manifest intermediate variables measuring the latent intermediate variable.

Covariates

What are the covariates in such a single-unit trial? Again, the answer depends on the choice of the cause. If it is the treatment variable X, then each attribute of the unit at the onset of treatment is a covariate pertaining to X. Examples are *sex*, *race*, and *educational status*. Note that the set of covariates of X is the same irrespective of the choice of the outcome variable. In this example we may choose the original outcome variable Y; however, we may also choose the intermediate variable M as an outcome variable in order to study the effects of X on M.

Focusing M as a cause, brings additional covariates into play, namely all those variables that are *in between* treatment and the assessment of the intermediate variable. This could be critical life events, additional drugs taken after treatment and before the assessment of the intermediate variable, or an additional treatment to which the unit is exposed and which may or may not be manipulated by the experimenter.

2.6 Experiments With Latent Outcome Variables

We may also consider single-unit trials of experiments with a *latent outcome variable*. The basic goal of such experiments is to investigate the effect of the treatment variable X on a *latent* outcome variable, say η. This is of interest, for example, where a quantitative outcome variable can only be measured by qualitative observations such as solving or not solving certain items indicating the (latent) ability. However, it can also be of interest if the manifest measures are *linearly* related to the latent variable such as in models of classical test theory (see, e. g., Steyer, 2001) or in models of latent state-trait theory (see, e. g., Steyer, Mayer, Geiser, & Cole, 2015). If, e. g., there are three manifest variables Y_1, Y_2, and Y_3 measuring a single latent variable η, we may ask if there is just one single effect of the treatment on the latent outcome variable η — which transmits these effects to the manifest variables Y_1, Y_2, and Y_3 — instead of three separate effects of X on each variable Y_i. Hence, the latent variable may also be considered to be a mediator variable. Showing that all effects of X on the variables Y_i are indirect, i. e., mediated by η is one of the research efforts that aims at establishing construct validity of the latent variable η.

In the simplest case with a single latent variable, we consider the following single-unit trial:

(a) Sampling a person u out of a set of persons (the population of persons),
(b) assigning the unit or observing its selection to one of several experimental conditions (represented by the value x of the treatment variable X),

(d) recording the numerical values $y_1, \ldots, y_m$ of the manifest outcome variables $Y_1, \ldots, Y_m$.

In this single-unit trial, the values u of the observational-unit variable U again represent the observational *unit at the onset of treatment*, while the latent outcome variable η represents some attribute of the unit at the time point at which the outcome of the treatment is assessed. Clearly, this time point is *after* treatment and *prior* to the observation of the manifest outcome variables Y_i, at least as long as we preclude change in the latent variable during the process of assessing the manifest outcome variables. If this cannot be precluded, we would have to consider the time sequence in assessing the manifest outcome variables (e. g., of the items to be solved) as well.

Covariates

What are the covariates in such a single-unit trial? Again, the answer depends on the cause considered. If it is the treatment variable X, then each attribute of the *unit at the onset of treatment* is a covariate (with respect to X). Obviously, this again includes variables such as *sex*, *race*, and *educational status*. Note that in this kind of experiments, the set of covariates of X is the same irrespective of the choice of the outcome variable. Remember, we may not only consider the *latent* outcome variable η but also the *manifest* outcome variables Y_i, e. g., in order to study whether or not the effects of X on these manifest outcome variables are perfectly transmitted (or mediated) through the latent variable η.

Choosing the latent outcome variable η as a cause of the manifest outcomes variables Y_i brings additional covariates into play, for instance, all those variables that are *in between* treatment and the assessment of η. If, e. g., we consider an experiment studying the effects of different teaching methods, these additional covariates are critical life events (such as father or mother leaving the family), or additional lessons taken after treatment and before outcome assessment, for instance.

2.7 Summary and Conclusions

In this chapter we described a number of random experiments in informal terms. The purpose was to get a first idea which kind of empirical phenomena causal theories and hypotheses refer to. We focused on single-unit trials, which are the kinds of empirical phenomena we are interested in, both in theory and practice. We emphasized that a single-unit trial is a random experiment and discussed several kinds of random variables playing a crucial role in the theory of causal effects. We also mentioned that there is a certain *time order* among these random variables, e. g., saying that the covariates are 'prior' or 'simultaneous' to the treatment variable, which itself is 'prior' to the outcome variable. Furthermore, for each single-unit trial and each cause in such a single-unit trial, we discussed the

Box 2.1 Glossary of New Concepts

Random experiment	The kind of empirical phenomenon that events, random variables, and their dependencies refer to.
Single-unit trial	A particular kind of random experiment that consists of sampling a unit from a set of observational units and observing the values of one or more random variables related to this unit.
Cause	A random variable. Its effect on an outcome variable is considered.
Outcome variable	A random variable. Its dependency on a cause is considered.
Covariate of a cause	A random variable that can never be affected by the cause. It is prior or simultaneous to the cause. It might be correlated with the cause and the outcome variable.
Fallible covariate	A covariate that is assessed with measurement error.
Latent covariate	A covariate that is not directly observed. Instead it is defined by a set of manifest variables and a measurement model describing the dependencies of the manifest variables on the latent covariate.
Intermediate Variable	A variable that might mediate (transmit) the effect of the cause on the outcome variable. The cause is always prior to an intermediate variable and an intermediate variable is always prior to the outcome variable. An intermediate variable is not *necessarily* affected by the cause and it does not *necessarily* have an effect on the outcome variable.
Mediator	An intermediate variable on which X has a causal effect and which itself has a causal effect on the outcome variable Y.

Note that all these terms are still of an informal nature. Their mathematical treatment starts in chapter 3.

covariates involved. We emphasized that each cause considered in such a single-unit trial has its own set of covariates.

Other Single-Unit Trials

The single-unit trials discussed in this chapter are just a small selection of single-unit trials in which causal effects and causality of stochastic dependencies are of interest. We might also consider single-unit trials with latent covariates *and* la-

tent outcome variables *and* manifest and/or latent intermediate variables, but also single-unit trials with multiple mediation. Furthermore, we could also consider single-unit trials of growth curve models (see, e. g., Biesanz, Deeb-Sossa, Aubrecht, Bollen, & Curran, 2004; Bollen & Curran, 2006; Meredith & Tisak, 1990; Singer & Willett, 2003; Tisak & Tisak, 2000), latent change models (see, e. g., McArdle, 2001; Steyer, Eid, & Schwenkmezger, 1997; Steyer, 2005), or cross-lagged panel models (see, e. g., Kenny, 1975; Rogosa, 1980b; Watkins, Lei, & Canivez, 2007; Wolf, Chandler, & Spies, 1981). Causality is also an issue in uni- and multivariate time-series analysis as well as in stochastic processes with continuous time. However, in this book our examples will usually deal with experiments and quasi-experiments, including latent covariates and outcome variables as well as intermediate variables.

Outlook

Steyer and Nagel (in press) study how random experiments and the dependencies between events and random variables can be represented in terms of probability theory. In chapter 3 we extend the mathematical structure so that we can also meaningfully talk about time order between events and random variables and distinguish between *covariates* and intermediate variables. This will provide the mathematical framework in which causal effects can be meaningfully discussed.

2.8 Exercises

▷ **Exercise 2-1** Imagine that the probabilities of a crash for a flight with Airline A is ten times smaller than with Airline B. Which airline would you choose?

▷ **Exercise 2-2** Why does the theory of causal effects refer to single-unit trials?

▷ **Exercise 2-3** Why is it important to know which random experiment we are talking about?

▷ **Exercise 2-4** Which type of random experiment did we refer to in Simpson's paradox and in the nonorthogonal ANOVA example described in chapter 1?

▷ **Exercise 2-5** Why is it important to emphasize that, in simple experiments and quasi-experiments (see section 2.1), the observational-unit variable U represents the observational units *at the onset of treatment*?

▷ **Exercise 2-6** What is the basic idea of a covariate pertaining to a cause?

▷ **Exercise 2-7** Which kinds of causal effects can be considered in the simple experiment or quasi-experiment in which no *fallible* covariate and no intermediate variable is assessed?

▷ **Exercise 2-8** Which are the covariates pertaining to an intermediate variable if it is considered a cause of the outcome variable?

Solutions

▷ **Solution 2-1** Of course, B. Note that we apply these probabilities to the random experiment of flying *once* with A or B, even if these probabilities have been estimated in a sample.

▷ **Solution 2-2** Within such a single-unit trial, the various concepts of causal effects can be defined and we can study how to identify these causal effects from the parameters describing the joint distribution of the random variables considered. In such a single-unit trial, there usually is a clear time order which helps to disentangle the possible causal relationships between the random variables considered.

▷ **Solution 2-3** Different random experiments are different empirical phenomena. Although the names of the variables in different random experiments might be the same, the variables themselves are different entities, implying that the dependencies and effects between these variables might be different in different random experiments.

▷ **Solution 2-4** The type of random experiment we refer to in these examples is the single-unit trial of simple experiments and quasi-experiments described in section 2.1.

▷ **Solution 2-5** In the social sciences, units are often persons, and persons can change over time. If, in a simple experiment or quasi-experiment, a value u of U represents the observational unit sampled *at the onset of treatment*, each covariate will be a function of U. If, in contrast, U would represent the *observational unit at the assessment of a fallible covariate* (see section 2.2), which is some time prior to the onset of treatment, there can be other covariates in between assessment of the fallible covariate and the onset of treatment. We have to consider these additional covariates both in the definition of causal effects and in data analysis.

▷ **Solution 2-6** A covariate pertaining to a cause is a variable that is prior or simultaneous to the cause.

▷ **Solution 2-7** If the treatment has just two values, say *treatment* and *control*, there are different kinds of causal effects of the treatment variable on the outcome variable Y, such as the average total treatment effect, the conditional total treatment effects given a value of a covariate Z, and the *individual total effect* of X on Y given an observational unit u. Aside from these treatment effects, we may also consider the causal effects of a covariate Z on the treatment variable X, but also on the outcome variable Y. Among the causal effects of such a covariate are its conditional direct effects on Y given the different values of the treatment variable, the average of these X-conditional direct effects, and its indirect effect mediated by X.

▷ **Solution 2-8** Covariates pertaining to such an intermediate variable M are all variables representing attributes of the observational units at the onset of treatment, all variables that are simultaneous to treatment, including X itself, all other variables that are in between treatment and the intermediate variable.

Part II
Basic Concepts

Chapter 3
Causality Space

In chapter 2 we described some random experiments and discussed several kinds of random variables playing a crucial role in the theory of causal effects. In this discussion we referred to the time order among these random variables, e. g., saying that covariates are 'prior' or 'simultaneous' to the cause, which itself is prior to the outcome variable. We also said that intermediate variables are 'in between' cause and outcome variables. Furthermore, for each kind of those random experiments, we also discussed the set of covariates pertaining to a cause, again drawing on the time order between the variables involved. Finally, we emphasized that even within the same random experiment different causes requires different sets of covariates.

A random experiment is represented by a *probability space*. Referring to such a probability space we can consider *events*, *random variables*, their *distributions*, *conditional expectations*, and *conditional distributions*, which can be used to describe various kinds of stochastic dependencies between random variables. In this chapter we presume that the reader is familiar with these fundamental concepts of probability theory, including the following concepts: σ-*algebra* , σ-*algebra generated by a set system*, σ-*algebra generated by a mapping*, *product of sets*, *product* σ-*algebra*, *measurable space*, *measure space*, *measurability of a mapping with respect to a* σ-*algebra*, and *measurability of a mapping with respect to another mapping*. These concepts are introduced in Steyer and Nagel (in press), but also in Bauer (1996), Feller (1968, 1971), Georgii (2008), Klenke (2013), Loève (1977, 1978), and many other textbooks on measure and probability theory.

Note that neither events nor random variables refer to a *process* with respect to which we can say that a cause is *prior* to the outcome variable and *not prior* to a covariate. Furthermore, the distinction between covariates and intermediate variables does not yet have a mathematical foundation in the terms of measure and probability theory mentioned above. Therefore, in this chapter, we introduce additional mathematical concepts allowing to introduce the priority and simultaneity relations between events, between sets of events, and between random variables. The additional mathematical structure also allows us to introduce a mathematical definition of covariates and intermediate variables in the next chapter.

While the concepts introduced in this chapter distinguish a *potential* causal dependence from an ordinary stochastic dependence, the causality conditions treated in chapters 6 to 9 make the distinction between a *potential* and an *ac-*

tual causal dependence. With these causality conditions we postulate certain relationships between the covariates on one hand, and the focused cause or the outcome variable on the other hand. Note that the mathematical concepts introduced in this chapter are not restricted to experiments and quasi-experiments. Instead, they are of fundamental importance whenever causal effects and causal dependencies are considered, even in processes with continuous time.

Overview

We start with the concept of a *filtration*, which we use for defining the *priority* and *simultaneity* relations between sets (events), sets of events (sets of events), and measurable mappings (random variables). We also introduce the concept of a *causality space*, which consists of all structural components that are necessary for the definition of causal effects and for raising the question if a stochastic dependence of a random variable on another one has a causal interpretation.

3.1 Filtration

The fundamental conceptual tool for introducing the priority relation mentioned above is the concept of a *filtration*, which is well-known in the theory of stochastic processes (see, e. g., Bauer, 1996; Klenke, 2013). In the following definition we refer to a *measurable space* $(\Omega, \mathscr{A})$, which is simply a pair consisting of a nonempty set Ω and a σ-algebra $\mathscr{A}$ on Ω (see, e.g., Def. 1.1 of Steyer & Nagel, in press, in the sequel abbreviated SN).

Definition 3.1 (Filtration)
Let $(\Omega, \mathscr{A})$ be a measurable space and $T \subset \mathbb{R}$. A family $(\mathscr{F}_t, t \in T)$ of σ-algebras $\mathscr{F}_t \subset \mathscr{A}$ is called a $filtration$ in $\mathscr{A}$, if $\mathscr{F}_s \subset \mathscr{F}_t$, for all $s, t \in T$ with $s \leq t$.

Remark 3.2 (Finite Index Set) Oftentimes it will be sufficient to consider a filtration with a finite index set $T = \{1, 2, \ldots, n\}$, $n \in \mathbb{N}$. If necessary, we may as well consider a subset $T \subset \mathbb{R}$. The important point is that the index set T is endowed with the relations $=$, $<$ and $\leq$. In applications, the elements of the set T oftentimes represent time points. ◁

Example 3.3 (Joe and Ann With Self-Selection) We illustrate the concept of a filtration by the numerical example presented in Table 3.1. This table refers to the following random experiment: *First*, we sample a unit u from the set $\Omega_U := \{Joe, Ann\}$. *Second*, each unit receives (*yes*) or does not receive a treatment (*no*), and *third* it is observed whether (+) or not (−) a success criterion is reached some appropriate time after treatment. Defining $\Omega_X := \{yes, no\}$ and $\Omega_Y := \{+, -\}$, the set of possible outcomes ω of this random experiment is

Table 3.1. Joe and Ann With Self-Selection to Treatment Conditions

Outcomes ω				Observables		
Unit	Treatment	Success	$P(\{\omega\})$	Person variable U	Treatment variable X	Outcome variable Y
(*Joe*, *no*, −)			.144	*Joe*	0	0
(*Joe*, *no*, +)			.336	*Joe*	0	1
(*Joe*, *yes*, −)			.004	*Joe*	1	0
(*Joe*, *yes*, +)			.016	*Joe*	1	1
(*Ann*, *no*, −)			.096	*Ann*	0	0
(*Ann*, *no*, +)			.024	*Ann*	0	1
(*Ann*, *yes*, −)			.228	*Ann*	1	0
(*Ann*, *yes*, +)			.152	*Ann*	1	1

$$\Omega := \Omega_U \times \Omega_X \times \Omega_Y = \big\{ (Joe, no, -), (Joe, no, +), \ldots, (Ann, yes, +) \big\}. \tag{3.1}$$

In this example, the set Ω has eight elements, the triples (*Joe*, *no*, −), (*Joe*, *no*, +), …, (*Ann*, *yes*, +) (see the first column of Table 3.1 for a complete list of these elements). Furthermore, we define $\mathscr{A} := \mathscr{P}(\Omega)$. Finally, because each nonempty element $A \in \mathscr{A}$ is a union of the singletons $\{\omega\}$, $\omega \in \Omega$, and because a measure is additive, the probability measure $P\colon \mathscr{A} \to [0,1]$ is uniquely defined by the second column of Table 3.1 [see Box 4.1 (x) of SN]. Hence, the probability space $(\Omega, \mathscr{A}, P)$ is completely specified.

Remember, by definition, a measurable mapping $X\colon (\Omega, \mathscr{A}) \to (\Omega'_X, \mathscr{A}'_X)$ is also called a random variable on the probability space $(\Omega, \mathscr{A}, P)$ with values in $(\Omega'_X, \mathscr{A}'_X)$ if P is a probability measure on $\mathscr{A}$ (see Def. 5.1 of SN). In this example, we also consider three random variables: the *observational-unit variable* $U\colon (\Omega, \mathscr{A}) \to [\Omega_U, \mathscr{P}(\Omega_U)]$, the *treatment variable* $X\colon (\Omega, \mathscr{A}) \to (\Omega'_X, \mathscr{A}'_X)$, and the *outcome variable* $Y\colon (\Omega, \mathscr{A}) \to (\Omega'_Y, \mathscr{A}'_Y)$, where $\Omega'_X = \{0,1\}$, $\mathscr{A}'_X = \mathscr{P}(\Omega'_X) = \big\{\Omega', \emptyset, \{0\}, \{1\}\big\}$, $\Omega'_Y = \{0,1\}$, and $\mathscr{P}(\Omega'_Y) = \big\{\Omega', \emptyset, \{0\}, \{1\}\big\}$. Table 3.1 shows how each of these random variables assigns one of its values to each of the eight elements $\omega \in \Omega$.

In this example, we consider a filtration $(\mathscr{F}_t, t \in T)$ with three σ-algebras (see Fig. 3.1). The first one is

$$\mathscr{F}_1 := U^{-1}\big[\mathscr{P}(\Omega_U)\big],$$

i. e., $\mathscr{F}_1$ is the σ-algebra generated by the observational-unit variable U and the power set $\mathscr{P}(\Omega_U)$ of the set $\Omega_U := \{Joe, Ann\}$ (see Def. 2.26 of SN). Hence, the σ-algebra $\mathscr{F}_1$ has only four elements, the event that *Joe is drawn*,

$$U^{-1}(\{Joe\}) = \{(Joe, no, -), (Joe, no, +), (Joe, yes, -), (Joe, yes, +)\},$$

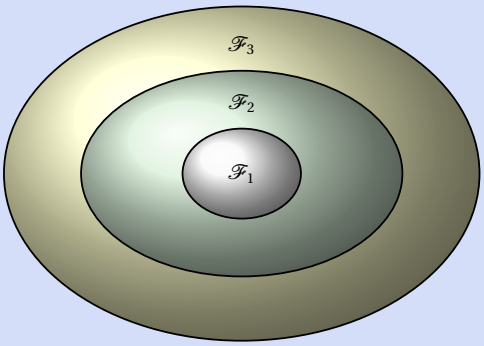


Figure 3.1. Venn-diagram of a filtration with $T = \{1,2,3\}$

the event that *Ann is drawn*,

$$U^{-1}(\{Ann\}) = \{(Ann, no, -), (Ann, no, +), (Ann, yes, -), (Ann, yes, +)\},$$

the *sure event* Ω, and the *impossible event* $\emptyset$.

Furthermore, we define

$$\mathscr{F}_2 := \sigma\left[\mathscr{F}_1 \cup X^{-1}(\mathscr{A}'_X)\right]$$

to be the σ-algebra generated by the union of the σ-algebras $\mathscr{F}_1$ and $X^{-1}(\mathscr{A}'_X)$, where $X^{-1}(\mathscr{A}'_X)$ represents the σ-algebra that is generated by X and $\mathscr{A}'_X = \mathscr{P}(\Omega'_X)$ (see Def. 1.13 of SN). The σ-algebra $\mathscr{F}_2$ has $2^4 = 16$ elements. For example, it includes the event that Joe is drawn, $U^{-1}(\{Joe\})$, the event that Ann is drawn, $U^{-1}(\{Ann\})$, the event that the *person drawn is treated*, $X^{-1}(\{1\})$, the event that the *person drawn is not treated*, $X^{-1}(\{0\})$, the event that *Joe is drawn and treated*, $(U,X)^{-1}(\{(Joe,1)\})$, and the event that *Ann is drawn and not treated*, $(U,X)^{-1}(\{(Ann,0)\})$.

Finally, the σ-algebra $\mathscr{F}_3$ is the power set of Ω. It has $2^8 = 256$ elements and it contains all possible events that might occur in this random experiment, including the elementary events such as *Joe is drawn, treated and successful*. Most important, the three σ-algebras $\mathscr{F}_t$ are constructed such that

$$\mathscr{F}_1 \subset \mathscr{F}_2 \subset \mathscr{F}_3$$

holds for the family $(\mathscr{F}_t, t \in T)$, where $T = \{1,2,3\}$ (see Fig. 3.1 and again Def. 1.13 of SN). ◁

Example 3.4 (Simple Experiments) How can we construct the filtration $(\mathscr{F}_t, t \in T)$ in general for the single-unit trial of a simple experiment or quasi-experiment in which no *fallible* covariates are observed (see section 2.1)?

In this kind of random experiment the probability space $(\Omega, \mathscr{A}, P)$ has the same structure as described in Example 3.3, except for the sets Ω_X, Ω_Y, Ω'_X, and

Ω'_Y, which may contain more than just two elements. Of course, the associated σ-algebras are different as well. Furthermore, the random variables U, X, and Y are also defined in the same way as in the example. While Ω_X and Ω'_X are finite or countable, the sets Ω_Y and Ω'_Y may be subsets of $\mathbb{R}^n$, $n \in \mathbb{N}$. In this case, the σ-algebra $\mathscr{A}'_Y$ will be the Borel σ-algebra $\mathscr{B}_n$ (see, e. g., section 1.2.2 of SN).

Note that the σ-algebra $\mathscr{A}$ on Ω is not necessarily the power set of Ω. If, e. g., we consider a real-valued outcome variable $Y\colon (\Omega,\mathscr{A}) \to (\mathbb{R},\mathscr{B})$, then $\mathscr{A}$ can be defined to be the product σ-algebra $\mathscr{P}(\Omega_U) \otimes \mathscr{P}(\Omega_X) \otimes Y^{-1}(\mathscr{B})$ (see, e. g., Def. 1.31 of SN), where $\mathscr{B}$ denotes the Borel σ-algebra on $\mathbb{R}$. In contrast to Example 3.3, in empirical applications, the probability measure P on $(\Omega,\mathscr{A})$ is unknown or only partly known.

In a simple experiment, the random variable $U\colon (\Omega,\mathscr{A}) \to [\Omega_U,\mathscr{P}(\Omega_U)]$ indicates with its value which observational unit u is drawn. Hence, we can define

$$\mathscr{F}_1 := U^{-1}\big[\mathscr{P}(\Omega_U)\big], \tag{3.2}$$

i. e., $\mathscr{F}_1$ is the σ-algebra generated by the observational-unit variable U [and the power set $\mathscr{P}(\Omega_U)$ of the set Ω_U of all units considered]. The number of elements of $\mathscr{F}_1$ is 2^n, where n denotes the number of elements of the set Ω_U. In Example 3.3 we considered two units. Hence, in this example, $\mathscr{P}(\Omega_U)$ has $2^2 = 4$ elements.

Random variables that are mappings of U such as *sex*, *race*, *socio-economic status*, *educational status*, etc. are measurable with respect to $\mathscr{F}_1$. This means that if there is a measurable mapping $f\colon [\Omega_U,\mathscr{P}(\Omega_U)] \to (\Omega'_Z,\mathscr{A}'_Z)$ such that $Z = f(U)$, then $Z^{-1}(\mathscr{A}'_Z) \subset \mathscr{F}_1$ (see, e. g., Lemma 2.52 of SN). Hence, in simple experiments and quasi-experiments, all covariates are measurable with respect to U.

Using the concept of a σ-algebra generated by a set of events (see 1.13 of SN), we define

$$\mathscr{F}_2 := \sigma\big[\mathscr{F}_1 \cup X^{-1}(\mathscr{A}'_X)\big], \tag{3.3}$$

where $X^{-1}(\mathscr{A}'_X)$ represents the σ-algebra generated by the treatment variable $X\colon (\Omega,\mathscr{A}) \to (\Omega'_X,\mathscr{A}'_X)$. In this most simple example, it is assumed that X is finite or countable and that there are no random variables on $(\Omega,\mathscr{A},P)$ that vary simultaneously to X. In particular this means that there are no other treatment variables aside from X.

Finally, we define

$$\mathscr{F}_3 := \sigma\big[\mathscr{F}_2 \cup Y^{-1}(\mathscr{A}'_Y)\big], \tag{3.4}$$

where $Y^{-1}(\mathscr{A}'_Y)$ is the σ-algebra generated by $Y\colon (\Omega,\mathscr{A}) \to (\Omega'_Y,\mathscr{A}'_Y)$, the outcome variable. Note that Y is not necessarily numerical.

In such a simple experiment, the filtration $(\mathscr{F}_t, t \in T)$, $T = \{1,2,3\}$, (see Fig. 3.1) allows to define a covariate of X to be a random variable on the probability space $(\Omega,\mathscr{A},P)$ that is measurable with respect to $\mathscr{F}_1$ (see section **??** for a general definition of a covariate). In a randomized experiment, $\mathscr{F}_1$ and X (and therefore also U and X) are stochastically independent. In applications, this is often secured by a assigning the unit to a treatment condition depending only on the outcome of a coin flip. ◁

Example 3.5 (Experiments With Fallible Covariates) Which is the filtration in the single-unit trial of experiments and quasi-experiments if we *do observe* at least one fallible covariate (see section 2.2)? Again, we start with the set of possible outcomes ω of this random experiment. Now we assume that Ω can be written

$$\Omega \;=\; \Omega_U \times \Omega_Z \times \Omega_X \times \Omega_Y. \tag{3.5}$$

Compared to Equation (3.1), Ω now involves an additional set Ω_Z. This random experiment consists of (a) drawing a unit from Ω_U, (b) observing an element ω_Z of Ω_Z based on which the covariate Z assigns a value $z \in \Omega'_Z$ to $\omega \in \Omega$, (c) assigning the unit or observing its selection to one of the experimental conditions (represented by the value x of the treatment variable X), and (d) recording the numerical value y of the outcome variable Y. Correspondingly, the σ-algebra $\mathscr{A}$ on Ω is now defined to be a fourfold product σ-algebra, involving appropriate σ-algebras on each of the four sets involved in the Cartesian product $\Omega_U \times \Omega_Z \times \Omega_X \times \Omega_Y$. Furthermore, now we have an additional random variable $Z\colon (\Omega, \mathscr{A}) \to (\Omega'_Z, \mathscr{A}'_Z)$, the fallible covariate. Note that this covariate may also be multivariate consisting of several univariate covariates.

Specifying the filtration $(\mathscr{F}_t, t \in T)$, the first of these σ-algebras can again be defined by

$$\mathscr{F}_1 \;:=\; U^{-1}\big[\mathscr{P}(\Omega_U)\big], \tag{3.6}$$

i. e., $\mathscr{F}_1$ is the σ-algebra generated by the observational-unit variable U. Note, however, that $\mathscr{F}_1$ is now a σ-algebra on the set $\Omega_U \times \Omega_Z \times \Omega_X \times \Omega_Y$.

Second, we define

$$\mathscr{F}_2 \;:=\; \sigma\big[\mathscr{F}_1 \cup Z^{-1}(\mathscr{A}'_Z)\big], \tag{3.7}$$

where Z is the (possibly multivariate) covariate that is assessed with measurement error. Hence, Z may consist of *fallible measures* of attributes of the units such as self-rated *motivation for therapy* as well as *personality* or *ability test-score variables*. Given a particular unit u, the values of all these variables still have a positive variance, due to measurement error. Because Z is fallible, there is no mapping $f\colon [\Omega_U, \mathscr{P}(\Omega_U)] \to (\Omega'_Z, \mathscr{A}'_Z)$ such that $Z = f(U)$. In other words, fallible covariates have $(U{=}u)$-conditional distributions with positive variances. This includes the fallible measures of a latent covariate, say ξ, which itself is, by definition, a mapping of U (see, e. g., Steyer et al., 2015). Hence, the latent variable ξ is measurable with respect to $\mathscr{F}_1$. Furthermore, the σ-algebra $\mathscr{F}_2$ is defined such that all random variables that are prior to treatment are measurable with respect to $\mathscr{F}_2$. This includes mappings of U, the fallible measures of attributes of the units, but also all measurable mappings of these two classes of variables.

Third, we define

$$\mathscr{F}_3 \;:=\; \sigma\big[\mathscr{F}_2 \cup X^{-1}(\mathscr{A}'_X)\big], \tag{3.8}$$

where $X^{-1}(\mathscr{A}'_X)$ is the σ-algebra generated by the treatment variable X. All variables that are measurable with respect to $\mathscr{F}_2$ are also measurable with respect to $\mathscr{F}_3$. Furthermore, the product of X and Z, if both are numerical, and the product of an indicator (with values 0 and 1) of sex — a function of U — and an indicator of a treatment condition are also measurable with respect to $\mathscr{F}_3$, for instance.

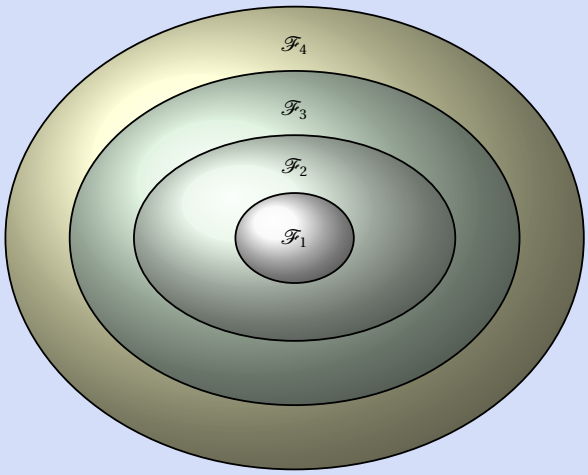


Figure 3.2. Venn-diagram of a filtration with $T = \{1,2,3,4\}$

Fourth and finally, we define

$$\mathscr{F}_4 := \sigma\left[\mathscr{F}_3 \cup Y^{-1}(\mathscr{A}'_Y)\right], \tag{3.9}$$

where $Y^{-1}(\mathscr{A}'_Y)$ is the σ-algebra generated by the outcome variable $Y\colon (\Omega,\mathscr{A}) \to (\Omega'_Y,\mathscr{A}'_Y)$, which might be a multivariate random variable that can also be qualitative. Note that $(\mathscr{F}_t, t \in T)$, $t = 1,\ldots,4$, is defined such that

$$\mathscr{F}_1 \subset \mathscr{F}_2 \subset \mathscr{F}_3 \subset \mathscr{F}_4.$$

This is illustrated by Figure 3.2.

This filtration $(\mathscr{F}_t, t \in T)$ with $T = \{1,\ldots,4\}$ is sufficient for many discussions of causal effects in experiments and quasi-experiments. For other purposes, we may consider more than these four σ-algebras, e. g., if models with intermediate variables are considered, or if we consider also a fallible outcome variable (see sections of 2.5 and 2.6.) Note that the concept of a filtration even applies if T is an uncountable subset of $\mathbb{R}$.

◁

3.2 Priority Relation

Utilizing the concept of a filtration, now we introduce the priority relation of sets (events), of set systems (sets of subsets, sets of events), and of measurable mappings (random variables). We define the term "Z is prior to Y" in such a way that also difference variables $Y - Z$, e. g., differences between post- and pre-tests

(change-score variables), can be ordered with respect to time even though pre- and post-tests refer to different time points. The analysis of such difference variables is sometimes used in the analysis of causal effects (see section 12.1 for the conditions under which such an analysis yields unbiased causal effects).

Definition 3.6 (Priority Relation of Set Systems)
Let $(\Omega, \mathscr{A})$ be a measurable space, $(\mathscr{F}_t, t \in T)$ a filtration in $\mathscr{A}$, and $\mathscr{C}, \mathscr{D} \subset \mathscr{A}$. Then we say that $\mathscr{C}$ is prior to $\mathscr{D}$ in $(\mathscr{F}_t, t \in T)$, if:

(a) there is an $s \in T$ with $\mathscr{C} \subset \mathscr{F}_s$, $\mathscr{D} \not\subset \mathscr{F}_s$, and
(b) there is a $t \in T$, $s < t$, with $\mathscr{D} \subset \mathscr{F}_t$.

Remark 3.7 (Priority Relation of Sets) Let $(\Omega, \mathscr{A})$ be a measurable space, $(\mathscr{F}_t, t \in T)$ a filtration in $\mathscr{A}$, and $A_1, A_2 \in \mathscr{A}$. Then we say that A_1 is *prior to* A_2 *in* $(\mathscr{F}_t, t \in T)$, if (a) and (b) of Definition 3.6 hold for $\mathscr{C} = \{A_1\}$ and $\mathscr{D} = \{A_2\}$. Note that $\{A\} \subset \mathscr{F}_t$ if and only if $A \in \mathscr{F}_t$. This implies: A_1 is prior to A_2 in $(\mathscr{F}_t, t \in T)$ if and only if

(a) there is an $s \in T$ with $A_1 \in \mathscr{F}_s$, $A_2 \notin \mathscr{F}_s$, and
(b) there is a $t \in T$, $t > s$, with $A_2 \in \mathscr{F}_t$.

◁

Remark 3.8 (Priority Relation of Measurable Mappings) Let $(\Omega, \mathscr{A})$ be a measurable space, $(\mathscr{F}_t, t \in T)$ a filtration in $\mathscr{A}$, and $X_i \colon (\Omega, \mathscr{A}) \to (\Omega'_i, \mathscr{A}'_i)$, $i = 1, 2$, measurable mappings. Then we say that X_1 is *prior to* X_2 *in* $(\mathscr{F}_t, t \in T)$, if (a) and (b) of Definition 3.6 hold with $\mathscr{C} = \sigma(X_1)$ and $\mathscr{D} = \sigma(X_2)$. ◁

Remark 3.9 (Priority Relation of σ-Algebras) Let $(\Omega, \mathscr{A})$ be a measurable space, $(\mathscr{F}_t, t \in T)$ a filtration in $\mathscr{A}$, and $\mathscr{C}, \mathscr{D} \subset \mathscr{A}$. Because $\mathscr{F}_s$ and $\mathscr{F}_t$, $s, t \in T$, are σ-algebras, we can conclude: $\mathscr{C}$ is prior to $\mathscr{D}$ in $(\mathscr{F}_t, t \in T)$ if and only if

(a) there is an $s \in T$ with $\sigma(\mathscr{C}) \subset \mathscr{F}_s$, $\sigma(\mathscr{D}) \not\subset \mathscr{F}_s$, and
(b) there is a $t \in T$, $s < t$, with $\sigma(\mathscr{D}) \subset \mathscr{F}_t$.

◁

Remark 3.10 (Comparing Sets to σ-Algebras) Note that $\sigma(\{A\}) = \{\Omega, \emptyset, A, A^c\}$, where A^c denotes the complement of A. Therefore, according to Definition 3.6 and Remark 3.9, priority of sets, of set systems, and of measurable mappings always refer to σ-algebras. As mentioned before, this allows us to compare sets to measurable mappings, measurable mappings to σ-algebras, etc. Hence, we say that the set A is prior in $(\mathscr{F}_t, t \in T)$ to the measurable mapping X, if the σ-algebra generated by $\{A\}$ is prior in $(\mathscr{F}_t, t \in T)$ to the σ-algebra generated by X, etc. ◁

Example 3.11 (Joe and Ann With Self-Selection – continued) The observational-unit variable U is prior to the treatment variable X with respect to the filtration presented in Example 3.3. Furthermore, X is prior to the outcome variable Y. Similarly, $\{Joe\} \times \Omega_X \times \Omega_Y$, i. e., the event that *Joe is drawn*, is prior to X in this filtration (see also Exercise 3-5). ◁

Some Properties of the Priority Relation

Now we study some properties of the priority relation introduced above. First of all, we ascertain that the priority relation is asymmetric and transitive.

Theorem 3.12 (Asymmetry and Transitivity of the Priority Relation)
Let $(\Omega, \mathscr{A})$ be a measurable space, $(\mathscr{F}_t, t \in T)$ a filtration in $\mathscr{A}$, and $\mathscr{C}, \mathscr{D} \subset \mathscr{A}$.

(i) If $\mathscr{C}$ is prior to $\mathscr{D}$ in $(\mathscr{F}_t, t \in T)$, then $\mathscr{D}$ is not prior to $\mathscr{C}$ (asymmetry).
(ii) Let $\mathscr{E} \subset \mathscr{A}$. If $\mathscr{C}$ is prior to $\mathscr{D}$ and $\mathscr{D}$ is prior to $\mathscr{E}$ in $(\mathscr{F}_t, t \in T)$, then $\mathscr{C}$ is also prior to $\mathscr{E}$ in $(\mathscr{F}_t, t \in T)$ (transitivity).

(Proof p. 94)

Remark 3.13 (Asymmetry and Transitivity) Because priority of measurable mappings is defined by their generated σ-algebras, Propositions (i) and (ii) of Theorem 3.12 also hold for measurable mappings X_i, $i = 1,2,3$, taking the role of the σ-algebras $\mathscr{C}$, $\mathscr{D}$, and $\mathscr{E}$, respectively, presuming of course that $(\mathscr{F}_t, t \in T)$ is a filtration in $\mathscr{A}$. Similarly, these propositions also hold for sets $A_i \in \mathscr{A}$, $i = 1,2,3$, taking the role of these σ-algebras. ◁

Remark 3.14 (Difference Variables) If X_1 is prior to X_2 with respect to $(\mathscr{F}_t, t \in T)$, then X_1 is also prior to $X_1 - X_2$ with respect to $(\mathscr{F}_t, t \in T)$ (see Exercise 3-6). ◁

Remark 3.15 (Constant Difference) If $X_1 - X_2 = \alpha$, $\alpha \in \mathbb{R}$, is a constant, then X_1 is not prior to X_2, because then the σ-algebras generated by X_1 and X_2 are identical. Hence, in this case the premise of Remark 3.14 does not hold. Note that X_1 can be prior to X_2 in $(\mathscr{F}_t, t \in T)$ even if $X_1 \underset{P}{=} X_2$ (see Example 3.16). ◁

Example 3.16 (Joe and Ann With Perfect Dependence of Y on X) Table 3.2 displays an example that is very similar to Example 3.3. The measurable space $(\Omega, \mathscr{A})$, and the measurable mappings U, X, and Y are unchanged, and we construct the filtration $(\mathscr{F}_t, t \in T)$ in the same way as in Example 3.3. Hence, X is again prior to Y in $(\mathscr{F}_t, t \in T)$. However, now the dependence of Y on X is perfect. More precisely, $X \underset{P}{=} Y$. The last column of Table 3.2 shows that the difference variable $X - Y$ is not a constant, i. e., $X - Y \neq 0$, although $X - Y \underset{P}{=} 0$, which means $P(X \neq Y) = 0$. This illustrates that the priority relation depends on the definition of the measurable mappings (random variables) and on the construction of the filtration, but not on the probability measure. ◁

Remark 3.17 (Priority Relation Among Events) Sets (events) can be stretched over several time points as well. If we consider again Example 3.3, then the events A_1 that *Joe is sampled and treated* and A_2 that *Joe is sampled, treated, and successful* are examples in case. According to our definition, A_1 is prior to A_2, because the σ-algebra generated by $\{A_1\}$ is a subset of $\mathscr{F}_2$, whereas the σ-algebra generated by $\{A_2\}$ is not a subset of $\mathscr{F}_2$, but a subset of $\mathscr{F}_3$ (see Exercise 3-8). These and similar properties of the priority relation are stated in formal terms in the following theorem ◁

Table 3.2. Joe and Ann With Perfect Dependence of Y on X

Outcomes ω				Observables			
Unit	Treatment	Success	$P(\{\omega\})$	Person variable U	Treatment variable X	Outcome variable Y	$X-Y$
(*Joe*, *no*, −)			.48	*Joe*	0	0	0
(*Joe*, *no*, +)			.00	*Joe*	0	1	−1
(*Joe*, *yes*, −)			.00	*Joe*	1	0	1
(*Joe*, *yes*, +)			.02	*Joe*	1	1	0
(*Ann*, *no*, −)			.12	*Ann*	0	0	0
(*Ann*, *no*, +)			.00	*Ann*	0	1	−1
(*Ann*, *yes*, −)			.00	*Ann*	1	0	1
(*Ann*, *yes*, +)			.38	*Ann*	1	1	0

Theorem 3.18 (Implications of the Priority Relation)
Let $(\Omega, \mathscr{A})$ be a measurable space, $(\mathscr{F}_t, t \in T)$ a filtration in $\mathscr{A}$, and $\mathscr{C}, \mathscr{D} \subset \mathscr{A}$.

(i) If $\mathscr{C}$ is prior to $\mathscr{D}$ in $(\mathscr{F}_t, t \in T)$, then $\mathscr{C}$ is also prior to $\mathscr{C} \cup \mathscr{D}$ and to $\sigma(\mathscr{C} \cup \mathscr{D})$.

(ii) Let $\mathscr{E} \subset \mathscr{A}$. If $\mathscr{C}$ and $\mathscr{D}$ are prior to $\mathscr{E}$ in $(\mathscr{F}_t, t \in T)$, then $\mathscr{C} \cup \mathscr{D}$ and $\sigma(\mathscr{C} \cup \mathscr{D})$ are also prior to $\mathscr{E}$.

(iii) If $\mathscr{C}$ is prior to $\mathscr{D}$ and to $\mathscr{E}$ in $(\mathscr{F}_t, t \in T)$, then $\mathscr{C}$ is also prior to $\mathscr{D} \cup \mathscr{E}$ and to $\sigma(\mathscr{D} \cup \mathscr{E})$.

(Proof p. 94)

Remark 3.19 (Implications on Priority Among Measurable Mappings) This theorem also applies to priority among measurable mappings (random variables). Hence, if the assumptions of Theorem 3.18 hold, and if $Y\colon (\Omega, \mathscr{A}) \to (\Omega'_Y, \mathscr{A}'_Y)$ and $Z\colon (\Omega, \mathscr{A}) \to (\Omega'_Z, \mathscr{A}'_Z)$ are also measurable mappings, then the following propositions hold:

(i) If X_1 is prior to X_2 in $(\mathscr{F}_t, t \in T)$ and $\sigma(Y) \subset \sigma(X_1, X_2)$ and $\sigma(Y) \not\subset X_1$, then X_1 is also prior to Y.

(ii) If both X_1 and X_2 are prior to Z in $(\mathscr{F}_t, t \in T)$ and $\sigma(Y) \subset \sigma(X_1, X_2)$, then Y is prior to Z.

Examples in case for measurable mappings of X_1 and X_2 mentioned above are $Y = \alpha_1 X_1 \cdot \alpha_2 X_2$ and $Y = \alpha_1 X_1 - \alpha_2 X_2$, where $\alpha_1, \alpha_2 \in \mathbb{R}$, $\alpha_1 \neq \alpha_2$. ◁

3.3 Simultaneity Relation

In chapter 2 we already discussed that random variables can be *simultaneous* to each other. As an example we mentioned a variable Z representing a second treatment that can be given (or not given) at the same time as the first treatment represented by X. As another example, consider studying the effects of M_1 = *amount of antibodies at time t*. In such an application, we might also consider a second variable, say M_2 = *amount of leucocytes at time t* referring to the same time point. Then M_1 and M_2 would be simultaneous to each other.

In the definition of the simultaneity relation we have to presume that the index set T is finite or, if this is not the case, that the filtration $(\mathscr{F}_t, t \in T)$ is right-continuous. A filtration $(\mathscr{F}_t, t \in T)$ in a σ-algebra $\mathscr{A}$ is *right-continuous*, if

$$\forall s \in T: \quad \mathscr{F}_s = \bigcap_{s<t} \mathscr{F}_t. \tag{3.10}$$

Remark 3.20 (An Implication of Right-Continuousness) Let $(\Omega, \mathscr{A})$ be a measurable space, $(\mathscr{F}_t, t \in T)$ a filtration in $\mathscr{A}$, and $\mathscr{C}, \mathscr{D} \subset \mathscr{A}$. If $(\mathscr{F}_t, t \in T)$ is finite or right-continuous and $\mathscr{C}$ is prior to $\mathscr{D}$ with respect to $(\mathscr{F}_t, t \in T)$, then there is a $t_0 \in T$ such that $\mathscr{D} \subset \mathscr{F}_{t_0}$ and $\mathscr{D} \not\subset \mathscr{F}_t$, for all $t < t_0, t \in T$.

Proof: If T is finite, the proof is trivial. Hence, we only proof the proposition for $(\mathscr{F}_t, t \in T)$ being right-continuous. Define $S := \{s \in T : \mathscr{D} \subset \mathscr{F}_s\}$. The set S is nonempty because there is a $t \in T$ such that $\mathscr{D} \subset \mathscr{F}_t$. Let $t_0 := \inf(S)$. Because there is an $s \in T$ such $\mathscr{D} \not\subset \mathscr{F}_s$ we can conclude $t_0 > -\infty$. For all $t > t_0$, $\mathscr{D} \subset \mathscr{F}_t$. Using right-continuity we can conclude: $\mathscr{D} \subset \mathscr{F}_{t_0} = \bigcap_{t>t_0} \mathscr{F}_t$. This proves the proposition, because $\mathscr{D} \not\subset \mathscr{F}_r$ for all $r < t_0$. ◁

Using the concept of right-continuousness, the simultaneity relation of set systems can be defined as follows:

Definition 3.21 (Simultaneity Relation of Set Systems)
Let $(\Omega, \mathscr{A})$ be a measurable space, $(\mathscr{F}_t, t \in T)$ a filtration in $\mathscr{A}$, and $\mathscr{C}, \mathscr{D} \subset \mathscr{A}$. Furthermore, assume that T is finite or $(\mathscr{F}_t, t \in T)$ is right-continuous. Then we say that $\mathscr{C}$ and $\mathscr{D}$ are simultaneous in $(\mathscr{F}_t, t \in T)$, if:

(a) there is a $t \in T$ with $\mathscr{C}, \mathscr{D} \subset \mathscr{F}_t$
(b) there is no $s \in T$, $s < t$, with $\mathscr{C} \subset \mathscr{F}_s$ or $\mathscr{D} \subset \mathscr{F}_s$.

Remark 3.22 (The σ-Algebra of $(\mathscr{F}_t, t \in T)$ Simultaneous to $\mathscr{C}$) Let the assumptions of Definition 3.21 hold. Then the σ-algebra $\mathscr{F}_t$ satisfying

(a) there is a $t \in T$ with $\mathscr{C} \subset \mathscr{F}_t$
(b) there is no $s \in T$, $s < t$, with $\mathscr{C} \subset \mathscr{F}_s$,

is also denoted $\mathscr{F}_{t_\mathscr{C}}$ and called the σ-*algebra of* $(\mathscr{F}_t, t \in T)$ *simultaneous to* $\mathscr{C}$. Hence, $t_\mathscr{C}$ is that element of T that satisfies (a) and (b) with $t = t_\mathscr{C}$. ◁

Remark 3.23 (The σ-Algebra of $(\mathscr{F}_t, t \in T)$ Simultaneous to X) If $X\colon (\Omega, \mathscr{A}) \to (\Omega'_X, \mathscr{A}'_X)$ is a measurable mapping, then the σ-algebra $\mathscr{F}_t$ satisfying (a) and (b) of Remark 3.22 with $\mathscr{C} = \sigma(X)$ and $t = t_X$ is also denoted $\mathscr{F}_{t_X}$ and called the σ-*algebra of* $(\mathscr{F}_t, t \in T)$ *simultaneous to* X. ◁

Remark 3.24 (Simultaneity Relation of Sets) Let $(\Omega, \mathscr{A})$ be a measurable space, $(\mathscr{F}_t, t \in T)$ a filtration in $\mathscr{A}$, and $A_1, A_2 \in \mathscr{A}$. Then we say that A_1 and A_2 are *simultaneous in* $(\mathscr{F}_t, t \in T)$, if (a) and (b) of Definition 3.21 hold with $\mathscr{C} = \{A_1\}$ and $\mathscr{D} = \{A_2\}$. Because $\{A\} \subset \mathscr{F}_t$ if and only if $A \in \mathscr{F}_t$, we can conclude that A_1 and A_2 are simultaneous if and only if

(a) there is a $t \in T$ with $A_1, A_2 \in \mathscr{F}_t$, and
(b) there is no $s \in T$, $s < t$, with $A_1 \in \mathscr{F}_s$ or $A_2 \in \mathscr{F}_s$.

◁

Remark 3.25 (Simultaneity Relation of Measurable Mappings) Let $(\Omega, \mathscr{A})$ be a measurable space, $(\mathscr{F}_t, t \in T)$ a filtration in $\mathscr{A}$, and $X_i\colon (\Omega, \mathscr{A}) \to (\Omega'_i, \mathscr{A}'_i)$, $i = 1, 2$, measurable mappings. Then we say that X_1 and X_2 are *simultaneous in* $(\mathscr{F}_t, t \in T)$, if (a) and (b) of Definition 3.21 hold with $\mathscr{C} = \sigma(X_1)$ and $\mathscr{D} = \sigma(X_2)$. ◁

Remark 3.26 (Simultaneity of σ-Algebras) Let $(\Omega, \mathscr{A})$ be a measurable space, $(\mathscr{F}_t, t \in T)$ a filtration in $\mathscr{A}$, and $\mathscr{C}, \mathscr{D} \subset \mathscr{A}$. Because $\mathscr{F}_s$ and $\mathscr{F}_t$ are σ-algebras, Proposition 1.11 of SN implies that $\mathscr{C}$ and $\mathscr{D}$ are simultaneous in $(\mathscr{F}_t, t \in T)$ if and only if

(a) there is a $t \in T$ with $\sigma(\mathscr{C})$, $\sigma(\mathscr{D}) \subset \mathscr{F}_t$, and
(b) there is no $s \in T$, $s < t$, with $\sigma(\mathscr{C}) \subset \mathscr{F}_s$ or $\sigma(\mathscr{D}) \subset \mathscr{F}_s$.

◁

Remark 3.27 (Comparing Sets to Set Systems) We say that the set A and the measurable mapping X are *simultaneous in* $(\mathscr{F}_t, t \in T)$, if $\{A\}$ and $\sigma(X)$ are simultaneous. Similarly, a set system $\mathscr{E}$ and a random variable X are called simultaneous in $(\mathscr{F}_t, t \in T)$, if $\mathscr{E}$ and $\sigma(X)$ are simultaneous. Finally, a set system $\mathscr{E}$ and a set A are simultaneous in $(\mathscr{F}_t, t \in T)$, if $\mathscr{E}$ and $\{A\}$ are simultaneous. ◁

Example 3.28 (Joe and Ann With Self-Selection – continued) In Example 3.3, the observational-unit variable U, the random variable $Z := \textit{sex}$, and the event $\{\textit{Joe}\} \times \Omega_X \times \Omega_Y$ that Joe is sampled are simultaneous in the filtration $(\mathscr{F}_t, t \in T)$ specified in Example 3.3. In this specific example, in which U only takes on the values *Joe* and *Ann*, the σ-algebras generated by U, by Z, and by the set $\big\{\{\textit{Joe}\} \times \Omega_X \times \Omega_Y\big\}$ are identical. Even if we consider an example, in which there is at least one more person in the set Ω_U, then the three σ-algebras are still simultaneous, because the σ-algebras generated by Z and by the event $\big\{\{\textit{Joe}\} \times \Omega_X \times \Omega_Y\big\}$ that Joe is sampled are subsets of the σ-algebra generated by U and because the first σ-algebra $\mathscr{F}_1$ in the filtration has been defined to be the σ-algebra generated by U. Hence, conditions (a) and (b) of Definition 3.21 hold for the σ-algebras generated by Z, by U, and by the event $\big\{\{\textit{Joe}\} \times \Omega_X \times \Omega_Y\big\}$. ◁

Properties of the Simultaneity Relation

Now we study some elementary properties of the simultaneity relation. First of all, we show that the simultaneity relation is reflexive, symmetric, and transitive.

Theorem 3.29 (Reflexivity, Symmetry and Transitivity)
Let $(\Omega, \mathcal{A})$ be a measurable space, $(\mathcal{F}_t, t \in T)$ a filtration in $\mathcal{A}$, and $\mathcal{C}, \mathcal{D} \subset \mathcal{A}$.

(i) If there is a $t \in T$ with $\mathcal{C} \subset \mathcal{F}_t$ and no $s \in T$, $s < t$ with $\mathcal{C} \subset \mathcal{F}_s$, then $\mathcal{C}$ is simultaneous to itself (reflexivity).
(ii) If $\mathcal{C}$ and $\mathcal{D}$ are simultaneous in $(\mathcal{F}_t, t \in T)$, then $\mathcal{D}$ and $\mathcal{C}$ are simultaneous (symmetry).
(iii) Let $\mathcal{E} \subset \mathcal{A}$. If $\mathcal{C}$ and $\mathcal{D}$ as well as $\mathcal{D}$ and $\mathcal{E}$ are simultaneous in $(\mathcal{F}_t, t \in T)$, then $\mathcal{C}$ and $\mathcal{E}$ are simultaneous in $(\mathcal{F}_t, t \in T)$ (transitivity).

(Proof p. 95)

Remark 3.30 (Simultaneity of Measurable Mappings) Because simultaneity of measurable mappings (random variables) is defined by their generated σ-algebras, Remark 3.26 implies that propositions (i) to (iii) also hold for measurable mappings X_i, $i = 1,2,3$, taking the role of the set systems $\mathcal{C}$, $\mathcal{D}$, and $\mathcal{E}$, respectively. Remark 3.26 also implies that propositions (i) to (iii) also hold for the sets (events) $A_i \subset \mathcal{A}$, $i = 1,2,3$, taking the role of the set systems $\mathcal{C}$, $\mathcal{D}$, and $\mathcal{E}$, respectively. ◁

Similar to what has been said in Remarks 3.14 to 3.17, one of the virtues of Definition 3.21 is that it also applies to difference and product variables. The following theorem is the foundation for propositions on simultaneity of such variables.

Theorem 3.31 (Implications of Simultaneity)
Let $(\Omega, \mathcal{A})$ be a measurable space, $(\mathcal{F}_t, t \in T)$ a filtration in $\mathcal{A}$, and $\mathcal{C}, \mathcal{D} \subset \mathcal{A}$.

(i) If $\mathcal{C}$ and $\mathcal{D}$ are simultaneous in $(\mathcal{F}_t, t \in T)$, then $\mathcal{C}$ and $\mathcal{C} \cup \mathcal{D}$ as well as $\mathcal{C}$ and $\sigma(\mathcal{C} \cup \mathcal{D})$ are simultaneous in $(\mathcal{F}_t, t \in T)$.
(ii) Let $\mathcal{E} \subset \mathcal{A}$. If $\mathcal{C}, \mathcal{D}$, and $\mathcal{E}$ are simultaneous in $(\mathcal{F}_t, t \in T)$, then $\mathcal{C} \cup \mathcal{D}$ and $\mathcal{E}$ as well as $\sigma(\mathcal{C} \cup \mathcal{D})$ and $\mathcal{E}$ are simultaneous in $(\mathcal{F}_t, t \in T)$.

(Proof p. 95)

Remark 3.32 (Implications on Simultaneity of Measurable Mappings)
Similar propositions also apply to simultaneity of measurable mappings. Hence, if the assumptions of Theorem 3.31 hold, and if $Y\colon (\Omega, \mathcal{A}) \to (\Omega'_Y, \mathcal{A}'_Y)$ and $Z\colon (\Omega, \mathcal{A}) \to (\Omega'_Z, \mathcal{A}'_Z)$ are measurable mappings, then the following propositions hold:

(i) If X_1 and X_2 are simultaneous in $(\mathcal{F}_t, t \in T)$ and $\sigma(Y) \subset \sigma(X_1, X_2)$, then Y is prior or simultaneous to X_1 in $(\mathcal{F}_t, t \in T)$.
(ii) If X_1, X_2, and Z are simultaneous in $(\mathcal{F}_t, t \in T)$ and $\sigma(Y) \subset \sigma(X_1, X_2)$, then Y is prior or simultaneous to Z in $(\mathcal{F}_t, t \in T)$.

Note that a constant is also a measurable mapping of (X_1, X_2). As an example, suppose that X_1 and X_2 are simultaneous in $(\mathscr{F}_t, t \in T)$ and consider $Y = X_1 - X_2$, which is a measurable function of X_1 and X_2 (see Example 2.61 of SN). If $Y = X_1 - X_2 = \alpha$, $\alpha \in \mathbb{R}$, then the σ-algebra generated by Y is $\{\Omega, \emptyset\}$. This σ-algebra is prior to X_1 and to X_2 unless X_1 and X_2 are simultaneous to the first σ-algebra $\mathscr{F}_1$ in the filtration $(\mathscr{F}_t, t \in T)$. ◁

Remark 3.33 (Product Variables) If X_1 is prior to X_2 in $(\mathscr{F}_t, t \in T)$, then the product $X_1 \cdot X_2$ is prior or simultaneous to X_2 (see Exercise 3-7). ◁

To summarize, a filtration $(\mathscr{F}_t, t \in T)$ does not only represent the process we refer to when we talk about causal effects, but it also allows to introduce the priority and simultaneity relations with respect to a filtration. In our context, these relations apply to random variables, events, and sets of events.

3.4 Causality Space

Now we summarize the assumptions under which we can define causal effects and meaningfully ask if conditional expected values such as $E(Y \mid X{=}x)$ or $E(Y \mid X{=}x, Z{=}z)$ and/or conditional distributions such as $P_{Y|X=x}$ or $P_{Y|X=x,Z=z}$ describe causal dependencies in a specific application. Just as a probability space $(\Omega, \mathscr{A}, P)$ is the mathematical framework of every probabilistic model and of every proposition about the dependence between events A and B or between random variables X and Y, a causality space is the mathematical framework of causal probabilistic models, propositions about causal probabilistic effects, and causal probabilistic dependencies.

Compared to 'ordinary' stochastic models involving two random variables X and Y, a causality space has an additional component, the filtration $(\mathscr{F}_t, t \in T)$ in $\mathscr{A}$ that serves to define priority and simultaneity of events, random variables, and σ-algebras. Aside from this additional component we presume that the cause X is prior to Y with respect to $(\mathscr{F}_t, t \in T)$.

BAUSTELLE

Definition 3.34 (Potential confounder σ-algebra of X with respect to t) *Let $(\mathscr{C}_t, t \in T)$ and $(\mathscr{F}_t, t \in T)$ be filtrations on $(\Omega, \mathscr{A})$ (with identical index set T) and $X\colon (\Omega, \mathscr{A}) \to (\Omega'_X, \mathscr{A}'_X)$ a measurable mapping with $\sigma(X) \not\subset \mathscr{C}_t$, for all $t \in T$. Furthermore, assume that there is a $t_X \in T$ with*

(a) $\forall\, t < t_X\colon\ \mathscr{C}_t = \mathscr{F}_t$.
(b) $\forall\, t \geq t_X\colon\ \mathscr{F}_t = \sigma\big(\sigma(X) \cup \mathscr{C}_t\big)$.

Then $\mathscr{C}_t$, $t \in T$, $t \geq t_X$ is called the potential confounder σ-algebra of X with respect to t.

BAUSTELLE

Definition 3.35 (Causality Space)
Let $X\colon (\Omega, \mathscr{A}, P) \to (\Omega'_X, \mathscr{A}'_X)$ *and* $Y\colon (\Omega, \mathscr{A}, P) \to (\Omega'_Y, \mathscr{A}'_Y)$ *be random variables and let* $(\mathscr{F}_t, t \in T)$ *be a filtration in* $\mathscr{A}$*. Then* $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ *is called a* causality space*, if* X *is prior to* Y *in* $(\mathscr{F}_t, t \in T)$*.*

Remark 3.36 (Structural Prerequisites) Note that the dependence of Y on X in a causality space, which might be described by the conditional distribution $P_{Y|X}$, or, if Y is numerical, by the conditional expectation $E(Y|X)$, does not necessarily have a causal interpretation. So far, we only dealt with the *structural prerequisites* that (a) make the question about causal effects and dependencies meaningful, (b) allow us to define covariates and intermediate variables, and (c) allow us to define causal effects and causal dependencies in the chapters to come. Also note that, in Definition 3.35, neither X nor Y have to be numerical. Furthermore, the index set T does not *necessarily* refer to a time set, although this will often be the case. ◁

Example 3.37 (Joe and Ann With Self-Selection – continued) All components of a causality space have already been illustrated by the examples presented in Table 3.1 and in Table 4.1. In Exercise 3-9 we summarize the components of a causality space for the example of Table 3.1. ◁

3.5 Summary and Conclusions

In this chapter we set the stage for defining causal effects and causal probabilistic dependencies. We specified the structures and formulated the assumptions under which we can define causal effects and meaningfully raise the question if conditional expectations such as $E(Y|X{=}x)$ or $E(Y|X{=}x, Z{=}z)$ can be used to describe causal effects or if conditional distributions such as $P_{Y|X=x}$ or $P_{Y|X=x,Z=z}$ can be used to define (probabilistic) causal dependencies. The structures and assumptions of a causality space do not presume that the cause X is a treatment variable. X representing treatments, interventions, or expositions are just a typical applications. In other applications, the cause may also be an intermediate variable, a latent variable, or even an attribute of the observational-units, for instance. In the latter case, however, there will be no individual causal effects and no manipulability of the cause on the individual level.

Filtration

We used the concept of a filtration for representing time order between sets (events), measurable mappings (random variables), and σ-algebras (sets of events) that *may* be causally related. Such a filtration is not only used for the definition of the priority and simultaneity relations of random variables, events, and sets of events, but also for the definition of covariates and intermediate variables.

Box 3.1 Glossary of New Concepts

$(\mathscr{F}_t, t \in T)$	*Filtration in* $\mathscr{A}$. Let $(\Omega, \mathscr{A})$ be a measurable space. Then a filtration in $\mathscr{A}$ is a family of σ-algebras $\mathscr{F}_t \subset \mathscr{A}$ with $\mathscr{F}_s \subset \mathscr{F}_t$, if $s \leq t$, where $s, t \in T$. It represents the process that allows to define priority and simultaneity of events, random variables, and σ-algebras. It also allows to make the distinction between covariates of X and intermediate variables of X and Y.
Priority relation	A random variable X is *prior in* $(\mathscr{F}_t, t \in T)$ to another random variable Y, if there is an $s \in T$ such that the σ-algebra generated by X is a subset of $\mathscr{F}_s$, the σ-algebra generated by Y is not a subset of $\mathscr{F}_s$, and there is a $t \in T$, $s < t$, such that the σ-algebra generated by Y is a subset of $\mathscr{F}_t$.
Simultaneity relation	A random variable X is *simultaneous* to a random variable Y in $(\mathscr{F}_t, t \in T)$, if there is a $t \in T$ such that the σ-algebras generated by X and by Y, respectively, are subsets of $\mathscr{F}_t$, but there is no $s \in T$, $s < t$, such that the σ-algebra generated by X or the σ-algebra generated by Y is a subset of $\mathscr{F}_s$.
$\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$	*Causality space.* It summarizes the mathematical structures and assumptions under which we can define causal effects and raise the question if the dependence of Y on X has a causal meaning. It consists of a probability space $(\Omega, \mathscr{A}, P)$, a filtration $(\mathscr{F}_t, t \in T)$, the focused cause X and outcome variable Y. It is called *numerical* if Y is numerical with finite second moment $E(Y^2)$. It is assumed that X is prior to Y.
$\mathscr{F}_{t_X}$	The σ-algebra of the filtration $(\mathscr{F}_t, t \in T)$ that is simultaneous to X.

Priority Relation

In the simple single-unit trials of experiments and quasi-experiments described in chapter 2, the observational-unit variables and their their functions are always prior to the treatment variable, which itself is always prior to the outcome variable considered. The priority relation allows to represent this asymmetry of causal dependencies between random variables or between events. The basic idea is to see if, in the filtration considered, the σ-algebra generated by one random variable comes first and the σ-algebra generated by the other one comes later. An important virtue of this conception is that events and random variables

can be stretched over several time points and the concept of priority will still apply. A similar argument applies to simultaneity.

We will use the asymmetry of the priority relation to represent the asymmetry of a causal dependence, a necessary but not sufficient condition for a dependence of Y on X to have a causal meaning. (Sufficient conditions will be introduced in chs. 6 to 9.) The asymmetry implies that X cannot be causally dependent on Y, if Y is causally dependent on X. Hence, reciprocal causality *between random variables* will be excluded. Note, however, that this does not preclude the idea of reciprocal causality altogether. It only means that reciprocal causality does not apply to *random variables*. We do *not* preclude reciprocal causality between two *stochastic processes*. A stochastic process $(X_t, t \in T)$ is a *family* of random variables. Hence, X_1 (e. g., *anger expression* of Jim at time 1) may cause Y_2 (e. g., *anger expression* of Jane at time 2) and Y_1 (e. g., *anger expression* of Jane at time 1) may cause X_2 (e. g., *anger expression* of Jim at time 2), etc. (see, e. g., Kenny & Judd, 1996). Note that X_1 and X_2 are different random variables representing anger expression of Jim at time 1 and time 2, respectively. Similarly, Y_1 and Y_2 are different random variables representing anger expression of Jane at time 1 and time 2. Furthermore, the asymmetry of the priority relation does *not exclude* that manipulating *motivation* leads to higher *achievement* in a first experiment and that manipulating *achievement* leads to a higher *motivation* in a second experiment. While this example refers to random variables in two *different* random experiments, the priority relation refers to random variables within the *same* random experiment.

Causality Space

A causality space summarizes the mathematical structure and the assumptions under which we can meaningfully define causal effects and raise the question if the dependence of Y on X describes a causal dependence. Aside from the probability space $(\Omega, \mathscr{A}, P)$ representing the random experiment considered, the cause X and the outcome variable Y, it also consists of the filtration $(\mathscr{F}_t, t \in T)$ in $\mathscr{A}$.

Outlook

Based on the notion of a numerical causality space, in chapter 4 we will introduce the concepts of a *total-effect true-outcome variable*, a *true total-effect variable*, *true direct-effectvariable*, and *true indirect-effect variable*. In chapter 5 we then define average and various kinds of conditional total, direct, and indirect effects. In chapter 6, we will introduce *unbiasedness*, a first causality condition. Chapters 7 to 9 are devoted to a number of other causality conditions that imply unbiasedness.

3.6 Proofs

Proof of Theorem 3.12

(i) If $\mathscr{C}$ is prior to $\mathscr{D}$ with respect to $(\mathscr{F}_t, t \in T)$, then there is an $s \in T$ such that $\mathscr{C} \subset \mathscr{F}_s$, $\mathscr{D} \not\subset \mathscr{F}_s$. If we assume that $\mathscr{D}$ is prior to $\mathscr{C}$, then there is an $r \in T$ such that $\mathscr{D} \subset \mathscr{F}_r$, $\mathscr{C} \not\subset \mathscr{F}_r$. Because $(\mathscr{F}_t, t \in T)$ is a filtration and $\mathscr{D} \not\subset \mathscr{F}_s$, $\mathscr{D} \subset \mathscr{F}_r$, we can conclude $s < r$. Similarly, $\mathscr{C} \not\subset \mathscr{F}_r$, $\mathscr{C} \subset \mathscr{F}_s$ implies $r < s$. This is a contradiction to our assumption.

(ii) If $\mathscr{C}$ is prior to $\mathscr{D}$, then

(a) there is an $r \in T$ with $\mathscr{C} \subset \mathscr{F}_r$ and $\mathscr{D} \not\subset \mathscr{F}_r$,

and if $\mathscr{D}$ is also prior to $\mathscr{E}$, then

(b) there is a $s \in T,\ r < s$, with $\mathscr{D} \subset \mathscr{F}_s$ and $\mathscr{E} \not\subset \mathscr{F}_s$, and

(c) there is a $t \in T,\ s < t$, with $\mathscr{E} \subset \mathscr{F}_t$.

(a) to (c) imply that there is an $r \in T$ with $\mathscr{C} \subset \mathscr{F}_r$, $\mathscr{E} \not\subset \mathscr{F}_r$, and there is a $t \in T,\ r < t$, with $\mathscr{E} \subset \mathscr{F}_t$.

Proof of Theorem 3.18

(i) If $\mathscr{C}$ is prior to $\mathscr{D}$ with respect to $(\mathscr{F}_t, t \in T)$, then:

(a) there is an $s \in T$ with $\mathscr{C} \subset \mathscr{F}_s$ and $\mathscr{D} \not\subset \mathscr{F}_s$,

(b) there is a $t \in T,\ s < t$, with $\mathscr{D} \subset \mathscr{F}_t$.

Because $(\mathscr{F}_t, t \in T)$ is a filtration, it follows that

(c) $\mathscr{C} \cup \mathscr{D}$, $\sigma(\mathscr{C} \cup \mathscr{D}) \subset \mathscr{F}_t$, and

(d) $\mathscr{C} \cup \mathscr{D}$, $\sigma(\mathscr{C} \cup \mathscr{D}) \not\subset \mathscr{F}_s$.

Now, (a), (c), and (d) imply proposition (i).

(ii) If $\mathscr{C}$ and $\mathscr{D}$ are prior to $\mathscr{E}$ with respect to $(\mathscr{F}_t, t \in T)$, then:

(a) there is an $r \in T$ with $\mathscr{C} \subset \mathscr{F}_r$ and $\mathscr{E} \not\subset \mathscr{F}_r$,

(b) there is an $s \in T$ with $\mathscr{D} \subset \mathscr{F}_s$ and $\mathscr{E} \not\subset \mathscr{F}_s$,

(c) there is a $t \in T,\ r, s < t$, with $\mathscr{E} \subset \mathscr{F}_t$.

Without loss of generality, we can assume $r \le s$, which implies $\mathscr{C}, \mathscr{D} \subset \mathscr{F}_s$, because $(\mathscr{F}_t, t \in T)$ is a filtration. However, if $\mathscr{C}, \mathscr{D} \subset \mathscr{F}_s$, then:

(d) $\mathscr{C} \cup \mathscr{D}$, $\sigma(\mathscr{C} \cup \mathscr{D}) \subset \mathscr{F}_s$.

Now (b), (c), and (d) imply proposition (ii).

(iii) If $\mathscr{C}$ is prior to both $\mathscr{D}$ and $\mathscr{E}$ with respect to $(\mathscr{F}_t, t \in T)$, then

(a) $\exists\, r \in T$ such that $\mathscr{C} \subset \mathscr{F}_r$ and $\mathscr{D} \not\subset \mathscr{F}_r$,

(b) $\exists\, s \in T$ such that $\mathscr{D} \subset \mathscr{F}_s$, $\mathscr{E} \not\subset \mathscr{F}_s$

(c) $\exists\, t \in T,\ t > r, s$ such that $\mathscr{D}, \mathscr{E} \subset \mathscr{F}_t$, because $(\mathscr{F}_t, t \in T)$ is a filtration.

Without loss of generality can assume $r \le s$. Because $\mathscr{E} \subset \mathscr{F}_s$ implies $\mathscr{E} \subset \mathscr{F}_r$, and therefore

(d) $\mathscr{D} \cup \mathscr{E}$, $\sigma(\mathscr{D} \cup \mathscr{E}) \subset \mathscr{F}_t \not\subset \mathscr{F}_r$,

proposition (d) implies

(e) $\exists\, t > r$ such that $\mathscr{D} \cup \mathscr{E}$, $\sigma(\mathscr{D} \cup \mathscr{E}) \subset \mathscr{F}_t$

Now (a), (d), and (e) imply proposition (iii).

Proof of Theorem 3.29

(i) Note that not every set system $\mathscr{C} \subset \mathscr{A}$ necessarily occurs in $(\mathscr{F}_t, t \in T)$, i. e., we do not presume that there is a $t \in T$ such that $\mathscr{C} \subset \mathscr{F}_t$. Therefore we have to make the assumption "if there is a $t \in T \ldots$" .
(ii) is trivial.
(iii) If $\mathscr{C}$ and $\mathscr{D}$ are simultaneous, then

(a) there is a $t \in T$ with $\mathscr{C}, \mathscr{D} \subset \mathscr{F}_t$ and no $s \in T,\ s < t$, with $\mathscr{C} \subset \mathscr{F}_s$ or $\mathscr{D} \subset \mathscr{F}_s$.

If $\mathscr{D}$ and $\mathscr{E}$ are also simultaneous, then this implies $\mathscr{E} \subset \mathscr{F}_t$ and that

(b) there is no $s \in T,\ s < t$, with $\mathscr{D} \subset \mathscr{F}_s$ or $\mathscr{E} \subset \mathscr{F}_s$.

However, this implies that there is a $t \in T$ with $\mathscr{C}, \mathscr{E} \subset \mathscr{F}_t$, and that there is no $s \in T,\ s < t$, with $\mathscr{C} \subset \mathscr{F}_s$ or $\mathscr{E} \subset \mathscr{F}_s$. Hence, $\mathscr{C}$ and $\mathscr{E}$ are simultaneous as well.

Proof of Theorem 3.31

(i) If $\mathscr{C}$ and $\mathscr{D}$ are simultaneous with respect to $(\mathscr{F}_t, t \in T)$, then:

(a) there is a $t \in T$ with $\mathscr{C}, \mathscr{D} \subset \mathscr{F}_t$,

(b) there is no $s \in T,\ s < t$, with $\mathscr{C} \subset \mathscr{F}_s$ or $\mathscr{D} \subset \mathscr{F}_s$.

$(\mathscr{F}_t, t \in T)$ being a filtration implies:

(c) $\mathscr{C} \cup \mathscr{D}$, $\sigma(\mathscr{C} \cup \mathscr{D}) \subset \mathscr{F}_t$, and

(d) there is no $s \in T,\ s < t$, with $\mathscr{C} \cup \mathscr{D}$, $\sigma(\mathscr{C} \cup \mathscr{D}) \subset \mathscr{F}_s$.

Now (a) to (d) imply proposition (i) .
(ii) If $\mathscr{C}, \mathscr{D}$, and $\mathscr{E}$ are simultaneous with respect to $(\mathscr{F}_t, t \in T)$, then:

(a) there is a $t \in T$ with $\mathscr{C}, \mathscr{D}, \mathscr{E} \subset \mathscr{F}_t$,

(b) there is no $s \in T,\ s < t$, with $\mathscr{C} \subset \mathscr{F}_s$, $\mathscr{D} \subset \mathscr{F}_s$, or $\mathscr{E} \subset \mathscr{F}_s$.

$(\mathscr{F}_t, t \in T)$ being a filtration implies:

(c) $\mathscr{C} \cup \mathscr{D}$, $\sigma(\mathscr{C} \cup \mathscr{D}) \subset \mathscr{F}_t$,

(d) there is no $s \in T,\ s < t$, with $\mathscr{C} \cup \mathscr{D}$, $\sigma(\mathscr{C} \cup \mathscr{D}) \subset \mathscr{F}_s$.

Now (a) to (d) imply proposition (ii).

3.7 Exercises

▷ **Exercise 3-1** What are the additional components distinguishing a causality space from a probability space $(\Omega, \mathscr{A}, P)$?

▷ **Exercise 3-2** What is a filtration $(\mathscr{F}_t, t \in T)$ in a σ-algebra $\mathscr{A}$?

▷ **Exercise 3-3** What is the basic idea of the priority relation between random variables?

▷ **Exercise 3-4** Construct a filtration $(\mathscr{F}_t, t \in T)$ for the random experiment of flipping a coin two times and define two random variables X and Y such that X is prior to Y with respect to $(\mathscr{F}_t, t \in T)$.

▷ **Exercise 3-5** Show that $Z :=$ *sex* is prior to X in Example 3.4.

▷ **Exercise 3-6** Prove the proposition of Remark 3.14.

▷ **Exercise 3-7** Prove the proposition of Remark 3.33.

▷ **Exercise 3-8** Consider Example 3.3 and Remark 3.17 as well as the events A_1 that *Joe is sampled and treated* and A_2 that *Joe is sampled, treated, and successful*. Show that the σ-algebra generated by the set system $\{A_1\}$ is a subset of $\mathscr{F}_2$, whereas the σ-algebra generated by the set system $\{A_2\}$ is not a subset of $\mathscr{F}_2$, but a subset of $\mathscr{F}_3$.

▷ **Exercise 3-9** Define a filtration and the σ-algebra $\mathscr{F}_{t_X}$ that is simultaneous to X for the example 'Joe and Ann With Self-Selection' (see Table 3.1, p. 49).

Solutions

▷ **Solution 3-1** First, there is a filtration $(\mathscr{F}_t, t \in T)$ in the σ-algebra $\mathscr{A}$ of the probability space $(\Omega, \mathscr{A}, P)$, representing the different phases of the causal process in which the events and random variables occur. Second, there are two random variables on the probability space, say X and Y, where X represents the cause and Y the outcome variable considered.

▷ **Solution 3-2** A filtration $(\mathscr{F}_t, t \in T)$ in $\mathscr{A}$ consists of a set T on which there are relations $<$, $=$, and $\leq$, and σ-algebras $\mathscr{F}_s$ and $\mathscr{F}_t$ with $\mathscr{F}_s \subset \mathscr{F}_t$ if $s \leq t$, where $s, t \in T$.

▷ **Solution 3-3** The basic idea is to see if the σ-algebra generated by one random variable comes first in the filtration $(\mathscr{F}_t, t \in T)$ and the σ-algebra generated by the other one comes later. More formally speaking, if $X\colon (\Omega, \mathscr{A}, P) \to (\Omega'_X, \mathscr{A}'_X)$ and $Y\colon (\Omega, \mathscr{A}, P) \to (\Omega'_Y, \mathscr{A}'_Y)$ are two random variables, then X *is prior to* Y if (a) there is an $s \in T$ with $X^{-1}(\mathscr{A}'_X) \subset \mathscr{F}_s$ and $Y^{-1}(\mathscr{A}'_Y) \not\subset \mathscr{F}_s$, and (b) there is a $t \in T$, $s < t$, with $Y^{-1}(\mathscr{A}'_Y) \subset \mathscr{F}_t$.

▷ **Solution 3-4** The set of possible outcomes is $\Omega = \{(h,h),(h,t),(t,h),(t,t)\}$, where, e. g., (h,t) represents the outcome of flipping 'heads' at the first flip and 'tails' at the second flip. As the σ-algebra $\mathscr{A}$ on Ω we choose the power set of Ω and the probability measure on $\mathscr{A}$ is defined by assigning the probabilities $1/4$ to each elementary event $\{\omega\}$, $\omega \in \Omega$. This uniquely defines the probabilities of all other events $A \in \mathscr{A}$.

Next, we define $\mathscr{F}_1 := \{\Omega, \emptyset, \{(h,h),(h,t)\}, \{(t,h),(t,t)\}\}$ containing, aside from Ω and $\emptyset$, the event $\{(h,h),(h,t)\}$ to flip 'heads' at the first trial and the event $\{(t,h),(t,t)\}$ to flip 'tails' at the first trial. As a second σ-algebra we define $\mathscr{F}_2 := \mathscr{A}$. Then $(\mathscr{F}_t, t \in T)$, $T = \{1,2\}$, is a filtration, because $\mathscr{F}_1 \subset \mathscr{F}_2$.

Finally, we define X to take on the value 1, if we flip 'heads' at the first toss and 0 otherwise. Similarly, we define Y to take on the value 1, if we flip 'heads' at the second flip and 0 otherwise. Then X is prior to Y, because the σ-algebra generated by X is $\mathscr{F}_1$, which is a subset of itself, whereas $\{\Omega, \emptyset, \{(h,h),(t,h)\}, \{(h,t),(t,t)\}\}$ is the σ-algebra generated by Y, and this σ-algebra is not a subset of $\mathscr{F}_1$, but of $\mathscr{F}_2$ (see Def. 3.1).

▷ **Solution 3-5** The first σ-algebra $\mathscr{F}_1$ in the filtration $(\mathscr{F}_t, t \in T)$, $T = \{1,2,3\}$, specified in Example 3.4 is the σ-algebra generated by the observational-unit variable. Presuming that the observational units are persons, $Z :=$ *sex* is measurable with respect to $\mathscr{F}_1$, whereas X is not. However, X is measurable with respect to $\mathscr{F}_2$.

▷ **Solution 3-6** If X_1 is prior to X_2 with respect to $(\mathscr{F}_t, t \in T)$, then
X_1 is also prior to $X_1 - X_2$ with respect to $(\mathscr{F}_t, t \in T)$.

▷ **Solution 3-7** Suppose that X_1 is prior to X_2 with respect to $(\mathscr{F}_t, t \in T)$. Then:

(a) $\exists s \in T\colon \sigma(X_1) \subset \mathscr{F}_s$, $\sigma(X_2) \not\subset \mathscr{F}_s$

(b) $\exists t \in T\colon \sigma(X_2) \subset \mathscr{F}_t$.

Let t be the smallest element of T with $\sigma(X_2) \subset \mathscr{F}_t$ (see Remark 3.20 for its existence).
Case 1: $\exists r < t, r \in T : \sigma(X_1 \cdot X_2) \subset \mathscr{F}_r$. This implies that $X_1 \cdot X_2$ is prior to X_2.
Case 2: $\forall\, r < t, r \in T : \sigma(X_1 \cdot X_2) \subset \mathscr{F}_r$ with

(a) $\forall\, r < t, r \in T : \sigma(X_2) \not\subset \mathscr{F}_r$ (see Remark 3.20,

(b) $\sigma(X_1 \cdot X_2) \subset \mathscr{F}_t$ (see Th. 2.57 of SN)

Hence, in this case $X_1 \cdot X_2$ is simultaneous to X_2.

▷ **Solution 3-8**

$$A_1 \;=\; \big\{(\mathit{Joe}, \mathit{yes}, -),\ (\mathit{Joe}, \mathit{yes}, +)\big\} \;=\; \{U{=}\mathit{Joe}\} \cap \{X{=}1\},$$

where

$$\{U{=}\mathit{Joe}\} \;=\; \big\{(\mathit{Joe}, \mathit{yes}, -),\ (\mathit{Joe}, \mathit{yes}, +),\ (\mathit{Joe}, \mathit{no}, -),\ (\mathit{Joe}, \mathit{no}, +)\big\}$$

and

$$\{X{=}1\} \;=\; \big\{(\mathit{Joe}, \mathit{yes}, -),\ (\mathit{Joe}, \mathit{yes}, +),\ (\mathit{Ann}, \mathit{yes}, -),\ (\mathit{Ann}, \mathit{no}, +)\big\}.$$

Now $\mathscr{F}_2 = \sigma[\mathscr{F}_1 \cup X^{-1}(\mathscr{A}'_X)] = \sigma[\sigma(U) \cup \sigma(X)]$. The definitions of $\sigma(U)$ and $\sigma(X)$ imply $\{U{=}\mathit{Joe}\} \in \sigma(U)$ and $\{X{=}1\} \in \sigma(X)$, and the definition of the σ-algebra $\mathscr{F}_2 = \sigma\big[\mathscr{F}_1 \cup X^{-1}(\mathscr{A}'_X)\big]$ implies $\{U{=}\mathit{Joe}\} \in \mathscr{F}_2$ and $\{X{=}1\} \in \mathscr{F}_2$. Finally, the definition of a σ-algebra implies $A_1 = \{U{=}\mathit{Joe}\} \cap \{X{=}1\} \in \mathscr{F}_2$ [see Eq. (1.7) of SN], $A_1^c \in \mathscr{F}_2$, $\Omega \in \mathscr{F}_2$, and $\emptyset \in \mathscr{F}_2$ [see Def. 1.1 of SN]. This proves that $\sigma(\{A_1\}) \subset \mathscr{F}_2$.

By definition,

$$\mathscr{F}_2 \;=\; \sigma\big(\mathscr{F}_1 \cup X^{-1}(\mathscr{A}'_X)\big) \;=\; \sigma\big(\big\{\Omega, \emptyset, \{U{=}\mathit{Joe}\}, \{U{=}\mathit{Ann}\}, \{X{=}0\}, \{X{=}1\}\big\}\big).$$

The definition of a σ-algebra implies that the intersections $B_1 = \{U{=}\mathit{Joe}\} \cap \{X{=}0\}$, $B_2 = \{U{=}\mathit{Joe}\} \cap \{X{=}1\}$, $B_3 = \{U{=}\mathit{Ann}\} \cap \{X{=}0\}$, and $B_4 = \{U{=}\mathit{Ann}\} \cap \{X{=}1\}$ are elements of $\mathscr{F}_2$ [see again Eq. (1.7) of SN]. However, $\mathscr{E} = \big\{B_1, B_2, B_3, B_4\big\}$ is a partition of Ω and

$$\mathscr{F}_2 \;=\; \sigma(\mathscr{E}).$$

According to Lemma 1.20 of SN, each element of $\mathscr{F}_2$ is a union of elements of $\mathscr{E}$, except for $\emptyset$. Because there are no elements of $\mathscr{E}$ such that $A_2 = \big\{(\mathit{Joe}, \mathit{yes}, +)\big\}$ is the union of these elements, we can conclude $A_2 \notin \mathscr{F}_2$ and therefore $\sigma(\{A_2\}) \not\subset \mathscr{F}_2$. However, $A_2 \in \mathscr{F}_3$ and therefore $\{A_2\} \subset \mathscr{F}_3$, because $\mathscr{F}_3$ is the power set of Ω.

▷ **Solution 3-9** In this random experiment, the set of possible outcomes can be written $\Omega = \Omega_U \times \Omega_X \times \Omega_Y$. The σ-algebra $\mathscr{A}$ is defined to be the power set of Ω. The filtration is specified as follows: $\mathscr{F}_1$ is the σ-algebra generated by U, $\mathscr{F}_2$ is generated by (X, U), and $\mathscr{F}_3$ is generated by (X, U, Y), which is also the power set of Ω. Aside from Ω and the empty set $\emptyset$, the σ-algebra $X^{-1}(\mathscr{A}'_X)$ generated by X consists of the sets $X^{-1}(\{1\}) = \Omega_U \times \{\mathit{yes}\} \times \Omega_Y$ that the person drawn is treated and $X^{-1}(\{0\}) = \Omega_U \times \{\mathit{no}\} \times \Omega_Y$ that the person drawn is not treated. The X-concurrent σ-algebra $\mathscr{F}_{t_X}$ is $\mathscr{F}_2$.

Chapter 4
True-Outcome Variables and Atomic Effect Variables

In chapter 1, it has been shown that the conditional expectation values $E(Y \mid X{=}x)$ of an outcome variable Y and their differences $E(Y \mid X{=}x) - E(Y \mid X{=}x')$, the *prima facie effects*, can be misleading in evaluating the effects of a treatment variable X on an outcome variable Y. In chapter 2 we described random experiments of various research designs in which causal effects are of interest. In chapter 3 we dealt with the additional structural components that distinguish a causality space from a probability space. The crucial additional component is a filtration, which allows to define the priority and simultaneity relations between random variables, events, and sets of events.

This chapter starts with the definitions of the potential-confounder σ-algebra, a covariate and a intermediate variable of a cause with respect to a filtration. Based on the concept of a potential-confounder σ-algebra, we define true-outcome variables *with respect to total effects* and *with respect to direct effects*. In contrast to the original outcome variables Y, these true-outcome variables are constructed such that they are purged from bias. Using these true-outcome variables we can then introduce *atomic total effect variables*, *atomic direct effect variables*, and *atomic indirect effect variables*. In chapter 5 we use these concepts to define various kinds of causal effects by expectations and conditional expectation values of the atomic effect variables.

4.1 Potential-Confounder σ-Algebra and Covariates

In chapters 2 and 3 we already discussed different kinds of covariates and their role a potential confounders in single-unit trials. There we already mentioned that a cause cannot be prior to one of its covariates, because we consider all variables that are prior or at least simultaneous to X as covariates (potential confounders) that may lead to biased prima-facie effects of X.

A potential confounder is each random variable that is measurable with respect to the σ-algebra $\mathscr{C}_X$, the σ-algebra of potential confounders of X, or more briefly, the *potential-confounder σ-algebra of* X. This σ-algebra can be conceived of as the set of all events that are prior or simultaneous to the cause$X$, except for the events represented by X itself.

Neuer Versuch zur Def der Globalen Covariate

Definition 4.1 (Global Covariate of X With Respect to t)
Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a causality space, let $t \in T$ with $t_X \le t < t_Y$, and let $\Omega_{\le t}$ and $\Omega_{>t}$ be sets such that

$$\Omega \;=\; \Omega_{\le t} \times \Omega_X \times \Omega_{>t}\,. \tag{4.1}$$

Furthermore, let $C_{X,t}\colon (\Omega, \mathscr{A}) \to (\Omega_{\le t}, \mathscr{A}_{\le t})$ be a projection mapping satisfying

(a) $\{\omega_{\le t}\} \in \mathscr{A}_{\le t}, \quad \forall\, \omega_{\le t} \in \Omega_{\le t}$
(b) $\sigma(X, C_{X,t_X}) = \mathscr{F}_{t_X}$
(c) $\sigma(X, C_{X,t}) \subset \mathscr{F}_t$
(d) $\sigma(X) \not\subset \sigma(C_{X,t})$.

Then $C_{X,t}$ is called the global covariate of X with respect to t (or the global t-covariate of X), and $C_X := C_{X,t_X}$ the global covariate of X.

Remark 4.2 (Uniqueness) Note that, in Definition 4.1, the set Ω and the random variable X are fixed, and this implies that the sets $\Omega_{\le t}$ and $\Omega_{>t}$ are fixed as well. Therefore, the projection mapping $C_{X,t}\colon (\Omega, \mathscr{A}) \to (\Omega_{\le t}, \mathscr{A}_{\le t})$ is uniquely defined. Its values are

$$C_{X,t}(\omega) \;=\; C_{X,t}(\omega_{\le t}, \omega_X, \omega_{>t}) \;=\; \omega_{\le t}, \quad \forall\, \omega \in \Omega\,. \tag{4.2}$$

◁

Definition 4.3 (Potential-Confounder σ-Algebra of X)
Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a causality space and $\mathscr{F}_{t_X}$ the σ-algebra of $(\mathscr{F}_t, t \in T)$ that is simultaneous to X. Then

$$\mathscr{C}_X \;:=\; \bigcap_{i \in I} \mathscr{C}_i, \tag{4.3}$$

where I is the index set of all σ-algebras $\mathscr{C}_i \subset \mathscr{F}_{t_X}$ satisfying

(a) $\sigma\big(\mathscr{C}_i \cup \sigma(X)\big) = \mathscr{F}_{t_X}$
(b) $\mathscr{F}_t \subset \mathscr{C}_i, \quad$ *for all* $t < t_X, t \in T$.

The σ-algebra $\mathscr{C}_X$ is called the potential-confounder σ-algebra of X with respect to $(\mathscr{F}_t, t \in T)$.

Hence, the potential-confounder σ-algebra $\mathscr{C}_X$ does not only refer to a specified random variable X, but also to a specified filtration and, of course, to a specified probability space $(\Omega, \mathscr{A}, P)$.

Remark 4.4 (Some Properties of the Potential-Confounder σ-Algebra) Note that there can be several σ-algebras satisfying conditions (a) and (b) of Definition 4.3. However, defining $\mathscr{C}_X$ as the smallest of these σ-algebras [see Eq. (4.3)] secures that $\mathscr{C}_X$ is uniquely defined. Furthermore, Definition 4.3 implies

(i) $\mathscr{C}_X$ is prior or simultaneous to X
(ii) $\sigma(X) \not\subset \mathscr{C}_X$
(iii) All σ-algebras $\mathscr{F}_t$, $t < t_X$, of the filtration $(\mathscr{F}_t, t \in T)$ are subsets of $\mathscr{C}_X$
(iv) If Z is a random variable on $(\Omega, \mathscr{A}, P)$ that is prior to X with respect to $(\mathscr{F}_t, t \in T)$, then $\sigma(Z) \subset \mathscr{C}_X$.

(see Exercise 4-1). ◁

Definition 4.5 (Covariate of X With Respect to a Filtration)
Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a causality space, $\mathscr{C}_X$ the potential-confounder σ-algebra of X, and $Z\colon (\Omega, \mathscr{A}, P) \to (\Omega'_Z, \mathscr{A}'_Z)$ a random variable. Then Z is called a covariate of X w.r.t. to $(\mathscr{F}_t, t \in T)$ if $\sigma(Z) \subset \mathscr{C}_X$.

Remark 4.6 (Some Propositions About Covariates) The definition of a covariate implies the following propositions:

(i) A cause X can never be prior to a covariate of X.
(ii) All random variables on $(\Omega, \mathscr{A}, P)$ that are prior to X are covariates of X.
(iii) If Z is a covariate of X, then every random variable on $(\Omega, \mathscr{A}, P)$ that is measurable with respect to Z is a covariate of X.

◁

Remark 4.7 (Two Causes) If we consider as causes two different random variables X_1 and X_2 on the same causality space, and X_1 is prior to X_2, then we also have to consider two different potential-confounder σ-algebras $\mathscr{C}_{X_1}$ and $\mathscr{C}_{X_2}$. If X_2 is a cause and X_1 is prior to X_2 with respect to $(\mathscr{F}_t, t \in T)$, then X_1 is a covariate of X_2 (see Exercise 4-2). ◁

4.1.1 Examples

Example 4.8 (Simple Experiments) In Example 3.4, we specified the filtration for the single-unit trial of experiments and quasi-experiments in which no, or at least no fallible covariates are observed. In these cases, $\mathscr{F}_1$ is the σ-algebra generated by the observational-unit variable U, whereas $\mathscr{F}_2$ is the σ-algebra generated by U and the treatment variable X, and $\mathscr{F}_3$ is the σ-algebra generated by U, X, and the outcome variable Y.

In this example, $\sigma(U)$ is the potential-confounder σ-algebra of X with respect to $(\mathscr{F}_t, t \in T)$. The definition of $\mathscr{F}_2$ is such that it is identical to $\mathscr{F}_{t_X}$ [see condition (a) of Definition 4.3]. Furthermore, $\mathscr{F}_1$ and $\sigma(U)$ are identical by definition so that condition (b) of Definition 4.3 is satisfied as well. Finally, in this example, $\sigma(U)$ is the smallest σ-algebra satisfying conditions (a) and (b) of Definition 4.3 (see Exercise 4-3). Hence, $\mathscr{C}_X = \sigma(U)$ [see Eq. (4.3)].

According to Definition 4.5, each random variable on $(\Omega, \mathscr{A}, P)$ that is measurable with respect to U is a covariate. *Sex*, *race*, and *educational status* are examples in case, as well as all other random variables representing an attribute of the

Table 4.1. Joe With Two Independent Treatments

Outcomes ω			Random variables		
Individual therapy	Group therapy	Success	Probabilities of elementary events $P(\{\omega\})$	Treatment variable X	Treatment variable Z	Outcome variable Y
(*no*,	*no*,	−)	.09	*no*	*no*	0
(*no*,	*no*,	+)	.21	*no*	*no*	1
(*no*,	*yes*,	−)	.04	*no*	*yes*	0
(*no*,	*yes*,	+)	.16	*no*	*yes*	1
(*yes*,	*no*,	−)	.24	*yes*	*no*	0
(*yes*,	*no*,	+)	.06	*yes*	*no*	1
(*yes*,	*yes*,	−)	.16	*yes*	*yes*	0
(*yes*,	*yes*,	+)	.04	*yes*	*yes*	1

observational units that is observed *without measurement error*. The important point is that all these other random variables are measurable with respect to U, which itself is measurable with respect to $\mathscr{F}_{t_X}$, the σ-algebra of $(\mathscr{F}_t, t \in T)$ that is simultaneous to X. ◁

Example 4.9 (Joe With Two Independent Treatments) Table 4.1 presents a random experiment in which a given unit, say Joe, may simultaneously receive either:

(a) no treatment at all,
(b) not a first treatment (e. g., no individual therapy) but a second one (e. g., group therapy),
(c) a first treatment (individual therapy) but no second one (no group therapy),
(d) both treatments (individual therapy and group therapy).

Furthermore, it is registered if a success criterion is reached (+) or not (−). The set Ω of all possible outcomes of the random experiment consists of the eight possible outcomes listed in the first column of the table. We choose $\mathscr{A} = \mathscr{P}(\Omega)$. The probabilities of all $2^8 = 256$ events are determined by the probabilities of the elementary events displayed in the first column of the table. The table also shows how the random variables X, Y, and Z are defined, i.e., how they assign their values to each $\omega \in \Omega$.

The first σ-algebra $\mathscr{F}_1$ in the filtration $(\mathscr{F}_t, t \in T)$ is chosen to be the σ-algebra generated by the two treatment variables X and Z, i. e,

$$\mathscr{F}_1 := \sigma(X, Z).$$

This σ-algebra has $2^4 = 16$ elements, the four events

$$\begin{array}{ll}
\{(no, no, -), (no, no, +)\} & \textit{'neither individual nor group therapy'}, \\
\{(no, yes, -), (no, yes, +)\} & \textit{'no individual but group therapy'}, \\
\{(yes, no, -), (yes, no, +)\} & \textit{'individual but no group therapy'}, \\
\{(yes, yes, -), (yes, yes, +)\} & \textit{'individual and group therapy'},
\end{array}$$

all (pairwise, triple-wise, quadruple-wise) unions of these four events, and the impossible event Ø. The first set $\{(no, no, -), (no, no, +)\}$, e. g., is the event that Joe neither receives treatment 1 nor treatment 2, whereas the second set $\{(no, yes, -), (no, yes, +)\}$ is the event that Joe does not receive treatment 1 (individual therapy) but receives treatment 2 (group therapy).

The second σ-algebra $\mathscr{F}_2$ in the filtration is chosen to be the power set of Ω, i. e.,

$$\mathscr{F}_2 := \mathscr{P}(\Omega).$$

This power set has $2^8 = 256$ elements. It consists of the eight elementary events

$$\{(no, no, -)\}, \{(no, no, +)\}, \ldots, \{(yes, yes, +)\}$$

(see the first column of Table 4.1, p. 72), all unions of these events, and the impossible event Ø. In this example, the filtration $(\mathscr{F}_t, t \in T)$ only consists of $\mathscr{F}_1$ and $\mathscr{F}_2$.

Now we focus on the first treatment variable X (receiving or not receiving individual therapy) as a cause. The second treatment variable $Z\colon (\Omega, \mathscr{A}, P) \to (\Omega'_Z, \mathscr{A}'_Z)$ (receiving or not receiving group therapy), with

$$\Omega'_Z = \{no, yes\} \qquad \text{and} \qquad \mathscr{A}'_Z := \mathscr{P}(\Omega'_Z) = \big\{\{no\}, \{yes\}, \Omega'_Z, \emptyset\big\},$$

is a covariate of X with respect to $(\mathscr{F}_t, t \in T)$.

The σ-algebra $Z^{-1}(\mathscr{A}'_Z) = \sigma(Z)$ generated by Z has four elements, the two events:

$$A := \{(no, yes, -), (no, yes, +), (yes, yes, -), (yes, yes, +)\} \qquad \textit{'group therapy'},$$

and

$$A^c := \{(no, no, -), (no, no, +), (yes, no, -), (yes, no, +)\} \qquad \textit{'no group therapy'},$$

the sure event Ω, and the impossible event Ø. It is easily seen that Z satisfies the requirement of Definition 4.5, because $\sigma(Z) \subset \mathscr{F}_{t_X} = \mathscr{F}_1$ and $\sigma(Z) \not\subset \sigma(X)$. In this example, $t_X = t_1$ and $\mathscr{F}_1$ is the σ-algebra of $(\mathscr{F}_t, t \in T)$ that is simultaneous to X. ◁

4.1.2 Intermediate Variable

In Definition 4.5 we introduced the concept of a covariate. In order to distinguish between total and direct effects, we need the concept of an *intermediate variable* with respect to a cause X and an outcome Y. A synonym for 'intermediate variable' is 'potential mediator'.

Definition 4.10 (Intermediate Variable)
Let $(\Omega, \mathscr{A}, P)$ be a probability space and $(\mathscr{F}_t, t \in T)$ a filtration in $\mathscr{A}$. Furthermore, let X, Y, and M be random variables on $(\Omega, \mathscr{A}, P)$ such that X is prior to Y in $(\mathscr{F}_t, t \in T)$. Then M is called an intermediate variable of X and Y with respect to $(\mathscr{F}_t, t \in T)$, if

(a) X is prior to M in $(\mathscr{F}_t, t \in T)$
(b) M is prior to Y in $(\mathscr{F}_t, t \in T)$.

Hence, whereas a cause X can not be prior to a covariate with respect to the filtration $(\mathscr{F}_t, t \in T)$, the cause X *has* to be prior to M, if M is an intermediate variable of X and Y. Furthermore, M *has* to be prior to the outcome variable Y with respect to $(\mathscr{F}_t, t \in T)$. Note that a intermediate variable is not necessarily a mediator variable, because in Definition 4.10 we do not presume that X has a nonzero causal effect on M and that M has a nonzero causal effect on Y.

Example 4.11 (Experiments With Intermediate Variables) In Example 3.4, we already specified the filtration for the single-unit trial of experiments and quasi-experiments if we *do* observe a (possibly multivariate) fallible covariate. If we also include two intermediate variables, say M_1 and M_2, the set of possible outcomes ω can be written

$$\Omega = \Omega_U \times \Omega_Z \times \Omega_X \times \Omega_{M_1} \times \Omega_{M_2} \times \Omega_Y. \tag{4.4}$$

Compared to Equation (3.5), Ω involves the additional sets Ω_{M_1} and Ω_{M_2}. Now, the random experiment consists of drawing an observational unit from Ω_U, observing an element ω_Z of Ω_Z (based on which the fallible covariate Z assigns a value $z \in \Omega'_Z$ to $\omega \in \Omega$), assigning the unit or observing its selection to one of the experimental conditions (represented by the value x of the treatment variable X), observing an element ω_{M_1} of Ω_{M_1} (based on which the intermediate variable M_1 assigns a value $m_1 \in \Omega'_{M_1}$ to $\omega \in \Omega$), observing an element ω_{M_2} of Ω_{M_2} (based on which the intermediate variable M_2 assigns a value $m_2 \in \Omega'_{M_2}$ to $\omega \in \Omega$), and recording the numerical value y of the outcome variable Y. Correspondingly, the σ-algebra $\mathscr{A}$ on Ω is now defined to be a sixfold product σ-algebra, involving appropriate σ-algebras on each of the sets involved in the Cartesian product $\Omega_U \times \Omega_Z \times \Omega_X \times \Omega_{M_1} \times \Omega_{M_2} \times \Omega_Y$. Furthermore, now we have two additional random variables $M_1 \colon (\Omega, \mathscr{A}, P) \to (\Omega'_{M_1}, \mathscr{A}'_{M_1})$ and $M_2 \colon (\Omega, \mathscr{A}, P) \to (\Omega'_{M_2}, \mathscr{A}'_{M_2})$, the intermediate variables.

The first three σ-algebras $\mathscr{F}_1$, $\mathscr{F}_2$, and $\mathscr{F}_3$ can be specified in the same way as in Example 3.4. However, note that now they are σ-algebras on the set $\Omega =$

$\Omega_U \times \Omega_Z \times \Omega_X \times \Omega_{M_1} \times \Omega_{M_2} \times \Omega_Y$. The fourth σ-algebra $\mathscr{F}_4 := \sigma[\mathscr{F}_3 \cup M_1^{-1}(\mathscr{A}'_{M_1})]$ is the σ-algebra generated by the union of $\mathscr{F}_3$ and the σ-algebra generated by the intermediate variable M_1. A fifth σ-algebra $\mathscr{F}_5 := \sigma[\mathscr{F}_4 \cup M_2^{-1}(\mathscr{A}'_{M_2})]$ is the σ-algebra generated by the union of $\mathscr{F}_4$ and the σ-algebra generated by the intermediate variable M_2. In this case we presume that M_1 is prior to M_2. (Otherwise we would define $\mathscr{F}_4 := \sigma[\mathscr{F}_3 \cup \sigma(M_1, M_2)]$. In this case M_1 and M_2 would be simultaneous to each other.) Finally, the σ-algebra $\mathscr{F}_6$ is the σ-algebra generated by the union of $\mathscr{F}_5$ and the σ-algebra generated by the outcome variable Y, i. e., $\mathscr{F}_6 := \sigma[\mathscr{F}_5 \cup Y^{-1}(\mathscr{A}'_Y)]$.

With respect to this filtration $(\mathscr{F}_t, t \in T)$, we know that each mapping of U and each function of Z [see Remark 4.6 (iii)] are covariates of X. (Lemma 2.52 and Corollary 2.53 of SN give the exact mathematical background.)

◁

4.2 True-Outcome Variable and Atomic Effect Variables

True-outcome variables with respect to *total effects* and their analogs, true-outcome variables with respect to *direct effects* (see section 4.2.4), are the fundamental concepts of the theory of causal effects, at least in the case in which the cause X is discrete and in which, for each value x of X, there is a chance for all observational units to be observed under x. If X represents a treatment variable, then this means that each observational unit must have a positive probability to be in treatment x. If this requirement is not met, then atomic causal effect variables are not defined and a causal dependence has to be defined in a different way (see ch. ????).

Because an intermediate variable is not considered to be a covariate of a cause X, we define the *true-outcome-variable with respect to total effects* to be the conditional expectation of the outcome variable Y given the X-concurrent σ-algebra $\mathscr{F}_{t_X}$ with respect to the conditional-probability measure $P^{X=x}$. In other words, we condition on $\{X{=}x\}$ and on *all* covariates of X.[1] Conditioning on all covariates of X means that the total-effect true-outcome variables and their differences, the *atomic total effect variables*, is defined such that they are purged from bias. To emphasize, defining the atomic *total* effect variables, we do *not* condition on intermediate variables, and this is why we use the term 'total' in this context. Similarly, conditioning on all covariates of an intermediate variable M, we define *atomic direct effect variables*.

The intuitive background is as follows. Conditioning on all covariates, we share John Stuart Mill's idea already described in the preface. However, we make a slight modification, comparing the *conditional expectation values* of Y between treatment conditions. Suppose an observational unit can be treated by condition 1 or by condition 0, *everything else being invariant*. If there is a difference in the

[1] The total-effect true-outcome variables play a similar role as the potential outcome variables in Rubin's approach to causality (see, e.g., Rubin, 2005.)

conditional expectation values of the outcome considered, then this difference is due to the difference in the two treatment conditions. In our framework, 'everything else being invariant' means conditioning on a combination of values of all covariates of X. Hence, for each combination of covariate values we have an unbiased total treatment effect on the *most fine-grained* level, i. e. on the *atomic* level. In general, these atomic total treatment effects can differ for different combinations of covariate values, and these effects are the values of the *atomic total effect variables* that is defined in this chapter. Similarly, conditioning on all covariates of an intermediate variable M yields the *atomic direct effect variables*.

4.2.1 Conditional Expectation With Respect to $P^{X=x}$

In the following definition we refer to a value $x \in \Omega'_X$ of X with $P(X{=}x) > 0$. We consider the $\mathscr{C}_X$-conditional expectation $E^{X=x}(Y|\mathscr{C}_X)$ of Y with respect to the conditional-probability measure $P^{X=x}\colon \mathscr{A} \to [0,1]$ defined by

$$P^{X=x}(A) \;=\; P(A|X{=}x) \;=\; \frac{P(A \cap \{X{=}x\})}{P(X{=}x)} \quad \text{for all } A \in \mathscr{A}, \tag{4.5}$$

where $\{X{=}x\} = \{\omega \in \Omega\colon X(\omega) = x\}$ denotes the event that X takes on the value x. A version of $E^{X=x}(Y|\mathscr{C}_X)$ is always a random variable on the probability space $(\Omega, \mathscr{A}, P^{X=x})$. However, such a version is also a random variable on $(\Omega, \mathscr{A}, P)$ (see Exercise **??**). The conditional expectation $E^{X=x}(Y|\mathscr{C}_X)$ is called *P-unique* if

$$P\big(\{\omega \in \Omega\colon V_x(\omega) \neq V_x^*(\omega)\}\big) \;=\; 0$$

whenever V_x and V_x^* are two versions of $E^{X=x}(Y|\mathscr{C}_X)$ (see section 14.6 of SN). Hence, $E^{X=x}(Y|\mathscr{C}_X)$ being P-unique means that $E^{X=x}(Y|\mathscr{C}_X)$ uniquely defined almost surely with respect to measure P. Note, by definition, $E^{X=x}(Y|\mathscr{C}_X)$ is uniquely defined almost surely only with respect to conditional-probability measure $P^{X=x}$.

According to Theorem 14.48, P-uniqueness of $E^{X=x}(Y|\mathscr{C}_X)$ is equivalent to

$$P(X{=}x|\mathscr{C}_X) \;\underset{P}{\geqslant}\; 0, \tag{4.6}$$

which is a shortcut for

$$P\big(\{\omega \in \Omega\colon P(X{=}x|\mathscr{C}_X)(\omega) = 0\}\big) \;=\; 0. \tag{4.7}$$

Hence, Proposition (4.6) holds and $E^{X=x}(Y|\mathscr{C}_X)$ is P-unique, if the probability (with respect to the measure P) that $P(X{=}x|\mathscr{C}_X)$ takes on the value 0 is itself 0.

Note that if $E^{X=x}(Y|\mathscr{C}_X)$ is *not* P-unique, then the expectation $E[E^{X=x}(Y|\mathscr{C}_X)]$ is not uniquely defined. This means that the expectations $E(V_x)$ and $E(V_x^*)$ might differ if V_x and V_x^* are different versions of $E^{X=x}(Y|\mathscr{C}_X)$. More details and examples of the concept of a conditional expectation with respect to a conditional-probability measure $P^{X=x}$ are presented in chapter 14 of SN. That chapter also includes theorems on properties of this concept and sufficient conditions of P-uniqueness of conditional expectations with respect to $P^{X=x}$.

4.2.2 Total-Effect True-Outcome Variables

Total-effect true-outcome variables can now be defined as follows:

Definition 4.12 (Total-Effect True-Outcome Variables)
Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a causality space, $\mathscr{C}_X$ the potential-confounder σ-algebra of X, let $x \in \Omega'_X$ with $P(X{=}x) > 0$, and let $Y\colon (\Omega, \mathscr{A}, P) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$ be nonnegative or with finite expectation $E^{X=x}(Y)$. If $E^{X=x}(Y|\mathscr{C}_X)$ is P-unique, then a version of $E^{X=x}(Y|\mathscr{C}_X)$ is called a total-effect true-outcome variable of x. It is denoted by τ_x, i.e.,

$$\tau_x := E^{X=x}(Y|\mathscr{C}_X). \tag{4.8}$$

Remark 4.13 (A Caveat on the Notation) Hence, if x is a value of X, we use the notation τ_x for a version of $E^{X=x}(Y|\mathscr{C}_X)$ (see Def. 10.2 of SN). Of course, this shortcut notation is meaningful only if $E^{X=x}(Y|\mathscr{C}_X)$ is P-unique and the references to a specified outcome variable Y, a specified cause X, and the potential-confounder σ-algebra $\mathscr{C}_X$ are unambiguous. ◁

Remark 4.14 (Total-Effect True-Outcome Variable in Treatment x) Note that in Definition 4.12 we do *not* presume that X is a treatment variable in an experiment or quasi-experiment. If, however, X is a treatment variable, then we call τ_x a total-effect true-outcome variable of Y *of treatment* x. ◁

4.2.3 Atomic Total-Effect Variable

Now we consider two values $x, x' \in \Omega'_X$ of X with $P(X{=}x),\ P(X{=}x') > 0$, and the difference

$$E^{X=x}(Y|\mathscr{C}_X) - E^{X=x'}(Y|\mathscr{C}_X).$$

Remark 4.15 (Atomic Total-Effect Variable) If x and x' are two values of X and the conditional expectations $E^{X=x}(Y|\mathscr{C}_X)$ and $E^{X=x'}(Y|\mathscr{C}_X)$ are both P-unique, then the difference variable

$$\tau_x - \tau_{x'} = \delta_{xx'}$$

is also P-unique [see Box 14.1 (viii) of SN]. Hence, if $\delta_{xx'}$ and $\delta^*_{xx'}$ are different versions of $E^{X=x}(Y|\mathscr{C}_X) - E^{X=x'}(Y|\mathscr{C}_X)$, then $E(\delta_{xx'}) = E(\delta^*_{xx'})$ [see Box 6.1, (ix) of SN]. Furthermore, the values of $\delta_{xx'}$ are the atomic total effects of x vs. x'. ◁

Definition 4.16 (Atomic Total-Effect Variable)
Let the assumptions of Definition 4.12 be satisfied and let $x, x' \in \Omega'_X$ with $P(X{=}x), P(X{=}x') > 0$. If $E^{X=x}(Y|\mathscr{C}_X)$ and $E^{X=x'}(Y|\mathscr{C}_X)$ are P-unique, then

$$\delta_{xx'} := \tau_x - \tau_{x'} \tag{4.9}$$

is called an atomic total-effect variable of x vs. x'.

Remark 4.17 ($P^{Z=z}$-Uniqueness of the Atomic Total-Effect Variable) Suppose that $Z\colon (\Omega,\mathscr{A},P)\to(\Omega'_Z,\mathscr{A}'_Z)$ is another random variable on $(\Omega,\mathscr{A},P)$, that $z\in\Omega'_Z$ with $P(Z{=}z)>0$, and that $P^{Z=z}$ denotes the $(Z{=}z)$-conditional-probability measure on $(\Omega,\mathscr{A})$ [see Eq. (4.5)]. Then $(Z{=}z)$-conditional expectation values of different versions of the difference $E^{X=x}(Y|\mathscr{C}_X)-E^{X=x'}(Y|\mathscr{C}_X)$ are identical, provided that $E^{X=x}(Y|\mathscr{C}_X)$ as well as $E^{X=x'}(Y|\mathscr{C}_X)$ are $P^{Z=z}$-unique. Remember that the conditional expectation $E^{X=x}(Y|\mathscr{C}_X)$ is $P^{Z=z}$*-unique*, if

$$P^{Z=z}\big(\{\omega\in\Omega\colon V_x(\omega)\neq V^*_x(\omega)\}\big)\;=\;0,$$

whenever V_x and V^*_x are versions of $E^{X=x}(Y|\mathscr{C}_X)$. Hence, if $\delta_{xx'}$ and $\delta^*_{xx'}$ are different versions of $E^{X=x}(Y|\mathscr{C}_X)-E^{X=x'}(Y|\mathscr{C}_X)$, then $E^{Z=z}(\delta_{xx'})=E^{Z=z}(\delta^*_{xx'})=E(\delta_{xx'}\,|\,Z{=}z)=E(\delta^*_{xx'}\,|\,Z{=}z)$, provided that $E^{X=x}(Y|\mathscr{C}_X)$ as well as $E^{X=x'}(Y|\mathscr{C}_X)$ are $P^{Z=z}$-unique. ◁

Remark 4.18 (P-Uniqueness implies $P^{Z=z}$-Uniqueness) According to Corollary 5.22 of SN, P-uniqueness of $E^{X=x}(Y|\mathscr{C}_X)$ implies that $E^{X=x}(Y|\mathscr{C}_X)$ is $P^{Z=z}$-unique, provided, of course, that $P(Z{=}z)>0$. Note that $P^{Z=z}$-uniqueness of $E^{X=x}(Y|\mathscr{C}_X)$ does not imply that $E^{X=x}(Y|\mathscr{C}_X)$ is P-unique. ◁

Remark 4.19 (Expectation of True-Outcome Variables) In chapter 5, we define *average total effects* by the difference $E(\tau_x)-E(\tau_{x'})$ between the expectations of two such true-outcome variables.[2] Similarly, various kinds of *conditional total effects* are defined by the difference between the corresponding conditional expectation values of the total-effect true-outcome variables (see ch. 5). Each of these kinds of conditional total effects provides specific information that might be of interest in empirical causal research and evaluation studies.

◁

4.2.4 Direct-Effect True-Outcome Variables

Der unten stehende Text ist der nachfolgenden Def noch nicht angepasst (RST 22. Mai 2016).

Let $\langle(\Omega,\mathscr{A},P),(\mathscr{F}_t,t\in T),X,Y\rangle$ be a causality space and $t_M\in T$ with $t_X<t_M<t_Y$, where $t_M\in T$ is the index of the σ-algebra $\mathscr{F}_{t_M}$ in $(\mathscr{F}_t,t\in T)$ that is simultaneous to Y. Then the conditional expectations

$$\tau_{x,t_M}\;:=\;E^{X=x}(Y|\mathscr{F}_{t_M}),$$

[2] The expectation $E(\tau_x)$ corresponds to the term $E[Y|do(x)]$ in Pearl's and to $E(Y_x)$ in Rubin's terminologies (see, e. g., Pearl, 2009, p. 108 and Rubin, 2005, p. 323). Note, however, in contrast to Rubin we neither presume $P(U{=}u)=1/N$, where N represents the number of units in the population, nor do we assume that a value of τ_x is observed if the random experiment is actually run. Instead we observe a value of Y.

called $\mathscr{F}_{t_M}$*-direct-effect true-outcome variables*, are used to define true, average, and conditional *direct* and *indirect* effects. Furthermore, $E^{X=x}(Y|\mathscr{F}_{t_M})$ denotes the $\mathscr{F}_{t_M}$-conditional expectation of Y with respect to the conditional-probability measure $P^{X=x}$. Considering $E^{X=x}(Y|\mathscr{F}_{t_M})$, we condition on all variables that are measurable with respect to $\mathscr{F}_{t_M}$.

Note that talking about direct effects makes sense only if we refer to $\mathscr{F}_{t_M}$. Once we specified $t_M \in T$ and the filtration $(\mathscr{F}_t, t \in T)$, the σ-algebra $\mathscr{F}_{t_M}$ is specified as well. This is why we use the term $\mathscr{F}_{t_M}$*-direct effects*. Of course, if there is no ambiguity, we may also omit the explicit reference to $\mathscr{F}_{t_M}$.

Definition 4.20 (True Direct-Effect and True Indirect-Effect Variables) *Let $\langle(\Omega,\mathscr{A},P),(\mathscr{F}_t,t\in T),X,Y\rangle$ be a causality space, let Y be real-valued and nonnegative or with finite expectation $E(Y)$, let X be discrete, and x a value of X with $P(X=x)>0$, and let $t\in T$ with $t_X \le t < t_Y$.*

(i) Then $E^{X=x}(Y|\mathscr{C}_{X,t})$ is called a t-direct-effect true-outcome variable. It is denoted by $\tau_{x,t}$, i. e.,

$$\tau_{x,t} := E^{X=x}(Y|\mathscr{C}_{X,t}).$$

(ii) Furthermore,

$$\delta_{xx',t} := \tau_{x,t} - \tau_{x',t}$$

is called an atomic t-direct-effect variable of x vs. x'.

(iii) Finally

$$\delta_{xx'} - \delta_{xx',t}$$

is called an atomic t-indirect-effect variable of x vs. x'.

Remark 4.21 (Assumption of P-Uniqueness) Because $\mathscr{F}_{t_X} \subset \mathscr{F}_{t_M}$, P-uniqueness of $E^{X=x}(Y|\mathscr{F}_{t_M})$ implies that $E^{X=x}(Y|\mathscr{C}_X)$ is P-unique as well (see Corollary PR-**??**). This is the reason why, in the definition of the indirect effect variable, we only have to assume that $E^{X=x}(Y|\mathscr{F}_{t_M})$ and $E^{X=x'}(Y|\mathscr{F}_{t_M})$ are P-unique. According to Corollary PR-**??**, we can conclude that the differences δ_{xx',t_M} and $\delta_{xx'}-\delta_{xx',t_M}$ are also P-unique. ◁

Remark 4.22 (Meaning of True Indirect Effects) A true $\mathscr{F}_{t_M}$-indirect effect is simply that component of the true total effect of x vs. x' on Y that is *not direct* with respect to $\mathscr{F}_{t_M}$. ◁

Remark 4.23 (Conditioning on M and a Global Covariate of M) What has been said in Remark **??** also applies to $E^{X=x}(Y|\mathscr{F}_{t_M})$. Hence, if M denotes an intermediate variable of X and Y, and C_{X,t_M} is a global covariate of M, then conditioning on M, C_{X,t_M}, *and* X is equivalent to conditioning on M and C_{X,t_M} alone, *provided that we also condition on a fixed value x of X*, i. e.,

$$E^{X=x}(Y|\mathscr{F}_{t_M}) \underset{P^{X=x}}{=} E^{X=x}(Y|X,M,C_{X,t_M}) \underset{P^{X=x}}{=} E^{X=x}(Y|M,C_{X,t_M}). \tag{4.10}$$

◁

Again, because there are several global covariates of M, we may raise the question if there are, for a given value x of X, also several $\mathscr{F}_{t_M}$-direct effect true-outcome variables that are not P-equivalent. However, according to theorem **??**, if C_{X,t_M} and C^*_M denote two global covariates of M, then

$$E^{X=x}(Y|M,C_{X,t_M}) \underset{P^{X=x}}{=} E^{X=x}(Y|M,C^*_M). \tag{4.11}$$

Hence, if $E^{X=x}(Y|\mathscr{F}_{t_M})$ is P-unique, then the $\mathscr{F}_{t_M}$-direct-effect true-outcome variable τ_{x,t_M} does not depend on the choice of the particular global covariate C_{X,t_M}. According to Definition **??**, each global covariate of M generates, together with M, the same σ-algebra. In this case, $\sigma(M,C_{X,t_M}) = \mathscr{F}_{t_M}$. Hence, if $E^{X=x}(Y|\mathscr{F}_{t_M})$ is P-unique, then

$$\tau_{x,t_M} = E^{X=x}(Y|\mathscr{F}_{t_M}) \underset{P}{=} E^{X=x}(Y|M,C_{X,t_M}). \tag{4.12}$$

Note that the total-effect true-outcome variable can be written $\tau_x = E^{X=x}(Y|\mathscr{C}_X) \underset{P}{=} E^{X=x}(Y|C_X)$. However, unlike X, we can *not* omit M in the term $E^{X=x}(Y|M,C_{X,t_M})$ unless C_{X,t_M} is chosen such that M is C_{X,t_M}-measurable. The reason is that M is not necessarily constant if $X(\omega) = x$.

4.3 Numerical Examples

Now we treat some examples illustrating true-outcome variables. In the first two examples, the X-concurrent σ-algebra $\mathscr{F}_{t_X}$ is generated by X and the observational-unit variable U, i. e., $\mathscr{F}_{t_X} = \sigma(U,X)$. In these cases, U is also a global covariate of X. In the third example, we treat a two-factorial experiment exemplifying (a) that the X-concurrent σ-algebra can be generated by U and *two* treatment variables and (b) that there can be bias at the level of the observational units. The fourth example illustrates a direct-effect true-outcome variable.

4.3.1 Joe and Ann With Random Assignment

The random experiment considered is the same as in Example 3.3. Hence, the set Ω of possible outcomes, the σ-algebra $\mathscr{A}$ on Ω, and the filtration $(\mathscr{F}_t, t \in T)$ remain unchanged. However, this time, Joe and Ann are both assigned to treatment condition $X{=}1$ *with probability* $P(X{=}1\,|\,U{=}Joe) = P(X{=}1\,|\,U{=}Ann) = .40$, which means that now we have a different probability measure P on $(\Omega,\mathscr{A})$ than in Example 3.3.

Again, we consider the observational-unit variable U, the treatment variable X, as well as the outcome variable Y, and we assume $\mathscr{F}_{t_X} = \sigma(X,U)$. Note that although the rules of assigning the values of U and X to the elements $\omega \in \Omega$ remain unchanged, these random variables now have a different joint distribution as compared to Example 3.3.

In this new experiment, the prima facie effect $E(Y|X=1)-E(Y|X=0) = .50 - .45 = .05$ is the average of the two conditional prima facie effects $E(Y|X=1,U=Joe) - E(Y|X=0,U=Joe) = .10$ and $E(Y|X=1,U=Ann)-E(Y|X=0,U=Ann) = .00$. In more formal terms, the prima facie effect $E(Y|X=1)-E(Y|X=0)$ is exactly the expectation of the prima facie effect function $PFE_{10;U} = E^{X=1}(Y|U)-E^{X=0}(Y|U)$. In order to show this, let us specify the true-outcome variables $\tau_x = E^{X=x}(Y|\mathscr{C}_X) = E^{X=x}(Y|U)$ [see Eq. (**??**)] for the two treatment conditions 0 and 1.

The Total-Effect True-Outcome Variables

The *total-effect true-outcome variable* $\tau_0 = E^{X=0}(Y|U)$ of the control condition is defined by

$$\tau_0 := g_0(U), \tag{4.13}$$

where $U\colon \Omega \to \Omega_U$ with $U(\omega) = u$ for all $\omega \in \Omega$ and $g_0\colon \Omega_U \to \mathbb{R}$ with

$$g_0(u) := E^{X=0}(Y|U=u) = E(Y|X=0,U=u) \quad \text{for all } u \in \Omega_U. \tag{4.14}$$

Hence, in order to assign a value of τ_0 to an outcome $\omega \in \Omega$ of the random experiment, we first have to assign to ω a value u (*Joe* or *Ann*) of the observational-unit variable U, and then assign to u the corresponding conditional expected value $E(Y|X=0,U=u)$.

If, for instance, $\omega_3 = (Joe, yes, -)$ (see the third row in Table 4.2), then $U(\omega_3) = Joe$, and the value of τ_0 is

$$\tau_0(\omega_3) = g_0(Joe) = E(Y|X=0,U=Joe) = P(Y=1|X=0,U=Joe) = .70,$$

even though $X(\omega_3) = 1$ and the value of the regression $E(Y|X,U)$ is

$$E(Y|X,U)(\omega_3) = E(Y|X=1,U=Joe) = P(Y=1|X=1,U=Joe) = .80.$$

Hence, the true-outcome variable τ_0 of treatment 0 takes on a well-defined value for ω_3 *even though* the unit drawn receives treatment 1.

Because the outcome variable Y is dichotomous with values 0 and 1, the conditional expected value $E(Y|X=0,U=u)$ is also the conditional probability of success $P(Y=1|X=0,U=u)$, and because, in this example, U has only two values, *Joe* and *Ann*, the true-outcome variable τ_0 also has only two different values, the conditional probabilities $P(Y=1|X=0,U=Joe) = .70$ and $P(Y=1|X=0,U=Ann) = .20$.

Similarly, the *total-effect true-outcome variable* $\tau_1 = E^{X=1}(Y|U)$ of the treatment condition is specified by

$$\tau_1 := g_1(U), \tag{4.15}$$

where $g_1\colon \Omega_U \to \mathbb{R}$ is defined by

$$g_1(u) := E^{X=1}(Y|U=u) = E(Y|X=1,U=u) \quad \text{for all } u \in \Omega_U. \tag{4.16}$$

The random variable τ_1 takes on the two values

Table 4.2. Joe and Ann With Random Assignment to Treatment

Outcomes ω			Observables			Theoretical random variables						
Unit	Treatment	Success	Observational-unit variable U	Treatment variable X	Outcome variable Y	Regression $E(Y\|X,U)$	Regression $E(Y\|X)$	Conditional treatment probability $P(X=1\|U)$	Conditional regression $E^{X=0}(Y\|U) := \tau_0$	Conditional regression $E^{X=1}(Y\|U) := \tau_1$	True-total-effect variable $\delta_{10} := \tau_1 - \tau_0$	Probabilities of elementary events $P(\{\omega\})$
(*Joe*,	*no*,	−)	*Joe*	0	0	.70	.45	.40	.70	.80	.10	.09
(*Joe*,	*no*,	+)	*Joe*	0	1	.70	.45	.40	.70	.80	.10	.21
(*Joe*,	*yes*,	−)	*Joe*	1	0	.80	.50	.40	.70	.80	.10	.04
(*Joe*,	*yes*,	+)	*Joe*	1	1	.80	.50	.40	.70	.80	.10	.16
(*Ann*,	*no*,	−)	*Ann*	0	0	.20	.45	.40	.20	.20	.00	.24
(*Ann*,	*no*,	+)	*Ann*	0	1	.20	.45	.40	.20	.20	.00	.06
(*Ann*,	*yes*,	−)	*Ann*	1	0	.20	.50	.40	.20	.20	.00	.16
(*Ann*,	*yes*,	+)	*Ann*	1	1	.20	.50	.40	.20	.20	.00	.04

$$E(Y\,|\,X{=}1, U{=}\mathit{Joe}) \;=\; P(Y{=}1\,|\,X{=}1, U{=}\mathit{Joe}) \;=\; .80$$

and

$$E(Y\,|\,X{=}1, U{=}\mathit{Ann}) \;=\; P(Y{=}1\,|\,X{=}1, U{=}\mathit{Ann}) \;=\; .20.$$

Table 4.2 shows which values are assigned to each of the eight possible outcomes ω of the random experiment by τ_0 and τ_1.

The True-Total-Effect Variable

In this example, the total-effect true-outcome variables τ_0 and τ_1 and their values are uniquely defined for *all* outcomes $\omega \in \Omega$. Therefore, τ_1 and τ_0 are also P-unique and we can define the *atomic total effect variable*

$$\delta_{10} \;:=\; \tau_1 - \tau_0\,.$$

This is a random variable on $(\Omega, \mathscr{A}, P)$ that has the two different values, one value if Joe is drawn,

$$P(Y{=}1\,|\,X{=}1, U{=}\mathit{Joe}) - P(Y{=}1\,|\,X{=}0, U{=}\mathit{Joe}) \;=\; .80 - .70 \;=\; .10$$

and another value if Ann is drawn,

$$P(Y{=}1\,|\,X{=}1, U{=}\mathit{Ann}) - P(Y{=}1\,|\,X{=}0, U{=}\mathit{Ann}) \;=\; .20 - .20 \;=\; .00.$$

Hence, there is a function $g_{10}\colon \Omega_U \to \mathbb{R}$ such that

$$\delta_{10} \;=\; g_{10}(U). \tag{4.17}$$

The values of g_{10} are $g_{10}(\mathit{Joe}) = .10$ and $g_{10}(\mathit{Ann}) = .00$. These are the true total effects of the treatment compared to control given the two observational units *Joe* and *Ann*, respectively. Again, Table 4.2 shows which values the random variable δ_{10} assigns to each of the eight possible outcomes ω of the random experiment.

Note, however, that, in the definitions of the true-outcome variables τ_0, τ_1, and their difference $\delta_{10} = \tau_1 - \tau_0$, we only condition on the units u. What we ignore altogether are intermediate variables that might transmit the effect of the treatment to the outcome variable Y. This is why τ_x is called the *total-effect* true-outcome variable and δ_{10} the *atomic total effect* variable.

Average Total Effect

Taking the expectation of δ_{10} with respect to the distribution of U yields the *average total effect* of treatment compared to control:

$$\begin{aligned}
E(\delta_{10}) \;&=\; E(\tau_1) - E(\tau_0) \;=\; E_U(g_1) - E_U(g_0) \;=\; \sum_u g_1(u)\,P(U{=}u) - \sum_u g_0(u)\,P(U{=}u) \\
&= P(Y{=}1 \mid X{=}1, U{=}\mathit{Joe}) \cdot P(U{=}\mathit{Joe}) \;+\; P(Y{=}1 \mid X{=}1, U{=}\mathit{Ann}) \;\cdot P(U{=}\mathit{Ann}) \\
&\quad - \big[P(Y{=}1 \mid X{=}0, U{=}\mathit{Joe}) \cdot P(U{=}\mathit{Joe}) \;+\; P(Y{=}1 \mid X{=}0, U{=}\mathit{Ann}) \cdot P(U{=}\mathit{Ann})\big] \\
&= .80 \cdot .50 \;+\; .20 \cdot .50 \;-\; [.70 \cdot .50 \;+\; .20 \cdot .50] \;=\; .05,
\end{aligned}$$

using the transformation theorem (see Th. PR-6.13 and Remark PR-**??**).

Figure 4.1 illustrates various conditional probabilities of success. The points marked by the dashed line represent the probabilities of success for Joe given that he is treated and given that he is not treated, respectively. Similarly, the two points marked by the solid line indicate the probabilities of success for Ann given that she is treated and not treated, respectively. The points marked by the dotted line represent the conditional probabilities $P(Y{=}1 \mid X{=}1)$ and $P(Y{=}1 \mid X{=}0)$ of success given treatment and given control, respectively. The size of the area of the dotted circles is proportional to the probabilities of the four unit-treatment combinations $(X{=}x, U{=}u)$.

Note that not only X, Y, and U are random variables on the underlying probability space $(\Omega, \mathscr{A}, P)$, but also the regressions $E(Y \mid X, U)$ and $E(Y \mid X)$. Other random variables on $(\Omega, \mathscr{A}, P)$ are the total-effect true-outcome variables τ_0, τ_1, and the atomic total effect variable δ_{10}. This means that τ_0 and τ_1, for instance, have a joint distribution and may be correlated. While, in this example, the regression $E(Y \mid X, U)$ is a function of X and U, the random variables τ_0, τ_1, and δ_{10} are functions of U alone [see Eqs. (4.13) to (4.17)].

To emphasize, if we conduct the random experiment described in this example, then the total-effect true-outcome variable τ_0 of the control condition takes on a well-defined value *even if the unit drawn is treated*, i. e., even if $X(\omega) = 1$. Correspondingly, τ_1 takes on a well-defined value even if the unit drawn is *not* treated, i. e., even if $X(\omega) = 0$. In this example, the values $\tau_0(\omega)$ and $\tau_1(\omega)$ of both true-outcome variables solely depend on the unit drawn.

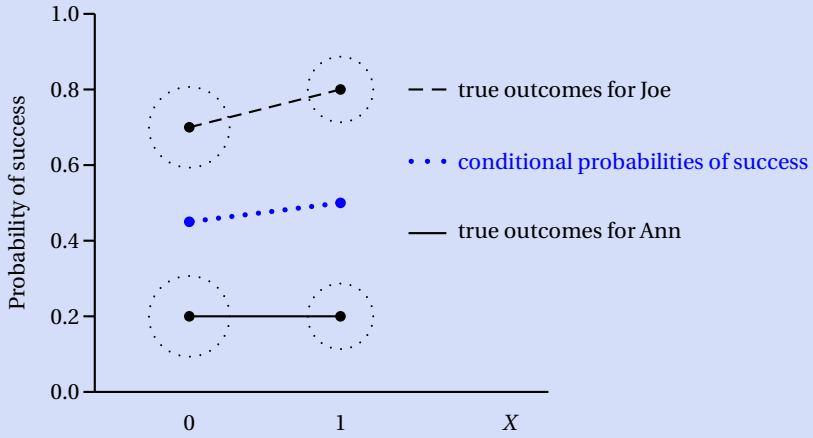


Figure 4.1. Conditional probabilities of success given treatment

4.3.2 No Treatment for Joe

Let us use a second example with Joe and Ann in order to illustrate that the true-outcome variables τ_x are not necessarily P-unique (see also ch. PR-14). The structure of the random experiment, i. e., the set Ω of possible outcomes, the σ-algebra $\mathscr{A}$ on Ω, the filtration $(\mathscr{F}_t, t \in T)$, the random variables U, X, Y, and the global covariate $C_X = U$ remain unchanged. However, in this new example, there are other probabilities of the elementary events (see Table 4.3). Now the probability that Joe receives treatment is zero, which implies that the total-effect true-outcome variable $\tau_1 = E^{X=1}(Y|U)$ is not P-unique; its values are not uniquely defined for $\omega \in \{U{=}Joe\} = \{\omega \in \Omega\colon U(\omega) = Joe\}$, although Joe has a positive probability to be drawn, i. e., although $P(\{U{=}Joe\}) > 0$. This also implies that the values of the atomic total effect variable $\delta_{10} = \tau_1 - \tau_0 = E^{X=1}(Y|U) - E^{X=0}(Y|U)$ are arbitrary for $\omega \in \{U{=}Joe\}$ as well. In contrast, the values of τ_1 and τ_0 are well-defined for $\omega \in \{U{=}Ann\}$. In other words, $\tau_0 = E^{X=0}(Y|U)$ and $\tau_1 = E^{X=1}(Y|U)$ are $P^{U=Ann}$-*unique*, i. e.,

$$P^{U=Ann}\big(\{\omega \in \Omega\colon V_x(\omega) \neq V_x'(\omega)\}\big) \;=\; 0,$$

whenever V_x and V_x' are two versions of $E^{X=0}(Y|U)$ or two versions of $E^{X=1}(Y|U)$.

Individual Treatment Regression

Of course, changing the probabilities of the elementary events also change the regressions $E(Y|X)$ and $E(Y|X,U)$ (see Exercises 4-14 and 4-15) as well as the conditional regressions $E^{X=0}(Y|U)$ (see Exercise 4-16). Two values of the regression $E(Y|X,U)$ are now arbitrary, namely its values for $\omega_3 = (Joe, yes, -)$ and $\omega_4 = (Joe, yes, +)$, because now $P(\{\omega_3\}) = P(\{\omega_4\}) = 0$. In order to make this vis-

Table 4.3. No Treatment for Joe

Outcomes ω			Observables			Regressions					
Unit	Treatment	Success	Observational-unit variable U	Treatment variable X	Outcome variable Y	Regression $E(Y\|X,U)$	Regression $E(Y\|X)$	Conditional probability $P(X=1\|U) = E(1_{X=1}\|U)$	Conditional regression $E^{X=0}(Y\|U) =: \tau_0$	Conditional regression $E^{X=1}(Y\|U) =: \tau_1$	Probabilities of elementary events $P(\{\omega\})$
(*Joe*,	*no*,	−)	*Joe*	0	0	.696	.60	0	.696	9	.152
(*Joe*,	*no*,	+)	*Joe*	0	1	.696	.60	0	.696	9	.348
(*Joe*,	*yes*,	−)	*Joe*	1	0	9	.40	0	.696	9	0
(*Joe*,	*yes*,	+)	*Joe*	1	1	9	.40	0	.696	9	0
(*Ann*,	*no*,	−)	*Ann*	0	0	.20	.60	.76	.20	.40	.096
(*Ann*,	*no*,	+)	*Ann*	0	1	.20	.60	.76	.20	.40	.024
(*Ann*,	*yes*,	−)	*Ann*	1	0	.40	.40	.76	.20	.40	.228
(*Ann*,	*yes*,	+)	*Ann*	1	1	.40	.40	.76	.20	.40	.152

ible, we assigned the value 9 of $E(Y|X,U)$ to each of these two possible outcomes. Because these two values are arbitrary, there are now different versions of $E(Y|X,U)$. A second version is obtained, e. g., if we replace 9 by 0 or by any other real number. If V and V^* are two such versions of $E(Y|X,U)$, then they are almost surely equivalent with respect to the measure P. In other words,

$$P\big(\{\omega\in\Omega\colon V(\omega)\neq V^*(\omega)\}\big) \;=\; 0.$$

Hence, the regression $E(Y|X,U)$ is not unique any more. Instead it is uniquely defined up to almost sure equivalence with respect to the probability measure P, i. e., it is P-unique.

(X=x)-Conditional Regressions

In contrast, the conditional regression $E^{X=1}(Y|U)$ is not P-unique, although it *is* $P^{X=1}$-unique. Hence, if V_1 and V_1^* are two different versions of $E^{X=1}(Y|U)$, and

$$\begin{aligned} A_1 \;&:=\; \big\{\omega\in\Omega\colon V_1(\omega)\neq V_1^*(\omega)\big\} \\ &=\; \big\{(\textit{Joe}, \textit{no}, -), (\textit{Joe}, \textit{no}, +), (\textit{Joe}, \textit{yes}, -), (\textit{Joe}, \textit{yes}, +)\big\}. \end{aligned}$$

then

$$P^{X=1}(A_1) \;=\; P(A_1\,|\,X=1) \;=\; \sum_{\omega\in A_1} P^{X=1}(\{\omega\}) \;=\; 0+0+0+0 \;=\; 0.$$

In contrast,

$$P(A_1) \;=\; \sum_{\omega \in A_1} P(\{\omega\}) \;=\; .152 + .348 + 0 + 0 \;=\; .50$$

(see the last column of Table 4.3). Hence, if V_1 and V_1^* are two different versions of the conditional regression $E^{X=1}(Y \mid U)$, then these versions are not almost surely equivalent with respect to the measure P.

The values of the conditional regression $E^{X=1}(Y \mid U)$ with respect to the measure $P^{X=1}$ are the conditional expectation values $E^{X=1}(Y \mid U{=}\mathit{Joe})$ and $E^{X=1}(Y \mid U{=}\mathit{Ann})$. Because $E^{X=1}(Y \mid U{=}u) = P^{X=1}(Y{=}1 \mid U{=}u)$,

$$P^{X=1}(Y{=}1 \mid U{=}\mathit{Joe}) \;=\; \frac{P^{X=1}(Y{=}1, U{=}\mathit{Joe})}{P^{X=1}(U{=}\mathit{Joe})},$$

and $P^{X=1}(U{=}\mathit{Joe}) = 0$, the conditional expected value $E^{X=1}(Y \mid U{=}u)$ is undefined and $E^{X=1}(Y \mid U)(\omega)$ can be any number if $\omega \in \{U{=}\mathit{Joe}\}$. In contrast,

$$P^{X=1}(Y{=}1 \mid U{=}\mathit{Ann}) \;=\; \frac{P^{X=1}(Y{=}1, U{=}\mathit{Ann})}{P^{X=1}(U{=}\mathit{Ann})} \;=\; \frac{.152\,/\,.38}{(.152 + .228)\,/\,.38} \;=\; .40,$$

and $E^{X=1}(Y \mid U)(\omega) = .40$ if $\omega \in \{U{=}\mathit{Ann}\}$. Hence, in this example we have shown that the true-outcome variable $\tau_1 = E^{X=1}(Y \mid U)$ is not P-unique.

4.3.3 Jim and Jane

Now we treat an example in which there are two treatment variables. Such a two-factorial experiment has already been discussed at an informal level in section 2.3. Here we use this example to exemplify (a) that the global covariate can be a bivariate random variable consisting of two univariate random variables and (b) that there can be bias at the level of the observational units.

The Random Experiment

The parameters displayed in Table 4.4 refer to an experiment in which a person is drawn from a set of two persons, each of which may or may not receive a first treatment, say *group therapy* (represented by Z), and, at the same time, receive or not receive a second treatment, say *individual therapy* (represented by X). Furthermore, both treatments affect the outcome variable Y (e. g., *well-being, life satisfaction*, or a score on a *symptom checklist*).

The set of possible outcomes of the random experiment is

$$\Omega \;:=\; \Omega_U \times \Omega_Z \times \Omega_X \times \Omega_Y,$$

where $\Omega_U := \{\mathit{Jim}, \mathit{Jane}\}$,

$$\Omega_Z := \{\textit{group therapy}, \textit{no group therapy}\},$$

$$\Omega_X := \{\textit{individual therapy}, \textit{no individual therapy}\},$$

and Ω_Y is the set of possible observations based on which the score of the outcome variable Y is computed. If Y is discrete, then we may choose $\mathscr{A} = \mathscr{P}(\Omega)$, where $\mathscr{P}(\Omega)$ denotes the power set of Ω. However, if Y is continuous, then $\mathscr{A}$ is the product σ-algebra $\mathscr{A} = \mathscr{P}(\Omega_U) \otimes \mathscr{P}(\Omega_Z) \otimes \mathscr{P}(\Omega_X) \otimes \mathscr{B}$, where $\mathscr{B}$ denotes the Borel σ-algebra on $\mathbb{R}$ (see section PR-1.2.3).

In this example, the *treatment probabilities* are as follows: Jim receives group therapy (Z=1) with probability $P(Z=1\,|\,U=Jim) = 1/2$, or no group therapy (Z=0) with probability $P(Z=0|U=Jim) = 1/2$. Furthermore, Jim receives individual therapy (X=1) with probability $P(X=1\,|\,U=Jim, Z=0) = 3/4$, if he does not receive group therapy (Z=0), and he receives individual therapy (X=1) with probability $P(X=1\,|\,U=Jim, Z=1) = 1/4$ if he receives group therapy (Z=1) as well. Similarly, Jane receives individual therapy (X=1) with probability $P(X=1\,|\,U=Jane, Z=0) = 3/4$, if she does not receive group therapy (Z=0), and she receives individual therapy (X=1) with probability $P(X=1\,|\,U=Jane, Z=1) = 1/4$ if she also receives group therapy (Z=1).

Various Total Causal Effects

On the *individual level*, there are *several* total causal effects we might look at. In principle, we might be interested in

- (a_1) the individual total effect of group therapy (Z=1) [compared to no group therapy (Z=0)] on Y given that Jim (Jane) also receives individual therapy (X=1),
- (b_1) the individual total effect of group therapy (Z=1) [compared to no group therapy (Z=0)] on Y given that Jim (Jane) does *not* receive individual therapy (X=0), and
- (c_1) the average of these individual total effects, averaging over the two values of X (individual therapy).

Of course, the latter effect is certainly less informative than the two conditional effects. Similarly, we may also be interested in

- (a_2) the individual total effect of individual therapy (X=1) [compared to no individual therapy (X=0)] on Y given that Jim (Jane) also receives group therapy (Z=1),
- (b_2) the individual total effect of individual therapy (X=1) [compared to no individual therapy (X=0)] on Y given that Jim (Jane) does *not* receive group therapy (Z=0), and
- (c_2) the average of these individual total effects of individual therapy (X=1), averaging over the two values of Z (group therapy).

Looking at the effects (a_1) to (c_1), we consider *individual therapy* a (qualitative) covariate and *group therapy* to be the treatment variable asking for the conditional effects of group therapy given individual therapy and their average, the 'main effect' of group therapy. In contrast, looking at the effects (a_2) to (c_2), we consider *group therapy* to be a (qualitative) covariate and *individual therapy* to

Table 4.4. Jim and Jane: An Example With Bias at the Individual Level

		Fundamental parameters					Derived parameters		
Observational-unit u	Sampling probability $P(U=u)$	Covariate (second treatment variable) *group therapy* Z	$P(Z=z\,\|\,U=u)$	Conditional expectations of outcomes under control $E(Y\,\|\,X=0,Z=z,U=u)$	Conditional expectations of outcomes under treatment $E(Y\,\|\,X=1,Z=z,U=u)$	Treatment probability $P(X=1\,\|\,U=u,Z=z)$	$E(Y\,\|\,X=1,Z=z,U=u)$ – $E(Y\,\|\,X=0,Z=z,U=u)$	$P(Z=z\,\|\,X=0,U=u)$	$P(Z=z\,\|\,X=1,U=u)$
Jim	1/2	0	1/2	68	82	3/4	14	1/4	3/4
		1	1/2	96	100	1/4	4	3/4	1/4
Jane	1/2	0	1/2	80	98	3/4	18	1/4	3/4
		1	1/2	104	106	1/4	2	3/4	1/4

be the treatment variable. In the example presented in Table 4.4, we made the second choice. There, we denoted *group therapy* by Z and *individual therapy* by X. Note that this choice is arbitrary and would be changed if we would ask for the effects of *group therapy* given *individual therapy*.

The Filtration and the Global Covariate

In this example, the filtration $(\mathscr{F}_t, t \in T)$ consists of three σ-algebras: $\mathscr{F}_1 := \sigma(U)$, $\mathscr{F}_2 := \sigma(U,Z,X)$, and $\mathscr{F}_3 := \sigma(U,Z,X,Y)$. Hence, U is pre-ordered to Z, X, and Y. Furthermore, Z and X are equi-ordered, and Z and X are both pre-ordered to Y. Both treatment variables, X and Z, also play the role of covariates, *depending on which treatment effects we are studying*. Note that we also consider *qualitative* variables as covariates, deviating from terminology traditionally used the analysis of covariance, where 'covariate' usually refers to a quantitative variable.

In this example, $t_X = t_2$ and $C_X = (U,Z)$ is a global covariate of X. This implies $\sigma(U,Z,X) = \mathscr{F}_{t_X}$ and from Equation (**??**) we can conclude that there is no other covariate that determines the expectation of Y in the treatment conditions over and above the observational-unit variable U and the second treatment variable Z. Hence, the total-effect true-outcome variables are

$$\tau_0 \;=\; E^{X=0}(Y\,|\,\mathscr{F}_{t_X}) \;=\; E^{X=0}(Y\,|\,U,Z)$$

and

$$\tau_1 \;=\; E^{X=1}(Y\,|\,\mathscr{F}_{t_X}) \;=\; E^{X=1}(Y\,|\,U,Z).$$

Therefore, Jim's true outcome under treatment ($X=1$) if he does not receive group therapy ($Z=0$) is $E(Y|X=1,Z=0,U=Jim) = 82$, and it is $E(Y|X=1,Z=1,U=Jim) = 100$ if he does ($Z=1$). Furthermore, Jim's true outcome under control ($X=0$) if he does not receive group therapy is equal to $E(Y|X=0,Z=0,U=Jim) = 68$, and it is $E(Y|X=0,Z=1,U=Jim) = 96$ if he does.

Now consider the atomic total effect variable $\delta_{10} := \tau_1 - \tau_0$, which, in this example, is a function of the observational-unit variable U and the second treatment variable Z. Therefore, Jim's effect of the individual therapy is 14 if he does not receive group therapy ($Z=0$), whereas it is 4, if he does ($Z=1$). Similarly, Jane's true effect of the individual therapy is 18 if she does not receive group therapy ($Z=0$), whereas it is 2, if she does ($Z=1$) (see Table 4.4). Hence, all true effects are positive.

Computing Individual Effects

In this example, the conditional expected values of Jim, $E(Y|X=1,U=Jim)$ and $E(Y|X=0,U=Jim)$, as well as the corresponding two conditional expected values of Jane, $E(Y|X=1,U=Jane)$ and $E(Y|X=0,U=Jane)$, are strongly biased.[3] The same holds true for the differences $E(Y|X=1,U=Jim) - E(Y|X=0,U=Jim)$ and $E(Y|X=1,U=Jane) - E(Y|X=0,U=Jane)$.

In order to compute the conditional expected values given treatment and unit, we use the equation

$$E(Y|X=x,U=u) = \sum_z E(Y|X=x,U=u,Z=z) \cdot P(Z=z|X=x,U=u), \qquad (4.18)$$

which is always true if $P(X=x,U=u) > 0$ and Z is discrete with $P(X=x,U=u,Z=z) > 0$ for all values of Z [see Rule (ii) in Box PR-9.2, p. PR-286]. Both kinds of parameters occurring on the right-hand side of this equation are displayed in Table 4.4 (p. 88). The conditional expectations $E(Y|X=x,U=u,Z=z)$ are among the fundamental parameters, whereas the probabilities $P(Z=z|X=x,U=u)$ have been computed from the fundamental parameters via:

$$P(Z=z|X=x,U=u) = \frac{P(X=x|U=u,Z=z) \cdot P(Z=z|U=u)}{\sum_z P(X=x|U=u,Z=z) \cdot P(Z=z|U=u)} \qquad (4.19)$$

(see Exercises 4-6 and 4-7).

Hence, using Equation (4.18), the $(X=x)$-conditional expected values for Jim are

$$E(Y|X=1,U=Jim) = 82 \cdot \frac{3}{4} + 100 \cdot \frac{1}{4} = 86.5,$$

$$E(Y|X=0,U=Jim) = 68 \cdot \frac{1}{4} + 96 \cdot \frac{3}{4} = 89,$$

[3] Note that, unlike the bias concept of statistics that refers to estimators, our concept of bias applies to conditional expected values and their differences. Chapter 6 is entirely devoted to the concepts of *bias with respect to total effects* and *bias with respect to direct effects*.

and his individual prima facie effect is

$$E(Y \mid X{=}1, U{=}\mathit{Jim}) - E(Y \mid X{=}0, U{=}\mathit{Jim}) \;=\; 86.5 - 89 \;=\; -2.5\,.$$

This shows that, in this example, the individual prima facie effect of *individual therapy* compared to *no individual therapy* is seriously biased yielding the negative effect −2.5 although both $(Z{=}z)$-conditional individual effects are positive, namely 14 for $Z{=}0$ (*no group therapy*) and 4 for $Z{=}1$ (*group therapy*). The reason for this bias is that, in this example, the covariate is correlated both, with the treatment and the outcome variables even conditioning on the observational-unit variable U. Hence, this example shows that there is bias even *on the individual level*, i. e., even if we condition on the observational-unit variable U.

Example 4.24 (Joe and Ann With Random Assignment – continued) Table 4.5 extends the example 'Joe and Ann With Random Assignment' by including an intermediate variable M. For simplicity, we again assume that this intermediate variable has only two values, 0 and 1, indicating if *motivation after treatment* is low or high, respectively. The two treatment conditions might be two methods of instruction or two therapy methods, and Y is again some criterion of success.

??? Vielleicht doch t_M-direct ???

The example is constructed such that omitting the intermediate variable yields exactly Table 4.2 (p. 82). Hence, the regressions $E(Y|X)$, $E(Y|X,U)$, the total-effect true-outcome variables, τ_0 and τ_1, and their difference, the atomic total effect variable δ_{10}, have already been treated in detail. Therefore, we can now concentrate on the true-outcome variables $\tau_{0,\,t_M}$ and $\tau_{1,\,t_M}$, their difference, the true-$\mathscr{F}_{t_M}$-direct-effect variable $\delta_{10,\,t_M}$, and the true-$\mathscr{F}_{t_M}$-indirect-effect variable $\delta_{10} - \delta_{10,\,t_M}$.

In this example, the filtration $(\mathscr{F}_t, t \in T)$ can be specified as follows: $\mathscr{F}_1 = \sigma(U)$, $\mathscr{F}_2 = \sigma(U,X)$, $\mathscr{F}_3 = \sigma(U,X,M)$, and $\mathscr{F}_4 = \sigma(U,X,M,Y)$. Furthermore, $C_X = U$ and $C_{X,t_M} = (U,X)$ are global covariates of X and of M, respectively. Finally, $\mathscr{F}_{t_X} = \mathscr{F}_2$ is the X-concurrent σ-algebra and $\mathscr{F}_{t_M} = \mathscr{F}_3$ is the M-concurrent σ-algebra. Because,

$$E^{X=0}(Y \mid M, U) \underset{P^{X=0}}{=} E^{X=0}(Y \mid X, M, U),$$

the true-outcome variable $\tau_{0,\,t_M}$ is the conditional regression $E^{X=0}(Y \mid M, U)$, and its value for $\omega_1 = (\mathit{Joe}, \mathit{no}, \mathit{lo}, -)$ is

$$\begin{aligned} E^{X=0}(Y \mid M{=}0, U{=}\mathit{Joe}) \;&=\; P^{X=0}(Y{=}1 \mid M{=}0, U{=}\mathit{Joe}) \;=\; \frac{P(Y{=}1, X{=}0, M{=}0, U{=}\mathit{Joe})}{P(X{=}0, M{=}0, U{=}\mathit{Joe})} \\ &=\; \frac{.042}{.027 + .042} \;\approx\; .609. \end{aligned}$$

The other values of $\tau_{0,\,t_M}$ can be computed correspondingly from the joint probabilities of all four variables U, X, M, and Y, and the same is true for the values of $\tau_{1,\,t_M}$. The results are displayed in Table 4.5.

Once, these values have been computed, the difference between $\tau_{1,\,t_M}$ and $\tau_{0,\,t_M}$ yields the true-$\mathscr{F}_{t_M}$-direct-effect variable $\delta_{10,\,t_M}$. Hence, $\delta_{10,\,t_M}(\omega_1) = \tau_{1,\,t_M}(\omega_1) -$

Table 4.5. An Example With an Intermediate Variable

Outcomes ω				Observables				Theoretical random variables										
Unit	Treatment	Intermediate	Outcome	Observational-unit variable U	Treatment variable X	Intermediate variable M	Outcome variable Y	Regression $E(Y\|X,M)$	Regression $E(Y\|X,M,U)$	Regression $E(Y\|X,U)$	Regression $E(Y\|X)$	Conditional regression $E^{X=0}(Y\|U)=\tau_0$	Conditional regression $E^{X=1}(Y\|U)=\tau_1$	True-total-effect variable $\delta_{10}=\tau_1-\tau_0$	Conditional regression $E^{X=0}(Y\|M,U)=\tau_{0,t_M}$	Conditional regression $E^{X=1}(Y\|M,U)=\tau_{1,t_M}$	True-direct-effect variable $\delta_{0,t_M}=\tau_{1,t_M}-\tau_{0,t_M}$	True-indirect-effect variable $\delta_{10}-\delta_{0,t_M}$	Probabilities of elementary events $P(\{\omega\})$
(*Joe*,	*no*,	lo,	+)	*Joe*	0	0	0	.286	.609	.70	.45	.70	.80	.10	.609	.667	.058	.042	.027
(*Joe*,	*no*,	lo,	−)	*Joe*	0	0	1	.286	.609	.70	.45	.70	.80	.10	.609	.667	.058	.042	.042
(*Joe*,	*no*,	hi,	+)	*Joe*	0	1	0	.587	.727	.70	.45	.70	.80	.10	.727	.818	.091	.009	.063
(*Joe*,	*no*,	hi,	−)	*Joe*	0	1	1	.587	.727	.70	.45	.70	.80	.10	.727	.818	.091	.009	.168
(*Joe*,	*yes*,	lo,	+)	*Joe*	1	0	0	.265	.667	.80	.50	.70	.80	.10	.609	.667	.058	.042	.008
(*Joe*,	*yes*,	lo,	−)	*Joe*	1	0	1	.265	.667	.80	.50	.70	.80	.10	.609	.667	.058	.042	.016
(*Joe*,	*yes*,	hi,	+)	*Joe*	1	1	0	.621	.818	.80	.50	.70	.80	.10	.727	.818	.091	.009	.032
(*Joe*,	*yes*,	hi,	−)	*Joe*	1	1	1	.621	.818	.80	.50	.70	.80	.10	.727	.818	.091	.009	.144
(*Ann*,	*no*,	lo,	+)	*Ann*	0	0	0	.286	.176	.20	.45	.20	.20	.00	.176	.179	.003	−.003	.168
(*Ann*,	*no*,	lo,	−)	*Ann*	0	0	1	.286	.176	.20	.45	.20	.20	.00	.176	.179	.003	−.003	.036
(*Ann*,	*no*,	hi,	+)	*Ann*	0	1	0	.587	.250	.20	.45	.20	.20	.00	.250	.227	−.023	.023	.072
(*Ann*,	*no*,	hi,	−)	*Ann*	0	1	1	.587	.250	.20	.45	.20	.20	.00	.250	.227	−.023	.023	.024
(*Ann*,	*yes*,	lo,	+)	*Ann*	1	0	0	.265	.179	.20	.50	.20	.20	.00	.176	.179	.003	−.003	.092
(*Ann*,	*yes*,	lo,	−)	*Ann*	1	0	1	.265	.179	.20	.50	.20	.20	.00	.176	.179	.003	−.003	.020
(*Ann*,	*yes*,	hi,	+)	*Ann*	1	1	0	.621	.227	.20	.50	.20	.20	.00	.250	.227	−.023	.023	.068
(*Ann*,	*yes*,	hi,	−)	*Ann*	1	1	1	.621	.227	.20	.50	.20	.20	.00	.250	.227	−.023	.023	.020

$\tau_{0,t_M}(\omega_1) = .667 - .609 = .058$, for instance, where $\omega_1 = (Joe, no, lo, -)$. Similarly, the value of the true-$\mathscr{F}_{t_M}$-indirect-effect variable $\delta_{10} - \delta_{10,t_M}$ for ω_1 is obtained by $\delta_{10}(\omega_1) - \delta_{10,t_M}(\omega_1) = .10 - .058 = .042$. Also note that the true-$\mathscr{F}_{t_M}$-direct-effect variable δ_{10,t_M} and the true-$\mathscr{F}_{t_M}$-indirect-effect variable $\delta_{10} - \delta_{10,t_M}$ always add up to the atomic total effect variable δ_{10}.

◁

4.4 Summary and Conclusions

In this chapter, we introduced two kinds of true-outcome variables τ_x and τ_{x,t_M}, the first *with respect to total effects* and the second *with respect to $\mathscr{F}_{t_M}$-direct effects*. Because we condition on the global covariate C_X and C_{X,t_M}, respectively, both are defined such that they are purged from bias.

True Total Effects

If τ_x and $\tau_{x'}$ are the total-effect true-outcome variables of x and x', then the atomic total effect variable $\delta_{xx'}$ is the difference between τ_x and $\tau_{x'}$. The true-outcome variable τ_x has been defined such that its values are the expectations of the outcome variable Y holding constant the global covariate. This implies that we condition on *all* covariates, ignoring intermediate variables. By their definition, the atomic total effect variables $\delta_{xx'}$ and the true-outcome variables τ_x are purged from bias. However, ignoring intermediate variables implies that the values of the variables $\delta_{xx'}$ are 'only' the true *total* effects.

True-Direct and True-Indirect-Effect Variables

While the values of $\delta_{xx'}$ are the true total effects, the values of the difference $\delta_{xx',t_M} := \tau_{x,t_M} - \tau_{x',t_M}$ are the true *$\mathscr{F}_{t_M}$-direct* effects of x compared to x'. By definition, τ_{x,t_M} and τ_{x',t_M} are unbiased as well. However, they represent the effects of x vs. x' that are *not transmitted* through the intermediate variable M. Therefore, the values of the difference variable $\delta_{xx'} - \delta_{xx',t_M}$ are the true *indirect* effects of x vs. x' that might be transmitted through M. Note, however, that defining the true-indirect-effect variable $\delta_{xx'} - \delta_{xx',t_M}$, we do not need a true-outcome variable or an atomic effectvariable for M. The true-indirect-effect variable $\delta_{xx'} - \delta_{xx',t_M}$ is simply the difference between the atomic total effect variable and the atomic direct effect variable.

Note that true-outcome variables do not only apply to treatment variables. Instead, they are also defined for other variables X taking the role of a cause. However, a limitation of the true-outcome variables is that they can only be defined for values x of X that have a positive probability.

Box 4.1 Glossary of New Concepts

Covariate of X	A random variable on $(\Omega, \mathscr{A}, P)$ that is measurable w.r.t. $\mathscr{F}_{t_X}$.
Intermediate variable M of X and Y	A random variable on $(\Omega, \mathscr{A}, P)$ such that X is prior to M which itself is prior to the outcome variable Y.
τ_x	*Short notation of the* $(X{=}x)$*-conditional regression* $E^{X=x}(Y\|\mathscr{C}_X)$. If $E^{X=x}(Y\|\mathscr{C}_X)$ is P-unique, then τ_x is also called the *total-effect true-outcome variable* pertaining to a value x of X. Because C_X denotes a global covariate, a true-outcome variable is constructed such that it is unbiased.
$\delta_{xx'}$	*True-total-effect variable.* Presuming that τ_x and $\tau_{x'}$ are P-unique, it is defined by $\delta_{xx'} := \tau_x - \tau_{x'}$. Its values are the true *total* effects of x vs. x' given the values of the global covariate C_X.
τ_{x,t_M}	*Short notation of the* $(X{=}x)$*-conditional regression* $E^{X=x}(Y\|M, C_X)$. If $E^{X=x}(Y\|M, C_X)$ is P-unique, then τ_{x,t_M} is also called the $\mathscr{F}_{t_M}$*-direct effect true-outcome variable* pertaining to a value x of X. If X is a treatment variable, then τ_{x,t_M} is the regression of Y on M and C_X in treatment x.
δ_{xx',t_M}	*True-*$\mathscr{F}_{t_M}$*-direct-effect variable.* Presuming that $E^{X=x}(Y\|M, C_X)$ and $E^{X=x'}(Y\|M, C_X)$ are P-unique, it is defined to be the difference $\delta_{xx',t_M} := \tau_{x,t_M} - \tau_{x',t_M}$. Its values are the true $\mathscr{F}_{t_M}$*-direct* effects of x vs. x'. These direct effects of x vs. x' on Y are *not* transmitted through the intermediate variable M.
$\delta_{xx'} - \delta_{xx',t_M}$	*True-indirect-effect variable w.r.t. M.* Its values are the true *indirect* effects of x vs. x', provided that $E^{X=x}(Y\|M, C_X)$ and $E^{X=x'}(Y\|M, C_X)$ are P-unique. These indirect effects of x vs. x' on Y might be transmitted through the intermediate variable M. A causal dependence of Y on M is not presumed.

Outlook

In chapter 5, we use the true-outcome variables with respect to total and with respect to direct effects to define various kinds of average and conditional total, direct, and indirect effects, presuming that X is discrete. In chapters 6 to 9, we study various causality conditions, i. e., conditions under which we can identify these causal effects from empirically estimable parameters.

While the theory of causal effects based on the concepts of true-outcome and effect variables just confirms what is already known about randomized experiments, it leads to designs and analysis techniques in quasi-experiments (see ch. 13) that were unknown before the first approaches to causal effects were developed in the last century. Furthermore, it provides the theoretical foundation for analyzing causal effects also beyond quasi-experiments. Last but not least it

should be emphasized that the theory of causal effects is also essential when it comes to analyzing *direct* and *indirect* effects. Although randomization guarantees that the treatment regression $E(Y|X)$ is unbiased, randomization does neither guarantee unbiasedness of the regression $E(Y|X,M)$ nor of $E(Y|X,M,Z)$, where Z denotes a covariate and M an intermediate variable (see Example 3 in ch. 1).

4.5 Proofs

Proof of Theorem 3.12

(i) If $\mathscr{C}$ is prior to $\mathscr{D}$ with respect to $(\mathscr{F}_t, t \in T)$, then there is an $s \in T$ such that $\mathscr{C} \subset \mathscr{F}_s$, $\mathscr{D} \not\subset \mathscr{F}_s$. If we assume that $\mathscr{D}$ is prior to $\mathscr{C}$, then there is an $r \in T$ such that $\mathscr{D} \subset \mathscr{F}_r$, $\mathscr{C} \not\subset \mathscr{F}_r$. Because $(\mathscr{F}_t, t \in T)$ is a filtration and $\mathscr{D} \not\subset \mathscr{F}_s$, $\mathscr{D} \subset \mathscr{F}_r$, we can conclude $s < r$. Similarly, $\mathscr{C} \not\subset \mathscr{F}_r$, $\mathscr{C} \subset \mathscr{F}_s$ implies $r < s$. This is a contradiction to our assumption.

(ii) If $\mathscr{C}$ is prior to $\mathscr{D}$, then

(a) there is an $r \in T$ with $\mathscr{C} \subset \mathscr{F}_r$ and $\mathscr{D} \not\subset \mathscr{F}_r$,

and if $\mathscr{D}$ is also prior to $\mathscr{E}$, then

(b) there is a $s \in T,\ r < s$, with $\mathscr{D} \subset \mathscr{F}_s$ and $\mathscr{E} \not\subset \mathscr{F}_s$, and

(c) there is a $t \in T,\ s < t$, with $\mathscr{E} \subset \mathscr{F}_t$.

(a) to (c) imply that there is an $r \in T$ with $\mathscr{C} \subset \mathscr{F}_r$, $\mathscr{E} \not\subset \mathscr{F}_r$, and there is a $t \in T,\ r < t$, with $\mathscr{E} \subset \mathscr{F}_t$.

Proof of Theorem 3.18

(i) If $\mathscr{C}$ is prior to $\mathscr{D}$ with respect to $(\mathscr{F}_t, t \in T)$, then:

(a) there is an $s \in T$ with $\mathscr{C} \subset \mathscr{F}_s$ and $\mathscr{D} \not\subset \mathscr{F}_s$,

(b) there is a $t \in T,\ s < t$, with $\mathscr{D} \subset \mathscr{F}_t$.

Because $(\mathscr{F}_t, t \in T)$ is a filtration, it follows that

(c) $\mathscr{C} \cup \mathscr{D}$, $\sigma(\mathscr{C} \cup \mathscr{D}) \subset \mathscr{F}_t$, and

(d) $\mathscr{C} \cup \mathscr{D}$, $\sigma(\mathscr{C} \cup \mathscr{D}) \not\subset \mathscr{F}_s$.

Now, (a), (c), and (d) imply proposition (i).

(ii) If $\mathscr{C}$ and $\mathscr{D}$ are prior to $\mathscr{E}$ with respect to $(\mathscr{F}_t, t \in T)$, then:

(a) there is an $r \in T$ with $\mathscr{C} \subset \mathscr{F}_r$ and $\mathscr{E} \not\subset \mathscr{F}_r$,

(b) there is an $s \in T$ with $\mathscr{D} \subset \mathscr{F}_s$ and $\mathscr{E} \not\subset \mathscr{F}_s$,

(c) there is a $t \in T,\ r, s < t$, with $\mathscr{E} \subset \mathscr{F}_t$.

Without loss of generality, we can assume $r \leq s$, which implies $\mathscr{C}, \mathscr{D} \subset \mathscr{F}_s$, because $(\mathscr{F}_t, t \in T)$ is a filtration. However, if $\mathscr{C}, \mathscr{D} \subset \mathscr{F}_s$, then:

(d) $\mathscr{C} \cup \mathscr{D}$, $\sigma(\mathscr{C} \cup \mathscr{D}) \subset \mathscr{F}_s$.

Now (b), (c), and (d) imply proposition (ii).

(iii) If $\mathscr{C}$ is prior to both $\mathscr{D}$ and $\mathscr{E}$ with respect to $(\mathscr{F}_t, t \in T)$, then

(a) $\exists r \in T$ such that $\mathscr{C} \subset \mathscr{F}_r$ and $\mathscr{D} \not\subset \mathscr{F}_r$,

(b) $\exists s \in T$ such that $\mathscr{D} \subset \mathscr{F}_s$, $\mathscr{E} \not\subset \mathscr{F}_s$

(c) $\exists t \in T$, $t > r, s$ such that $\mathscr{D}, \mathscr{E} \subset \mathscr{F}_t$, because $(\mathscr{F}_t, t \in T)$ is a filtration.

Without loss of generality can assume $r \leq s$. Because $\mathscr{E} \subset \mathscr{F}_s$ implies $\mathscr{E} \subset \mathscr{F}_r$, and therefore

(d) $\mathscr{D} \cup \mathscr{E}$, $\sigma(\mathscr{D} \cup \mathscr{E}) \subset \mathscr{F}_t \not\subset \mathscr{F}_r$,

proposition (d) implies

(e) $\exists t > r$ such that $\mathscr{D} \cup \mathscr{E}$, $\sigma(\mathscr{D} \cup \mathscr{E}) \subset \mathscr{F}_t$

Now (a), (d), and (e) imply proposition (iii).

Proof of Theorem 3.29

(i) Note that not every set system $\mathscr{C} \subset \mathscr{A}$ necessarily occurs in $(\mathscr{F}_t, t \in T)$, i. e., we do not presume that there is a $t \in T$ such that $\mathscr{C} \subset \mathscr{F}_t$. Therefore we have to make the assumption "if there is a $t \in T \ldots$" .
(ii) is trivial.
(iii) If $\mathscr{C}$ and $\mathscr{D}$ are simultaneous, then

(a) there is a $t \in T$ with $\mathscr{C}, \mathscr{D} \subset \mathscr{F}_t$ and no $s \in T$, $s < t$, with $\mathscr{C} \subset \mathscr{F}_s$ or $\mathscr{D} \subset \mathscr{F}_s$.

If $\mathscr{D}$ and $\mathscr{E}$ are also simultaneous, then this implies $\mathscr{E} \subset \mathscr{F}_t$ and that

(b) there is no $s \in T$, $s < t$, with $\mathscr{D} \subset \mathscr{F}_s$ or $\mathscr{E} \subset \mathscr{F}_s$.

However, this implies that there is a $t \in T$ with $\mathscr{C}, \mathscr{E} \subset \mathscr{F}_t$, and that there is no $s \in T$, $s < t$, with $\mathscr{C} \subset \mathscr{F}_s$ or $\mathscr{E} \subset \mathscr{F}_s$. Hence, $\mathscr{C}$ and $\mathscr{E}$ are simultaneous as well.

Proof of Theorem 3.31

(i) If $\mathscr{C}$ and $\mathscr{D}$ are simultaneous with respect to $(\mathscr{F}_t, t \in T)$, then:

(a) there is a $t \in T$ with $\mathscr{C}, \mathscr{D} \subset \mathscr{F}_t$,

(b) there is no $s \in T$, $s < t$, with $\mathscr{C} \subset \mathscr{F}_s$ or $\mathscr{D} \subset \mathscr{F}_s$.

$(\mathscr{F}_t, t \in T)$ being a filtration implies:

(c) $\mathscr{C} \cup \mathscr{D}$, $\sigma(\mathscr{C} \cup \mathscr{D}) \subset \mathscr{F}_t$, and

(d) there is no $s \in T$, $s < t$, with $\mathscr{C} \cup \mathscr{D}$, $\sigma(\mathscr{C} \cup \mathscr{D}) \subset \mathscr{F}_s$.

Now (a) to (d) imply proposition (i) .
(ii) If $\mathscr{C}$, $\mathscr{D}$, and $\mathscr{E}$ are simultaneous with respect to $(\mathscr{F}_t, t \in T)$, then:

(a) there is a $t \in T$ with $\mathscr{C}, \mathscr{D}, \mathscr{E} \subset \mathscr{F}_t$,

(b) there is no $s \in T$, $s < t$, with $\mathscr{C} \subset \mathscr{F}_s$, $\mathscr{D} \subset \mathscr{F}_s$, or $\mathscr{E} \subset \mathscr{F}_s$.

$(\mathscr{F}_t, t \in T)$ being a filtration implies:

(c) $\mathscr{C} \cup \mathscr{D}$, $\sigma(\mathscr{C} \cup \mathscr{D}) \subset \mathscr{F}_t$,

(d) there is no $s \in T$, $s < t$, with $\mathscr{C} \cup \mathscr{D}$, $\sigma(\mathscr{C} \cup \mathscr{D}) \subset \mathscr{F}_s$.

Now (a) to (d) imply proposition (ii).

4.6 Exercises

▷ **Exercise 4-1** Propositions (i) to (iv) of Remark 4.4.

▷ **Exercise 4-2** Prove: If X_2 is a cause and X_1 is prior to X_2 with respect to $(\mathscr{F}_t, t \in T)$, then X_1 is a covariate of X_2.

▷ **Exercise 4-3** Consider Example 4.8 and show that, in this example, $\sigma(U)$ is the smallest of all σ-algebras $\mathscr{C}_i$ satisfying conditions (a) and (b) of Definition 4.3.

▷ **Exercise 4-4** How are the true total-effect variable and the average total effect defined?

▷ **Exercise 4-5** Use Theorem PR-4.25 to compute the probability $P(X{=}1)$ for the example displayed in Table 4.4 (p. 88).

▷ **Exercise 4-6** Compute the probabilities $P(Z{=}z \,|\, X{=}x, U{=}u)$ in Table 4.4 (p. 88).

▷ **Exercise 4-7** Compute the probability $P(U{=}\mathit{Jane}, Z{=}0 \,|\, X{=}1)$ in the example of Table 4.4 (p. 88).

▷ **Exercise 4-8** What are the values of the total-effect true-outcome variable τ_1 and the regression $E(Y|X,U)$ for $\omega_4 = (\mathit{Joe}, \mathit{yes}, +)$ in the example presented in Table 4.2 (p. 82)?

▷ **Exercise 4-9** What are the elements of the σ-algebra generated by Z in Example 4.3.3?

▷ **Exercise 4-10** What are the elements of the σ-algebra $\mathscr{F}_1$ in the filtration mentioned in Example 4.3.1?

▷ **Exercise 4-11** What are the elements of the σ-algebra $\mathscr{F}_2$ in the filtration mentioned in Example 4.3.1?

▷ **Exercise 4-12** Name at least two elements of the σ-algebra $\mathscr{F}_3$ in the filtration mentioned in Example 4.3.1 that are not already elements of $\mathscr{F}_2$.

▷ **Exercise 4-13** What is an $(X{=}x)$-conditional regression $E^{X=x}(Y|Z)$?

▷ **Exercise 4-14** Compute the values of the regression $E(Y|X)$ in the example presented in Table 4.3 (p. 85).

▷ **Exercise 4-15** Compute the values of the regression $E(Y|X,U)$ in the example presented in Table 4.3 (p. 85).

▷ **Exercise 4-16** Compute the values of the conditional regression $E^{X=0}(Y|U)$ in the example presented in Table 4.3 (p. 85).

▷ **Exercise 4-17** Why is the conditional expected value $E(Y \,|\, X{=}1, U{=}\mathit{Joe})$ not uniquely defined in the example presented in Table 4.3 (p. 85)?

▷ **Exercise 4-18** What is the *atomic total effect variable* $\delta_{xx'}$ of x vs. x' with respect to the outcome variable Y?

▷ **Exercise 4-19** Why is $\delta_{xx'}$ called the true-*total*-effect variable?

▷ **Exercise 4-20** What does it mean when we assume that the true-outcome variable τ_x is uniquely defined up to P-equivalence?

Solutions

▷ **Solution 4-1** Still missing.

(i) $\mathscr{C}_X$ is prior or simultaneous to X

(ii) $\sigma(X) \not\subset \mathscr{C}_X$

(iii) All σ-algebras $\mathscr{F}_t$, $t < t_X$, $t \in T$, of the filtration $(\mathscr{F}_t, t \in T)$ are measurable with respect to $\mathscr{C}_X$

(iv) If Z is a random variable on $(\Omega, \mathscr{A}, P)$ that is prior to X with respect to $(\mathscr{F}_t, t \in T)$, then $\sigma(Z) \subset \mathscr{C}_X$.

Condition (a) of Definition 4.3 implies (i), and condition (b) of Definition 4.3 implies (ii).

▷ **Solution 4-2** Still missing.

▷ **Solution 4-3** In this example, $\mathscr{F}_1 = \sigma(U)$, $\mathscr{F}_2 = \sigma(U, X) = \mathscr{F}_{t_X}$, and $\sigma(U)$ is the smallest of all σ-algebras $\mathscr{C}_i$ satisfying conditions (a) and (b) of Definition 4.3. This can be shown by contraposition. Suppose, there is a σ-algebra $\mathscr{C}_0$ satisfying these conditions and Equation (4.3). Then $\sigma(U)$ would be a subset of $\mathscr{C}_0$, which shows that $\mathscr{C}_0$ cannot be smaller than $\sigma(U)$. Hence, $\sigma(U)$ is the potential-confounder σ-algebra of X.

▷ **Solution 4-4** The true total-effect variable is defined to be the difference $E^{X=x}(Y|\mathscr{C}_X)$ $- E^{X=x'}(Y|\mathscr{C}_X)$ between the $(X=x)$- and $(X=x')$-conditional regressions of Y on a global covariateC_X. If we can assume that $E^{X=x}(Y|\mathscr{C}_X)$ and $E^{X=x'}(Y|\mathscr{C}_X)$ are P-unique, then the average total effects are defined by the difference

$$E(\tau_x) - E(\tau_{x'}) \;=\; E[E^{X=x}(Y|\mathscr{C}_X)] - E[E^{X=x'}(Y|\mathscr{C}_X)]$$

between the expectations over the distribution of C_X. Similarly, various kinds of *conditional total effects* can be defined by the difference between various conditional expectation values of the total-effect true-outcome variables (for details, see ch. 5).

▷ **Solution 4-5** Note that the four pairs (u, z) of values of U and Z are disjoint and all these pairs of values have positive probabilities. Hence we can apply the theorem of total probability:

$$\begin{aligned} P(X=1) \;&=\; \sum_u \sum_z P(X=1\,|\,U=u, Z=z) \cdot P(U=u, Z=z) \\ &= \left(\frac{3}{4} + \frac{1}{4} + \frac{3}{4} + \frac{1}{4}\right) \cdot \frac{1}{4} \;=\; \frac{1}{2}. \end{aligned}$$

Another way to compute $P(X=1)$ is to use $P(X=1) = E(I_{X=1}) = E[E(I_{X=1}\,|\,U, Z)]$ [see Rule (iv) in Box PR-10.2], the fact that the probabilities $P(X=1\,|\,U=u, Z=z)$ are the values of the regression $E(I_{X=1}\,|\,U, Z)$, and then Equation PR-(6.15). Although the resulting formula is the same as above, this second way makes clear that the unconditional probability $P(X=1)$ is the expectation of the conditional probabilities $P(X=1\,|\,U=u, Z=z)$ over the joint distribution of the two random variables U and Z.

▷ **Solution 4-6** Note that the complementary probability to $P(X=1\,|\,U=u, Z=0) = 3/4$ is $P(X=0\,|\,U=u, Z=0) = 1/4$ for both units u. Similarly the complementary probability to $P(X=1\,|\,U=u, Z=1) = 1/4$ is $P(X=0\,|\,U=u, Z=1) = 3/4$. Now we can use the equation

$$P(Z=z\,|\,X=x, U=u) \;=\; \frac{P(X=x\,|\,U=u, Z=z) \cdot P(Z=z\,|\,U=u)}{\sum_z P(X=x\,|\,U=u, Z=z) \cdot P(Z=z\,|\,U=u)}.$$

For example, the probability of not receiving group therapy ($Z{=}0$), if Jane is drawn ($U{=}\mathit{Jane}$) and does not receive individual therapy ($X{=}0$), is

$$P(Z{=}0\,|\,X{=}0, U{=}\mathit{Jane}) \;=\; \frac{P(X{=}0\,|\,U{=}\mathit{Jane}, Z{=}0)\cdot P(Z{=}0\,|\,U{=}\mathit{Jane})}{\sum_z P(X{=}0\,|\,U{=}\mathit{Jane}, Z{=}z)\cdot P(Z{=}z\,|\,U{=}\mathit{Jane})},$$

where $P(X{=}0\,|\,U{=}\mathit{Jane}, Z{=}0) = 1/4$, $P(Z{=}0\,|\,U{=}\mathit{Jane}) = 1/2$, and

$$\begin{aligned}&\sum_z P(X{=}0\,|\,U{=}\mathit{Jane}, Z{=}z)\cdot P(Z{=}0\,|\,U{=}\mathit{Jane})\\ &= P(X{=}0\,|\,U{=}\mathit{Jane}, Z{=}0)\cdot P(Z{=}0\,|\,U{=}\mathit{Jane}) \;+\; P(X{=}0\,|\,U{=}\mathit{Jane}, Z{=}1)\cdot P(Z{=}1\,|\,U{=}\mathit{Jane})\\ &= \frac{1}{4}\cdot\frac{1}{2} + \frac{3}{4}\cdot\frac{1}{2} \;=\; \frac{1}{2}.\end{aligned}$$

Inserting this results yields the probability $P(Z{=}0\,|\,X{=}0, U{=}\mathit{Jane}) = 1/4$. Using the same procedure for all values of U, X, and Z leads to the probabilities displayed in Table 4.4 (p. 88).

▷ **Solution 4-7** We have to use Equation

$$P(U{=}u, Z{=}z\,|\,X{=}x) \;=\; \frac{P(X{=}x, U{=}u, Z{=}z)}{P(X{=}x)} \;=\; \frac{P(X{=}x\,|\,U{=}u, Z{=}z)\cdot P(U{=}u, Z{=}z)}{P(X{=}x)}.$$

For $U{=}\mathit{Jane}$, $Z{=}0$ and $X{=}1$ this equation yields

$$\begin{aligned}P(U{=}\mathit{Jane}, Z{=}0\,|\,X{=}1) \;&=\; \frac{P(X{=}1\,|\,U{=}\mathit{Jane}, Z{=}0)\cdot P(U{=}\mathit{Jane}, Z{=}0)}{P(X{=}1)}\\ &=\; \frac{3/4\cdot 1/4}{1/2} \;=\; \frac{3}{8}.\end{aligned}$$

▷ **Solution 4-8** If $\omega_4 = (\mathit{Joe}, \mathit{yes}, +)$, then $U(\omega_4) = \mathit{Joe}$ and the value of the total-effect true-outcome variable τ_1 is $E(Y\,|\,X{=}1, U{=}\mathit{Joe}) = .80$ (see the fourth row in Table 4.2, p. 82). The value $E(Y\,|\,X,U)(\omega_4)$ of the regression $E(Y\,|\,X,U)$ is also $E(Y\,|\,X{=}1, U{=}\mathit{Joe}) = .80$.

▷ **Solution 4-9** The set of possible outcomes of the random experiment is

$$\Omega := \Omega_U \times \Omega_Z \times \Omega_X \times \Omega_Y,$$

where $\Omega_Z := \{\mathit{group\ therapy}, \mathit{no\ group\ therapy}\}$. Hence, the σ-algebra generated by Z and $\mathscr{A}'_Z = \big\{\{0\}, \{1\}, \{0,1\}, \emptyset\big\}$ is

$$Z^{-1}(\mathscr{A}'_Z) = \big\{\Omega,\ \emptyset,\ \Omega_U \times \{\mathit{group\ therapy}\} \times \Omega_X \times \Omega_Y,\ \Omega_U \times \{\mathit{no\ group\ therapy}\} \times \Omega_X \times \Omega_Y\big\}.$$

Aside from Ω and the empty set Ø, this σ-algebra contains the event

$$\Omega_U \times \{\mathit{group\ therapy}\} \times \Omega_X \times \Omega_Y$$

that the person drawn receives group therapy, and the event

$$\Omega_U \times \{\mathit{no\ group\ therapy}\} \times \Omega_X \times \Omega_Y$$

that the person drawn does *not* receive group therapy.

▷ **Solution 4-10** The σ-algebra $\mathscr{F}_1$ has four elements. Aside from Ω and Ø, these are the event

$$\textit{Joe is drawn} := \{\textit{Joe}\} \times \Omega_X \times \Omega_Y,$$

consisting of the four elements $(\textit{Joe}, \textit{no}, -)$, $(\textit{Joe}, \textit{yes}, -)$, $(\textit{Joe}, \textit{no}, +)$, and $(\textit{Joe}, \textit{yes}, +)$, and the event

$$\textit{Ann is drawn} := \{\textit{Ann}\} \times \Omega_X \times \Omega_Y,$$

consisting of the four elements $(\textit{Ann}, \textit{no}, -)$, $(\textit{Ann}, \textit{yes}, -)$, $(\textit{Ann}, \textit{no}, +)$, and $(\textit{Ann}, \textit{yes}, +)$.

▷ **Solution 4-11** The σ-algebra $\mathscr{F}_2$ has $2^4 = 16$ elements. Aside from the four events mentioned in Exercise 4-10, these are the events

$$A_5 := \textit{The person drawn is treated} := \Omega_U \times \{\textit{yes}\} \times \Omega_Y,$$

$$A_6 := \textit{The person drawn is not treated} := \Omega_U \times \{\textit{no}\} \times \Omega_Y,$$

but also

$$A_7 := \textit{Joe is drawn and treated} := \{\textit{Joe}\} \times \{\textit{yes}\} \times \Omega_Y,$$

$$A_8 := \textit{Joe is drawn and is not treated} := \{\textit{Joe}\} \times \{\textit{no}\} \times \Omega_Y,$$

$$A_9 := \textit{Ann is drawn and treated} := \{\textit{Ann}\} \times \{\textit{yes}\} \times \Omega_Y,$$

$$A_{10} := \textit{Ann is drawn and is not treated} := \{\textit{Ann}\} \times \{\textit{no}\} \times \Omega_Y,$$

as well as

$$\textit{Joe is drawn and treated} \text{ or } \textit{Ann is drawn and is not treated} := A_7 \cup A_{10},$$

$$\textit{Joe is drawn and not treated} \text{ or } \textit{Ann is drawn and treated} := A_8 \cup A_9,$$

and

$$\textit{Joe is drawn} \text{ or } \textit{Ann is drawn and treated} := A_7 \cup A_8 \cup A_9,$$

$$\textit{Joe is drawn} \text{ or } \textit{Ann is drawn and not treated} := A_7 \cup A_8 \cup A_{10},$$

$$\textit{Joe is drawn and treated} \text{ or } \textit{Ann is drawn} := A_7 \cup A_9 \cup A_{10},$$

$$\textit{Joe is drawn and not treated} \text{ or } \textit{Ann is drawn} := A_8 \cup A_9 \cup A_{10}.$$

▷ **Solution 4-12** Two such elements of $\mathscr{F}_3$ are

$$\textit{The person drawn is successful} := \Omega_U \times \Omega_X \times \{+\},$$

and

$$\textit{The person drawn is not successful} := \Omega_U \times \Omega_X \times \{-\}.$$

Others are

$$\textit{Joe is successful} := \{\textit{Joe}\} \times \Omega_X \times \{+\}.$$

or

$$\textit{Ann is treated and is not successful} := \{\textit{Ann}\} \times \{\textit{yes}\} \times \{-\}.$$

▷ **Solution 4-13** The $(X{=}x)$-conditional regression $E^{X=x}(Y|Z)$ is the regression of Y on Z with respect to the probability measure $P^{X=x}$. The values of $E^{X=x}(Y|Z)$ are identical with the conditional expectation values $E(Y|X{=}x, Z{=}z)$. In more formal terms, $E^{X=x}(Y|Z)(\omega) = E(Y|X{=}x, Z{=}z)$, if $\omega \in (X,Z)^{-1}(\{(x,z)\})$.

▷ **Solution 4-14** The values of the regression $E(Y|X)$ are the two conditional expectation values $E(Y|X{=}0)$ and $E(Y|X{=}1)$. Because $E(Y|X{=}x) = P(Y{=}1\,|\,X{=}x)$, they can be computed as follows:

$$P(Y{=}1\,|\,X{=}0) \;=\; \frac{P(Y{=}1, X{=}0)}{P(X{=}0)} \;=\; \frac{.348+.024}{.152+.348+.096+.024} \;=\; .60,$$

and

$$P(Y{=}1\,|\,X{=}1) \;=\; \frac{P(Y{=}1, X{=}1)}{P(X{=}1)} \;=\; \frac{0+.152}{0+0+.228+.152} \;=\; .40.$$

▷ **Solution 4-15** The values of the regression $E(Y|X,U)$ are the four conditional expectation values $E(Y\,|\,X{=}x, U{=}u)$. Because $E(Y\,|\,X{=}x, U{=}u) = P(Y{=}1\,|\,X{=}x, U{=}u)$, they can be computed as follows:

$$P(Y{=}1\,|\,X{=}0, U{=}\mathit{Joe}) \;=\; \frac{P(Y{=}1, X{=}0, U{=}\mathit{Joe})}{P(X{=}0, U{=}\mathit{Joe})} \;=\; \frac{.348}{.152+.348} \;=\; .696,$$

$$P(Y{=}1\,|\,X{=}0, U{=}\mathit{Ann}) \;=\; \frac{P(Y{=}1, X{=}0, U{=}\mathit{Ann})}{P(X{=}0, U{=}\mathit{Ann})} \;=\; \frac{.024}{.096+.024} \;=\; .20,$$

$$P(Y{=}1\,|\,X{=}1, U{=}\mathit{Ann}) \;=\; \frac{P(Y{=}1, X{=}1, U{=}\mathit{Ann})}{P(X{=}1, U{=}\mathit{Ann})} \;=\; \frac{.152}{.228+.152} \;=\; .40.$$

The conditional expectation $E(Y\,|\,X{=}1, U{=}\mathit{Joe}) = P(Y{=}1\,|\,X{=}1, U{=}\mathit{Joe})$ is undefined, because $P(X{=}1, U{=}\mathit{Joe}) = 0$. Choosing any number (such as 9) as a value of $E(Y|X,U)$ for $\omega_3 = (\mathit{Joe}, \mathit{yes}, -)$ and $\omega_4 = (\mathit{Joe}, \mathit{yes}, +)$ yields a version of $E(Y|X,U)$. Different versions of $E(Y|X,U)$ are identical almost surely with respect to the measure P.

▷ **Solution 4-16** The values of the conditional regression $E^{X=0}(Y\,|\,U)$ are the two conditional expectation values $E^{X=0}(Y\,|\,U{=}\mathit{Joe})$ and $E^{X=0}(Y\,|\,U{=}\mathit{Ann})$ with respect to the measure $P^{X=0}$. Because $E^{X=0}(Y\,|\,U{=}u) = P^{X=0}(Y{=}1\,|\,U{=}u)$, they can be computed as follows:

$$P^{X=0}(Y{=}1\,|\,U{=}\mathit{Joe}) \;=\; \frac{P^{X=0}(Y{=}1, U{=}\mathit{Joe})}{P^{X=0}(U{=}\mathit{Joe})} \;=\; \frac{.348\,/\,.62}{(.152+.348)\,/\,.62} \;=\; .696$$

and

$$P^{X=0}(Y{=}1\,|\,U{=}\mathit{Ann}) \;=\; \frac{P^{X=0}(Y{=}1, U{=}\mathit{Ann})}{P^{X=0}(U{=}\mathit{Ann})} \;=\; \frac{.024\,/\,.62}{(.096+.024)\,/\,.62} \;=\; .20.$$

▷ **Solution 4-17** In this example, $P(X{=}1, U{=}\mathit{Joe}) = 0$. This implies that the conditional probabilities $P(Y{=}y\,|\,X{=}1, U{=}\mathit{Joe})$ that are used in the definition of $E(Y|X{=}1, U{=}\mathit{Joe})$ (see Def. PR-9.2) are not defined.

▷ **Solution 4-18** The atomic total effect variable $\delta_{xx'}$ is that random variable whose values are the differences between the $(X{=}x)$- and $(X{=}x')$-conditional regressions of the outcome variable Y on the global covariate C_X. For this interpretation of the values of $\delta_{xx'}$, we have to presume that $E^{X=x}(Y|\mathscr{C}_X)$ and $E^{X=x'}(Y|\mathscr{C}_X)$ are P-unique.

▷ **Solution 4-19** Assuming that $E^{X=x}(Y|\mathscr{C}_X)$ and $E^{X=x'}(Y|\mathscr{C}_X)$ are P-unique, $\delta_{xx'}$ is defined to be the difference $\tau_x - \tau_{x'}$, where $\tau_x := E^{X=x}(Y|\mathscr{C}_X)$ and C_X is a global covariate of X. Conditioning on C_X means ignoring altogether intermediate variables that might transmit the effect to the outcome variable Y. Hence, the values of $\delta_{xx'}$ are only *total effects* comprising *indirect effects* and *direct effects*. Indirect effects might be transmitted through an intermediate variable M, whereas direct effects *are not* transmitted through M.

▷ **Solution 4-20** This means that there may be different versions of the true-outcome variable. However, under P-uniqueness, two versions are identical with probability 1.

Chapter 5

Causal Effects

In chapter 4, we defined the total-effect *true-outcome variables* τ_x and the *true-total-effect variables* $\delta_{xx'} := \tau_x - \tau_{x'}$, where x and x' denote two values of the cause X. These variables have been constructed such that they are purged from bias with respect to total effects. Although, in empirical applications, these effects are hard to estimate, the expectation $E(\delta_{xx'})$ as well as conditional expected values $E(\delta_{xx'} \mid V=v)$ *can* be estimated under realistic assumptions. In this chapter, we define the *average total effect* of x vs. x' as the expectation $E(\delta_{xx'})$. Similarly, we define the $(V=v)$*-conditional total effect* of treatment x vs. treatment x' as the conditional expectation value $E(\delta_{xx'} \mid V=v)$.

While the average and conditional total effects mentioned above are built on the total-effect true-outcome variables τ_x and their differences, the various concepts of *direct effects* are built on the *t-direct-effect true-outcome variables* $\tau_{x,t}$ and on the *atomic t-direct-effect variables* $\delta_{xx',t} := \tau_{x,t} - \tau_{x',t}$ that also have been defined in chapter 4. Taking the expectation $E(\delta_{xx',t})$ or the conditional expectation values $E(\delta_{xx',t} \mid V=v)$ yields various kinds of direct effects, which are *not* transmitted by intermediate variables that are in between times t_X and t or simultaneous to tm, where $t_X \leq t < t_Y$. Last but not least we also consider various kinds of *indirect effects* that rest on the difference variable $\delta_{xx'} - \delta_{xx',t}$. Taking the expectation $E(\delta_{xx'} - \delta_{xx',t})$ or the conditional expectation values $E(\delta_{xx'} - \delta_{xx',t} \mid V=v)$ yields various kinds of indirect effects.

Overview

We start defining the *average total effect* as the expectation of the true-total-effect variable. Similarly, we define the *conditional total effect given the value v of a variable V* and discuss its meaning for various choices of V, e. g., a pre-treatment variable Z, the observational-unit variable U, the putative cause X, and combinations of these variables. The same task is then tackled for various kinds of *direct* and *indirect* effects. The only difference is that the true-total effect variable $\delta_{xx'}$ is replaced by the variables $\delta_{xx',t}$ and $\delta_{xx'} - \delta_{xx',t}$, respectively.

5.1 Average Total Effect

In the following definition, we presume that the conditional expectation $E^{X=x}(Y|\mathscr{C}_X)$ of Y given the potential-confounder σ-algebra $\mathscr{C}_X$ with respect to the conditional-probability measure $P^{X=x}$ and the conditional expectation $E^{X=x'}(Y|\mathscr{C}_X)$ with respect to $P^{X=x'}$ are P-unique. This assumption ensures that the difference $\delta_{xx'} := E^{X=x}(Y|\mathscr{C}_X) - E^{X=x'}(Y|\mathscr{C}_X)$ between two versions of these conditional expectations is itself P-unique and that the expectation of $\delta_{xx'}$ is a uniquely defined number.

Definition 5.1 (Average Total Effect)
Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a causality space, let Y be numerical and nonnegative or with finite expectation $E(Y)$, let x and x' be two values of X with $P(X=x)$, $P(X=x') > 0$, and let $\tau_x = E^{X=x}(Y|\mathscr{C}_X)$ and $\tau_{x'} = E^{X=x'}(Y|\mathscr{C}_X)$ be P-unique. Then the expectation $E(\delta_{xx'})$ of the atomic total-effect variable $\delta_{xx'} := \tau_x - \tau_{x'}$ is defined to be the average total effect of x vs. x' on Y.

Remark 5.2 (Substantive Meaning) If X represents a treatment variable, then $E(\delta_{xx'})$ is also called the average total effect of treatment x vs. treatment x'. Sometimes it is also called the 'average causal effect' or the 'average treatment effect', which is unambiguous as long as no direct and/or indirect treatment effects are considered. The average total effects are among the parameters that one might want to estimate analyzing sample data. Although the average total effect is defined as the expectation of the true-total-effect variable $\delta_{xx'}$, we will show that there *are* ways to identify the average total effect *without knowing the values of* $\delta_{xx'}$. However, identification of the average total effect rests on assumptions. Such assumptions, also called *causality conditions*, will be treated in chapters 6 to 9. ◁

Remark 5.3 (Expectations of True-Outcome Variables and Adjusted Means)
Because

$$E(\delta_{xx'}) \;=\; E(\tau_x - \tau_{x'}) \;=\; E(\tau_x) - E(\tau_{x'}),$$

the expectations of the total-effect true-outcome variables τ_x are of interest as well. These expectations are estimated, e. g., by the sample group means in a randomized experiment and by the *adjusted means* in analysis of covariance and its generalizations, as well as in procedures based on propensity scores, *provided that certain assumptions hold* (for more details see ch. 13).

Note again that each version of τ_x is a random variable on $(\Omega, \mathscr{A}, P)$ and that the expectation $E(\tau_x)$ of this random variable is with respect to the probability measure P. If $\tau_x = E^{X=x}(Y|\mathscr{C}_X)$ is P-unique, then the expectation $E(\tau_x)$ does not depend on the choice of the version of τ_x. In general, $E(\tau_x) \neq E(Y|X=x)$, i. e., the expectations of the true-outcome variables τ_x are *not* equal to the $(X=x)$-conditional expectations $E(Y|X=x)$ of the outcome variable Y. In other words, in general, the conditional expectation values $E(Y|X=x)$ are *biased*, and this bias carries over to the prima facie effects $E(Y|X=x) - E(Y|X=x')$ which, in general, are

not equal to the average total effects $E(\delta_{xx'}) = E(\tau_x) - E(\tau_{x'})$ (for more details, see ch. 6). ◁

Remark 5.4 (Average Total Effect vs. Main Effect) In a perfect randomized experiment with a treatment factor and a second factor representing a qualitative covariate (such as sex), the average total effects are what is tested as the *main effect* of the 'treatment factor' in analysis of variance. This is true if there is no interaction, but it is also true in cases *with* interaction between the two factors, in the sense that the effects of the treatment factor depend on the values of the second factor. However, the definition of an average total effect is much more general. It also holds beyond the randomized experiment. The average total effect is defined without reference to a *particular* covariate (or second factor). The average total effect *does* refer, however, to a specified random experiment, a filtration $(\mathscr{F}_t, t \in T)$, a focused cause X, and an outcome variable Y (see ch. 3). ◁

5.1.1 Numerical Example

Example 5.5 (Jim and Jane – continued) In the example presented in Table 4.4 (see p. 88), assuming $\mathscr{C}_X = \sigma(U,Z)$ implies that δ_{10} is $\mathscr{C}_X$-measurable. Hence, there is a measurable function $f_{10}\colon (\Omega_U, \Omega'_Z) \to (\mathbb{R}, \mathscr{B})$ such that $\delta_{10} = f_{10}(U,Z)$ is the composition of (U,Z) and f_{10} (see Lemma 2.52 of SN). The values of δ_{10} are: $\delta_{10}(\omega) = f_{10}[U(\omega), Z(\omega)] = f_{10}(u,z)$, where

$$f_{10}(u,z) \;=\; E(Y|X{=}1, U{=}u, Z{=}z) \;-\; E(Y|X{=}0, U{=}u, Z{=}z). \tag{5.1}$$

According to Table 4.4, the true-total effect variable δ_{10} has the four different values 14, 4, 18, and 2. Hence, the average total effect of *individual therapy* $(X{=}1)$ vs. *no individual therapy* $(X{=}0)$ can be computed by

$$\begin{aligned} E(\delta_{10}) \;&=\; \sum_u \sum_z \big[E(Y|X{=}1, U{=}u, Z{=}z) - E(Y|X{=}0, U{=}u, Z{=}z)\big] \cdot P(U{=}u, Z{=}z) \\ &=\; 14 \cdot \frac{1}{4} + 4 \cdot \frac{1}{4} + 18 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} \;=\; 9.5, \end{aligned}$$

using Equation (6.15) of SN, and

$$P(U{=}u, Z{=}z) = P(Z{=}z|U{=}u) \cdot P(U{=}u) = 1/2 \cdot 1/2 = 1/4,$$

for all values (u,z) of U and Z. ◁

5.2 Conditional Total Effect

Now we introduce the concept of a *conditional total effect*, conditioning on the value v of a random variable V. Examples for V are the observational-unit variable U, a mapping $f(U)$ of U (such as *sex*), but also the treatment variable X.

The conditional total effect given the value v of V is the $(V{=}v)$-conditional expectation value of the true-total-effect variable $\delta_{xx'}$. If v represents a subpopulation, i. e., a subset of Ω_U such as 'males', then we simply take the expectation of $\delta_{xx'}$ *within the subpopulation* of males. The expectation $E(\delta_{xx'}\,|\,V{=}v)$ is well-defined if we can assume that $\tau_x = E^{X=x}(Y|\mathscr{C}_X)$ and $\tau_{x'} = E^{X=x'}(Y|\mathscr{C}_X)$ are at least $P^{V=v}$-unique (see Remark 4.15).

Definition 5.6 (Conditional Total Effect)
Let $\langle\,(\Omega,\mathscr{A},P),(\mathscr{F}_t,t\in T),X,Y\,\rangle$ be a causality space, let Y be numerical and nonnegative or with finite expectation $E(Y)$, let x and x' be two values of X with $P(X{=}x)$, $P(X{=}x')>0$, and let V be a random variable on $(\Omega,\mathscr{A},P)$.

(i) Let v be a value of V with $P(V{=}v)>0$. If τ_x and $\tau_{x'}$ are $P^{V=v}$-unique, then the $(V{=}v)$-conditional expectation value

$$E(\delta_{xx'}\,|\,V{=}v) \tag{5.2}$$

is defined to be the $(V{=}v)$-conditional total effect of x vs. x' on the (expectation of the) outcome variable Y.

(ii) If τ_x and $\tau_{x'}$ are P-unique, then the conditional expectation

$$E(\delta_{xx'}\,|\,V) \tag{5.3}$$

is defined to be the V-conditional total-effect function of x vs. x'.

Example 5.7 (Joe and Ann Self-Selected: No Treatment for Joe) In the example of Table 4.3 (see p. 85), the true-outcome variables are $\tau_1 = E^{X=1}(Y|\mathscr{C}_X) = E^{X=1}(Y|U)$ and $\tau_0 = E^{X=0}(Y|\mathscr{C}_X) = E^{X=0}(Y|U)$. In this example, we cannot define the average total effect, because $E^{X=1}(Y|U)$ is not P-unique; its values are not uniquely defined if Joe is sampled, although Joe has a positive probability to be drawn. This implies that the values of $E^{X=1}(Y|U)-E^{X=0}(Y|U)$ are also arbitrary for $\omega\in U^{-1}(\{Joe\})$. However, the values of $E^{X=1}(Y|U)$ are well-defined for $\omega\in U^{-1}(\{Ann\})$; in other words, $\tau_0 = E^{X=0}(Y|U)$ and $\tau_1 = E^{X=1}(Y|U)$ are $P^{U=Ann}$-unique. Hence, we can define the $(U{=}Ann)$-conditional total effect of treatment 1 vs. treatment 0. In this example, this effect is .20. It will also be called the *individual total effect* of Ann. ◁

Remark 5.8 (Conditional Expectation Value) If $P(V{=}v)>0$ and $\tau_x = E^{X=x}(Y|\mathscr{C}_X)$ as well as $\tau_{x'} = E^{X=x'}(Y|\mathscr{C}_X)$ are both $P^{V=v}$-unique, then

$$E(\delta_{xx'}\,|\,V{=}v) \;=\; E^{V=v}(\delta_{xx'}),$$

where $E^{V=v}(\delta_{xx'})$ denotes the expectation of $\delta_{xx'} = \tau_x - \tau_{x'}$ with respect to the measure conditional-probability measure $P^{V=v}$. ◁

Remark 5.9 (Conditional vs. Average Total Effects) Conditional total effects are more informative than the average total effect. If V is a function of the observational-unit variable U such as $V :=$ *sex* or $V :=$ *educational status*, then the $(V=v)$-conditional total effect is the average total effect *in the subpopulation* associated with the value v of V. (In Example 5.7, this subpopulation consists of a single observational-unit, namely Ann.) In other cases, such a conditional total effect is simply the conditional total effect *given the value v of V*.

Also note that the concept of a $(V=v)$-conditional total effect is not restricted to a univariate variable V. Instead, V may also be a vector of variables $V_1,\ldots,V_m$ such that a value $v = (v_1,\ldots,v_m)$ of V is a vector of values of the variables $V_1,\ldots,V_m$. ◁

Remark 5.10 (Expectation of the Conditional-Total-Effect Function) The conditional total effects given a value v of a variable V are the values of the *conditional total-effect function* $E(\delta_{xx'}\,|\,V)$. It is easy to see that the average total effect is also the expectation of the conditional total effects, i. e.,

$$E\big[E(\delta_{xx'}\,|\,V)\big] \;=\; E(\delta_{xx'}) \qquad \text{[Box 10.2 (iv) of SN].} \tag{5.4}$$

However, considering the conditional expectation $E(\delta_{xx'}\,|\,V)$ and its expectation, we have to assume that $\tau_x = E^{X=x}(Y|\mathscr{C}_X)$ and $\tau_{x'} = E^{X=x'}(Y|\mathscr{C}_X)$ are P-unique. ◁

5.2.1 Conditional Total Effects given a Covariate

If the variable V in the definition of a conditional total effect is a pre-test that measures the 'same' attribute (e. g., *live satisfaction*) as the outcome variable Y (the post-test), but prior to the onset of the treatment, then studying the conditional total effects $E(\delta_{xx'}\,|\,V=v)$ allows to investigate if the effect of x vs. x' depends on the values of this pre-test. V may also represent an attribute of the observational unit such as *sex* or *educational status*. Furthermore, we may also consider a second treatment variable taking the role of V. Also note that V may be multivariate, consisting of several univariate variables. In all these cases we might be interested in comparing the $(V=v)$-conditional total effects $E(\delta_{xx'}|V=v)$ for different values v of V. (See ch. 2 for a discussion of covariates and ch. 3 for their mathematical definition and their role in a causality space).

Example 5.11 (Jim and Jane – continued) Because $\mathscr{C}_X = \sigma(U,Z)$ in the example presented in Table 4.4 (see p. 88), the conditional total effects of *individual therapy* $(X=1)$ vs. *no individual therapy* $(X=0)$ given $Z=0$ (*no group therapy*) can be computed by

$$\begin{aligned}
&E(\delta_{10}\,|\,Z=0)\\
&= \sum_u \sum_z \big[E(Y\,|\,X=1,U=u,Z=z) - E(Y\,|\,X=0,U=u,Z=z)\big]\cdot P(U=u,Z=z\,|\,Z=0)\\
&= (82-68)\cdot\frac{1}{2} + (100-96)\cdot 0 + (98-80)\cdot\frac{1}{2} + (106-104)\cdot 0 \;=\; 16
\end{aligned}$$

[see Eq. (5.1) and Eq. (9.19) of SN]. Similarly, for $Z{=}1$ we receive

$$\begin{aligned} & E(\delta_{10}\,|\,Z{=}1) \\ &= \sum_u \sum_z \big[E(Y\,|\,X{=}1,U{=}u,Z{=}z) - E(Y\,|\,X{=}0,U{=}u,Z{=}z)\big]\cdot P(U{=}u,Z{=}z\,|\,Z{=}1) \\ &= (82-68)\cdot 0 + (100-96)\cdot\frac{1}{2} + (98-80)\cdot 0 + (106-104)\cdot\frac{1}{2} \;=\; 3. \end{aligned}$$

According to Equation (5.4), taking the expectation

$$E\big[E(\delta_{10}\,|\,Z)\big] \;=\; \sum_z E(\delta_{10}\,|\,Z{=}z)\cdot P(Z{=}z) \;=\; 16\cdot\frac{1}{2} \;+\; 3\cdot\frac{1}{2} \;=\; 9.5 \tag{5.5}$$

again yields the average total effect. In this equation we used the theorem of total probability in order to compute $P(Z{=}z) = \sum_u P(Z{=}z|U{=}u)\cdot P(U{=}u)$ (see Th. 4.25 of SN), which yields $P(Z{=}0) = P(Z{=}1) = 1/2$. ◁

5.2.2 Individual Total Effects

If we consider the single-unit trial of simple experiments and quasi-experiments described in section 2.1, or the single-unit trials described in sections 2.2 or 2.5, we can also condition on the observational-unit variable U. An *individual total effect* is a special case of a conditional total effect, in which we condition on the observational-unit variable U, i. e., we may define the *individual total effect of unit* of x vs. x' for unit u by

$$E(\delta_{xx'}\,|\,U{=}u), \tag{5.6}$$

assuming that $E^{X=x}(Y|\mathscr{C}_X)$ and $E^{X=x'}(Y|\mathscr{C}_X)$ are (at least) $P^{U=u}$-unique. Defining the *individual total-effect variable* by

$$E(\delta_{xx'}\,|\,U), \tag{5.7}$$

we presume that $E^{X=x}(Y|\mathscr{C}_X)$ and $E^{X=x'}(Y|\mathscr{C}_X)$ are P-unique, implying that $\delta_{xx'}$ is P-unique as well [see Remark 2.76 (ii) of SN)].

Remark 5.12 (Expectation of the Conditional-Total-Effect Function) Assume that $\tau_x = E^{X=x}(Y|\mathscr{C}_X)$ and $\tau_{x'} = E^{X=x'}(Y|\mathscr{C}_X)$ are P-unique. Then according to Equation (5.4),

$$E\big[E(\delta_{xx'}\,|\,U)\big] \;=\; E(\delta_{xx'}). \tag{5.8}$$

Hence, the average total effect is the expectation of the individual total effects. ◁

Remark 5.13 (Individual Effects vs. Average Effects) Individual total effects are more informative than the average total effect and usually more informative than conditional total effects given a value of a pre-test. However, note again that individual total effects are not necessarily the most fine-grained total effects (the true effects), as has been shown in Example 4.3.3. In that example, there is a second treatment variable, denoted Z, that contributes to the variation of the outcome

variable Y beyond the individual level. Furthermore, in empirical applications, individual effects can only be estimated under very restrictive assumptions, e. g., that a group of individuals that is represented by the value v of a covariate V have identical individual effects. ◁

Example 5.14 (Jim and Jane – continued) Because $\mathscr{C}_X = \sigma(U,Z)$ holds in the example presented in Table 4.4 (see p. 88), the individual total effects of *individual therapy* ($X{=}1$) vs. *no individual therapy* ($X{=}0$) can be computed by

$$\begin{aligned} &E(\delta_{10}\,|\,U{=}\mathit{Jim}) \\ &= \sum_u \sum_z \big[E(Y\,|\,X{=}1,U{=}u,Z{=}z) - E(Y\,|\,X{=}0,U{=}u,Z{=}z)\big]\cdot P(U{=}u,Z{=}z\,|\,U{=}\mathit{Jim}) \\ &= (82-68)\cdot\frac{1}{2} + (100-96)\cdot\frac{1}{2} + (98-80)\cdot 0 + (106-104)\cdot 0 \;=\; 9 \end{aligned}$$

for Jim [see again Eq. (5.1) and Eq. (9.19) of SN], and by

$$\begin{aligned} &E(\delta_{10}\,|\,U{=}\mathit{Jane}) \\ &= \sum_u \sum_z \big[E(Y\,|\,X{=}1,U{=}u,Z{=}z) - E(Y\,|\,X{=}0,U{=}u,Z{=}z)\big]\cdot P(U{=}u,Z{=}z\,|\,U{=}\mathit{Jane}) \\ &= (82-68)\cdot 0 + (100-96)\cdot 0 + (98-80)\cdot\frac{1}{2} + (106-104)\cdot\frac{1}{2} \;=\; 10 \end{aligned}$$

for Jane. Hence, in this example, the two individual total effects for Jim and Jane are both positive. In other examples, however, some individual effects might be positive, while others are negative.

According to Equation (5.4), the expectation of these individual effects

$$E\big[E(\delta_{10}\,|\,U)\big] \;=\; \sum_u E(\delta_{10}\,|\,U{=}u)\cdot P(U{=}u) \;=\; 9\cdot\frac{1}{2} \;+\; 10\cdot\frac{1}{2} \;=\; 9.5$$

is the average total effect $E(\delta_{10})$ of *individual therapy* vs. *no individual therapy*. ◁

5.2.3 Conditional Total Effects Given a Value of X

There is another kind of total effect that is sometimes of interest, the *conditional total effect of $X{=}x$ vs. $X{=}x'$ given $X{=}x^*$*. Again, instead of taking the (unconditional) expectation of the true effects, we take a conditional expectation value, this time, given a value x^* of X. Hence, if $E^{X=x}(Y\,|\,\mathscr{C}_X)$ and $E^{X=x'}(Y\,|\,\mathscr{C}_X)$ are $P^{X=x^*}$-unique, then

$$E(\delta_{xx'}\,|\,X{=}x^*) \tag{5.9}$$

is the *conditional total effect* of x vs. x' *given* $X{=}x^*$.

Remark 5.15 (Substantive Meaning) Suppose X represents a treatment variable in an experiment or in a quasi-experiment. According to Equation (5.9), if there

are two treatment conditions $X=1$ (treatment) and $X=0$ (control), we may consider $E(\delta_{10}|X=1)$, the conditional total effect (of treatment 1 vs. 0) *given treatment*, and $E(\delta_{10}|X=0)$ the conditional total effect (of treatment 1 vs. control) *given control*. These effects are also known as the 'average effect on the treated' and 'average effect on the untreated', respectively. ◁

Remark 5.16 (Pre-Facto Perspective) At first sight, the concept of a *conditional total effect* of x vs. x' *given* $X=x'$ seems strange. How can we talk about the average (or conditional) total treatment effect on the untreated? However, remember that we are not talking about data that resulted from an experiment. Instead we are considering a random experiment *that is still to be conducted*, i. e., we look at the random experiment from the *pre-facto perspective*. This is what stochastic theories are about: a random experiment that is not yet conducted. Talking about the probability of an event does not make sense for an event that already occurred, unless we *do as if* it did not yet occur, i. e., unless we take the pre-facto perspective. Hence, we can talk about an individual total effect although the individual is not yet treated and even if it will never be treated, just in the same way as we can talk about the probability of flipping 'heads', even if the coin is never tossed. Similarly, the conditional or unconditional expectation values of the true total effects can be considered even if the random experiment is never conducted. ◁

Remark 5.17 ($(X=x^*)$-Conditional Total Treatment Effects) The conditional total effects given a specific value x^* of the treatment variable X are often more informative than the average total effects, especially, if the conditional expectation values of the true-outcome variables *depend on* X. If, however, the expectations of the true outcomes do *not* depend on X, i. e., if $E(\tau_0|X) = E(\tau_0)$ and $E(\tau_1|X) = E(\tau_1)$, then the conditional total effects $E(\delta_{10}|X=1)$ and $E(\delta_{10}|X=0)$ do not differ between different treatment conditions 1 and 0, i. e., $E(\delta_{10}|X=0) = E(\delta_{10}|X=1) = E(\delta_{10})$. As will be shown in chapter 7, this will be the case, for instance, if X and the potential confounder σ-algebra $\mathscr{C}_X$ are independent, a condition that is created in the randomized experiment. ◁

Examples

Example 5.18 Suppose there are two treatment conditions, 'treatment' ($X=1$) and 'no treatment' ($X=0$). If the conditional total effect $E(\delta_{10}|X=1)$ is smaller than the conditional total effect $E(\delta_{10}|X=0)$, one may raise the question whether or not it would be worthwhile to change the regime of assigning units to treatment conditions. ◁

Example 5.19 Suppose we are interested in the effects of the educational program in school 1 (represented by $X=1$) vs. that of school 0 (represented by $X=0$) with respect to the outcome variable Y, say the *achievement* of its students. Because there will be no random assignment of units to schools, the students of school 1 may differ from school 0 in socio-economic status and educational sta-

tus of their parents. In this case, there might be large differences in the conditional total effects $E(\delta_{10}|X=1)$ of school 1 (vs. school 0) on its own students vs. the conditional total effects $E(\delta_{10}|X=0)$ that school 1 (vs. school 0) *would have on the students of school* 0.

Which effect should be optimized by school 1? Is it the average total effect $E(\delta_{10})$ in the total population of students of both schools or is it the conditional effect $E(\delta_{10}\,|\,X=1)$ on its own students? The answer is clear: If school 1 wishes to do the best to the kind of students it has, then it should optimize the latter. ◁

Numerical Example

Example 5.20 (Jim and Jane – continued) Consider again the example presented in Table 4.4. Because $\mathscr{C}_X = \sigma(U,Z)$, the conditional total effect of *individual therapy* $(X=1)$ vs. *no individual therapy* $(X=0)$ *given* $X=0$ can be computed by

$$\begin{aligned} &E(\delta_{10}\,|\,X=0) \\ &= \sum_u \sum_z \big[E(Y\,|\,X=1,U=u,Z=z) - E(Y\,|\,X=0,U=u,Z=z)\big] \cdot P(U=u,Z=z\,|\,X=0) \\ &= (82-68)\cdot\frac{1}{8} + (100-96)\cdot\frac{3}{8} + (98-80)\cdot\frac{1}{8} + (106-104)\cdot\frac{3}{8} \;=\; 6.25 \end{aligned}$$

[see again Eq. (5.1) and Eq. (9.19) of SN]. In contrast,

$$\begin{aligned} &E(\delta_{10}\,|\,X=1) \\ &= \sum_u \sum_z \big[E(Y\,|\,X=1,U=u,Z=z) - E(Y\,|\,X=0,U=u,Z=z)\big] \cdot P(U=u,Z=z\,|\,X=1) \\ &= (82-68)\cdot\frac{3}{8} + (100-96)\cdot\frac{1}{8} + (98-80)\cdot\frac{3}{8} + (106-104)\cdot\frac{1}{8} \;=\; 12.75 \end{aligned}$$

yields the conditional total effect of *individual therapy* $(X=1)$ vs. *no individual therapy* $(X=0)$ *given* $X=1$. In these equations, we used

$$P(U=u,Z=z\,|\,X=x) \;=\; \frac{P(X=x\,|\,U=u,Z=z)\cdot P(U=u,Z=z)}{P(X=x)}. \tag{5.10}$$

According to Equation (5.4), taking the expectation

$$E\big[E(\delta_{10}\,|\,X)\big] \;=\; \sum_x E(\delta_{10}\,|\,X=x)\cdot P(X=x) \;=\; 6.25\cdot\frac{1}{2} \;+\; 12.75\cdot\frac{1}{2} \;=\; 9.5$$

yields the average total effect. In this equation, we again used the theorem of total probability, i. e., $P(X=x) = \sum_u \sum_z P(X=x\,|\,U=u,Z=z)\cdot P(U=u,Z=z)$ (see Th. 4.25 of SN), which yields $P(X=0) = P(X=1) = 1/2$.

According to these results, the assignment regime as described by the individual treatment probabilities in Table 4.4 seems to be reasonable, because the conditional total effect $E(\delta_{10}\,|\,X=1) = 12.75$ [of treatment $(X=1)$ vs. control $(X=0)$] given treatment 1 is *greater* than the corresponding conditional total effect $E(\delta_{10}\,|\,X=0) = 6.25$ given treatment 0. ◁

5.2.4 Conditional Total Effects Given Values of X and Z

If X represents a treatment variable and Z is a mapping of the observational-unit variable U, this allows us, for instance, to ask for the effects of a treatment x vs. treatment x' given treatment x^* *in specific subpopulations* represented by the values z of Z. Hence, in this case

$$E(\delta_{xx'} \mid X{=}x^*, Z{=}z) \tag{5.11}$$

is the *conditional total effect* of x vs. x' *given* $X{=}x^*$ and $Z{=}z$, provided that $P(X{=}x^*, Z{=}z) > 0$ and that $E^{X=x}(Y|\mathscr{C}_X)$ and $E^{X=x'}(Y|\mathscr{C}_X)$ are at least $P^{X=x^*,Z=z}$-unique.

Suppose that X represents a treatment variable. If there are two treatment conditions $X{=}1$ and $X{=}0$, then, according to Equation (5.11), we may consider both, $E(\delta_{xx'} \mid X{=}1, Z{=}z)$, the *conditional total effect given treatment and* $Z{=}z$, as well as $E(\delta_{xx'} \mid X{=}0, Z{=}z)$ the *conditional total effect given control and* $Z{=}z$. If, e. g., Z is the covariate *sex*, then $E(\delta_{xx'} \mid X{=}1, Z{=}m)$ is the $(X{=}1)$-conditional total treatment effect in the *male* subpopulation, whereas $E(\delta_{xx'} \mid X{=}1, Z{=}f)$ is the $(X{=}1)$-conditional total treatment effect in the *female* subpopulation.

Example 5.21 (Jim and Jane – continued) Because $\mathscr{C}_X = \sigma(U, Z)$ in the example presented in Table 4.4, the conditional total effects of *individual therapy* $(X{=}1)$ vs. *no individual therapy* $(X{=}0)$ *given* $X{=}x$ and $Z{=}z$ can be computed by

$$\begin{aligned} E(\delta_{10} \mid X{=}x, Z{=}z) \;=\; \sum_u \sum_{z'} \big[& E(Y \mid X{=}1, U{=}u, Z{=}z') - E(Y \mid X{=}0, U{=}u, Z{=}z') \big] \\ & \cdot P(U{=}u, Z{=}z' \mid X{=}x, Z{=}z). \end{aligned}$$

If $z' \neq z$, then the conditional probabilities $P(U{=}u, Z{=}z' | X{=}x, Z{=}z)$ are zero. Otherwise they can be computed via

$$P(U{=}u, Z{=}z \mid X{=}x, Z{=}z) \;=\; \frac{P(X{=}x \mid U{=}u, Z{=}z) \cdot P(U{=}u, Z{=}z)}{P(X{=}x \mid Z{=}z) \cdot P(Z{=}z)}, \tag{5.12}$$

where

$$P(X{=}x \mid Z{=}z) \;=\; \sum_u P(X{=}x \mid U{=}u, Z{=}z) \cdot P(U{=}u \mid Z{=}z), \tag{5.13}$$

with $P(U{=}u|Z{=}z) = P(Z{=}z|U{=}u) \cdot P(U{=}u) / P(Z{=}z)$. In this example, $P(U{=}u \mid Z{=}z) = 1/2$, for all values of U and Z. Hence, for $X{=}0$ and $Z{=}0$, Equation (5.13) yields $P(X{=}0 \mid Z{=}0) = 1/4 \cdot 1/2 + 1/4 \cdot 1/2 = 1/4$, and using Equation (5.12) we receive:

$$P(U{=}u, Z{=}0 \mid X{=}0, Z{=}0) \;=\; \frac{1/4 \cdot 1/4}{1/4 \cdot 1/2} \;=\; \frac{1}{2},$$

for $U{=}Jim$ and for $U{=}Jane$. Hence, the equation for $E(\delta_{10} \mid X{=}x, Z{=}z)$ yields

$$E(\delta_{10} \mid X{=}0, Z{=}0) \;=\; (82-68) \cdot \frac{1}{2} + (100-96) \cdot 0 + (98-80) \cdot \frac{1}{2} + (106-104) \cdot 0$$

$$= 16$$

for $X=0$ and $Z=0$. For $X=1$ and $Z=0$, we receive

$$\begin{aligned} E(\delta_{10}\,|\,X=1,Z=0) &= (82-68)\cdot\frac{1}{2}+(100-96)\cdot 0+(98-80)\cdot\frac{1}{2}+(106-104)\cdot 0 \\ &= 16, \end{aligned}$$

for $X=0$ and $Z=1$, we receive

$$\begin{aligned} E(\delta_{10}\,|\,X=0,Z=1) &= (82-68)\cdot 0+(100-96)\cdot\frac{1}{2}+(98-80)\cdot 0+(106-104)\cdot\frac{1}{2} \\ &= 3, \end{aligned}$$

and for $X=1$ and $Z=1$:

$$\begin{aligned} E(\delta_{10}\,|\,X=1,Z=1) &= (82-68)\cdot 0+(100-96)\cdot\frac{1}{2}+(98-80)\cdot 0+(106-104)\cdot\frac{1}{2} \\ &= 3. \end{aligned}$$

Hence, in this special example, the conditional total effects $E(\delta_{10}|Z=z)$ and $E(\delta_{10}|X=x,Z=z)$ are identical.

According to Equation (5.4), taking the expectation

$$\begin{aligned} E\big[E(\delta_{10}\,|\,X,Z)\big] &= \sum_x\sum_z E(\delta_{10}\,|\,X=x,Z=z)\cdot P(X=x,Z=z) \\ &= 16\cdot\frac{1}{4}+3\cdot\frac{1}{4}+16\cdot\frac{1}{4}+3\cdot\frac{1}{4}=9.5, \end{aligned}$$

yields the average total effect. In this equation, we again used the theorem of total probability, i. e., $P(X=x,Z=z)=\sum_u P(X=x,Z=z\,|\,U=u)\cdot P(U=u)$ (see Th. 4.25 of SN), which yields $P(X=x,Z=z)=1/4$, for all values of X and Z. ◁

5.3 Average and Conditional Direct and Indirect Effects

Once we have identified the average and/or conditional total effect of X on Y, we may raise questions about the causal mechanisms leading to this effect. Neither the average nor the conditional total effects are informative for a number of important questions. Which are the intermediate variables transmitting the total effects from X to Y? Is there a *direct* effect of X on Y that is *not* transmitted by a specified intermediate variable M? In other words, is there still a causal effect of X on Y if, aside from the potential confounder σ-algebra $\mathscr{C}_X$, we also control for M and all other variables that are in between X and M or simultaneous to M?

Remark 5.22 (Atomic Direct-Effect and Atomic Indirect-Effect Variables) Instead of constructing the concepts of direct effect on the total-effect true-outcome variables τ_x, we will now built our concepts on the *t-direct-effect true-outcome variables*

$$\tau_{x,t} := E^{X=x}(Y|\mathscr{C}_{X,t}), \tag{5.14}$$

where $t \in T$ with $t_X \leq t < t_Y$. Just like the total-effect true-outcome variables τ_x, the variables $\tau_{x,t}$ are unbiased by definition, this time, unbiased with respect to t-direct effects, because we control for *all* covariates that are comprised in $\mathscr{C}_{X,t}$ (see section 4.2.4). The atomic t-direct effects, i. e., the values of

$$\delta_{xx',t} = \tau_{x,t} - \tau_{x',t}$$

may be different for different combinations of values of the covariates and the intermediate variables, and the same applies to the *atomic t-indirect-effect variable*

$$\delta_{xx'} - \delta_{xx',t},$$

which has also been introduced in Definition 4.20.

Just like the atomic total-effect variable $\delta_{xx'}$, the atomic direct-effect and atomic indirect-effect variables are of a purely theoretical nature. Estimating their values is not possible in many applications, because of the 'fundamental problem of causality' that we cannot observe the same units under different values x and x' of X (cf. Holland, 1986). Nevertheless, under certain assumptions, it *is* possible to estimate their expectations or conditional expectation values that are considered in the following definition. In point (ii) of the following definition we relax the assumption that $\tau_{x,t} = E^{X=x}(Y|\mathscr{C}_{X,t})$ and $\tau_{x',t} = E^{X=x'}(Y|\mathscr{C}_{X,t})$ are P-unique. ◁

Definition 5.23 (Average and Conditional Direct and Indirect Effects)
Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a causality space, let Y be numerical and nonnegative or with finite expectation $E(Y)$, let x and x' be two values of X with $P(X=x)$, $P(X=x') > 0$, and let $t \in T$ with $t_X \leq t < t_Y$.

(i) If $\tau_{x,t}$ and $\tau_{x',t}$ are P-unique, then the average t-direct effect of x vs. x' on Y is defined by

$$E\big(\delta_{xx',t}\big), \tag{5.15}$$

where $\delta_{xx',t} := \tau_{x,t} - \tau_{x',t}$. Furthermore, the average t-indirect effect of x vs. x' on Y is defined by

$$E\big(\delta_{xx'} - \delta_{xx',t}\big), \tag{5.16}$$

where $\delta_{xx'} = \tau_x - \tau_{x'}$.

(ii) Let V be a random variable on $(\Omega, \mathscr{A}, P)$, let v be a value of V with $P(V=v) > 0$, and let $\tau_{x,t}$ and $\tau_{x',t}$ be $P^{V=v}$-unique. Then the $(V=v)$-conditional t-direct effect of x vs. x' on Y is defined by

$$E\big(\delta_{xx',t} \,\big|\, V=v\big), \tag{5.17}$$

and the $(V=v)$-conditional t-indirect effect of x vs. x' on Y by

$$E\big(\delta_{xx'} - \delta_{xx',t} \,\big|\, V=v\big). \tag{5.18}$$

Remark 5.24 (Implication of P-Uniqueness) Note that P-uniqueness of $\tau_{x,t} = E^{X=x}(Y|\mathscr{C}_{X,t})$ implies P-uniqueness of $E^{X=x}(Y|\mathscr{C}_X)$ (see Remark 4.21). Similarly, $P^{V=v}$-uniqueness of $E^{X=x}(Y|\mathscr{C}_{X,t})$ implies $P^{V=v}$-uniqueness of $E^{X=x}(Y|\mathscr{C}_X)$ for each value x of X with $P(X=x) > 0$. Therefore, we only have to assume $P^{V=v}$-uniqueness of $E^{X=x}(Y|\mathscr{C}_{X,t})$ in Definition 5.23 (ii). ◁

Remark 5.25 (Examples) The variable V can be any random variable on the probability space $(\Omega, \mathscr{A}, P)$. Hence, we may consider, the *conditional t-direct effect*

$$E\big(\delta_{xx',t}\,|\,Z=z\big) \tag{5.19}$$

of x vs. x' given the value z of a *pre-treatment variable* Z, the *individual t-direct effect*

$$E\big(\delta_{xx',t}\,|\,U=u\big) \tag{5.20}$$

of x vs. x' given the observational unit u, the *$(M=m)$-conditional t-direct effect*

$$E\big(\delta_{xx',t}\,|\,M=m\big) \tag{5.21}$$

of x vs. x' given the value m of an *intermediate variable* M, etc., and the same can be done for the *conditional t-indirect effects*, simply replacing the variable $\delta_{xx',t}$ by the difference $\delta_{xx'} - \delta_{xx',t}$. Unless we can assume P-uniqueness of $E^{X=x}(Y|\mathscr{C}_{X,t})$ and $E^{X=x'}(Y|\mathscr{C}_{X,t})$ we have to make the appropriate uniqueness assumption in each of these specific cases. Instead of listing all these effects, we pick out some and discuss some general issues. ◁

Remark 5.26 (Average t-Direct and t-Indirect Effects) As mentioned before, the true t-direct effects may be different for different combinations of values of covariates and intermediate variables. Hence, taking the conditional or unconditional expectation values of $\delta_{xx',t}$ may be meaningful for at least two reasons:

(a) to reduce complexity, and
(b) to make possible reliable estimation in (relatively) small samples.

Which level of aggregation we choose is, in principle, at our disposal. However, in empirical applications, sample sizes oftentimes set close limits.

The strongest simplification is made if we consider the average t-direct effect $E\big(\delta_{xx',t}\big)$, which is one single number: the expectation of the true t-direct effects. Because it is an expectation, in general, this number gives no information about the conditional t-direct effects given values of M and/or values of a covariate, for instance. ◁

Remark 5.27 (Decomposition of the Average Total Effect) The average t-indirect effect of x vs. x' is a single number as well. Obviously, if $E^{X=x}(Y|\mathscr{C}_X)$ and $E^{X=x'}(Y|\mathscr{C}_X)$ as well as $E^{X=x}(Y|\mathscr{C}_{X,t})$ and $E^{X=x'}(Y|\mathscr{C}_{X,t})$ are P-unique, then

$$E\big(\delta_{xx'} - \delta_{xx',t}\big) \;=\; E\big(\delta_{xx'}\big) - E\big(\delta_{xx',t}\big)\,,$$

which shows that the average total effect $E\big(\delta_{xx'}\big)$ is the sum of the average t-direct effect $E\big(\delta_{xx',t}\big)$ and the average t-indirect effect, i. e.,

$$E(\delta_{xx'}) \;=\; E(\delta_{xx',t}) + E(\delta_{xx'} - \delta_{xx',t}). \tag{5.22}$$

As mentioned before, the idea of decomposing a total effect into the sum of a direct effect and an indirect effect goes back to the early papers of Sewall Wright (1918, 1921, 1923)) on path analysis. In chapter 10 we show how to identify average t-direct effects and t-indirect effects. ◁

Remark 5.28 (Conditional t-Direct and t-Indirect Effects) If M is an intermediate variable and we consider the conditional t-direct effect $E(\delta_{xx',t} \mid M{=}m)$, there might be a different t-direct effect for each value m of the intermediate variable M. (Note that we have to presume that $E^{X=x}(Y|\mathscr{C}_{X,t})$ and $E^{X=x'}(Y|\mathscr{C}_{X,t})$ are $P^{M=m}$-unique for each of the values of M considered.) Hence, if X is a dichotomous treatment variable that may affect the intermediate variable M *post-treatment motivation to learn*, then we might be interested in comparing the t-direct effect of X on Y *(aptitude)* when post-treatment motivation is high ($M{=}m_1$) to the t-direct effect of X on Y when post-treatment motivation is low ($M{=}m_2$). Similarly, considering $E(\delta_{xx',t} \mid Z{=}z)$, there might be a different t-direct effect for each value z of the covariate Z *(pre-treatment motivation to learn)*. (Note again that we have to presume that $E^{X=x}(Y|\mathscr{C}_{X,t})$ and $E^{X=x'}(Y|\mathscr{C}_{X,t})$ are $P^{Z=z}$-unique for each of the values of Z considered.) In this case we might be interested in comparing the t-direct effect of X on Y when pre-treatment motivation is high ($Z{=}z_1$) to the t-direct effect of X on Y when pre-treatment motivation is low ($Z{=}z_2$). Finally, the conditional t-direct effect $E(\delta_{xx',t} \mid X{=}x^*)$, informs about the conditional t-direct effect of X given the value x^* of X. Hence, it might be interesting to compare $E(\delta_{xx',t} \mid X{=}x)$ to $E(\delta_{xx',t} \mid X{=}x')$, for instance. ◁

Remark 5.29 (Decomposition of Conditional Total Effects) Assume that $E^{X=x}(Y|\mathscr{C}_{X,t})$ and $E^{X=x'}(Y|\mathscr{C}_{X,t})$ are $P^{V=v}$-unique. Then, for a given value v of V, the conditional t-indirect effect $E(\delta_{xx'} - \delta_{xx',t} | V{=}v)$ of x vs. x' is a single number. Obviously,

$$E(\delta_{xx'} - \delta_{xx',t} \mid V{=}v) \;=\; E(\delta_{xx'} \mid V{=}v) \;-\; E(\delta_{xx',t} \mid V{=}v)\,.$$

This equation shows that the conditional total effect $E(\delta_{xx'} \mid V{=}v)$ is the sum of the conditional t-direct effect $E(\delta_{xx',t} \mid V{=}v)$ and the conditional t-indirect effect $E(\delta_{xx'} - \delta_{xx',t} \mid V{=}v)$. ◁

Remark 5.30 (Other Conditional t-Direct and t-Indirect Effects) In Remark 5.28 we discussed the most important conditional t-direct effects. Note, however, that we may also condition on the values of the multidimensional variables (M,Z), (X,Z), (X,M), and (X,M,Z). In the example used in Remark 5.28, conditioning, for instance, on the values (m,z) of (M,Z) would inform about the conditional direct of X on Y for a specific value m (say 'high') of post-treatment motivation and a specific value z (say 'low') of pre-treatment motivation. Furthermore, note that the conditional expectations such as $E(\delta_{xx',t} \mid M)$ and $E(\delta_{xx',t} \mid Z)$ represent conditional t-direct-effect functions whose values have been discussed above. If we assume that $E^{X=x}(Y|\mathscr{C}_{X,t})$ and $E^{X=x'}(Y|\mathscr{C}_{X,t})$ are P-unique, then taking the expectations $E\big[E(\delta_{xx',t} \mid M)\big]$ and $E\big[E(\delta_{xx',t} \mid Z)\big]$ yields the average t-direct effect $E(\delta_{xx',t})$. ◁

Box 5.1 Glossary of New Concepts

The effects and effect functions listed below are only defined under appropriate uniqueness assumptions. All effects and effect functions for total effects are well-defined if we assume P-uniqueness of the conditional expectations $\tau_x = E^{X=x}(Y|\mathscr{C}_X)$ for $x = 0, 1, \ldots, J$. All effects and effect functions for t-direct effects are well-defined if we assume P-uniqueness of the conditional expectations $\tau_{x;t} = E^{X=x}(Y|\mathscr{C}_{X,t})$ for $x = 0, 1, \ldots, J$. For the $(V=v)$-conditional effects it suffices to assume $P^{V=v}$-uniqueness of τ_x and $\tau_{x;t}$, respectively, for $x = 0, 1, \ldots, J$.

	Total Effects
$E(\delta_{xx'})$	*Average total effect of x vs. x'.*
$E(\delta_{xx'}\,\vert\,V=v)$	*$(V=v)$-conditional total effect of x vs. x'.*
$E(\delta_{xx'}\,\vert\,V)$	*V-conditional total-effect function of x vs. x'.*
	Direct Effects
$E\big(\delta_{xx';t}\big)$	*Average t-direct effect of x vs. x'.*
$E\big(\delta_{xx';t}\,\big\vert\,V=v\big)$	*$(V=v)$-conditional t-direct effect of x vs. x'.*
$E\big(\delta_{xx';t}\,\big\vert\,V\big)$	*V-conditional t-direct-effect function of x vs. x'.*
	Indirect Effects
$E\big(\delta_{xx'}-\delta_{xx';t}\big)$	*Average t-indirect effect of x vs. x'.*
$E\big(\delta_{xx'}-\delta_{xx';t}\,\big\vert\,V=v\big)$	*$(V=v)$-conditional t-indirect effect of x vs. x'.*
$E\big(\delta_{xx'}-\delta_{xx';t}\,\big\vert\,V\big)$	*V-conditional t-indirect-effect function of x vs. x'.*

5.4 Summary and Conclusions

In this chapter we defined several kinds of total effects based on the *true total-effect variable* $\delta_{xx'} := \tau_x - \tau_{x'}$, where τ_x is the total-effect true-outcome variable defined in chapter 4. The term 'total' is used in order to convey the idea that this effect variable can be decomposed into a true direct-effect variable and a true indirect-effect variable. The *average total effect* was then defined as the expectation $E(\delta_{xx'})$ of the atomic total-effect variable. Similarly, we also defined the *conditional total effect* given a value v of a random variable V. Examples of V are the observational-unit variable U, a pre-treatment variable Z, the treatment variable X, and an intermediate variable M. All these conditional total effects are defined as conditional expectation values of the atomic total-effect variable $\delta_{xx'}$.

The definitions of the average and conditional total effects rest on the presumption that there are not only a cause X and an outcome variable Y, but also a filtration $(\mathscr{F}_t, t \in T)$ of σ-algebras with respect to which the variables considered are prior or simultaneous to each other. If these components of the causality space are specified, the theory of total effects is also applicable for causal modeling beyond experiments and quasi-experiments. For instance, the different kinds

of total effects are also defined for an intermediate variable M taking the role of X, provided that M is discrete.

Average Total Effects

Often we have to content ourselves with the average total effect or one kind of the conditional total effects discussed. Note, however, that an average total effect may not apply to any unit at all. There might very well be cases in which half of the units have positive individual total effects and the other half negative ones. The average total effect can then be zero. This is not a paradox but the nature of an average. Also remember that an average total effect is already much more informative for causal inference than ordinary true mean differences, the prima facie effects $E(Y \mid X=1) - E(Y \mid X=0)$ considered in chapter 1. These prima facie effects have no causal interpretation at all, unless specific assumptions are made, which will be studied in detail in chapters 6 to 9.

Main Effects vs. Conditional Effects

Conceptually, the *average total effect* is what is tested as the *main effect* of the 'treatment factor' in orthogonal analysis of variance, provided that the data are sampled in a perfect randomized experiment. Note that the definition of the average total effect is unique even if there are inter-individual differences in the individual total effects, and even if there is interaction between X and a covariate Z in the sense that the effect of X depends on the values of Z. The average total effect is uniquely defined even if X and Z are correlated and/or stochastically dependent. If Z is a qualitative covariate, it is considered a second 'factor' in analysis of variance, and the average total effects are what we test as the *main effect* of the 'treatment factor'. However, only in the randomized experiment we can be sure that the main effects in analysis of variance estimate the average total effects. In other cases, the main effects will be different for different covariates, and this is true even in orthogonal designs.

Of course, the conditional effects given the values of a covariate are usually more informative than their average, i. e., than the average total effect; but sometimes averaging is useful in order to avoid information overload, and sometimes we may be able to estimate precisely enough only the average effect, but not the conditional effects, e. g., because of small sample sizes. If the covariate is a non-numerical random variable with a few number of values, it is often considered a second factor in analysis of variance. In this case, the $(Z=z)$-conditional total effects are often called the 'simple main effects' (see, e. g., Woodward & Bonett, 1991).

Pre-Facto vs. Counterfactual Perspective

Note that our definitions of the various kinds of total effects just use concepts of probability theory. No concepts had to be borrowed from philosophy or any

other science — although the basic idea goes back at least to Mill (1843/1865). We did not take a counterfactual but a *pre-facto perspective*, which is the perspective taken in *every* probabilistic model. In our theory, total effects are parameters, just in the same way as the probability of flipping 'heads' is a parameter about which we can talk *before* the coin is tossed and even if the coin is *never* tossed. Therefore, it is also meaningful to talk about the individual effects of a treatment for a unit which is actually never treated, and about the *average effect of a treatment* including also those that are not treated. It is even meaningful to talk about the *conditional effect of a treatment given control*.

Direct and Indirect Effects, and the Decomposition of Total Effects

While the total effects mentioned above are built on the true-outcome variables $\tau_x := E^{X=x}(Y|\mathscr{C}_X)$ with respect to total effects, the various concepts of *direct effects* with respect to an intermediate variable M have been built on the t-direct-effect true-outcome variables $\tau_{x,t} := E^{X=x}(Y|\mathscr{C}_{X,t})$ and their differences $\delta_{xx',t} = \tau_{x,t} - \tau_{x',t}$. The definition of $\tau_{x,t}$ shows that we not only condition on the potential confounder σ-algebra $\mathscr{C}_X$ of X — thus controlling for all covariates — but also on all intermediate variables that are in between X and an intermediate variable M and all variables that are simultaneous to M.

We also considered *indirect effects* with respect to time t. The difference $\delta_{xx'} - \delta_{xx',t}$ between the true total-effect variables and the true direct-effect variables has been defined to be the *true indirect-effect variable* with respect to t. It is that part of the true total-effect variable $\delta_{xx'}$ that is not due to the true direct effects, the values of $\delta_{xx',t}$. Taking the expectation $E(\delta_{xx'} - \delta_{xx',t})$ yields the average indirect effect. It is that part of the average total effect $E(\delta_{xx'})$ that we receive after subtracting the average direct effect $E(\delta_{xx',t})$. Taking the conditional expectation values $E(\delta_{xx'} - \delta_{xx',t} \,|\, V=v)$ yields various kinds of conditional t-indirect effects. Again, their meaning depends on the variable V and, of course, on the time point t.

Outlook

Note that all concepts introduced in this chapter such as the *true total effects*, *average total effects*, *conditional total effects*, *average direct effects*, *conditional direct effects*, etc. are of a purely theoretical nature. They explicate what exactly we are looking for when we ask for the effects, e. g., of a treatment variable or of another discrete cause. This also applies to the examples treated in this chapter. For example, Tables 4.2 to 4.5 do not show data that might be obtained in a sample. Instead they contain the theoretical parameters we would like to estimate from sample data. In terms of the metaphor presented in the preface, they are the *size* of the invisible man. In contrast, in chapter 1, we only dealt with the prima facie effects: (a) ordinary conditional expectation values $E(Y|X=x)$ of an outcome variable Y given treatment x, (b) conditional expectation values $E(Y|X=x, Z=z)$ of the outcome variable given treatment x and value z of the covariate Z, (c) differ-

ences between these (conditional) expectation values, the (conditional) *prima facie effects*, and (d) averages over these conditional prima facie effects. The conditional expectation values $E(Y \mid X{=}x)$, $E(Y \mid X{=}x, Z{=}z)$, $E(Y \mid X{=}x, M{=}m)$, and also $E(Y \mid X{=}x, Z{=}z, M{=}m)$, are easily estimated under the usual assumptions made for a sample, such as the assumption of independent, identically distributed observations. However, they are only like the length of the invisible man's shadow; depending on the angle of the sun, they can be seriously biased if mistaken for the size of the invisible man itself.

But which is the relationship between the prima facie effects (the shadow) and the causal effects (the size of the invisible man)? Is it possible to make inferences from the conditional and unconditional prima facie effects to the conditional and unconditional total, direct, and indirect effects? And if 'yes', which are the necessary assumptions? These are the questions dealt with in the chapters 6 to 9.

5.5 Exercises

▷ **Exercise 5-1** Suppose X is a treatment variable and Y an outcome variable. Why are the conditional expectation values $E(Y \mid X{=}x)$ and their differences $E(Y \mid X{=}x) - E(Y \mid X{=}x')$, the prima facie effects, often useless in the evaluation of treatment effects?

▷ **Exercise 5-2** Suppose X is a treatment variable and Y an outcome variable. If the conditional expectation values $E(Y \mid X{=}x)$ and their differences $E(Y \mid X{=}x) - E(Y \mid X{=}x')$ do not represent the treatment effects we are interested in, then what *are* the treatment effects we would like to study?

▷ **Exercise 5-3** What is the *average total effect* $E(\delta_{xx'})$ of x vs. x' on outcome variable Y?

▷ **Exercise 5-4** What is the *conditional total effect* of x vs. x' on outcome variable Y *given the value z of a covariate Z*?

▷ **Exercise 5-5** What is the *conditional total effect* of x vs. x' on outcome variable Y *given treatment x^**?

▷ **Exercise 5-6** Suppose the value 0 of X denotes 'no treatment'. Which is the meaning of the term $E(\delta_{10} \mid X{=}0)$?

▷ **Exercise 5-7** What is the *conditional total effect* of x vs. x' on the outcome variable Y *given treatment x^* and value z of the covariate Z*?

▷ **Exercise 5-8** Suppose $E(\delta_{10} \mid U{=}u) = 10$ and $E(\delta_{20} \mid U{=}u) = 7$ for unit u. Which is the individual total effect $E(\delta_{12} \mid U{=}u)$?

▷ **Exercise 5-9** Compute the average total effect $E(\delta_{10})$ for the random experiment presented in Table 4.4.

▷ **Exercise 5-10** Compute the conditional total effect $E(\delta_{10} \mid Z{=}0)$ given *no group therapy* for the random experiment presented in Table 4.4.

▷ **Exercise 5-11** Let Z represent *sex* with values m (males) and f (females). Furthermore, suppose $E(\delta_{20} \mid Z{=}m) = 11$, $E(\delta_{20} \mid Z{=}f) = 5$, $P(Z{=}m) = 1/3$, and $P(Z{=}f) = 2/3$. Which is the average total effect $E(\delta_{20})$?

Solutions

▷ **Solution 5-1** Certain other variables, the covariates, may determine both the probability of being treated *and* the conditional expectation values of the outcome variable. This implies that the conditional expectation values $E(Y \mid X{=}x)$ and their differences $E(Y \mid X{=}x) - E(Y \mid X{=}x')$ are biased, and this means that they do not represent the treatment effects to be studied. An example of such a covariate is *severity of the symptoms*. If there is self-selection or if there is systematic selection to treatment by experts that is also determined by the severity of the symptoms, then the variable *severity of the symptoms* will both affect the treatment probability and the conditional expectation values of the outcome variable (e. g., *severity of the symptoms after treatment*). Simpson's paradox presented in chapter 1 is another example.

▷ **Solution 5-2** The basic idea is to consider the treatment effects in the most fine-grained strata — called the *true treatment effects* — defined by controlling for *all* covariates and then taking the expectation over the distribution of all covariates. Taking the expectation is useful in order to reduce the complexity involved considering *all* covariates. The mathematical version of this idea is to define the average total treatment effect by $E(\tau_x - \tau_{x'})$, assuming that τ_x and $\tau_{x'}$ are P-unique. Instead of (or in addition to) taking the unconditional expectation of the treatment effects in the most fine-grained strata, we may also consider various kinds of *conditional* expectations of these true treatment effects.

▷ **Solution 5-3** The average total effect $E(\delta_{xx'})$ of x vs. x' on the outcome variable Y is the expectation of the true-total-effect variable $\delta_{xx'}$, i. e. $E(\delta_{xx'})$.

▷ **Solution 5-4** The conditional total effect of x vs. x' on the outcome variable Y given the value z of the covariate Z is the $(Z{=}z)$-conditional expectation value of the true-total-effect variable $\delta_{xx'}$, i. e., $E(\delta_{xx'} \mid Z{=}z)$. If z represents a subpopulation such as males, then the conditional expectation value $E(\delta_{xx'} \mid Z{=}z)$ is the average total effect in this subpopulation.

▷ **Solution 5-5** The conditional total effect of x vs. x' on the outcome variable Y given treatment x^* is the $(X{=}x^*)$-conditional expectation value of the true-total-effect variable $\delta_{xx'}$, i. e., $E(\delta_{xx'} \mid X{=}x^*)$. If X represents a treatment variable, $x{=}x^*$, and $X{=}0$ represents an untreated control group, then $E(\delta_{x0} \mid X{=}x)$ is the average total effect of treatment x vs. control *given treatment* x. If $x^*{=}0$, then $E(\delta_{x0} \mid X{=}0)$ is the average total effect of treatment x vs. control *given control* (see also Exercise 5-6).

▷ **Solution 5-6** The term $E(\delta_{10} \mid X{=}0)$ denotes the $(X{=}0)$-conditional total effect of treatment 1 vs. the control. Although this sounds paradoxical, this term is meaningful and well-defined, because the true-total-effect variables and their (conditional) expectations refer to a random experiment to be conducted *in the future*. This means that they are well-defined, even if the experiment is not yet conducted, or is never conducted (see section 5.2.3 for more details).

▷ **Solution 5-7** The conditional total effect of x vs. x' on the outcome variable Y given treatment x^* and value z of the covariate Z is the $(X{=}x^*, Z{=}z)$-conditional expectation value of the true effects, i. e., $E(\delta_{xx'} \mid X{=}x^*, Z{=}z)$. If X represents a treatment variable, $x^*{=}x$, the value 0 of X represents an untreated control group, and m is the value 'male' of the covariate *sex*, then $E(\delta_{x0} \mid X{=}x, Z{=}m)$ is the average total effect of treatment x vs. control given treatment x in the male subpopulation. If $x^*{=}0$, then $E(\delta_{x0} \mid X{=}0, Z{=}m,)$ is the average total effect of treatment x vs. control given control in the male subpopulation.

▷ **Solution 5-8**

$$E(\delta_{10}\,|\,U{=}u) \;=\; E(\tau_1-\tau_0\,|\,U{=}u) = E(\tau_1\,|\,U{=}u) - E(\tau_0\,|\,U{=}u) \;=\; 10,$$

$$E(\delta_{20}\,|\,U{=}u) \;=\; E(\tau_2\,|\,U{=}u) - E(\tau_0\,|\,U{=}u) \;=\; 7,$$

and

$$E(\delta_{12}\,|\,U{=}u) \;=\; E(\tau_1\,|\,U{=}u) - E(\tau_2\,|\,U{=}u) \;=\; E(\delta_{10}\,|\,U{=}u) - E(\delta_{20}\,|\,U{=}u) \;=\; 10-7 \;=\; 3.$$

▷ **Solution 5-9**

$$\begin{aligned} E(\delta_{10}) \;&=\; \sum_u \sum_z \big[E(Y\,|\,X{=}1,U{=}u,Z{=}z) - E(Y\,|\,X{=}0,U{=}u,Z{=}z)\big]\cdot P(U{=}u,Z{=}z) \\ &=\; \big[(82-68)+(100-96)+(98-80)+(106-104)\big]\cdot\frac{1}{4} \;=\; 9.5. \end{aligned}$$

▷ **Solution 5-10**

$$\begin{aligned} E(\delta_{10}\,|\,Z{=}0) \;&=\; \sum_u \sum_z \big[E(Y\,|\,X{=}1,U{=}u,Z{=}z) - E(Y\,|\,X{=}0,U{=}u,Z{=}z)\big]\cdot P(U{=}u,Z{=}z\,|\,Z{=}0) \\ &=\; (82-68)\cdot\frac{1}{2} + (100-96)\cdot 0 + (98-80)\cdot\frac{1}{2} + (106-104)\cdot 0 \;=\; 16. \end{aligned}$$

▷ **Solution 5-11** Using Equation (5.4), we can compute the average total effect as follows:

$$E(\delta_{20}) \;=\; E(\delta_{20}\,|\,Z{=}m)\cdot\frac{1}{3} + E(\delta_{20}\,|\,Z{=}f)\cdot\frac{2}{3} \;=\; 11\cdot\frac{1}{3} + 5\cdot\frac{2}{3} \;=\; 7.$$

Chapter 6

Unbiasedness

In chapter 5 we defined various kinds of average and conditional total, direct, and indirect effects. In this chapter we will define *unbiasedness* of various conditional expectation values and prima facie effects (a) *with respect to total effects* and (b) *with respect to direct effects*. Furthermore, we will study how these prima facie effects are related to their corresponding causal effects. The *unbiasedness conditions* are the first of several kinds of causality conditions, which, together with the structural components listed in a causality space, distinguish causal stochastic dependencies from ordinary stochastic dependencies.

Overview

We start defining unbiasedness of the conditional expectation values $E(Y|X{=}x)$ and $E(Y|X{=}x,Z{=}z)$ *with respect to total effects (τ_x-unbiasedness)*, presuming that X is discrete. Next, we illustrate these concepts by some numerical examples. Then we show how these conditional expectation values and prima facie effects are related to the corresponding conditional expectation values of the true-outcome variables with respect to total effects. Next we study the implications of these relationships for the prima facie effects and the corresponding causal effects. Then we show that unbiasedness can be accidental, presenting an example in which the conditional expectation values $E(Y|X{=}x)$ are τ_x-unbiased, whereas the conditional expectation values $E(Y|X{=}x,Z{=}z)$ are not. Then we introduce the concept of unbiasedness of conditional expectation values *with respect to t_M-direct effects (τ_{x,t_M}-unbiasedness)* and illustrate it by another example.

6.1 Unbiasedness With Respect to Total Effects

In this section we will introduce the concepts of unbiasedness *with respect to total effects* for conditional expectation values and prima facie effects. In section 6.3, we will present the corresponding concepts *with respect to t_M-direct effects*, referring to an intermediate variable M.

6.1.1 τ_x-Unbiasedness of Conditional Expectations

Reading the following definition, remember that $\tau_x := E^{X=x}(Y|\mathscr{C}_X)$ denotes the conditional expectation of Y given the potential-confounder σ-algebra $\mathscr{C}_X$ with respect to the conditional-probability measure $P^{X=x}$. If, for example, X is a treatment variable, then $E^{X=x}(Y|\mathscr{C}_X)$ is the conditional expectation of Y given $\mathscr{C}_X$ in treatment x (for details, see ch. 14 of SN). Furthermore, note that we call a random variable $X\colon (\Omega,\mathscr{A},P) \to (\Omega'_X,\mathscr{A}'_X)$ *discrete* if $\{x\} \in \mathscr{A}'_X$ for all $x \in \Omega'_X$ and if there is a finite or countable set $\Omega'_0 \subset \Omega'_X$ such that $P_X(\Omega'_0) = 1$ (see Def. 5.56 of SN).

Definition 6.1 (τ_x-Unbiasedness of Conditional Expectations)
Let $\langle (\Omega,\mathscr{A},P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a causality space, let Y be numerical, non-negative or with finite expectation $E(Y)$, and let $x \in \Omega'_X$ be a value of X with $P(X{=}x) > 0$.

(i) $E(Y|X{=}x)$ is called τ_x-unbiased if $E^{X=x}(Y|\mathscr{C}_X)$ is P-unique and

$$E(Y|X{=}x) \;=\; E(\tau_x). \tag{6.1}$$

(ii) Let X be discrete. Then $E(Y|X)$ is called τ_x-unbiased if the conditional expectation values $E(Y|X{=}x)$ are τ_x-unbiased for all values $x \in \Omega'_X$ for which $P(X{=}x) > 0$.

Remark 6.2 (Unbiased Estimators vs. Unbiased Parameters) Unbiasedness in statistics usually refers to *estimators* of a parameter. However, if the expectation of the estimator is not identical to the parameter to be estimated, then the expectation of the estimator itself is also biased. Although the sample mean differences are unbiased estimators of $E(Y|X{=}x) - E(Y|X{=}x')$, they can be biased estimators of the corresponding average total effect $E(\delta_{xx'}) := E(\tau_x) - E(\tau_{x'})$. Hence, the difference $E(Y|X{=}x) - E(Y|X{=}x')$ is also biased in the sense that it is not identical to the parameter $E(\delta_{xx'})$ (see Theorem 6.27). (In this context we use the term *$\delta_{xx'}$-unbiasedness* see Def. 6.6.) ◁

Remark 6.3 (Expectations With Respect to $P^{X=x}$) Under the assumptions made in Definition 6.1,

$$E(Y|X{=}x) \;=\; E^{X=x}(Y) \tag{6.2}$$

(see Cor. 9.5 of SN). Hence the conditional expectation value $E(Y|X{=}x)$ is τ_x-unbiased if and only if the expectation $E^{X=x}(Y)$ of Y with respect to the conditional-probability measure $P^{X=x}$ is τ_x-unbiased. ◁

Before considering prima facie effects let us extend τ_x-unbiasedness to conditioning on a random variable Z. Although Z is usually a covariate, we do not have to make this assumption in the following definition. Note that $Z := (Z_1,\ldots,Z_m)$ is a random variable on $(\Omega,\mathscr{A},P)$ if and only if $Z_1,\ldots,Z_m$ are random variables on

$(\Omega, \mathscr{A}, P)$ (see Th. 2.38 and Def. 5.1 of SN). Also remember that two random variables V_1 and V_2 on a probability space $(\Omega, \mathscr{A}, P)$ are called *P-equivalent*, denoted $V_1 \underset{P}{=} V_2$, if $P\big(\{\omega \in \Omega\colon V_1(\omega) \neq V_2(\omega)\}\big) = 0$.

Definition 6.4 (τ_x-Unbiasedness of Conditional Expectations)
Let the assumptions of Definition 6.1 be true and let Z be a random variable on $(\Omega, \mathscr{A}, P)$.

(i) *$E^{X=x}(Y|Z)$ is called τ_x-unbiased if $E^{X=x}(Y|\mathscr{C}_X)$ is P-unique and*

$$E^{X=x}(Y|Z) \underset{P}{=} E(\tau_x|Z). \tag{6.3}$$

(ii) *Let X be discrete. Then $E(Y|X,Z)$ is called τ_x-unbiased if $E^{X=x}(Y|Z)$ is τ_x-unbiased for all values $x \in \Omega'_X$ for which $P(X{=}x) > 0$.*

(iii) *Furthermore, let z be a value of Z such that $P(X{=}x, Z{=}z) > 0$. Then the conditional expectation value $E(Y|X{=}x, Z{=}z)$ is called τ_x-unbiased if $E^{X=x}(Y|\mathscr{C}_X)$ is $P^{Z=z}$-unique and*

$$E(Y|X{=}x, Z{=}z) = E(\tau_x|Z{=}z). \tag{6.4}$$

Remark 6.5 (Unbiasedness of a Conditional Expectation Value) If $P(X{=}x, Z{=}z) > 0$, then

$$E(Y|X{=}x, Z{=}z) = E^{X=x}(Y|Z{=}z) \tag{6.5}$$

(see Exercise **??**). This equation implies that the $(X{=}x, Z{=}z)$-conditional expectation value $E(Y|X{=}x, Z{=}z)$ of Y is τ_x-unbiased if and only if $E^{X=x}(Y|Z{=}z)$, the $(Z{=}z)$-conditional expectation value of Y with respect to $P^{X=x}$, is τ_x-unbiased. ◁

6.1.2 $\delta_{xx'}$-Unbiasedness of Prima Facie Effects

If we consider two different values x and x' of X, then we can also define *$\delta_{xx'}$-unbiasedness* of the *prima facie effects*

$$PFE_{xx'} := E(Y|X{=}x) - E(Y|X{=}x'),$$

$\delta_{xx'}$-unbiasedness of the Z-conditional *prima-facie-effect functions*

$$PFE_{xx';Z} := E^{X=x}(Y|Z) - E^{X=x'}(Y|Z),$$

and *$\delta_{xx'}$-unbiasedness* of the *$(Z{=}z)$-conditional prima facie effects*

$$PFE_{xx';Z=z} := E(Y|X{=}x, Z{=}z) - E(Y|X{=}x', Z{=}z).$$

Definition 6.6 ($\delta_{xx'}$-Unbiasedness of Prima Facie Effects)
Let the assumptions of Definition 6.1 be true and let $\delta_{xx'} := E^{X=x}(Y|\mathscr{C}_X) - E^{X=x'}(Y|\mathscr{C}_X)$.

(i) *$PFE_{xx'}$ is called $\delta_{xx'}$-unbiased, if $E^{X=x}(Y|\mathscr{C}_X)$ and $E^{X=x'}(Y|\mathscr{C}_X)$ are P-unique and*

$$PFE_{xx'} = E(\delta_{xx'}). \tag{6.6}$$

(ii) *Let Z be a random variable on $(\Omega,\mathscr{A},P)$. Then $PFE_{xx';Z}$ is called $\delta_{xx'}$-unbiased if $E^{X=x}(Y|\mathscr{C}_X)$ and $E^{X=x'}(Y|\mathscr{C}_X)$ are P-unique and*

$$PFE_{xx';Z} \underset{P}{=} E(\delta_{xx'}|Z). \tag{6.7}$$

(iii) *Finally, let z be a value of Z such that $P(X=x,Z=z)$, $P(X=x',Z=z) > 0$. Then $PFE_{xx';Z=z}$ is called $\delta_{xx'}$-unbiased, if $E^{X=x}(Y\,|\,\mathscr{C}_X)$ and $E^{X=x'}(Y|\mathscr{C}_X)$ are $P^{Z=z}$-unique and*

$$PFE_{xx';Z=z} = E(\delta_{xx'}|Z=z). \tag{6.8}$$

Given τ_x-unbiasedness, we can concentrate on the conditional expectation values $E(Y|X=x)$ or $E(Y|X=x,Z=z)$ and ignore other random variables. Usually, Z will be a covariate. However, the definitions above apply to any random variable Z on $(\Omega,\mathscr{A},P)$. While the true-outcome variables τ_x and their conditional expectation values are not estimable in empirical studies, unless very strong assumptions are introduced, the conditional expectation values $E(Y\,|\,X=x)$, $E(Y|X=x,Z=z)$, and the conditional expectation $E^{X=x}(Y|Z)$ *can* be estimated under relatively weak assumptions, and the same is true for the corresponding conditional and unconditional prima facie effects.

Corollary 6.7 (Identification of Total Effects)
Let the assumptions of Definition 6.1 be true and let Z be a random variable on $(\Omega,\mathscr{A},P)$. If $E(Y|X=x)$ is τ_x-unbiased and $E(Y|X=x')$ is $\tau_{x'}$-unbiased, then the prima facie effect $PFE_{xx'}$ is $\delta_{xx'}$-unbiased [see Eq. (6.6)]. Similarly, if $E^{X=x}(Y\,|\,Z)$ is τ_x-unbiased and $E^{X=x'}(Y\,|\,Z)$ is $\tau_{x'}$-unbiased, then the Z-conditional prima-facie-effect function $PFE_{xx';Z}$ is $\delta_{xx'}$-unbiased [see Eq. (6.7)]. And finally, if $E(Y|X=x,Z=z)$ is τ_x-unbiased and $E(Y|X=x',Z=z)$ is $\tau_{x'}$-unbiased, then the $(Z=z)$-conditional prima facie effect $PFE_{xx';Z=z}$ is $\delta_{xx'}$-unbiased [see Eq. (6.8)].

(Proof p. 151)

Remark 6.8 (τ_x-Unbiasedness and Randomization) In chapter 7 we show that unbiasedness of the prima facie effects and the conditional prima facie effects can be created by randomized assignment of the observational unit to one of the treatment conditions. Unbiasedness with respect to total effects can also be strived for via covariate selection, that is, we may try to select covariates $Z_1,\ldots,Z_m$ such that τ_x-unbiasedness of the conditional expectations $E^{X=x}(Y|Z)$, $x=0,1,\ldots,n$, holds for the m-variate covariate $Z:=(Z_1,\ldots,Z_m)$. ◁

Remark 6.9 (Identifying Average From Conditional Total Effects) We can also identify the *average total effects* from a $\delta_{xx'}$-unbiased Z-conditional prima-facie-effect function, because

$$E(PFE_{xx';Z}) \;=\; E[E(\delta_{xx'}\,|\,Z)] \;=\; E(\delta_{xx'}) \tag{6.9}$$

[see Box 10.2 Rule (iv) of SN]. This will, among other things, be discussed in more detail in chapter 10. ◁

Remark 6.10 (τ_x-Unbiasedness and Covariate Selection) Unfortunately, unbiasedness with respect to total effects cannot be used as a criterion for covariate selection. The reason is that it cannot be tested empirically, because the definitions involve the total-effect true-outcome variables τ_x. These variables even cannot be estimated unless very restrictive assumptions are introduced. This has been discussed in some detail by (Holland, 1986) and has been called the ´´fundamental problem of causality" (see also our preface). However, in chapters 7 to 9 we introduce other causality conditions that *can* be tested empirically. ◁

6.2 Numerical Examples

Tables 6.1 (p. 128) to 6.3 (p. 130) show parameters pertaining to fictive *random experiments* such as the single-unit trials described in chapter 2. Among these parameters are the individual expectation values $E(Y\,|\,X{=}x, U{=}u)$ in the treatment conditions and the individual treatment probabilities $P(X{=}1\,|U{=}u)$. The parameters presented in the tables can be used to generate sample data that would result if the random experiments the tables refer to would be repeated independently n times.[1]

6.2.1 Assumptions in all Examples

For simplicity, we consider single-unit trials in which no fallible covariate is observed and in which there is neither a second treatment variable nor any other variable that is simultaneous to the treatment variable with respect to the filtration $(\mathscr{F}_t, t \in T)$ (see section 2.1) specified below. In this case, the set $\Omega = \Omega_U \times \Omega_X \times \mathbb{R}$ suffices to describe the set of possible outcomes of the random experiment. Furthermore, we consider the product σ-algebra $\mathscr{A} = \mathscr{P}(\Omega_U) \otimes \mathscr{P}(\Omega_X) \otimes \mathscr{B}$, where $\mathscr{B}$ denotes the Borel σ-algebra on $\mathbb{R}$ (see section 1.2.3 of SN). The probability measure P on $(\Omega, \mathscr{A})$ is only partly known.

In all examples, we consider the observational-unit variable $U\colon (\Omega, \mathscr{A}, P) \to [\Omega_U, \mathscr{P}(\Omega_U)]$, the treatment variable $X\colon (\Omega, \mathscr{A}, P) \to (\Omega'_X, \mathscr{A}'_X)$ with $\Omega'_X = \{0,1\}$, and

[1] Although the focus of this book is on theory and not on data analysis, we also provide sample data for each table at the home page of this book: www.causal-effects.de. These and other examples of this type as well as sample data generated by these examples can easily be created with the PC-program CausalEffectsExplorer also provided at www.causal-effects.de.

the outcome variable Y: $(\Omega, \mathscr{A}, P) \to (\mathbb{R}, \mathscr{B})$. Furthermore, the treatment variable X takes on the value 1 for treatment and 0 for control.

Finally, the filtration $(\mathscr{F}_t, t \in T)$ consists of three σ-algebras: $\mathscr{F}_1 := \sigma(U)$, $\mathscr{F}_2 := \sigma(U, X)$, and $\mathscr{F}_3 := \sigma(U, X, Y)$. Hence, U is prior to X and Y with respect to $(\mathscr{F}_t, t \in T)$. Furthermore, X is prior to Y with respect to $(\mathscr{F}_t, t \in T)$.

In such a simple experiment $\mathscr{C}_X = \sigma(U)$, that is, the potential-confounder σ-algebra $\mathscr{C}_X$ is identical to the σ-algebra generated by the observational-unit variable U. Because, in all examples presented in this chapter, $P(X{=}x, U{=}u) > 0$ for all pairs (x, u) of values of X and U,

$$\tau_x := E^{X=x}(Y|\mathscr{C}_X) = E^{X=x}(Y|U), \quad x = 0, 1, \tag{6.10}$$

and

$$P(X{=}1\,|\,\mathscr{C}_X) = P(X{=}1\,|\,U). \tag{6.11}$$

Note that the values $E^{X=x}(Y|U{=}u)$ of the conditional expectation $E^{X=x}(Y|U)$ of Y with respect to the conditional-probability measure $P^{X=x}$ are identical with the conditional expectation values $E(Y|X{=}x, U{=}u)$ [see Eq. (9.11) of SN].

Remark 6.11 (Conditional Expectation Values of the Outcome Variable) According to Equation (6.10), there are no covariates that determine the conditional expectation values of the outcomes in the treatment conditions over and above the observational-unit variable U. Therefore the total-effect true-outcome variable τ_x can also be written as a function of the observational-unit variable U, that is, $\tau_x = f_x(U)$, where $f_x\colon [\Omega_U, \mathscr{P}(\Omega_U)] \to (\mathbb{R}, \mathscr{B})$ is a measurable function defined by:

$$\tau_x(\omega) = f_x[U(\omega)] \qquad \text{for each } \omega \in \Omega, \tag{6.12}$$

with

$$f_x(u) = E(Y|X{=}x, U{=}u) \qquad \text{for each } u \in \Omega_U. \tag{6.13}$$

Note that (6.10) not only implies that the total-effect true outcomes and the individual expected outcomes $E(Y|X{=}x, U{=}u)$, but also the atomic total effects and the individual total effects $E(\delta_{xx'}|U{=}u)$ are identical. ◁

Remark 6.12 (Individual Treatment Probabilities) According to Equation (6.11), there are no covariates that determine the treatment probabilities over and above the observational-unit variable U. The examples will differ only in whether and how the units u determine the treatment probabilities.

In empirical applications, the assumption $\mathscr{C}_X = \sigma(U)$ is realistic if (a) there is neither a second treatment variable nor another variable varying simultaneously to X and if (b) no fallible covariate is observed. In this case, u signifies the observational unit *at the onset of treatment*. If, however, a fallible covariate is observed, then u represents the observational unit *at the time at which the covariate is assessed*. In this case there may very well be covariates that are not measurable with respect to U which affect the outcome variable Y and/or the treatment probability. Hence, in this case, $\mathscr{C}_X = \sigma(U)$ would not hold (see section 2.2).

◁

6.2.2 Description of the Examples

In the first example (see Table 6.1, p. 128), the treatment probabilities are *different for each and every unit*, and they strongly depend on the individual expectation values of the outcomes under control, that is, they strongly depend on the total-effect true-outcome variable τ_0. Therefore, there is not only bias of the unconditional prima facie effects with respect to total effects, that is,

$$PFE_{xx'} \neq E(\delta_{xx'}),$$

but also of the conditional prima facie effects for males ($Z{=}m$) and for females ($Z{=}f$), that is,

$$PFE_{xx';Z=z} \neq E(\delta_{xx'} \,|\, Z{=}z).$$

In the second example (see Table 6.2, p. 129), the treatment probabilities are *the same for all units*. Therefore, both the unconditional and the conditional prima facie effects are $\delta_{xx'}$-unbiased, that is, $PFE_{xx'} = E(\delta_{xx'})$ and $PFE_{xx';Z=z} = E(\delta_{xx'} \,|\, Z{=}z)$ for both values z of Z. In the third example (see Table 6.3, p. 130), the treatment probabilities are *different between males and females*. Furthermore, these two subpopulations (sets of units) also differ in the conditional expectation values of the true-outcome variable τ_0, that is, $E(\tau_0|Z{=}m) \neq E(\tau_0|Z{=}f)$. *Within* the sex subpopulations, however, the treatment probabilities do not *differ from each other*. Hence, in this example, the unconditional prima facie effect is biased with respect to total effects, whereas the ($Z{=}z$)-conditional prima facie effects are not.

Tables 6.1, 6.2, and 6.3 display the true outcomes and the true treatment probabilities. According to Equation (6.10), the true outcomes are also the individual conditional expectation values $E(Y \,|\, X{=}x, U{=}u)$, and, according to Equation (6.11), the true treatment probabilities, that is, the values of the conditional probability $P(X{=}1|\mathscr{C}_X) = P(X{=}1|U)$, are identical with the individual treatment probabilities $P(X{=}1 \,|\, U{=}u)$. The tables also display the values of the covariate $Z :=$ *sex*.

Some of the parameters appearing in these tables are called *fundamental parameters*. Other parameters are called *derived parameters* because they can be computed from the fundamental parameters. The true total effects and the conditional probabilities $P(U{=}u|X{=}1)$ of observational unit u in treatment condition x, for instance, are such derived parameters.[2]

Looking at the *fundamental parameters*, the three tables differ only in the treatment probabilities $P(X{=}1|U{=}u)$. All other entries, such as the true outcomes and the true effects, are the same. However, if we look at the *derived parameters*, the three tables differ in important aspects.

[2] One may also consider the probabilities $P(U{=}u|X{=}1)$ as fundamental and the treatment probabilities $P(X{=}1 \,|\, U{=}u)$ as derived. One can be computed as soon as the other one is given.

Table 6.1. Biased Treatment and Covariate-Treatment Regressions

		Fundamental parameters				Derived parameters		
Observational-unit variable U	Covariate sex Z	Sampling probability $P(U=u)$	True outcome variable under control τ_0	True outcome variable under treatment τ_1	True treatment probability $P(X=1\,\|\,U)$	True total effect variable $\delta_{10}=\tau_1-\tau_0$	$P(U=u\,\|\,X=0)$	$P(U=u\,\|\,X=1)$
u_1	m	1/6	68	81	6/7	13	1/21	6/21
u_2	m	1/6	78	86	5/7	8	2/21	5/21
u_3	m	1/6	88	100	4/7	12	3/21	4/21
u_4	m	1/6	98	103	3/7	5	4/21	3/21
u_5	f	1/6	106	114	2/7	8	5/21	2/21
u_6	f	1/6	116	130	1/7	14	6/21	1/21
	Expectations		92.333	102.333	1/2	10		

$E(Y\,\|\,X=0)$	$E(Y\,\|\,X=1)$	PFE_{10}
100.286	94.429	−5.857

	male	female
Conditional total effects	9.5	11
Conditional prima facie effects	2.278	7.879

6.2.3 Average Total Effect

Let us start looking at the prima facie effects and the average total effects. In the first example (see Table 6.1, p. 128), there is a strong negative prima facie effect (−5.857), although the average total effect is large and positive (10) and even though *each and every* individual total effect is positive. The average total effect and the prima facie effect are easy to compute from the parameters displayed in the table. The average total effect $E(\delta_{10})$ of treatment 1 compared to treatment 0 can be computed by $E(\delta_{10}) = E(\tau_1) - E(\tau_0)$, where the expectations $E(\tau_x)$, $x = 0,1$, of the two true-outcome variables are obtained from taking the expectations of the true outcomes using the *unconditional probabilities* $P(U=u)$ as weights (see the second column of Table 6.1). In contrast, the corresponding prima facie effect is obtained by $PFE_{10} = E(Y\,|\,X=1) - E(Y\,|\,X=0)$, where the conditional expectation values $E(Y\,|\,X=x)$, $x = 0,1$, are identical to the $(X=x)$-conditional expectation values of the true outcomes, using as weights the *conditional probabilities* $P(U=u\,|\,X=x)$. In order to keep things simple, we delay computational details to section 6.3.2.

Table 6.2. Unbiased Treatment and Covariate-Treatment Regressions

		Fundamental parameters				Derived parameters		
Observational-unit variable U	Covariate sex Z	Sampling probability $P(U=u)$	True outcome variable under control τ_0	True outcome variable under treatment τ_1	True treatment probability $P(X=1\,\vert\,U)$	True total effect variable $\delta_{10}=\tau_1-\tau_0$	$P(U=u\,\vert X=0)$	$P(U=u\,\vert X=1)$
u_1	m	1/6	68	81	3/4	13	1/6	1/6
u_2	m	1/6	78	86	3/4	8	1/6	1/6
u_3	m	1/6	88	100	3/4	12	1/6	1/6
u_4	m	1/6	98	103	3/4	5	1/6	1/6
u_5	f	1/6	106	114	3/4	8	1/6	1/6
u_6	f	1/6	116	130	3/4	14	1/6	1/6
	Expectations		92.333	102.333	3/4	10		

$E(Y\,\vert X=0)$	$E(Y\,\vert X=1)$	PFE_{10}
92.333	102.333	10

	male	female
Conditional total effects	9.5	11
Conditional prima facie effects	9.5	11

The discrepancy between the (negative) prima facie effect and the (positive) average total effect shows that the prima facie effect is strongly biased in our first example (see Table 6.1). If used for the evaluation of the total treatment effect, the prima facie effect would lead to completely wrong conclusions. Not only is the direction of the prima facie effect reversed as compared to the average total effect, but also as compared *to each and every individual total effect*. All individual total effects are positive in this example, ranging between 5 and 14. As we shown later on, this bias is due to strong inter-individual differences in the true outcomes under control and to the fact that the individual treatment probabilities $P(X=1\,\vert\,U=u)$ heavily depend on the true outcomes under control. For instance, unit u_1 has a true outcome under control of 68 and a treatment probability of 6/7, while unit u_6 has a true outcome under control of 116 and a treatment probability of 1/7. Such a constellation is to be expected under self-selection of subjects to treatments, if the subjects base their decisions to take treatment on the *severity of their dysfunction before treatment* and if *severity of their dysfunction after treatment* is assessed as the outcome variable.

Table 6.3. Unbiased Covariate-Treatment Regression

		Fundamental parameters				Derived parameters		
Observational-unit variable U	Covariate sex Z	Sampling probability $P(U=u)$	True outcome variable under control τ_0	True outcome variable under treatment τ_1	True treatment probability $P(X=1\,\|\,U)$	True total effect variable $\delta_{10}=\tau_1-\tau_0$	$P(U=u\,\|X=0)$	$P(U=u\,\|X=1)$
u_1	m	1/6	68	81	3/4	13	1/10	3/14
u_2	m	1/6	78	86	3/4	8	1/10	3/14
u_3	m	1/6	88	100	3/4	12	1/10	3/14
u_4	m	1/6	98	103	3/4	5	1/10	3/14
u_5	f	1/6	106	114	1/4	8	3/10	1/14
u_6	f	1/6	116	130	1/4	14	3/10	1/14
	Expectations		92.333	102.333	7/12	10		

$E(Y\,\|X=0)$	$E(Y\,\|X=1)$	PFE_{10}
99.8	96.714	−3.086

	male	female
Conditional total effects	9.5	11
Conditional prima facie effects	9.5	11

For the second example presented in Table 6.2, the situation is completely different. Here, the prima facie effect and the average total effect are equal to 10. Hence, in this example, the prima facie effect PFE_{10} is δ_{10}-*unbiased*. As shown later on, this is due to the fact that the individual treatment probabilities *do not depend on the units*. This constellation occurs in a perfect randomized experiment, in which the experimenter decides that each subject is in treatment (X=1) with probability $P(X=1)$ and in control (X=0) with probability $1-P(X=1)$. In our second example, $P(X=1)=3/4$. Note, however, that $P(X=1)$ could be *any* number between 0 and 1, exclusively. The only important point is that the individual treatment probabilities *do not differ between units*, that is, $P(X=1\,|\,U=u)=P(X=1)$ for all units $u\in\Omega_U$. Such a randomized assignment may be performed by drawing a ball from an urn with three black balls and one white ball, adopting the rule that the subject is treated if a black ball is drawn.

In the third example (see Table 6.3), the prima facie effect in the total population is biased again. Here, the prima facie effect is negative (−3.086), although the average total effect is positive (10). Hence, the prima facie effect is strongly biased in this example as well. However, in contrast to the first example, the conditional

prima facie effects in the subpopulations of males and females are δ_{10}-unbiased. In this example, the treatment probability is 3/4 *for all male units*, while it is 1/4 *for all female units*. The crucial point in this example is that these probabilities are the same for each unit *within each of the two subpopulations*, that is, $P(X{=}1 \mid Z{=}z, U{=}u) = P(X{=}1 \mid Z{=}z)$ for each unit u and both values z of the covariate Z. This constellation holds if there is a perfect randomized experiment *within each of the two subpopulations* of males and females.

6.2.4 Conditional Total Effect

Comparing the conditional prima facie effects to the conditional average total effects reveals that the conditional prima facie effects are still biased with respect to total effects in the random experiment presented in Table 6.1, but not in the examples displayed in Tables 6.2 and 6.3. Hence, in the first example, the conditional prima facie effect and the conditional total effect for males are *not* identical, while they *are* identical in the second and third examples, and the same applies to the corresponding prima facie effects in the subpopulation of females.

The bias of the conditional prima facie effects in the example presented in Table 6.1 is no surprise, because there are still individual differences within the two sex subpopulations with respect to (a) the true outcomes under treatment and under control, as well as (b) in the individual treatment probabilities $P(X{=}1 \mid U{=}u)$. In contrast, in the second and third examples, the individual treatment probabilities are all the same *within each of the two sex subpopulations*.

6.2.5 Computing Average Total Effect From Conditional Total Effects

In all three examples, the average over the sex-specific *total effects* is equal to the average total effect in the total population. However, only in the second and third examples the average of the conditional *prima facie effects* is equal to the average total effect. Because this is no coincidence, this fact can be used for causal inference even in those cases in which the *unconditional* prima facie effects are biased, provided that the *conditional* prima facie effects are δ_{10}-unbiased, that ist, provided that $PFE_{10;Z=z} = E(\delta_{10} \mid Z{=}z)$ for each value z of the covariate Z.

Whether or not an average total effect is meaningful if there are different conditional total effects — some of which may even be negative, while some are positive — needs substantive judgement in the specific applications considered. In some applications it might be meaningful, in others it might not. Clearly, conditional total effects give more specific information than the average total effect. However, there are also advantages of average total effects. First, they give a brief summary evaluation of a treatment in *a single number* and different treatments may be compared to each other with respect to this number. Second, in samples of limited size, average effects can be estimated with more accuracy than the plenitude of conditional effects. And third, one should keep in mind that even conditional effects are only average effects (see, e. g., section 5.2). Hence, it is al-

ways a matter of content-based judgement how fine-grained the analysis should be.

6.2.6 First Conclusions

The three examples show that conditioning on a covariate does not *necessarily* yield τ_x-unbiasedness within the subpopulations defined by the values of the covariate nor does it *necessarily* lead to a $\delta_{xx'}$-unbiased estimate of the average total effect. While there is no bias at all in the second example, the third example shows that conditioning *may* remove bias. Comparing Examples 2 and 3 to each other shows that τ_x-unbiasedness in *subpopulations* relies on specific conditions [here: equal individual treatment probabilities $P(X{=}1 \mid U{=}u)$] in a similar way as τ_x-unbiasedness in the *total population*. Such conditions implying unbiasedness are called *causality conditions*. Note, however, that there are several of such causality conditions that do *not* involve the treatment probabilities (see chapters 7 to 9).

6.3 Bias With Respect to Total Effects

In Theorem 6.13 we formulate the general relationship between the conditional expectation values of the outcome variable Y and the corresponding conditional expectation values of the total-effect true-outcome variables τ_x. In Corollary 6.16 we consider the implications of this theorem for the general relationship between prima facie effects and average total effects, and in Box 6.1, we summarize the simplifications of Theorem 6.13 if we can assume $\mathscr{C}_X = \sigma(U)$. Then we treat a theorem according to which the bias of prima facie effects with respect to total effects can be decomposed into the sum of the baseline bias and the effect bias. Finally, we illustrate the various biases by numerical examples.

6.3.1 Theory

Let us start with relationships between the conditional expectation values of the outcome variable Y and the corresponding conditional expectation values of the true-outcome variables τ_x that *always* hold. Note that in the following theorem we do *not* have to assume that the conditional expectation $E^{X=x}(Y|\mathscr{C}_X)$ of Y given $\mathscr{C}_X$ with respect to the conditional-probability measure $P^{X=x}$ is P-unique.

Theorem 6.13 (Bias of Conditional Expected Values and Regressions)
Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a causality space, let Y be numerical and nonnegative or with finite expectation $E(Y)$, let x be a value of X with $P(X{=}x) > 0$, and let $\tau_x := E^{X=x}(Y|\mathscr{C}_X)$.

(i) Then

$$E(Y|X{=}x) \;=\; E(\tau_x|X{=}x). \tag{6.14}$$

(ii) Furthermore, if Z is measurable w.r.t. $\mathscr{F}_{t_X}$, then

$$E^{X=x}(Y|Z) \underset{P^{X=x}}{=} E^{X=x}(\tau_x|Z). \tag{6.15}$$

(iii) If Z is measurable w.r.t. $\mathscr{F}_{t_X}$ and z is a value of Z such that $P(X{=}x, Z{=}z) > 0$, then

$$E(Y|X{=}x, Z{=}z) \;=\; E(\tau_x|X{=}x, Z{=}z). \tag{6.16}$$

(Proof p. 151)

Remark 6.14 (Bias) According to Equation (6.14), the conditional expectation value $E(Y|X{=}x)$ is τ_x-biased unless τ_x is P-unique and $E(\tau_x|X{=}x) = E(\tau_x)$ [see Eq. (6.1)]. Similarly, the conditional expectation $E^{X=x}(Y|Z)$ of Y given Z with respect to the measure $P^{X=x}$ is τ_x-biased unless τ_x is P-unique and $E^{X=x}(\tau_x|Z) \underset{P}{=} E(\tau_x|Z)$ [see Eq. (6.3)]. And finally, the conditional expectation value $E(Y|X{=}x, Z{=}z)$ is τ_x-biased unless τ_x is $P^{Z=z}$-unique and $E(\tau_x|X{=}x, Z{=}z) = E(\tau_x|Z{=}z)$ [see Eq. (6.4)]. This immediately implies the following corollary. ◁

Corollary 6.15 (Equivalent Conditions for τ_x-Unbiasedness)
Let the assumptions of Definition 6.1 be true.

(i) $E(Y|X{=}x)$ is τ_x-unbiased if and only if τ_x is P-unique and

$$E(\tau_x|X{=}x) \;=\; E(\tau_x). \tag{6.17}$$

(ii) Let Z be a random variable on $(\Omega, \mathscr{A}, P)$ that is measurable w.r.t. $\mathscr{F}_{t_X}$. Then $E^{X=x}(Y|Z)$ is τ_x-unbiased if and only if τ_x is P-unique and

$$E^{X=x}(\tau_x|Z) \underset{P}{=} E(\tau_x|Z). \tag{6.18}$$

(iii) Let z be a value of Z such that $P(X{=}x, Z{=}z) > 0$. Then $E(Y|X{=}x, Z{=}z)$ is τ_x-unbiased if and only if τ_x is P-unique and

$$E(\tau_x|X{=}x, Z{=}z) \;=\; E(\tau_x|Z{=}z). \tag{6.19}$$

Theorem 6.13 immediately implies the following corollary that shows the differences between prima facie effects and average total effects. Note that only in proposition (ii) of this corollary we have to assume that the conditional expectations $E^{X=x}(Y|Z)$ and $E^{X=x'}(Y|Z)$ are P-unique. According to Corollary 14.48 (a) and (c) this is equivalent to $P(X{=}x|Z) \underset{P}{>} 0$ and $P(X{=}x'|Z) \underset{P}{>} 0$, respectively.

Corollary 6.16 (Bias of Prima Facie Effects)
Let the assumptions of Theorem 6.13 be true.

(i) Then

$$\begin{aligned} PFE_{xx'} &:= E(Y|X{=}x) - E(Y|X{=}x') \\ &= E(\tau_x|X{=}x) - E(\tau_{x'}|X{=}x'). \end{aligned} \tag{6.20}$$

(ii) Furthermore, if Z is a covariate, and $E^{X=x}(Y|Z)$ as well as $E^{X=x'}(Y|Z)$ are P-unique, then

$$\begin{aligned} PFE_{xx';Z} &:= E^{X=x}(Y|Z) \;-\; E^{X=x'}(Y|Z) \\ &\overset{=}{P} E^{X=x}(\tau_x|Z) \;-\; E^{X=x'}(\tau_{x'}|Z). \end{aligned} \tag{6.21}$$

(iii) If z is a value of a covariate Z such that $P(X{=}x, Z{=}z) > 0$, then

$$\begin{aligned} PFE_{xx';Z=z} &:= E(Y|X{=}x, Z{=}z) \;-\; E(Y|X{=}x', Z{=}z) \\ &= E(\tau_x|X{=}x, Z{=}z) \;-\; E(\tau_{x'}|X{=}x', Z{=}z). \end{aligned} \tag{6.22}$$

(Proof p. 151)

Remark 6.17 (Prima Facie Effect vs. Average Total Effect) Note the difference between Equation (6.20) for the prima facie effect and the equation

$$E(\delta_{xx'}) \;=\; E(\tau_x) - E(\tau_{x'})$$

for the average total effect. While the average total effect is the difference between the *unconditional* expectations of the total-effect true-outcome variables, the prima facie effect is the difference between their *conditional* expectation values *given* $X{=}x$ and $X{=}x'$, respectively. Similarly, comparing Equation (6.22) to

$$E(\delta_{xx'}|Z{=}z) \;=\; E(\tau_x|Z{=}z) - E(\tau_{x'}|Z{=}z)$$

explains why the $(Z{=}z)$-conditional prima facie effects and the $(Z{=}z)$-conditional total effects can differ from each other. ◁

Remark 6.18 (Conditional Prima Facie Effect vs. Conditional Total Effect) Note that the prima facie effect is *neither* identical to the $(X{=}x)$-conditional total effect $E(\delta_{xx'}|X{=}x)$ nor to the $(X{=}x')$-conditional total effect $E(\delta_{xx'}|X{=}x')$. Instead, the prima facie effect $PFE_{xx'}$ [see Eq. (6.20)] *has no causal interpretation at all*, unless specific assumptions can be made, and the same applies to the $(Z{=}z)$-conditional prima facie effect $PFE_{xx';Z=z}$. ◁

6.3.2 Numerical Examples

Now we consider again the numerical examples presented in section 6.2. In these examples, $\mathscr{C}_X = \sigma(U)$, which implies some specific equations that will be useful for our computations.

Remark 6.19 (Specific Equations in the Examples) Equations

$$E(Y|X{=}x) \;=\; \sum_u E(Y|X{=}x,U{=}u)\cdot P(U{=}u|X{=}x) \tag{6.23}$$

and

$$E(Y|X{=}x,Z{=}z) \;=\; \sum_u E(Y|X{=}x,U{=}u,Z{=}z)\cdot P(U{=}u|X{=}x,Z{=}z), \tag{6.24}$$

are always true if U is discrete [see Eq. (iv) in Box 9.2 of SN]. In our numerical examples, $\mathscr{C}_X = \sigma(U)$ and Z is measurable with respect to U. Because, in these examples, $P(X{=}x,U{=}u) > 0$ for all pairs (x,u) of values of X and U, this implies $E(Y|X,\mathscr{C}_X) = E(Y|X,U) = E(Y|X,U,Z)$, and

$$E(Y|X{=}x,U{=}u,Z{=}z) = E(Y|X{=}x,U{=}u), \tag{6.25}$$

as well as

$$E(Y|X{=}x,Z{=}z) \;=\; \sum_u E(Y|X{=}x,U{=}u)\cdot P(U{=}u|X{=}x,Z{=}z).$$

Hence, in our examples, these equations can be used to compute the conditional expectation values $E(Y|X{=}x,Z{=}z)$ from the individual conditional expectation values $E(Y|X{=}x,U{=}u)$ displayed in Tables 6.1 (p. 128) to 6.3 (p. 130). ◁

Remark 6.20 (Bias Again) Remember $\tau_x = f_x(U)$ [see Eq. (6.12)] and note the difference between Equation (6.23) for the conditional expectation value $E(Y|X{=}x)$ and the equation

$$E(\tau_x) \;=\; E[f_x(U)] \;=\; \sum_u E(Y|X{=}x,U{=}u)\cdot P(U{=}u) \tag{6.26}$$

for the expectation of the true-outcome variable τ_x [see Eq. (6.13) and Eq. (6.15) of SN]. While $E(\tau_x)$ is the sum of the conditional expectation values $E(Y|X{=}x,U{=}u)$ weighted by the *unconditional* probabilities $P(U{=}u)$, the conditional expectation value $E(Y|X{=}x)$ is the sum of the conditional expectation values $E(Y|X{=}x,U{=}u)$ weighted by the *conditional* probabilities $P(U{=}u|X{=}x)$ [see Eq. (6.23)].

Similarly, the conditional expectation value $E(Y|X{=}x,Z{=}z)$ [see Eq. (**??**)] is the sum of the conditional expectation values $E(Y|X{=}x,U{=}u)$ weighted by the conditional probabilities $P(U{=}u|X{=}x,Z{=}z)$. In contrast, the $(Z{=}z)$-conditional expectation value

$$E(\tau_x|Z{=}z) \;=\; E[f_x(U)\,|\,Z{=}z] \;=\; \sum_u E(Y|X{=}x,U{=}u)\cdot P(U{=}u|Z{=}z) \tag{6.27}$$

of the true-outcome variable $\tau_x = f_x(U)$ [see Eq. (6.12)] is the sum (over the units u) of the conditional expectation values $E(Y|X{=}x,U{=}u)$ that are weighted by the conditional probabilities $P(U{=}u|Z{=}z)$ [see Eq. (9.19)]. These equations are summarized in Box 6.1. ◁

Box 6.1 Prima Facie Effects and Average Total Effects in the Examples

Let U denote the observational-unit variable in the examples presented in Tables 6.1 (p. 128) to 6.3 (p. 130), let Z be measurable w.r.t. U, let $P(X{=}x, Z{=}z) > 0$, and let τ_x denote the total-effect true-outcome variable of X, and let $\mathscr{C}_X = \sigma(U)$. Then:

$$E(Y|X{=}x) \;=\; \sum_u E(Y|X{=}x, U{=}u)\cdot P(U{=}u|X{=}x), \tag{i}$$

whereas

$$E(\tau_x) \;=\; \sum_u E(Y|X{=}x, U{=}u)\cdot P(U{=}u). \tag{ii}$$

Furthermore,

$$E(Y|X{=}x, Z{=}z) \;=\; \sum_u E(Y|X{=}x, U{=}u)\cdot P(U{=}u|X{=}x, Z{=}z), \tag{iii}$$

whereas

$$E(\tau_x|Z{=}z) \;=\; \sum_u E(Y|X{=}x, U{=}u)\cdot P(U{=}u|Z{=}z). \tag{iv}$$

Remark 6.21 (Computing Prima Facie Effects) In order to be able to compute the prima facie effect

$$PFE_{10} \;=\; E(Y|X{=}1) - E(Y|X{=}0) \tag{6.28}$$

from the parameters given in Tables 6.1 to 6.3, we need Equation (i) in Box 6.1 relating the conditional expectation values $E(Y|X{=}x)$ of Y in treatment x to the individual expected outcomes $E(Y|X{=}x, U{=}u)$ of Y in treatment x. Note that Equation (i) in Box 6.1 is always true if all pairs (x,u) of values of X and U have a non-zero probability. Also note that this equation involves the conditional probabilities $P(U{=}u|X{=}x)$, which occur in Tables 6.1 (p. 128) to 6.3 (p. 130) as 'derived parameters'. Remember that the conditional probabilities $P(U{=}u|X{=}x)$ occurring in Equation (i) in Box 6.1 are related to the individual treatment probabilities $P(X{=}x|U{=}u)$ by

$$P(U{=}u|X{=}x) \;=\; \frac{P(X{=}x|U{=}u)\cdot P(U{=}u)}{P(X{=}x)}.$$

◁

Remark 6.22 (Computing Conditional Prima Facie Effects) The conditional prima facie effects

$$PFE_{10;Z=z} \;=\; E(Y|X{=}1, Z{=}z) - E(Y|X{=}0, Z{=}z) \tag{6.29}$$

can be computed from the conditional expectation values $E(Y|X{=}x, Z{=}z)$, the values of the covariate-treatment conditional expectation $E(Y|X,Z)$. In order to

apply Equation (iii) in Box 6.1 to the examples displayed in Tables 6.1 (p. 128) to 6.3 (p. 130), we also need the formula

$$P(U{=}u|X{=}x,Z{=}z) \;=\; \frac{P(X{=}x|U{=}u)\cdot P(U{=}u,Z{=}z)}{P(X{=}x|Z{=}z)\cdot P(Z{=}z)}, \tag{6.30}$$

where

$$P(X{=}x|Z{=}z) \;=\; \frac{\sum\limits_u P(X{=}x|U{=}u)\cdot P(U{=}u,Z{=}z)}{P(Z{=}z)} \tag{6.31}$$

(see Exercise 6-16). All terms on the right-hand side of Equation (6.30) are displayed in Tables 6.1 to 6.3 or can be computed from the parameters displayed in these tables.[3] ◁

Example 6.23 (Prima Facie Effect) Let us compute the (unconditional) prima facie effect PFE_{10} for the example of Table 6.1 (see p. 128). Using the conditional probabilities $P(U{=}u|X{=}1)$ displayed in this table, we obtain:

$$\begin{aligned} E(Y|X{=}1) \;&=\; \sum_u E(Y|X{=}1,U{=}u)\cdot P(U{=}u|X{=}1) \\ &=\; 81\cdot\frac{6}{21}+86\cdot\frac{5}{21}+100\cdot\frac{4}{21}+103\cdot\frac{3}{21}+114\cdot\frac{2}{21}+130\cdot\frac{1}{21} \;\approx\; 94.429. \end{aligned}$$

Using the same procedure for $E(Y|X{=}0)$ yields (approximately) 100.286. Hence, the prima facie effect is

$$PFE_{10} \;=\; E(Y|X{=}1)-E(Y|X{=}0) \;\approx\; 94.429-100.286 \;=\; -5.857.$$

◁

Example 6.24 (Average Total Effect) The expectations $E(\tau_0)$ and $E(\tau_1)$ of the true outcomes and their difference, $E(\delta_{10}) = E(\tau_1) - E(\tau_0)$, are also easy to compute. For the example in Table 6.1 (p. 128) and treatment 1 we obtain

$$\begin{aligned} E(\tau_1) \;&=\; \sum_u E(Y|X{=}1,U{=}u)\cdot P(U{=}u) \\ &=\; (81+86+100+103+114+130)\cdot\frac{1}{6} \;\approx\; 102.333, \end{aligned}$$

and for $X{=}0$:

$$\begin{aligned} E(\tau_0) \;&=\; \sum_u E(Y|X{=}0,U{=}u)\cdot P(U{=}u) \\ &=\; (68+78+88+98+106+116)\cdot\frac{1}{6} \;\approx\; 92.333. \end{aligned}$$

Hence, the average total effect is $E(\delta_{10}) = E(\tau_1) - E(\tau_0) \approx 102.333 - 92.333 = 10$.

[3] An alternative is using the Causal Effects Explorer (Nagengast et al., 2007) provided at www.causal-effects.de, the home page of this book.

Because, according to the results obtained in Example 6.23, the prima facie effect is −5.857, the *bias of the prima facie effect* is:

$$PFE_{10} - E(\delta_{10}) \approx -5.857 - 10 = -15.857.$$

◁

Example 6.25 (Conditional Prima Facie Effect) Now we compute the conditional prima facie effect

$$PFE_{10;Z=m} = E(Y|X=1,Z=m) - E(Y|X=0,Z=m)$$

for males ($Z=m$) in the example displayed in Table 6.1 (p. 128). Using Equation (iii) in Box 6.1 presumes that the conditional probabilities $P(U=u|X=x,Z=z)$ are known; they are 6/18, 5/18, 4/18, and 3/18 for u_1, u_2, u_3, and u_4, respectively. The other two, $P(U=u_5|X=1,Z=m)$ and $P(U=u_6|X=1,Z=m)$ are 0 (see Exercise 6-10).

Using these results, Equation (iii) in Box 6.1 now yields:

$$\begin{aligned} E(Y|X=1,Z=m) &= \sum_u E(Y|X=1,U=u)\cdot P(U=u|X=1,Z=m) \\ &= 81\cdot\frac{6}{18}+86\cdot\frac{5}{18}+100\cdot\frac{4}{18}+103\cdot\frac{3}{18} + 114\cdot 0+130\cdot 0 \\ &\approx 90.278. \end{aligned}$$

Following the same procedure for $E(Y|X=0,Z=m)$ yields 88, and the difference $PFE_{10;Z=m} = E(Y|X=1,Z=m) - E(Y|X=0,Z=m) \approx 90.278-88 = 2.278$ is the prima facie effect for males. ◁

Example 6.26 (Conditional Total Effect for Males) Turning to the conditional expectation values $E(\tau_0|Z=z)$ and $E(\tau_1|Z=z)$, as well as their difference, we obtain

$$\begin{aligned} E(\tau_1|Z=m) &= \sum_u E(Y|X=1,U=u)\cdot P(U=u|Z=m) \\ &= (81+86+100+103)\cdot\frac{1}{4} + (114+130)\cdot 0 = 92.5, \end{aligned}$$

for $X=1$ and $Z=m$, whereas for $X=0$ and $Z=m$ we receive

$$\begin{aligned} E(\tau_0|Z=m) &= \sum_u E(Y|X=0,U=u)\cdot P(U=u|Z=m) \\ &= (68+78+88+98)\cdot\frac{1}{4} + (106+116)\cdot 0 = 83. \end{aligned}$$

Hence, the conditional total effect for males is $E(\delta_{10}|Z=m) = 92.5-83 = 9.5$. Because, according to the results obtained in Example 6.25, the conditional prima facie effect for males is 2.278, the *bias of the* $(Z=m)$*-conditional prima facie effect* is

$$PFE_{10;Z=m} - E(\delta_{10}|Z=m) \approx 2.278 - 9.5 \approx -7.222.$$

◁

6.3.3 Baseline Bias and Effect Bias

In the example above, we have seen that unconditional and conditional prima facie effects can both be biased if compared to the corresponding average and conditional total effects, respectively. Now we show that each (conditional and unconditional) prima facie effect is equal to the corresponding average total effect plus two kinds of biases (see also Winship & Morgan, 1999).

In the first part of the theorem we refer to the atomic total-effect variables $\delta_{xx'} := \tau_x - \tau_{x'} = E^{X=x}(Y|\mathscr{C}_X) - E^{X=x'}(Y|\mathscr{C}_X)$ and the prima facie effect $PFE_{xx'} := E(Y|X=x) - E(Y|X=x')$. In the second part, we refer to the conditional prima-facie-effect functions defined by $PFE_{xx';Z} := E^{X=x}(Y|Z) - E^{X=x'}(Y|Z)$ and to the conditional total-effect functions $E(\delta_{xx'}|Z) = E(\tau_x|Z) - E(\tau_{x'}|Z)$.

Theorem 6.27 (Baseline and Effect Biases)
Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a causality space, let Y be numerical and nonnegative or with finite expectation, and let x and x' be two values of X with $P(X=x), P(X=x') > 0$. Furthermore, assume that $E^{X=x}(Y|\mathscr{C}_X)$ and $E^{X=x'}(Y|\mathscr{C}_X)$ are P-unique.

(i) Then

$$PFE_{xx'} = E(\delta_{xx'}) + \textit{baseline bias}_{xx'} + \textit{effect bias}_{xx'},$$

where

$$\textit{baseline bias}_{xx'} := E(\tau_{x'}|X=x) - E(\tau_{x'}|X=x') \tag{6.32}$$

and

$$\textit{effect bias}_{xx'} := E(\delta_{xx'}|X=x) - E(\delta_{xx'}). \tag{6.33}$$

(ii) Let Z be measurable w.r.t. $\mathscr{C}_X$, let $E^{X=x}(\tau_{x'}|Z)$ denote the $(X=x)$-conditional expectation of $\tau_{x'}$ on Z, and let $PFE_{xx';Z} := E^{X=x}(Y|Z) - E^{X=x'}(Y|Z)$. Then:

$$PFE_{xx';Z} \underset{P}{=} E(\delta_{xx'}|Z) + \textit{baseline bias}_{xx';Z} + \textit{effect bias}_{xx';Z} \tag{6.34}$$

where

$$\textit{baseline bias}_{xx';Z} := E^{X=x}(\tau_{x'}|Z) - E^{X=x'}(\tau_{x'}|Z)$$

and

$$\textit{effect bias}_{xx';Z} := E^{X=x}(\delta_{xx'}|Z) - E(\delta_{xx'}|Z).$$

(iii) Let z be a value of Z such that $P(X=x, Z=z), P(X=x', Z=z) > 0$ and let $PFE_{xx';Z=z} := E(Y|X=x, Z=z) - E(Y|X=x', Z=z)$. Then:

$$PFE_{xx';Z=z} = E(\delta_{xx'}|Z=z) + \textit{baseline bias}_{xx';Z=z} + \textit{effect bias}_{xx';Z=z}$$

where

$$\textit{baseline bias}_{xx';Z=z} := E(\tau_{x'}|Z=z,X=x) - E(\tau_{x'}|Z=z,X=x')$$

and

$$\textit{effect bias}_{xx';Z=z} := E(\delta_{xx'}|Z=z,X=x) - E(\delta_{xx'}|Z=z).$$

(Proof p. 152)

Remark 6.28 (General Implications) According to this theorem, the true mean differences $PFE_{xx'} := E(Y|X=x) - E(Y|X=x')$ do not allow to draw any conclusions on the average total effects unless we have good reasons to assume that both biases are zero or cancel each other. This is true for the unconditional prima facie effects as well as for the conditional prima facie effects $PFE_{xx';Z=z} := E(Y|X=x,Z=z) - E(Y|X=x',Z=z)$. ◁

Remark 6.29 (Baseline Bias) Let us try to understand in which cases the prima facie effects are biased and in which cases the two biases vanish. Although we do not have to make this assumption in the theorem, let us, for the sake of simplicity, assume $\mathscr{C}_X = \sigma(U)$ and that X represents a treatment variable in an experiment or quasi-experiment. First of all, let us take a look at Equation (6.32) and note that a value of the true-outcome variable $\tau_{x'}$ represents a *(usually unknown) attribute of the observational unit*, namely the individual expected outcome $E(Y|X=x',U=u)$ if unit u *would* be in the treatment x'. Remember that this refers to an experiment to be conducted *in the future*. If the conditional expectation value of $\tau_{x'}$ given $X=x$ would differ from its conditional expectation value given $X=x'$, this would imply, for example, that a unit u with high $E(Y|X=x',U=u)$ would tend to be in treatment $X=x$ rather than in treatment x', or vice versa. In other words, in this case there would be a selection bias due to the individual expected outcomes $E(Y|X=x',U=u)$ in treatment x'. If we call the individual expected outcome under treatment x' — which is the reference treatment, and in some applications it might be the untreated control — the 'baseline', then calling the first kind of bias 'baseline bias' seems justified. ◁

Remark 6.30 (Effect Bias) Similarly, under the assumption made in Remark 6.29, the values of the atomic total effect variable $\delta_{xx'}$ represent an attribute of the observational unit u that exists already *before* the experiment is conducted, and even if the experiment is never conducted. If the conditional expectation value of $\delta_{xx'}$ given treatment $X=x$ would deviate from the average total effect $E(\delta_{xx'})$ [see Eq. (6.33)], this would mean, for example, that those tending to have a high individual total effect of treatment x compared to treatment x' would tend to be in treatment x rather than in treatment x'. In other words, in this case, there would be a selection bias due to the total effect *that would be expected* if the unit *would* be treated in condition x. If units are assigned to treatment x with a high probability, because one correctly expects a high individual total effect for this

unit — which is an important goal in medical, educational, and psychological assessment — one would induce such a selection bias. The corresponding interpretation also holds for the conditional case.

In contrast to our artificial numerical examples, it is *neither* possible to discover bias nor to test τ_x-unbiasedness in empirical applications. However, we present other causality conditions that *can* be tested empirically and that can guide us in designing and analyzing experiments and quasi-experiments (see chs. 7 to 9). These other causality conditions are also *sufficient conditions* for τ_x-unbiasedness. ◁

6.3.4 Numerical Examples

Consider again the example presented in Table 6.1 (p. 128). In this example, the baseline bias is

$$\textit{baseline bias}_{10} \;=\; E(\tau_0 \,|\, X{=}1) \;-\; E(\tau_0 \,|\, X{=}0) \;\approx\; -15.905, \tag{6.35}$$

and the effect bias is

$$\textit{effect bias}_{10} \;=\; E(\delta_{10} \,|\, X{=}1) \;-\; E(\delta_{10}) \;\approx\; 0.048 \tag{6.36}$$

(see Exercises 6-12 and 6-13 for computational details.) The strong baseline bias is due to the fact that the treatment probabilities $P(X{=}1 \,|\, U{=}u)$ strongly depend on the true outcomes under control. In contrast, the effect bias almost vanishes, which follows from the fact that the true treatment probabilities depend on the true total effects only to a small extent.

Turning to the $(Z{=}z)$-conditional effects, the *baseline bias for males* is

$$\textit{baseline bias}_{10;\,Z=m} \;=\; E(\tau_0 \,|\, X{=}1, Z{=}m) \;-\; E(\tau_0 \,|\, X{=}0, Z{=}m) \;\approx\; -7.778, \tag{6.37}$$

and the *effect bias for the males* is

$$\textit{effect bias}_{10;\,Z=m} \;=\; E(\delta_{10} \,|\, X{=}1, Z{=}m) \;-\; E(\delta_{10} \,|\, Z{=}m) \;\approx\; 0.556. \tag{6.38}$$

Again, the strong baseline bias for males is due to fact that the treatment probabilities $P(X{=}1 \,|\, U{=}u)$ of the males strongly depend on the true outcomes under control. In contrast, the effect bias for males is less severe, which is a consequence of the treatment probabilities $P(X{=}1 \,|\, U{=}u)$ depending on the true total effects only to a small extent. For *females* the *baseline bias* is

$$\textit{baseline bias}_{10;\,Z=f} \;=\; E(\tau_0 \,|\, X{=}1, Z{=}f) \;-\; E(\tau_0 \,|\, X{=}0, Z{=}f) \;\approx\; -2.121, \tag{6.39}$$

and the *effect bias* is

$$\textit{effect bias}_{10;\,Z=f} \;=\; E(\delta_{10} \,|\, X{=}1, Z{=}f) \;-\; E(\delta_{10} \,|\, Z{=}f) \;=\; -1. \tag{6.40}$$

In the example presented in Table 6.2 (p. 129), all biases, conditional and unconditional, are zero. This is due to the fact that the treatment probabilities

$P(X=1\,|\,U=u)$ neither dependent on the true outcomes under control nor on the true total effects. In fact, in this example, the true treatment probabilities are the same for all units.

In the example presented in Table 6.3 (p. 130), the prima facie effect is biased, that is, $PFE_{10} \neq E(\delta_{10})$, while the $(Z=z)$-conditional biases for $Z = sex$ are zero. This is due to the fact that the $(Z=z, U=u)$-conditional treatment probabilities $P(X=1\,|\,Z=z, U=u)$ are the same for all units u, implying that these treatment probabilities neither dependent on the true outcomes under control nor on the true total effects. In contrast, this independence does *not* hold for the $(U=u)$-conditional treatment probabilities $P(X=1\,|\,U=u)$.

6.3.5 Another Example

Now we treat an example showing that there can be $\delta_{xx'}$*-unbiasedness of the prima facie effects* and at the same time $\delta_{xx'}$*-biasedness of the* $(Z=z)$*-conditional prima facie effects*. This example shows that unbiasedness with respect to total effects can be accidental, that is, there are cases in which unbiasedness is not a logical consequence of experimental design but an 'accidence of numbers'. Later we will show that the experimental design technique of randomization *always* induces τ_x-unbiasedness and Z-conditional τ_x-unbiasedness for all covariates.[4]

Example 6.31 (Prima Facie Effects) Table 6.4 (p. 143) displays the relevant parameters. We assume that it is a simple experiment so that $\mathscr{C}_X = \sigma(U)$. In this specific example, the individual total effects are the same for all units, namely 5, implying that the average total effect is also 5. The prima facie effect can be computed from the difference between the two conditional expectation values $E(Y\,|\,X=0)$ and $E(Y\,|\,X=1)$. In this example, Equation (i) of Box 6.1 yields

$$\begin{aligned} E(Y\,|\,X=0) &= \sum_u E(Y\,|\,X=0, U=u)\cdot P(U=u\,|X=0) \\ &= 95\cdot\frac{3}{16} + 65\cdot\frac{1}{16} + 80\cdot\frac{7}{16} + 50\cdot\frac{5}{16} = 72.5 \end{aligned}$$

and

$$\begin{aligned} E(Y\,|\,X=1) &= \sum_u E(Y\,|\,X=1, U=u)\cdot P(U=u\,|X=1) \\ &= 100\cdot\frac{5}{16} + 70\cdot\frac{7}{16} + 85\cdot\frac{1}{16} + 55\cdot\frac{3}{16} = 77.5. \end{aligned}$$

Hence, the prima facie effect is $E(Y\,|\,X=1) - E(Y\,|\,X=0) = 5$, which is equal to the average total effect. ◁

Example 6.32 (Conditional Prima Facie Effects) Because the individual total effect is 5 for all units, the conditional total effect in both subpopulations, males

[4] Note that τ_x-unbiasedness does not refer to a sample and that there is no (successful) randomization if there is systematic attrition.

Table 6.4. Accidental Unbiasedness

		Fundamental parameters				Derived parameters		
Observational-unit variable U	Covariate sex Z	Sampling probability $P(U=u)$	True outcome variable under control τ_0	True outcome variable under treatment τ_1	True treatment probability $P(X=1\,\vert\,U)$	True effect variable $\delta_{xx'}$	$P(U=u\vert X=0)$	$P(U=u\vert X=1)$
u_1	m	1/4	95	100	5/8	5	3/16	5/16
u_2	m	1/4	65	70	7/8	5	1/16	7/16
u_3	f	1/4	80	85	1/8	5	7/16	1/16
u_4	f	1/4	50	55	3/8	5	5/16	3/16

$E(\tau_0)$	$E(\tau_1)$	$P(X=1)$	$E(\delta_{10})$
72.5	77.5	1/2	5

$E(Y\vert X=0)$	$E(Y\vert X=1)$	PFE_{10}
72.5	77.5	5

	male	female
Conditional total effects	5	5
Conditional prima facie effects	−5	−5

and females, is also 5. What about the prima facie effects in the two subpopulations? Remember that the prima facie effects in the two subpopulations can be computed from the difference between the two conditional expectation values $E(Y\,|\,X{=}1, Z{=}z)$ and $E(Y\,|\,X{=}0, Z{=}z)$, which themselves can be computed from Equation (iii) of Box 6.1. This equation holds, because, in this example, the random variable Z is measurable with respect to U. While the individual expected outcomes $E(Y\,|\,X{=}x, U{=}u)$ are displayed in Table 6.4, the conditional probabilities $P(U{=}u\,|\,X{=}x, Z{=}z)$ have to be computed via Equation (6.30) (see Exercise 6-16).

For males, Equation (iii) of Box 6.1 yields

$$E(Y\,|\,X{=}0, Z{=}m) \;=\; 95\cdot\frac{9}{12} + 65\cdot\frac{3}{12} \;+\; 80\cdot 0 + 50\cdot 0 \;=\; 87.5$$

and

$$E(Y\,|\,X{=}1, Z{=}m) \;=\; 100\cdot\frac{5}{12} + 70\cdot\frac{7}{12} \;+\; 85\cdot 0 \;+\; 55\cdot 0 \;=\; 82.5.$$

Hence, the prima facie effect in the subpopulation of males is the difference between the conditional expectation values $E(Y|X=1,Z=m)$ and $E(Y|X=0,Z=m)$, which is $=-5$. This difference is *not equal* to the conditional total effect in this subpopulation, which is 5.

For females, Equation (iii) of Box 6.1 yields

$$E(Y|X=0,Z=f) \;=\; 95\cdot 0+65\cdot 0 \;+\; 80\cdot 7/12+50\cdot 5/12 \;=\; 67.5$$

and

$$E(Y|X=1,Z=f) \;=\; 100\cdot 0+70\cdot 0 \;+\; 85\cdot 3/12+55\cdot 9/12 \;=\; 62.5.$$

Hence, also for females, the prima facie effect is $E(Y\,|\,X=1,Z=f)-E(Y|\,X=0,Z=f)$ $=-5$, which is *not equal* to the corresponding conditional total effect, which is also 5. ◁

Remark 6.33 (Methodological Implications) This example shows that τ_x-unbiasedness of the conditional expectations $E(Y|X=x)$ does not imply τ_x-unbiasedness of the conditional expectations $E(Y|X=x,Z=z)$, even if Z is a covariate. Hence, this example shows that *unbiasedness can be accidental*, that is, it may be a fortunate coincidence, an 'accidence of numbers', not a logical consequence of experimental design. In chapter 7, however, we will show that experimental design techniques such as randomized assignment of the unit to one of the treatment conditions *always* leads to τ_x-unbiasedness and to $(Z=z)$-conditional τ_x-unbiasedness if Z denotes a covariate. If Z is measurable with respect to the observational-unit variable, this implies that randomization always leads to τ_x-unbiasedness of the conditional expectations $E(Y|X=x)$ and $E(Y|X=x,Z=z)$. Note, however, that this beneficial implication of randomization *only applies to τ_x-unbiasedness*. Unfortunately, it does not apply to unbiasedness of the conditional expectations $E(Y|X=x,M=m)$ with respect to t_M-direct effects (τ_{x,t_M}-unbiasedness), where M denotes an intermediate variable. ◁

6.4 Unbiasedness With Respect to Direct Effects

The notion of unbiasedness also applies to comparing prima facie effects to direct and indirect effects with respect to a specified intermediate variable M. Again, we start considering the conditional expectation values and then turn to the corresponding prima facie effects.

6.4.1 $\tau_{x,t}$-Unbiasedness of Conditional Expectations

While the true-outcome variables $\tau_x := E^{X=x}(Y|\mathscr{C}_X)$ with respect to total effects are sufficient for introducing the concept of unbiasedness *with respect to total effects*, now we use the conditional expectation (with respect to the conditional-probability measure $P^{X=x}$) of the outcome variable Y given the potential-confounder σ-algebra $\mathscr{C}_{X,t}$ (with respect to time t),

$$\tau_{x,t} := E^{X=x}(Y|\mathscr{C}_{X,t}), \tag{6.41}$$

in order to define the concepts of unbiasedness *related to t-direct effects ($\tau_{x,t}$-unbiasedness*). The conditional expectations $E^{X=x}(Y|\mathscr{C}_{X,t})$ are also called *true-outcome variables with respect to t-direct effects* or *t-direct effect true-outcome variables* (see Def. 4.20).

Remember, an *intermediate variable* M with respect to X and Y is a random variable on $(\Omega,\mathscr{A},P)$ such that X is prior to M that itself is prior to Y with respect to the filtration $(\mathscr{F}_t, t\in T)$ (see section 4.2.4). Although, the random variable W occurring in the following definition is usually measurable with respect to $\mathscr{F}_t = \sigma(X,\mathscr{C}_{X,t})$, no such assumption is necessary in the definition of $\tau_{x,t}$-unbiasedness itself.

Definition 6.34 ($\tau_{x,t}$-Unbiasedness)
Let $\langle(\Omega,\mathscr{A},P),(\mathscr{F}_t,t\in T),X,Y\rangle$ be a causality space, let Y be numerical and nonnegative or with finite expectation $E(Y)$, and let x be a value of X with $P(X{=}x)>0$. Furthermore, let $t\in T$ with $t_X\le t<t_Y$ and let W be a random variable on $(\Omega,\mathscr{A},P)$.

(i) $E^{X=x}(Y|W)$ is called $\tau_{x,t}$-unbiased, if $\tau_{x,t}=E^{X=x}(Y\,|\,\mathscr{C}_{X,t})$ is P-unique and

$$E^{X=x}(Y|W) \underset{P}{=} E(\tau_{x,t}|W). \tag{6.42}$$

(ii) Let X be discrete. Then $E(Y\,|\,X,W)$ is called $\tau_{x,t}$-unbiased, if $E^{X=x}(Y|W)$ is $\tau_{x,t}$-unbiased for all values x of X for which $P(X{=}x)>0$.

(iii) Finally, let w be a value of W such that $P(X{=}x,W{=}w)>0$. Then $E(Y|X{=}x,W{=}w)$ is called $\tau_{x,t}$-unbiased, if $\tau_{x,t}=E^{X=x}(Y|\mathscr{C}_{X,t})$ is $P^{W=w}$-unique and

$$E(Y|X{=}x,W{=}w) = E(\tau_{x,t}|W{=}w). \tag{6.43}$$

Remark 6.35 (Special Cases) A typical random variable W for which we might consider $\tau_{x,t}$-unbiasedness is $W=(M,Z)$, where Z denotes a t_X-covariate of X and M an intermediate variable of X and Y. In this case, Equation (6.42) can also be written

$$E^{X=x}(Y|M,Z) \underset{P}{=} E(\tau_{x,t}|M,Z). \tag{6.44}$$

Other examples of W are $W=M$, $W=X$, $W=(X,Z)$, and $W=(X,M)$. ◁

6.4.2 $\delta_{xx',t}$-Unbiasedness of Prima Facie Effects

Remark 6.36 ($\delta_{xx',t}$-Unbiasedness of Conditional Direct Effects) Now consider the $(W{=}w)$-conditional prima facie effect

$$PFE_{xx';W=w} := E(Y|X{=}x,W{=}w) - E(Y|X{=}x',W{=}w). \tag{6.45}$$

If we assume that $E(Y|X=x,W=w)$ is $\tau_{x,t}$-unbiased and $E(Y|X=x',W=w)$ is $\tau_{x',t}$-unbiased, then this implies $\delta_{xx',t}$-unbiasedness of the $(W=w)$-conditional prima facie effect with respect to the t-direct effect of X, which is defined by

$$PFE_{xx';W=w} = E(\delta_{xx',t}|W=w) \tag{6.46}$$

and the assumption that $\tau_{x,t} = E^{X=x}(Y|\mathscr{C}_{X,t})$ and $\tau_{x',t} = E^{X=x'}(Y|\mathscr{C}_{X,t})$ are $P^{W=w}$-unique. ◁

Remark 6.37 ($\delta_{xx',t}$-Unbiasedness of a Conditional Direct-Effect Function) Similarly, consider the W-conditional prima facie effect function

$$PFE_{xx';W} := E^{X=x}(Y|W) - E^{X=x'}(Y|W). \tag{6.47}$$

If $E^{X=x}(Y|W)$ is $\tau_{x,t}$-unbiased and $E^{X=x'}(Y|W)$ is $\tau_{x',t}$-unbiased, then this implies $\delta_{xx',t}$-unbiasedness of the W-conditional prima-facie-effect function with respect to t-direct-effect function of X, which is defined by

$$PFE_{xx';W} = E(\delta_{xx',t}|W) \tag{6.48}$$

and the assumption that $\tau_{x,t} = E^{X=x}(Y|\mathscr{C}_{X,t})$ and $\tau_{x',t} = E^{X=x'}(Y|\mathscr{C}_{X,t})$ are P-unique. ◁

Remark 6.38 (Identifying Average From Conditional Direct Effects) If $PFE_{xx';W}$ is $\delta_{xx',t}$-unbiased, then we can also identify the *average direct effect* from a $\delta_{xx',t}$-unbiased W-conditional prima-facie-effect function, because

$$E(PFE_{xx';W}) = E\big[E(\delta_{xx',t}|W)\big] = E(\delta_{xx',t}) \tag{6.49}$$

[see Box 10.2 (iv) of SN]. This will be discussed in more detail in chapter 10. ◁

Remark 6.39 (A Caveat on Mediation in Randomized Experiments) Note that the terms on the left-hand sides of Equations (6.42) to (6.43) are estimable in concrete samples. These terms only involve the variables Y, X, and W that can be observed in empirical applications. However, if W is an intermediate variable of X and Y, then the assumptions of $\tau_{x,t}$-unbiasedness of $E(Y|X=x,W=w)$ and $\tau_{x',t}$-unbiasedness of $E(Y|X=x',W=w)$ with respect to t-direct effects will rarely hold, and this is true even for the randomized experiment in which we create independence of X and a potential-confounder σ-algebra $\mathscr{C}_X$ (see ch. 7). In contrast, $\tau_{x,t}$-unbiasedness of $E(Y|X=x,W=w)$ and $\tau_{x',t}$-unbiasedness of $E(Y|X=x',W=w)$ is much more realistic for $W=(M,Z)$, because we may appropriately select the possibly multivariate covariate $Z=(Z_1,\ldots,Z_m)$. In other words, we may select the covariates $Z_1,\ldots,Z_m$ such that $\tau_{x,t}$-unbiasedness of $E(Y|X=x,W=w)$ and $\tau_{x',t}$-unbiasedness of $E(Y|X=x',W=w)$ hold. ◁

Remark 6.40 (A Caveat on Covariate Selection) Unfortunately, $\tau_{x,t}$-unbiasedness itself cannot be used as a criterion for covariate selection, because it cannot

be tested empirically. The reason is that the definitions involve $\tau_{x,t}$, the true-outcome variable with respect to t-direct effects, which usually cannot be observed. It even cannot be estimated, because of the fundamental problem of causality discussed before. However, in chapters 7 to 9 we will introduce other causality conditions implying $\tau_{x,t}$-unbiasedness of a conditional expectation value $E(Y|X=x',W=w)$ that *can* be tested empirically, at least in the sense that it can be falsified. ◁

Theorem 6.41 (Bias w.r.t. Direct Effects)
Let $\langle(\Omega,\mathscr{A},P),(\mathscr{F}_t,t\in T),X,Y\rangle$ be a causality space, let Y be numerical and nonnegative or with finite expectation $E(Y)$, let x be a value of X with $P(X=x)>0$, let $t\in T$ with $t_X\le t<t_Y$ and let W be measurable w.r.t. $\mathscr{F}_t$. Then

$$E^{X=x}(Y|W) \underset{P^{X=x}}{=} E^{X=x}(\tau_{x,t}|W). \tag{6.50}$$

(Proof p. 152)

Remark 6.42 (Bias) If $W=M$, then comparing Equation (6.50) to Equation (6.42) shows that the conditional expectation $E^{X=x}(Y|M)$ is biased with respect to $\tau_{x,t}$ unless $\tau_{x,t}=E^{X=x}(Y|\mathscr{C}_{X,t})$ is P-unique and

$$E^{X=x}(\tau_{x,t}|M) \underset{P}{=} E(\tau_{x,t}|M). \tag{6.51}$$

Similarly, if $W=(M,Z)$, then comparing Equation (6.50) to Equation (6.44) shows that the conditional expectation $E^{X=x}(Y|M,Z)$ is biased with respect to $\tau_{x,t}$ unless $\tau_{x,t}=E^{X=x}(Y|\mathscr{C}_{X,t})$ is P-unique and

$$E^{X=x}(\tau_{x,t}|M,Z) \underset{P}{=} E(\tau_{x,t}|M,Z). \tag{6.52}$$

Hence, Theorem 6.41 immediately implies the following corollary: ◁

Corollary 6.43 (An Equivalent Condition of $\tau_{x,t}$-Unbiasedness)
Let the assumptions if Theorem 6.41 be true. Then $E^{X=x}(Y|W)$ is $\tau_{x,t}$-unbiased if and only if $\tau_{x,t}=E^{X=x}(Y|\mathscr{C}_{X,t})$ is P-unique and

$$E^{X=x}(\tau_{x,t}|W) \underset{P}{=} E(\tau_{x,t}|W). \tag{6.53}$$

In chapters 7 to 9 we study conditions under which Equation (6.53) holds.

Example 6.44 (Joe and Ann With Random Assignment – continued) Table 4.5 (p. 91) displays a numerical example with an intermediate variable M. In this example, we can check if the conditional expectation values $E(Y|X=x,M=m)$ are τ_{x,t_M}-unbiased, that is, if

$$E(Y|X=x,M=m) = E(\tau_{x,t_M}|M=m) \tag{6.54}$$

holds for all pairs of values of X and M. We start comparing the conditional expectation value $E(Y \mid X{=}0, M{=}0)$ to $E(\tau_{0,\,t_M} \mid M{=}0)$. The conditional expectation value

$$E(Y \mid X{=}0, M{=}0) \;=\; .286$$

is displayed in Table 4.5. It is the value of the conditional expectation $E(Y \mid X, M)$ for those outcomes ω of the random experiment in which $X(\omega) = 0$ and $M(\omega) = 0$.

In contrast, the conditional expectation value $E(\tau_{0,\,t_M} \mid M{=}0)$ can not directly be read from this table. However, it can be computed using

$$E(\tau_{0,\,t_M} \mid M{=}0) \;=\; E[E^{X=0}(Y \mid U, M) \mid M{=}0], \tag{6.55}$$

because the values of the conditional expectation $E^{X=0}(Y \mid U, M) = \tau_{0,\,t_M}$ *are* displayed in Table 4.5. Using the definition of a conditional expectation value (see Def. 9.2 of SN), we compute

$$\begin{aligned}
E(\tau_{0,\,t_M} \mid M{=}0) \;&=\; \sum_u \sum_m E^{X=0}(Y \mid U{=}u, M{=}m) \cdot P(U{=}u, M{=}m \mid M{=}0) \\
&= \sum_u \sum_m E^{X=0}(Y \mid U{=}u, M{=}m) \cdot \frac{P(U{=}u, M{=}m, M{=}0)}{P(M{=}0)} \\
&= .609 \cdot \frac{.168 + .036 + .072 + .040}{0.409} \;+\; .727 \cdot 0 \;+ \\
&\quad\; .176 \cdot \frac{.027 + .042 + .008 + .016}{0.409} \;+\; .250 \cdot 0 \\
&\approx\; .609 \cdot .773 \;+\; .176 \cdot .227 \;\approx\; .471 \;+\; .040 \;=\; .511,
\end{aligned}$$

where $P(M{=}0) = .027 + .042 + .008 + .016 + .168 + .036 + .072 + .040 = 0.409$. Because $E(\tau_{0,\,t_M} \mid M{=}0) \approx .511$ is not identical to $E(Y \mid X{=}0, M{=}0) = .286$, we can conclude that the conditional expectation value $E(Y \mid X{=}0, M{=}0)$ is biased with respect to t_M-direct effects and cannot be used to compute the t_M-direct effects of X on Y. In contrast, in this example the conditional expectation values $E(Y \mid X{=}x, U{=}u, M{=}m)$ are $\tau_{0,\,t_M}$-unbiased and *can* be used to compute various t_M-direct effects of X on Y. Hence, in this last example, (U, M) would take the role of W in Definition 6.34. ◁

6.5 Summary and Conclusions

In this chapter, we studied the relationship between the unconditional and conditional prima facie effects on one side and the average and conditional total and direct effects on the other side. We showed that the prima facie effects $E(Y \mid X{=}x) - E(Y \mid X{=}x')$ are always equal to the average total effects plus two kinds of biases: the 'baseline bias' and the 'effect bias'. The corresponding proposition also holds for the $(Z{=}z)$-conditional prima facie effects $E(Y \mid X{=}x, Z{=}z)$ $- E(Y \mid X{=}x', Z{=}z)$. The general insight of this chapter is that *mean differences*, for example, between treatment conditions — and this includes mean differences within subpopulations characterized by a given value z of a covariate Z

Box 6.2 Glossary of New Concepts

Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a causality space, let Y be numerical and non-negative or with finite expectation, and let X be discrete with $P(X \in \{0,1,\ldots,J\}) = 1$ and $P(X{=}x) > 0$ for all $x = 0,1,\ldots,J$. Furthermore, let $\tau_x := E^{X=x}(Y \mid \mathscr{C}_X)$ and $\tau_{x,t} := E^{X=x}(Y \mid \mathscr{C}_{X,t})$.

Unbiasedness of ...	**with respect to total effects**
$E(Y \mid X{=}x)$	τ_x is P-unique and $E(Y \mid X{=}x) = E(\tau_x)$
$PFE_{xx'}$	τ_x and $\tau_{x'}$ are P-unique and $PFE_{xx'} = E(\delta_{xx'})$
$E(Y \mid X)$	For each $x = 0,1,\ldots,J$: τ_x is P-unique and $E(Y \mid X{=}x) = E(\tau_x)$.

Let Z be a random variable on $(\Omega, \mathscr{A}, P)$.

$E(Y \mid X{=}x, Z{=}z)$	τ_x is $P^{Z=z}$-unique and $E(Y \mid X{=}x, Z{=}z) = E(\tau_x \mid Z{=}z)$
$PFE_{xx';Z=z}$	τ_x, $\tau_{x'}$ are $P^{Z=z}$-unique and $PFE_{xx';Z=z} = E(\delta_{xx'} \mid Z{=}z)$
$E^{X=x}(Y \mid Z)$	τ_x is P-unique and $E^{X=x}(Y \mid Z) \underset{P}{=} E(\tau_x \mid Z)$
$PFE_{xx';Z}$	$\tau_x, \tau_{x'}$ are P-unique and $PFE_{xx';Z} \underset{P}{=} E(\delta_{xx'} \mid Z)$
$E(Y \mid X, Z)$	For each $x = 0,1,\ldots,J$: τ_x is P-unique and $E^{X=x}(Y \mid Z) \underset{P}{=} E(\tau_x \mid Z)$.

Let M and W be random variables on $(\Omega, \mathscr{A}, P)$ and let M be an intermediate variable of X and Y.

Unbiasedness of ...	**with respect to direct effects**
$E(Y \mid X{=}x, W{=}w)$	$\tau_{x,t}$ is $P^{W=w}$-unique and $E(Y \mid X{=}x, W{=}w) = E(\tau_{x,t} \mid W{=}w)$
$PFE_{xx';W=w}$	$\tau_{x,t}$ and $\tau_{x',t}$ are $P^{W=w}$-unique and $PFE_{xx';W=w} = E(\delta_{xx',t} \mid W{=}w)$
$E^{X=x}(Y \mid W)$	$\tau_{x,t}$ is P-unique and $E^{X=x}(Y \mid W) \underset{P}{=} E(\tau_{x,t} \mid W)$
$PFE_{xx';W}$	$\tau_{x,t}$ and $\tau_{x',t}$ are P-unique and $PFE_{xx';W} \underset{P}{=} E(\delta_{xx',t} \mid W)$
$E(Y \mid X, W)$	For each $x = 0,1,\ldots,J$: $\tau_{x,t}$ is P-unique and $E^{X=x}(Y \mid W) \underset{P}{=} E(\tau_{x,t} \mid W)$.

— *are meaningless for the evaluation of treatment effects* unless they can also be causally interpreted, that is, unless they are equal to the average total effect or to one of the $(Z{=}z)$-conditional total effects. Comparing means does not allow to draw any conclusions on the effects of a treatment or intervention *unless we have good reasons to assume that the two biases are zero or cancel each other*. In more general terms, then the conditional expectation values such as $E(Y|X{=}x)$ or $E(Y|X{=}x,Z{=}z)$ and their differences do not describe any causal dependencies *unless they are unbiased*. They are like the shadow, which only is identical with the height of the invisible man (see the metaphor discussed in the preface). The same conclusion can also be drawn for direct and indirect effects with respect to time t.

Unbiasedness

The unbiasedness conditions are the most important kind of causality conditions, because they are the weakest kind of assumption under which we can identify total, direct, and indirect effects (see ch. 10). All other causality conditions imply unbiasedness (see chs. 7 to 9). Several concepts can be unbiased *with respect to total effects* and several other concepts can be unbiased *with respect to direct effects* related to a specified time point t. All these terms and their definitions of unbiasedness are summarized in Box 6.2.

Limitations of the Concept of Unbiasedness

Unbiasedness of the conditional expectation values $E(Y|X{=}x)$ and $E(Y|X{=}x,Z{=}z)$ and of the conditional expectations $E^{X=x}(Y|Z)$ is a first kind of causality conditions, which, together with the additional structural components listed in a causality space, distinguishes a causally interpretable conditional expectation from an ordinary conditional expectation. The same applies to the conditional expectation values $E(Y|X{=}x,M{=}m)$ and $E(Y|X{=}x,M{=}m,Z{=}z)$ and the conditional expectations $E^{X=x}(Y|M,Z)$. Unfortunately, unbiasedness cannot be tested empirically. However, as will be shown in chapters 7 to 9, the unbiasedness conditions are the weakest kind of causality conditions. In other words, they are implied by all other causality conditions, some of which, in contrast to unbiasedness itself, *are* empirically testable, at least in the sense of falsifiability.

Another drawback of unbiasedness has been exemplified by the numerical example displayed in Table 6.4 (p. 143). This example shows that the conditional expectation values $E(Y|X{=}x,Z{=}z)$ and the conditional prima facie effects — and this includes mean differences in subpopulations — can be biased even in cases in which the conditional expectation values $E(Y|X{=}x)$ are unbiased (see also Greenland & Robins, 1986). In contrast, the sufficient conditions for unbiasedness treated in chapters 7 to 9 are less volatile, that is, they generalize to conditioning on a covariate, and this implies that they generalize to all subpopulations. *Generalizability* and *empirical testability* are two important virtues of these alternative causality conditions.

Outlook

In the definitions of the unbiasedness conditions presented in this chapter we presumed that each value x of the cause X has a positive probability. Another limitation of this chapter is that we only identified average and conditional total effects, as well as *conditional* t-direct effects. We did not identify *average* direct effects. This will be tackled in chapter 10, where we will treat a number of procedures adjusting for bias, identifying conditional, and average causal effects. This not only includes identification of average and conditional *total* effects, but also average and conditional *direct* and *indirect* effects.

6.6 Proofs

Proof of Corollary 6.7

Fehlt noch

Proof of Theorem 6.13

According to Equation (9.11) in Corollary 9.5 of SN, Equation (6.14) can also be written

$$E^{X=x}(Y) \;=\; E^{X=x}(\tau_x),$$

which shows that (i) is a special case of (ii) for Z being a constant. Furthermore, (iii) follows from (ii) and Remark 14.39 of SN if $P(X{=}x, Z{=}z) > 0$. In this case, Equation (6.16) is equivalent to

$$E^{X=x}(Y \,|\, Z{=}z) \;=\; E^{X=x}(\tau_x \,|\, Z{=}z).$$

Hence, we only have to prove (ii). Proposition (ii) follows from:

$$\begin{aligned}
E^{X=x}(\tau_x \,|\, Z) &\underset{P^{X=x}}{=} E^{X=x}\big[\, E^{X=x}(Y|\mathscr{C}_X) \,\big|\, Z\big] && \text{[def. of } \tau_x\text{]} \\
&\underset{P^{X=x}}{=} E^{X=x}\big[E^{X=x}(Y \,|\mathscr{F}_{t_X}) \,\big|\, Z\big] && \big[\mathscr{F}_{t_X} = \sigma[\sigma(X) \cup \mathscr{C}_X],\ (14.72) \text{ of SN}\big] \\
&\underset{P^{X=x}}{=} E^{X=x}(Y \,|\, Z) && \text{[Box 10.2 (v) of SN].}
\end{aligned}$$

In the last equation we used the assumption that Z is measurable with respect to $\mathscr{F}_{t_X}$.

Proof of Corollary 6.16

Propositions (i) and (iii) are immediate implications of Theorem 6.13. In proposition (ii), we need the additional assumption that $E^{X=x}(Y \,|\, Z)$ and $E^{X=x'}(Y \,|\, Z)$ are P-unique, because this implies that not only (6.15) holds, but also

$$E^{X=x}(\tau_x \,|\, Z) \;\underset{P}{=}\; E^{X=x}(Y \,|\, Z).$$

and

$$E^{X=x'}(\tau_{x'} \,|\, Z) \;\underset{P}{=}\; E^{X=x'}(Y \,|\, Z).$$

Taking the differences between terms on each side of these equations yields (6.21).

Proof of Theorem 6.27

Proposition (i) is a special case of (ii) with Z being a constant. Furthermore, (iii) immediately follows from (ii). Hence, we only have to prove (ii). We assume that the conditional expectation $\tau_x := E^{X=x}(Y|\mathscr{C}_X)$ is P-unique. According to Box 14.1 (iv) of SN, this implies that $E^{X=x}(Y|Z)$ is P-unique as well, because Z is measurable with respect to $\mathscr{C}_X$. Hence, we can consider the difference

$$\begin{aligned} E^{X=x}(Y|Z) - E^{X=x'}(Y|Z) &\underset{P}{=} E^{X=x}(\tau_x|Z) - E^{X=x'}(\tau_{x'}|Z) \qquad [(6.15)] \\ &\underset{P}{=} E^{X=x}(\tau_{x'}|Z) + E^{X=x}(\tau_x|Z) - E^{X=x}(\tau_{x'}|Z) - E^{X=x'}(\tau_{x'}|Z) \\ &\underset{P}{=} E^{X=x}(\tau_{x'}|Z) + E^{X=x}(\delta_{xx'}|Z) - E^{X=x'}(\tau_{x'}|Z) \end{aligned}$$

using $E^{X=x}(\delta_{xx'}|Z) \underset{P}{=} E^{X=x}(\tau_x|Z) - E^{X=x}(\tau_{x'}|Z)$. Adding and subtracting $E(\delta_{xx'}|Z)$ on the right-hand side yields

$$\begin{aligned} E^{X=x}(Y|Z) - E^{X=x'}(Y|Z) \underset{P}{=}\ & E(\delta_{xx'}|Z) + E^{X=x}(\tau_{x'}|Z) - E^{X=x'}(\tau_{x'}|Z) \\ & + E^{X=x}(\delta_{xx'}|Z) - E(\delta_{xx'}|Z), \end{aligned}$$

which is Equation (6.34).

Proof of Theorem 6.41

The proof is analog to the proof of Theorem 6.13. Hence,

$$\begin{aligned} E^{X=x}(\tau_{x,t}|W) &\underset{P^{X=x}}{=} E^{X=x}\left[E^{X=x}(Y|\mathscr{C}_{X,t})\,\middle|\,W\right] && [\text{def. of } \tau_{x,t}] \\ &\underset{P^{X=x}}{=} E^{X=x}\left[E^{X=x}(Y|X,\mathscr{C}_{X,t})\,\middle|\,W\right] && [(14.72) \text{ of SN}] \\ &\underset{P^{X=x}}{=} E^{X=x}(Y|W) && [\text{ Box 10.2 (v) of SN}]. \end{aligned}$$

In the last equation, we used the assumption that W is measurable with respect to $\mathscr{F}_t = \sigma(X, \mathscr{C}_{X,t})$.

6.7 Exercises

▷ **Exercise 6-1** What is the difference between the two terms $E(\tau_x)$ and $E(Y|X{=}x)$?

▷ **Exercise 6-2** What is the difference between the average total effect $E(\delta_{xx'})$ and the prima facie effect $PFE_{xx'}$?

▷ **Exercise 6-3** Which are the two biases of the prima facie effect $PFE_{xx'}$?

▷ **Exercise 6-4** What does it mean that the prima facie effect $PFE_{xx'}$ is $\delta_{xx'}$-unbiased and why is it important?

▷ **Exercise 6-5** Compute the probabilities $P(Z{=}z)$ occurring in Equation (6.31) for both values of Z in the example displayed in Table 6.1 (p. 128) .

▷ **Exercise 6-6** Which are the probabilities $P(U{=}u_1, Z{=}m)$ and $P(U{=}u_5, Z{=}m)$ occurring in Equation (6.31) for the example displayed in Table 6.1 (p. 128).

▷ **Exercise 6-7** Compute the two $(Z{=}m)$-conditional probabilities $P(U{=}u_1 \,|\, Z{=}m)$ and $P(U{=}u_5 \,|\, Z{=}m)$ displayed in Table 6.1 (p. 128).

▷ **Exercise 6-8** Use Theorem 4.25 of SN to compute the probability $P(X{=}1)$ for the example displayed in Table 6.1 (p. 128).

▷ **Exercise 6-9** Compute the probabilities $P(U{=}u \,|\, X{=}0)$ and $P(U{=}u \,|\, X{=}1)$ for all six units in the example of Table 6.1 (p. 128).

▷ **Exercise 6-10** Compute the conditional probabilities $P(U{=}u \,|\, X{=}1, Z{=}m)$ occurring in Equation (iii) of Box 6.1.

▷ **Exercise 6-11** Compute the conditional expectation values $E(Y \,|\, X{=}0)$ and $E(Y \,|\, X{=}1)$ for the example in Table 6.3 (p. 128).

▷ **Exercise 6-12** Compute $\textit{baseline bias}_{10}$ for the example in Table 6.1 (p. 128).

▷ **Exercise 6-13** Compute the $\textit{effect bias}_{10}$ for the example in Table 6.1 (p. 128).

▷ **Exercise 6-14** Compute the $(Z{=}f)$-conditional expectation values $E(\tau_1 \,|\, Z{=}f)$ and $E(\tau_0 \,|\, Z{=}f)$ displayed in Table 6.1 (p. 128).

▷ **Exercise 6-15** Prove the proposition of Remark 6.5.

▷ **Exercise 6-16** Show that Equation (6.30) holds.

Solutions

▷ **Solution 6-1** The term $E(\tau_x)$ denotes the *expectation of the true-outcome variable* τ_x. It is these expectations $E(\tau_x)$ that are of interest in the empirical sciences, because their differences for different values x and x' yield the average total effect $E(\delta_{xx'})$. In applications, we often aim at estimating the expectations $E(\tau_x)$ and their differences. In contrast, the $(X{=}x)$-conditional expectation values $E(Y \,|\, X{=}x)$ of the outcomes are usually not of interest in the empirical sciences, because in general $E(Y \,|\, X{=}x) \neq E(\tau_x)$.

▷ **Solution 6-2** The average total effect $E(\delta_{xx'})$ of treatment x compared to treatment x' is the expectation of the true total effects of x compared to x'. It is these average total effects that are of interest in the empirical sciences. In empirical applications we aim at estimating the average total effects. In contrast, the prima facie effects $PFE_{xx'}$ of x compared to x' are usually not of interest because they can be biased. Both terms, differ from each other because $E(\delta_{xx'}) = E(\tau_x) - E(\tau_{x'})$, whereas $PFE_{xx'} := E(Y \,|\, X{=}x) - E(Y \,|\, X{=}x')$.

▷ **Solution 6-3** According to Theorem 6.27, the prima facie effect $PFE_{xx'}$ differs from the average total effect $E(\delta_{xx'})$ by the baseline bias and by the effect bias, that ist, $PFE_{xx'} = E(\delta_{xx'}) + \textit{baseline bias}_{xx'} + \textit{effect bias}_{xx'}$. Suppose that X represents a treatment variable. Then the baseline bias will be zero if there is no selection to treatment x due to the true outcomes under treatment x'.

According to Equation (6.32), $\textit{baseline bias}_{xx'} := E(\tau_{x'} \,|\, X{=}x) - E(\tau_{x'} \,|\, X{=}x')$. This implies that the baseline bias is zero, if the expectations of the true-outcome variable in

treatment x' (the 'baseline condition') do not depend on the treatment variable, or, vice versa, if the probabilities of getting treatment x and treatment x' do not depend on the true outcomes in treatment x'. The effect bias will be zero if there is no selection to treatment x due to the atomic total-effect variable $\delta_{xx'}$. According to Equation (6.33), $\textit{effect bias}_{xx'} := E(\delta_{xx'} \mid X{=}x) - E(\delta_{xx'})$. This implies that the effect bias is zero, if the expectation of the total-effect variable $\delta_{xx'}$ does not depend on the treatment variable, or, vice versa, if the probability of getting treatment x does not depend on the atomic total effects of treatment x vs. x'.

▷ **Solution 6-4** If the prima facie effect $PFE_{xx'}$ is $\delta_{xx'}$-unbiased, then it is equal to the average total effect $E(\delta_{xx'})$. If $PFE_{xx'}$ is $\delta_{xx'}$-unbiased, an estimate of this parameter is also an estimate of $E(\delta_{xx'})$. Note, however, that estimating the average total effect via the prima facie effect is only one of several ways to estimate the average total effect.

▷ **Solution 6-5** The events that U takes on one of its values $u_1,\ldots,u_6$ are disjoint. Therefore, we can use the theorem of total probability (see Th. 4.25 of SN):

$$\begin{aligned} P(Z{=}m) &= P(Z{=}m, U{=}u_1) + \ldots + P(Z{=}m, U{=}u_6) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + 0 + 0 = \frac{4}{6}. \end{aligned}$$

$$\begin{aligned} P(Z{=}f) &= P(Z{=}f, U{=}u_1) + \ldots + P(Z{=}f, U{=}u_6) \\ &= 0 + 0 + 0 + 0 + \frac{1}{6} + \frac{1}{6} = \frac{2}{6}. \end{aligned}$$

▷ **Solution 6-6** $P(U{=}u_1, Z{=}m) = \frac{1}{6}$ and $P(U{=}u_5, Z{=}m) = 0$.

▷ **Solution 6-7**

$$P(U{=}u_1 \mid Z{=}m) = \frac{P(U{=}u_1, Z{=}m)}{P(Z{=}m)} = \frac{1/6}{4/6} = \frac{1}{4}.$$

$$P(U{=}u_5 \mid Z{=}m) = \frac{P(U{=}u_5, Z{=}m)}{P(Z{=}m)} = \frac{0}{4/6} = 0.$$

▷ **Solution 6-8** Note that the values $u_1,\ldots,u_6$ of U are disjoint and all these values have positive probabilities. Hence we can apply the theorem of total probability:

$$\begin{aligned} P(X{=}1) &= P(X{=}1 \mid U{=}u_1) \cdot P(U{=}u_1) + \ldots + P(X{=}1 \mid U{=}u_6) \cdot P(U{=}u_6) \\ &= \frac{6}{7} \cdot \frac{1}{6} + \frac{5}{7} \cdot \frac{1}{6} + \frac{4}{7} \cdot \frac{1}{6} + \frac{3}{7} \cdot \frac{1}{6} + \frac{2}{7} \cdot \frac{1}{6} + \frac{1}{7} \cdot \frac{1}{6} \\ &= \frac{21}{42} = \frac{1}{2}. \end{aligned}$$

▷ **Solution 6-9** We have to use Equation

$$P(U{=}u \mid X{=}x) = \frac{P(X{=}x \mid U{=}u) \cdot P(U{=}u)}{P(X{=}x)},$$

which yields for $U{=}u_1$ and $X{=}1$:

$$\begin{aligned} P(U{=}u_1 \mid X{=}1) &= \frac{P(X{=}1 \mid U{=}u_1) \cdot P(U{=}u_1)}{P(X{=}1)} \\ &= \frac{6/7 \cdot 1/6}{1/2} = \frac{6}{21}. \qquad \text{[Exercise 6-8]} \end{aligned}$$

Using the same procedure, we obtain $5/21, 4/21, \ldots, 1/21$, the corresponding probabilities for the units $u_2, u_3 \ldots, u_6$, respectively. For $U{=}u_1$ and $X{=}0$, we obtain

$$\begin{aligned} P(U{=}u_1 \,|\, X{=}0) &= \frac{P(X{=}0 \,|\, U{=}u_1) \cdot P(U{=}u_1)}{P(X{=}0)} \\ &= \frac{1/7 \cdot 1/6}{1/2} = \frac{1}{21} \qquad \text{[Exercise 6 – 8]} \end{aligned}$$

Using the same procedure, we obtain $2/21,\ 43/21, \ldots, 6/21$ for the units $u_2, u_3 \ldots, u_6$, respectively.

▷ **Solution 6-10** According to Equation (6.30) we need the conditional treatment probabilities $P(X{=}1 \,|\, U{=}u)$ displayed in Table 6.1 (p. 128). The other probabilities occurring in this equation can be computed from the probabilities displayed in the table.

One of these other probabilities that needs some computation is $P(X{=}1 \,|\, Z{=}m)$. For $X{=}1$ and $Z{=}m$ in the example of Table 6.1 (p. 128), Equation (6.31) results in:

$$\begin{aligned} P(X{=}1 \,|\, Z{=}m) &= \frac{\sum\limits_u P(X{=}1 \,|\, U{=}u) \cdot P(U{=}u, Z{=}m)}{P(Z{=}m)} \\ &= \frac{(6/7) \cdot (1/6) + \ldots + (3/7) \cdot (1/6) + (2/7) \cdot 0 + (1/7) \cdot 0}{4/6} = \frac{27}{42}. \end{aligned}$$

Using this result, Equation (6.30) yields:

$$\begin{aligned} P(U{=}u_1 \,|\, X{=}1, Z{=}m) &= \frac{P(X{=}1 \,|\, U{=}u_1) \cdot P(U{=}u_1, Z{=}m)}{P(X{=}1 \,|\, Z{=}m) \cdot P(Z{=}m)} \\ &= \frac{(6/7) \cdot (1/6)}{(27/42) \cdot (4/6)} = \frac{6/7}{(27/42) \cdot 4} = \frac{6}{18}, \end{aligned}$$

as well as $5/18,\ 4/18$, and $3/18$ for the corresponding conditional probabilities for u_2, u_3, and u_4. The conditional probabilities $P(U{=}u_5 \,|\, X{=}1, Z{=}m)$ and $P(U{=}u_6 \,|\, X{=}1, Z{=}m)$ are zero.

▷ **Solution 6-11** According to Equation (i) of Box 6.1,

$$\begin{aligned} E(Y \,|\, X{=}0) &= \sum_u E(Y \,|\, X{=}0, U{=}u) \cdot P(U{=}u \,|\, X{=}0) \\ &= (68 + 78 + 88 + 98) \cdot \frac{1}{10} + (106 + 116) \cdot \frac{3}{10} \\ &= 33.20 + 66.6 = 99.80, \end{aligned}$$

and

$$\begin{aligned} E(Y \,|\, X{=}1) &= \sum_u E(Y \,|\, X{=}1, U{=}u) \cdot P(U{=}u \,|\, X{=}1) \\ &= (81 + 86 + 100 + 103) \cdot \frac{3}{14} + (114 + 130) \cdot \frac{1}{14} \\ &\approx 79.286 + 17.429 \approx 96.715. \end{aligned}$$

▷ **Solution 6-12** According to Equation (6.32) we have to take the difference between the two conditional expectation values $E(\tau_0 \,|\, X{=}1)$ and $E(\tau_0 \,|\, X{=}0)$. In this example, $E(Y \,|\, X, \mathscr{C}_X) = E(Y \,|\, X, U)$. Therefore,

$$E(\tau_0 \mid X{=}x) \;=\; \sum_u E(Y \mid X{=}0, U{=}u) \cdot P(U{=}u \mid X{=}x),$$

and we can use the probabilities $P(U{=}u \mid X{=}x)$ computed in Exercise 6-9. Using these probabilities, we receive

$$\begin{aligned} E(\tau_0 \mid X{=}1) \;&=\; \sum_u E(Y \mid X{=}0, U{=}u) \cdot P(U{=}u \mid X{=}1) \\ &=\; 68 \cdot \frac{6}{21} + 78 \cdot \frac{5}{21} + \ldots + 116 \cdot \frac{1}{21} \;=\; 84.38095. \end{aligned}$$

For $X{=}0$, we obtain:

$$\begin{aligned} E(\tau_0 \mid X{=}0) \;&=\; \sum_u E(Y \mid X{=}0, U{=}u) \cdot P(U{=}u \mid X{=}0) \\ &=\; 68 \cdot \frac{1}{21} + 78 \cdot \frac{2}{21} + \ldots + 116 \cdot \frac{6}{21} \;=\; 100.2857. \end{aligned}$$

The difference $E(\tau_0 \mid X{=}1) - E(\tau_0 \mid X{=}0) \approx -15.905$ is the *baseline bias*$_{10}$ [see Eq. (6.35)].

▷ **Solution 6-13** Because $E(\delta_{10} \mid X{=}1) = E(\tau_1 \mid X{=}1) - E(\tau_0 \mid X{=}1)$ and we can use $E(\tau_0 \mid X{=}1) = 84.38095$ obtained in Exercise 6-12, we only have to compute

$$\begin{aligned} E(\tau_1 \mid X{=}1) \;&=\; \sum_u E(Y \mid X{=}1, U{=}u) \cdot P(U{=}u \mid X{=}1) \\ &=\; 81 \cdot \frac{6}{21} + 86 \cdot \frac{5}{21} + \ldots + 130 \cdot \frac{1}{21} \;=\; 94.42857. \end{aligned}$$

Hence, $E(\delta_{10} \mid X{=}1) \;=\; 94.42857 - 84.38095 \;=\; 10.04762$. Subtracting $E(\delta_{10}) \;=\; 10$ yields $10.04762 - 10 = 0.04762$ [see Eq. (6.36)].

▷ **Solution 6-14** Remember again, in this example, $E(Y \mid X, \mathscr{C}_X) = E(Y \mid X, U)$. Hence,

$$\begin{aligned} E(\tau_0 \mid Z{=}f) \;&=\; \sum_u E(Y \mid X{=}0, U{=}u) \cdot P(U{=}u \mid Z{=}f) \\ &=\; 68 \cdot 0 + \ldots + 98 \cdot 0 + 106 \cdot \frac{1}{2} + 116 \cdot \frac{1}{2} \;=\; 111. \\ E(\tau_1 \mid Z{=}f) \;&=\; \sum_u E(Y \mid X{=}1, U{=}u) \cdot P(U{=}u \mid Z{=}f) \\ &=\; 81 \cdot 0 + \ldots + 103 \cdot 0 + 114 \cdot \frac{1}{2} + 130 \cdot \frac{1}{2} \;=\; 122. \end{aligned}$$

▷ **Solution 6-15** Fehlt noch

If $P(X{=}x, Z{=}z) > 0$, then $P(X{=}x), P(Z{=}z) > 0$ (see ??).

$$E(Y \mid X{=}x, Z{=}z) \;=\; E^{X=x}(Y \mid Z{=}z).$$

▷ **Solution 6-16** In our examples, $P(X{=}x, Z{=}z) > 0$. Hence, using the definition of a conditional probability several times, we can consider

$$\begin{aligned} P(U{=}u \mid X{=}x, Z{=}z) \;&=\; \frac{P(X{=}x, U{=}u, Z{=}z)}{P(X{=}x, Z{=}z)} \\ &=\; \frac{P(X{=}x \mid U{=}u, Z{=}z) \cdot P(U{=}u, Z{=}z)}{P(X{=}x \mid Z{=}z) \cdot P(Z{=}z)}, \end{aligned}$$

which holds if $P(U{=}u, Z{=}z) > 0$. In this case

$$P(X{=}x \mid U{=}u) \;=\; P(X{=}x \mid U{=}u, Z{=}z),$$

because, in our examples, Z (sex) is a function of U. Therefore, the equation for the conditional probability $P(U{=}u \mid X{=}x, Z{=}z)$ above can be simplified to Equation (6.30). If the joint probability $P(U{=}u, Z{=}z)$ is zero, then $P(X{=}x, U{=}u, Z{=}z) = 0$, which implies that $P(U{=}u \mid X{=}x, Z{=}z) = 0$. The same result is also obtained applying Equation (6.30).

Chapter 7

Independent Cause and Mean-Independent Outcome

In chapter 6 we have seen that the conditional expectation values $E(Y|X{=}x)$ and $E(Y|X{=}x,Z{=}z)$ and their differences, the prima facie effects, can be biased, where X denotes a cause and Z a covariate of X. We have defined *unbiasedness* of these conditional expectation values *with respect to total effects* as well as unbiasedness of the conditional expectations $E(Y|X)$ and $E(Y|X,Z)$ with respect to total effects assuming that the cause X is discrete.

Unbiasedness of $E(Y|X)$ and $E(Y|X,Z)$ are a first kind of causality conditions. If $E(Y|X)$ is unbiased with respect to total effects and X is discrete, then the differences $E(Y\,|\,X{=}x) - E(Y\,|\,X{=}x')$ are identical to the *average total effects* $E(\delta_{xx'})$. Similarly, unbiasedness with (respect to total effects) of the conditional expectation $E(Y|X,Z)$ implies that the differences $E(Y|X{=}x,Z{=}z) - E(Y|X{=}x',Z{=}z)$ are equal to the $(Z{=}z)$-*conditional total effect* $E(\delta_{xx'}\,|\,Z{=}z)$, provided that $P(X{=}x,Z{=}z),\ P(X{=}x',Z{=}z) > 0$. If M denotes an intermediate variable of X and Y, then similar propositions also hold for unbiasedness of the conditional expectations $E(Y|X,M)$ and $E(Y|X,M,Z)$ *with respect to t-direct effects*.

However, the unbiasedness conditions have two limitations. First, they cannot be tested empirically, which implies that they cannot be used for covariate selection. Second, they can be *accidental* in the sense that unbiasedness may hold for $E(Y|X)$ but not for $E(Y|X,Z)$, where Z denotes a covariate. Therefore, in this chapter, we study two other kinds of causality conditions that are less volatile. These conditions will be referred to as the *independent cause conditions* and the *mean-independent outcome conditions* respectively.

Overview

We start with the concepts of independence and conditional independence, and use them to introduce the independent cause conditions. Next we define the mean-independent outcome conditions and study the implications of both kinds of conditions on unbiasedness. Then we illustrate these causality conditions by some examples. Finally, we discuss their methodological implications. In particular, we discuss the role of randomization and conditional randomization as well as covariate selection in the analysis of total and direct effects.

7.1 Independent Cause Conditions

7.1.1 Independence and Conditional Independence

A general definition (including the case in which X is continuous) and many important properties of conditional independence are presented in section PR-16.1.1, which also includes conditional independence of sets of events and of more than two random variables. A number of useful properties of conditional independence are summarized in PR-Box 16.2. (See Dawid, 1979, for an overview of applications of conditional independence in statistics.)

The following corollary is a special case of Theorem PR-16.26 (see Rem. PR-16.27)

Corollary 7.1 (Independence and Conditional Independence: Discrete X) *Let X and V be random variables on the probability space $(\Omega, \mathscr{A}, P)$, let X be discrete with a finite number of values $x = 0, 1, \ldots, J$ and let $P(X{=}x) > 0$ for all $x = 0, 1, \ldots, J$.*

(i) Then X and V are independent (abbr.: $X \perp\!\!\!\perp_P V$) if and only if

$$P(X{=}x \,|\, V) \underset{P}{=} P(X{=}x) \quad \text{for all } x = 0, 1, \ldots, J. \tag{7.1}$$

(ii) If Z is a random variable on $(\Omega, \mathscr{A}, P)$ too, then X and V are Z-conditionally independent (abbr.: $X \perp\!\!\!\perp_P V \,|\, Z$) if and only if

$$P(X{=}x \,|\, Z, V) \underset{P}{=} P(X{=}x \,|\, Z) \quad \text{for all } x = 0, 1, \ldots, J. \tag{7.2}$$

7.1.2 Independent Cause Conditions for Total Effects

Now let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a causality space, let C_X denote a global covariate of X, and consider

$$P(X{=}x \,|\, C_X) \underset{P}{=} P(X{=}x), \quad \text{for all } x = 0, 1, \ldots, J, \tag{7.3}$$

which is abbreviated by $X \perp\!\!\!\perp_P C_X$. Next, let Z be a C_X-measurable random variable on $(\Omega, \mathscr{A}, P)$ and consider

$$P(X{=}x \,|\, C_X) \underset{P}{=} P(X{=}x \,|\, Z), \quad \text{for all } x = 0, 1, \ldots, J, \tag{7.4}$$

abbreviated by $X \perp\!\!\!\perp_P C_X \,|\, Z$. These two conditions will be referred to as the *independent cause conditions for total effects*. Note that Equation (7.3) is a special case of (7.4) for Z being a constant mapping, i. e., for $Z(\omega) = const$, for all $\omega \in \Omega$.

If Z is measurable with respect to C_X, i. e., if $\sigma(Z) \subset \sigma(C_X)$, then $\sigma(Z, C_X) = \sigma(C_X)$ and, therefore, $P(X{=}x \,|\, Z, C_X) \underset{P}{=} P(X{=}x \,|\, C_X)$ (see sections PR-2.3.2 and PR-10.1).

Remark 7.2 (Propensity) If Z is C_X-measurable, then the conditional probability

$$\pi_x := P(X{=}x \,|\, Z) \tag{7.5}$$

is also called the *Z-conditional propensity of x*. The propensity plays an important role in adjustment techniques, because Equations (7.4) and (7.5) imply

$$P(X{=}x \,|\, C_X) \underset{P}{=} P(X{=}x \,|\, \pi_x) \quad \text{for all } x = 0, 1, \ldots, J \tag{7.6}$$

(see Exercise 7-1). Hence, $X \perp\!\!\!\perp_P C_X \,|\, Z$ implies $\mathbb{1}_{X=x} \perp\!\!\!\perp_P C_X \,|\, \pi_x$, which means that π_x can be used in adjustment techniques instead of the original covariate Z, provided that X is dichotomous (for details, see ch. 11). ◁

7.1.3 Independent Cause Conditions for Direct Effects

Now we turn to direct effects with respect to time t. We consider a $C_{X,t}$-measurable random variable Z on $(\Omega, \mathscr{A}, P)$

$$P(X{=}x \,|\, C_{X,t}) \underset{P}{=} P(X{=}x \,|\, Z) \quad \text{for all } x = 0, 1, \ldots, J. \tag{7.7}$$

This condition is abbreviated by $X \perp\!\!\!\perp_P C_{X,t} \,|\, Z$ and called the *independent cause condition for t-direct effects*.

7.1.4 Falsifiability of the Independent Cause Conditions

The independent cause conditions are empirically falsifiable. More specifically, if W is measurable with respect to C_X, then $X \perp\!\!\!\perp_P C_X$ implies

$$P(X{=}x \,|\, W) \underset{P}{=} P(X{=}x) \quad \text{for all } x = 0, 1, \ldots, J. \tag{7.8}$$

Hence, in order to test $X \perp\!\!\!\perp_P C_X$, we simply have to select a C_X-measurable random variable W and see if Equation (7.8) holds.

Similarly, if W is measurable with respect to C_X, then $X \perp\!\!\!\perp_P C_X \,|\, Z$ implies

$$P(X{=}x \,|\, Z, W) \underset{P}{=} P(X{=}x \,|\, Z) \quad \text{for all } x = 0, 1, \ldots, J. \tag{7.9}$$

Hence, in order to test $X \perp\!\!\!\perp_P C_X \,|\, Z$, we can select a C_X-measurable random variable W and see if Equation (7.9) holds.

Finally, if W is measurable with respect to $C_{X,t}$ then $X \perp\!\!\!\perp_P C_{X,t} \,|\, Z$ implies

$$P(X{=}x \,|\, Z, W) \underset{P}{=} P(X{=}x \,|\, Z) \quad \text{for all } x = 0, 1, \ldots, J. \tag{7.10}$$

This shows: In order to test $X \perp\!\!\!\perp_P C_{X,t} \,|\, Z$, we can choose a $C_{X,t}$-measurable random variable W and see if Equation (7.10) holds. Note that all these tests can only *falsify* but not verify these independent cause conditions.

Example 7.3 (First Examples) A first example of independence of the cause X and a global covariate C_X has already been presented in Table 6.2 (p. 129). Furthermore, Table 6.3 (p. 130) displays an example for Z-conditional independence of X and C_X. In both examples U takes the role of a global covariate C_X. In section 7.4 we treat several examples in more detail. ◁

7.2 Mean-Independent Outcome Conditions

Before we study the implications of the independent cause conditions on unbiasedness (see section 7.3), we introduce a second kind of causality conditions, which we will call the *mean-independent outcome conditions*.

7.2.1 Conditional Mean Independence

Instead of conditional independence, we now use the concept of *conditional mean independence*. If X, Y, and Z are random variables on a probability space $(\Omega, \mathscr{A}, P)$ such that Y is nonnegative or with finite expectation $E(Y)$, then

$$E(Y|X,Z) \underset{P}{=} E(Y|X) \tag{7.11}$$

defines *X-conditional mean independence of Y from Z*. (For more details see section PR-10.6.)

7.2.2 Mean-Independent Outcome Conditions for Total Effects

Hence, if $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ is a causality space and Y is numerical nonnegative or with finite expectation, then

$$E(Y|X,C_X) \underset{P}{=} E(Y|X) \tag{7.12}$$

defines X-conditional mean independence of Y from C_X, which will be abbreviated by $Y \vdash C_X|X$. Furthermore, if Z is measurable with respect to C_X, then

$$E(Y|X,C_X) \underset{P}{=} E(Y|X,Z) \tag{7.13}$$

defines *(X,Z)-conditional mean independence of Y from C_X*, abbreviated by $Y \vdash C_X|X,Z$.

If $x \in \Omega'_X$ and $P(X{=}x) = 0$, then Equation (7.12) implies

$$E^{X=x}(Y|C_X) \underset{P^{X=x}}{=} E^{X=x}(Y) \tag{7.14}$$

[see PR-(14.82)], and (7.13) implies

$$E^{X=x}(Y|C_X) \underset{P^{X=x}}{=} E^{X=x}(Y|Z) \tag{7.15}$$

[see PR-(14.81)].

The two conditions $Y \vdash C_X \,|\, X$ and $Y \vdash C_X \,|\, X, Z$ will also be referred to as the *mean-independent outcome conditions for total effects*. Assuming that Z is measurable with respect to C_X implies $\sigma(X, Z, C_X) = \sigma(X, C_X)$, which in turn implies

$$E(Y \,|\, X, Z, C_X) \underset{P}{=} E(Y \,|\, X, C_X) \tag{7.16}$$

[see Eq. (10.1)].

Remark 7.4 (Mean-Independent Outcome Conditions Again) Referring to the filtration $(\mathscr{F}_t, t \in T)$, note that Equation (7.12) is equivalent to

$$E(Y \,|\, \mathscr{F}_{t_X}) \underset{P}{=} E(Y \,|\, X) \tag{7.17}$$

(see Rem. 3.23). According to this equation, there are no random variables other than X that are prior or simultaneous to X and determine the conditional expectation of Y, once we condition on X. Intuitively speaking, X is the only cause of Y, at least with respect to all potential causes that are prior or simultaneous to X. Hence, this condition does not exclude other causes of Y that are in between X and Y, that is posterior to X and prior to Y unless T is such that there is no $t \in T$ such that $t_X < t < t_Y$.

Similarly, Equation (7.13) is equivalent to

$$E(Y \,|\, \mathscr{F}_{t_X}) \underset{P}{=} E(Y \,|\, X, Z), \tag{7.18}$$

because $\sigma(X, C_X) = \mathscr{F}_{t_X}$ [see Def. 4.1 (b)], provided of course that Z is C_X-measurable, that is, provided that Z is a covariate of X. According to Equation (7.18), there are no random variables other than X and Z that are prior or simultaneous to X and determine the conditional expectation of Y, once we condition on X and Z. Intuitively speaking, conditioning on Z, the random variable X is the only cause of Y, at least with respect to all potential causes that are prior or simultaneous to X. $\lhd$

7.2.3 Mean-Independent Outcome Conditions for Direct Effects

Now we specify the mean-independent outcome conditions *for direct effects*. If Z is measurable with respect to $C_{X,t}$, then

$$E(Y \,|\, X, C_{X,t}) \underset{P}{=} E(Y \,|\, X, Z) \tag{7.19}$$

defines *Z-conditional mean independence of Y from $C_{X,t}$*. It is abbreviated by $Y \vdash C_{X,t} \,|\, X, Z$. This condition will be also be referred to as the *mean-independent outcome condition for t-direct effects*. If X is discrete with values $x = 0, 1, \ldots, J$ and $P(X{=}x) = 0$ for all $x = 0, 1, \ldots, J$, then Equation (7.13) implies

$$E^{X=x}(Y \,|\, C_{X,t}) \underset{P^{X=x}}{=} E^{X=x}(Y \,|\, Z) \tag{7.20}$$

Again, note that measurability of Z with respect to $C_{X,t}$ implies

$$E(Y \,|\, X, Z, C_{X,t}) \underset{P}{=} E(Y \,|\, X, C_{X,t}). \tag{7.21}$$

Example 7.5 (Implications if U is Measurable With Respect to C_X) Assume that X represents a treatment variable and C_X is a global covariate such that the observational-unit variable U is measurable with respect to C_X. Then $Y \vdash C_X \,|\, X$ implies that there are no individual differences between units with respect to their expectation of the outcome variable Y within each treatment condition x, i. e.,

$$E(Y \,|\, X{=}x, U{=}u) \;=\; E(Y \,|\, X{=}x) \quad \text{for all values } x \text{ of } X \text{ and all units } u,$$

provided that $P(X{=}x, U{=}u) > 0$. This equation may more conveniently be written

$$E(Y \,|\, X, U) \underset{P}{=} E(Y \,|\, X)\,.$$

Note that using this equation, we do not have to presume $P(X{=}x, U{=}u) > 0$.

Similarly, $Y \vdash C_X \,|\, X, Z$ implies that no covariate affects the expectations of Y over and above X and the covariate Z. And again, if X represents a treatment variable, this implies that there are no individual differences between units with respect to their conditional expectation values of the outcome variable Y within treatment conditions *given* Z, i. e.,

$$E(Y \,|\, X, Z, U) \underset{P}{=} E(Y \,|\, X, Z)\,.$$

If $P(X{=}x, Z{=}z, U{=}u) > 0$, this equation implies

$$E(Y \,|\, X{=}x, Z{=}z, U{=}u) \;=\; E(Y \,|\, X{=}x, Z{=}z)\,.$$

Table 7.1 (p. 175) displays an example. ◁

7.2.4 Falsifiability of the Mean-Independent Outcome Conditions

The mean-independent outcome conditions imply some propositions that lend themselves for falsification. Let $\langle\, (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \,\rangle$ be a causality space and let Y be numerical and nonnegative or with finite expectation. Then the following propositions hold:

(i) If W is measurable with respect to C_X, then $Y \vdash C_X \,|\, X$ implies

$$E(Y \,|\, X, W) \underset{P}{=} E(Y \,|\, X)\,. \tag{7.22}$$

(ii) If Z and W are measurable with respect to C_X, then $Y \vdash C_X \,|\, X, Z$ implies

$$E(Y \,|\, X, Z, W) \underset{P}{=} E(Y \,|\, X, Z)\,. \tag{7.23}$$

(iii) If Z and W are measurable with respect to $C_{X,t}$, then $Y \vdash C_{X,t} \,|\, X, Z$ implies

$$E(Y \,|\, X, Z, W) \underset{P}{=} E(Y \,|\, X, Z)\,. \tag{7.24}$$

Propositions (i) and (ii) immediately follow from measurability of W with respect to C_X, and Proposition (iii) follows from measurability of W with respect to $C_{X,t}$.

Hence, just as the independent cause conditions, the mean-independent outcome conditions can easily be tested empirically, again only in the sense of falsifiability. For example, in order to test $Y \vdash C_X \,|\, X$, we simply have to select a variable W that is measurable with respect to C_X and see if Equation (7.22) holds. If it does not, then the hypothesis $Y \vdash C_X \,|\, X$ is falsified. Similarly, if we want to test $Y \vdash C_X \,|\, X, Z$, we may select a variable W that is measurable with respect to C_X and see if Equation (7.23) holds. If it does not, then we can conclude that $Y \vdash C_X \,|\, X, Z$ is not satisfied. Finally, if we want to test $Y \vdash C_{X,t} \,|\, X, Z$, we may select a variable W that is measurable with respect to $C_{X,t}$ and see if Equation (7.24) holds. If it does not, then we can conclude that $Y \vdash C_{X,t} \,|\, X, W$ is not satisfied.

7.3 Implications on Unbiasedness

Now we present the most important implications of the causality conditions introduced above on unbiasedness of various conditional expectations and prima facie effect functions. The theorems and corollaries presented in this section summarize the most important points with respect to unbiasedness of the conditional expectations $E(Y \,|\, X)$ and $E(Y \,|\, X, Z)$. After treating some examples in section 7.4, we will discuss the methodological implications of our first two causality conditions in section 7.5.

7.3.1 Unbiasedness With Respect to Total Effects: No Covariates

In this section we study the implications of our first two causality conditions on unbiasedness *with respect to total effects*. Let us start without considering covariates.

Conditional Expectations

In our first theorem we consider a single value x of X with $P(X{=}x) > 0$ and the conditional expectation value $E(Y \,|\, X{=}x)$. Reading the following theorem, remember that τ_x denotes the conditional expectation $E^{X=x}(Y \,|\, C_X)$ (see Def. 4.12).

Theorem 7.6 (Unbiasedness of Conditional Expectation Values)
Let $\langle\, (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \,\rangle$ be a causality space, let Y be numerical and nonnegative or with finite expectation, and let $x \in \Omega'_X$ be a value of X with $P(X{=}x) > 0$.

(i) Then $X \underset{P}{\perp\!\!\!\perp} C_X$ implies $P(X{=}x \,|\, C_X) \underset{P}{>} 0$ and

$$E(Y \,|\, X{=}x) \;=\; E(\tau_x). \tag{7.25}$$

(ii) $P(X{=}x\,|\,C_X) \gtrsim_P 0$ *and* $Y \vdash C_X\,|\,X$ *imply Equation* (7.25).

(Proof p. 181)

Remark 7.7 (Randomized Experiment) Remember that Equation (7.25) defines τ_x-unbiasedness of $E(Y\,|\,X{=}x)$. The first sufficient condition for unbiasedness, the independent cause condition $X \perp\!\!\!\perp_P C_X$, can be created by randomized assignment of the unit to one of the treatment conditions. In this case, X and C_X are independent. In contrast, there is no design technique implying that $Y \vdash C_X\,|\,X$ holds. ◁

Remark 7.8 (Less Restrictive Conditions Implying Unbiasedness) If $P(X{=}x) > 0$, then $1_{X=x} \perp\!\!\!\perp_P C_X$ implies $E(Y\,|\,X{=}x) = E(\tau_x)$. Similarly, if $P(X{=}x\,|\,C_X) \gtrsim_P 0$, then $Y \vdash C_X\,|\,1_{X=x}$ implies $E(Y\,|\,X{=}x) = E(\tau_x)$. This immediately follows from Theorem 7.6 for $X := 1_{X=x}$. ◁

Prima Facie Effects

The following corollary is another immediate implication of Theorem 7.6. In this corollary we specify two conditions under which the prima facie effect

$$PFE_{xx'} := E(Y\,|\,X{=}x) - E(Y\,|\,X{=}x')$$

is $\delta_{xx'}$-unbiased. Reading this corollary, remember that $\delta_{xx'} := \tau_x - \tau_{x'}$ denotes the atomic total-effect variable (see Def. 4.12) and that $E(\delta_{xx'}) = E(\tau_x) - E(\tau_{x'})$.

Corollary 7.9 (Unbiasedness of Prima Facie Effects)
Let $\langle\,(\Omega,\mathscr{A},P),(\mathscr{F}_t,t\in T),X,Y\,\rangle$ *be a causality space, let* Y *be numerical and nonnegative or with finite expectation, and let* $x,x'\in\Omega'_X$ *be two values of* X.

(i) If $P(X{=}x)>0$, $P(X{=}x')>0$, *and* $X \perp\!\!\!\perp_P C_X$, *then*

$$P(X{=}x\,|\,C_X) \gtrsim_P 0 \quad \textit{and} \quad P(X{=}x'\,|\,C_X) \gtrsim_P 0 \tag{7.26}$$

and

$$PFE_{xx'} = E(\delta_{xx'}). \tag{7.27}$$

(ii) If (7.26) *holds and* $Y \vdash C_X\,|\,X$, *then Equation* (7.27) *holds as well.*

Conditional Expectation

Theorem 7.6 also implies the following corollary concerning the conditional expectation

$$E(Y\,|\,X) := \sum_{x=0}^{J} E(Y\,|\,X{=}x)\cdot 1_{X=x}. \tag{7.28}$$

Corollary 7.10 (Unbiasedness of the Conditional Expectation)
Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ *be a causality space, let* Y *be numerical and nonnegative or with finite expectation, and let* X *be discrete with values* $x = 0, 1, \ldots, J$.

(i) If $P(X{=}x) > 0$ *for* all $x = 0, 1, \ldots, J$ *and* $X \underset{P}{\perp\!\!\!\perp} C_X$, *then*

$$P(X{=}x \mid C_X) \underset{P}{>} 0 \quad \textit{for all } x = 0, 1, \ldots, J \tag{7.29}$$

and

$$E(Y \mid X{=}x) = E(\tau_x) \quad \textit{for all } x = 0, 1, \ldots, J. \tag{7.30}$$

(ii) If (7.29) *holds and* $Y \vdash C_X \mid X$, *then Equation* (7.30) *holds as well.*

Hence, if X is discrete with values $x = 0, 1, \ldots, J$, then $X \underset{P}{\perp\!\!\!\perp} C_X$ implies τ_x-unbiasedness of the conditional expectation $E(Y \mid X)$ [see Eq. (7.28)], provided that X is discrete with nonzero probabilities for all its values. Similarly, $Y \vdash C_X \mid X$ implies τ_x-unbiasedness of the conditional expectation $E(Y \mid X)$, provided that X is discrete and $P(X{=}x \mid C_X) \underset{P}{>} 0$ holds for all its values x. Remember that $E(Y \mid X)$ denotes the conditional expectation of Y given X, that is, the best fitting function of X, and *not* $Q_{lin}(Y \mid X)$, the best fitting *linear* function of X (see Def. PR-7.2).

7.3.2 Conditioning on a Covariate

Now we extend Theorem 7.6 and its corollaries to the case in which we condition on a covariate Z. Let us start considering the implications of the independent cause condition and mean-independent outcome condition on the conditional expectation $E^{X=x}(Y \mid Z)$ and the conditional expectation values $E(Y \mid X{=}x, Z{=}z)$. Remember, the term $E^{X=x}(Y \mid Z)$ denotes the Z-conditional expectation of Y with respect to the $(X{=}x)$-conditional-probability measure $P^{X=x}$. Hence, if X represents a treatment variable, then $E^{X=x}(Y \mid Z)$ is the Z-conditional expectation of Y in treatment x. Also remember that τ_x-*unbiasedness of* $E^{X=x}(Y \mid Z)$ is defined by

$$E^{X=x}(Y \mid Z) \underset{P}{=} E(\tau_x \mid Z), \tag{7.31}$$

whereas

$$E(Y \mid X{=}x, Z{=}z) = E(\tau_x \mid Z{=}z). \tag{7.32}$$

defines τ_x-*unbiasedness of* $E(Y \mid X{=}x, Z{=}z)$.

Theorem 7.11 (Unbiasedness of the Conditional Expectation $E^{X=x}(Y \mid Z)$)
Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ *be a causality space, let* Y *be numerical and nonnegative or with finite expectation, let* $x \in \Omega'_X$ *be a value of* X *with* $P(X{=}x \mid C_X) \underset{P}{>} 0$, *and let* Z *be measurable w.r.t.* C_X.

(i) *Then each of* $X \perp\!\!\!\perp_P C_X \mid Z$ *and* $Y \vdash C_X \mid X, Z$ *implies that* $E^{X=x}(Y \mid Z)$ *is* τ_x*-unbiased.*
(ii) *If* x, z *are two values of* X *and* Z *such that* $P(X{=}x, Z{=}z) > 0$, *then each of* $X \perp\!\!\!\perp_P C_X \mid Z$ *and* $Y \vdash C_X \mid X, Z$ *implies that* $E(Y \mid X{=}x, Z{=}z)$ *is* τ_x*-unbiased.*

(Proof p. 183)

Hence, presuming $P(X{=}x \mid C_X) \gtrsim_P 0$, the Z-conditional expectation $E^{X=x}(Y \mid Z)$ of Y is unbiased with respect to total effects if $X \perp\!\!\!\perp_P C_X \mid Z$ or $Y \vdash C_X \mid X, Z$.

Remark 7.12 (Common Support) In Theorem 7.11 we required $P(X{=}x \mid C_X) \gtrsim_P 0$. If Z is C_X-measurable, then this implies $P(X{=}x \mid Z) \gtrsim_P 0$, which in turn implies $P(X{=}x \mid Z{=}z) > 0$ for all values z of Z that have a positive probability $P(Z{=}z) > 0$. The condition $P(X{=}x \mid Z) \gtrsim_P 0$ may also be called the *common support condition* (cf., e. g., Lechner, 2000). Note that common support refers to conditional *probabilities* and not to relative frequencies. At the level of a concrete sample, we may use the term *data sparseness* if the relative frequencies estimating the conditional probabilities $P(X{=}x \mid Z{=}z)$ are zero. ◁

Remark 7.13 (Conditional Randomization) The first condition implying unbiasedness of the $E^{X=x}(Y \mid Z)$ with respect to total effects, $X \perp\!\!\!\perp_P C_X \mid Z$, can be created in an experiment either by (a) randomized assignment of units to treatment conditions (represented by X) or (b) randomized assignment of units to treatment conditions *conditional on the values* z *of a covariate* Z. However, it may also hold if the covariate Z is properly selected. In contrast, the condition $Y \vdash C_X \mid X, Z$ cannot be created by a design technique. Instead we can only try to find a (possibly multivariate) covariate Z such that $Y \vdash C_X \mid X, Z$ holds. In other words, $Y \vdash C_X \mid X, Z$ can only be used for covariate selection (see section 7.5 for more details). ◁

Prima-Facie-Effect Functions

Remember, the prima facie effect function is defined by $PFE_{xx';Z} := E^{X=x}(Y \mid Z) - E^{X=x'}(Y \mid Z)$. Hence, unbiasedness of the conditional expectations $E^{X=x}(Y \mid Z)$ and $E^{X=x'}(Y \mid Z)$ has implications on unbiasedness of the prima-facie-effect function $PFE_{xx';Z}$ and on the $(Z{=}z)$-conditional prima facie effect $PFE_{xx';Z=z}$.

Corollary 7.14 (Unbiasedness of Conditional Prima Facie Effects)
Let $\langle (\Omega, \mathcal{A}, P), (\mathcal{F}_t, t \in T), X, Y \rangle$ *be a causality space, let* Y *be numerical and nonnegative or with finite expectation, let* $x, x' \in \Omega'_X$ *be two values of* X *with* $P(X{=}x \mid C_X) \gtrsim_P 0$ *and* $P(X{=}x' \mid C_X) \gtrsim_P 0$, *let* Z *be measurable w.r.t.* C_X, *and let* $\delta_{xx'} := \tau_x - \tau_{x'}$.

(i) Each of the two conditions $X \perp\!\!\!\perp_P C_X \,|\, Z$ *and* $Y \vdash C_X \,|\, X, Z$ *implies*

$$PFE_{xx';Z} := E^{X=x}(Y|Z) - E^{X=x'}(Y|Z) \underset{P}{=} E(\delta_{xx'}|Z).$$

(ii) If $P(X{=}x, Z{=}z), P(X{=}x', Z{=}z) > 0$*, then each of the two conditions* $X \perp\!\!\!\perp_P C_X \,|\, Z$ *and* $Y \vdash C_X \,|\, X, Z$ *implies*

$$PFE_{xx';Z=z} := E(Y|X{=}x, Z{=}z) - E(Y|X{=}x', Z{=}z) = E(\delta_{xx'}|Z{=}z).$$

(Proof p. 184)

According to this corollary, the prima-facie-effect function $PFE_{xx';Z}$ is unbiased with respect to total effects if (a) $P(X{=}x|C_X) \underset{P}{>} 0$ and $P(X{=}x'|Z) \underset{P}{>} 0$ and (b) the independent cause condition $X \perp\!\!\!\perp_P C_X \,|\, Z$ or the mean-independent outcome condition $Y \vdash C_X \,|\, X, Z$, hold. Correspondingly, the prima facie effect $PFE_{xx';Z=z}$ is unbiased with respect to the $(Z{=}z)$-conditional total effect of x vs. x', provided that $P(X{=}x, Z{=}z) > 0$ and $P(X{=}x', Z{=}z) > 0$ and at least one of these two causality conditions holds.

Now let us turn to the implications of our first two causality conditions on the conditional expectation $E(Y|X,Z)$.

Corollary 7.15 (Unbiasedness of the Conditional Expectation $E(Y|X,Z)$**)** *Let* $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ *be a causality space, let* Y *be numerical and nonnegative or with finite expectation, let* X *be discrete with a finite number of values* $x = 0, 1, \ldots, J$*, let* Z *be measurable w.r.t.* C_X*, and let* $P(X{=}x|C_X) \underset{P}{>} 0$ *for all values* $x = 0, 1, \ldots, J$ *of* X*. Then each of the two conditions* $X \perp\!\!\!\perp_P C_X \,|\, Z$ *and* $Y \vdash C_X \,|\, X, Z$ *implies that* $E(Y|X,Z)$ *is* τ_x*-unbiased.*

(Proof p. 184)

7.3.3 Unbiasedness With Respect to Direct Effects

Now we study the implications of the independent cause and mean-independent outcome conditions on unbiasedness *with respect to* t*-direct effects*. Remember that $\tau_{x,t}$*-unbiasedness of* $E^{X=x}(Y|Z)$ is defined by

$$E^{X=x}(Y|Z) \underset{P}{=} E(\tau_{x,t}|Z), \tag{7.33}$$

whereas

$$E(Y|X{=}x, Z{=}z) = E(\tau_{x,t}|Z{=}z). \tag{7.34}$$

defines $\tau_{x,t}$*-unbiasedness of* $E(Y|X{=}x, Z{=}z)$.

Theorem 7.16 (Unbiasedness of the Conditional Expectation $E^{X=x}(Y|Z)$) *Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a causality space, let Y be numerical and nonnegative or with finite expectation, let $t \in T$ with $t_X \le t < t_Y$, let Z be measurable w.r.t. $C_{X,t}$, and let $x \in \Omega'_X$ be a value of X with $P(X=x|C_{X,t}) \gtrsim_P 0$.*

(i) Then each of $X \perp\!\!\!\perp_P C_{X,t}|Z$ and $Y \vdash C_{X,t}|X,Z$ implies

$$E^{X=x}(Y|Z) \eqsim_P E(\tau_{x,t}|Z). \tag{7.35}$$

(ii) If x and z are values of X and Z with $P(X=x, Z=z) > 0$, then each of $X \perp\!\!\!\perp_P C_{X,t}|Z$ and $Y \vdash C_{X,t}|X,Z$ implies

$$E^{X=x}(Y|Z=z) = E(\tau_{x,t}|Z=z). \tag{7.36}$$

(Proof p. 184)

Remark 7.17 Equation (7.35) means that $E^{X=x}(Y|Z)$ is $\tau_{x,t}$-unbiased. Because $P(X=x) > 0$,

$$E^{X=x}(Y|Z=z) = E(Y|X=x, Z=z) \tag{7.37}$$

(see Remark PR-14.37). Therefore Equation (7.36) is equivalent to

$$E(Y|X=x, Z=z) = E(\tau_{x,t}|Z=z). \tag{7.38}$$

◁

Now we consider the implications of the two causality conditions on the prima facie effect function $PFE_{xx';Z}$ and the prima facie effect $PFE_{xx';Z=z}$.

Corollary 7.18 (Unbiasedness of Conditional Prima Facie Effects) *Let the assumptions of Theorem 7.16 be true, let $x, x' \in \Omega'_X$ be two values of X with $P(X=x|C_{X,t}) \gtrsim_P 0$ and $P(X=x'|C_{X,t}) \gtrsim_P 0$, and let $\delta_{xx',t} := \tau_{x,t} - \tau_{x',t}$.*

(i) Then each of the conditions $X \perp\!\!\!\perp_P C_{X,t}|Z$ and $Y \vdash C_{X,t}|X,Z$ implies

$$PFE_{xx';Z} := E^{X=x}(Y|Z) - E^{X=x'}(Y|Z) \eqsim_P E(\delta_{xx',t}|Z).$$

(ii) If $P(X=x, Z=z)$, $P(X=x', Z=z) > 0$, then each of the two conditions $X \perp\!\!\!\perp_P C_{X,t}|Z$ and $Y \vdash C_{X,t}|X,Z$ implies

$$\begin{aligned} PFE_{xx';Z=z} &:= E(Y|X=x, Z=z) - E(Y|X=x', Z=z) \\ &\eqsim_P E(\delta_{xx',t}|Z=z). \end{aligned}$$

(Proof p. 185)

Hence, if the conditions specified in this corollary are met, then the prima facie effect function $PFE_{xx';Z}$ and the prima facie effect $PFE_{xx';Z=z}$ are $\delta_{xx',t}$-unbiased.

Now we turn to the implications of the two causality conditions on the conditional expectation $E(Y|X,Z)$, where Z is assumed to be measurable with respect to $C_{X,t}$. The following corollary immediately follows from Theorem 7.16.

Corollary 7.19 (Unbiasedness of the Conditional Expectation $E(Y|X,W)$) *Let $\langle(\Omega,\mathscr{A},P),(\mathscr{F}_t,t\in T),X,Y\rangle$ be a causality space, let Y be numerical and nonnegative or with finite expectation, let $t\in T$ with $t_X \le t < t_Y$, let Z be measurable w.r.t. $C_{X,t}$, let X be discrete with a finite number of values $x=0,1,\dots,J$, and assume that $P(X{=}x\,|\,C_{X,t}) \gneq_P 0$ for all $x=0,1,\dots,J$. Then each of $X \perp\!\!\!\perp_P C_{X,t}\,|\,Z$ and $Y \vdash C_{X,t}\,|\,X,Z$ implies*

$$E^{X=x}(Y\,|\,Z) \underset{P}{=} E(\tau_{x,t}\,|\,Z) \quad \textit{for all } x=0,1,\dots,J. \tag{7.39}$$

Equation (7.39) means that $E(Y|X,Z)$ is $\tau_{x,t}$-unbiased [see Def. 6.34 (ii)].

7.4 Examples

Now we illustrate the two causality conditions treated in this chapter by some examples.

7.4.1 Examples for the Independent Cause Conditions

In chapter 4 we introduced a first example with independence of X and a global covariate C_X (see Table 4.2, p. 82). A second example has already been presented in chapter 6 (see Table 6.2, p. 129), and in the same chapter (see Table 6.3, p. 130), there is also an example with Z-conditional independence of X and C_X. In all these examples, the global covariate is identical to the observational-unit variable U, i. e., $C_X = U$, implying

$$E(Y\,|\,X,C_X) \underset{P}{=} E(Y\,|\,X,U) \tag{7.40}$$

and

$$P(X{=}1\,|\,C_X) \underset{P}{=} P(X{=}1\,|\,U) \tag{7.41}$$

for the individual treatment probability function $P(X{=}1|U)$, whose values are the individual treatment probabilities $P(X{=}1|U{=}u)$. If the values u of U represent the observational units *at the onset of treatment*, $C_X = U$ will hold in empirical applications if (a) no fallible covariate is assessed and (b) there is no other variable that is simultaneous with respect to $(\mathscr{F}_t, t\in T)$ to the treatment variable X (such as a second treatment variable).

Example 7.20 (Independent Treatment) Table 6.2 (p. 129) displays an example in which $X \perp\!\!\!\perp_P C_X$ holds, where $C_X = U$. As is easily seen in this table, the individual treatment probabilities are the same for all units, namely

$$P(X{=}1\,|\,U) \;=\; P(X{=}1) \;=\; \frac{3}{4}.$$

$X \perp\!\!\!\perp U$ implies that X and *all* random variables that are measurable with respect to U are independent.

In the same example, $X \perp\!\!\!\perp_P C_X\,|\,Z$ holds as well, where $Z := \mathit{sex}$, because the $Z{=}z$-conditional treatment probabilities also do not depend on the units, i. e.,

$$P(X{=}1\,|\,U,Z) \;=\; P(X{=}1\,|\,Z) \;=\; P(X{=}1) \;=\; \frac{3}{4}.$$

$X \perp\!\!\!\perp_P U\,|\,Z$ implies that X and all random variables that are measurable with respect to U are also Z-conditionally independent. This is no coincidence but an implication of the fact that $X \perp\!\!\!\perp_P U$ and measurability of Z with respect to U implies $X \perp\!\!\!\perp_P U\,|\,Z$, which in turn implies Z-conditional independence of X and all random variables that are measurable with respect to U [see PR-Box 16.3 (vi)].

If we assume $C_X = U$, then, according to Corollary 7.10, independence of X and U implies unbiasedness of the treatment-conditional expectation $E(Y\,|\,X)$ with respect to average total effects. Furthermore, because in this example $P(X{=}0) > 0$ and $P(X{=}1) > 0$, independence of X and U also implies τ_x-unbiasedness of the conditional expectation values $E(Y\,|\,X{=}0)$ and $E(Y\,|\,X{=}1)$ and as well as δ_{10}-unbiasedness of the prima facie effect

$$\mathit{PFE}_{10} \;:=\; E(Y\,|\,X{=}1) - E(Y\,|\,X{=}0) \;\approx\; 102.333 - 92.333 \;\approx\; 10$$

(see Th. 7.6 and Cor. 7.9). In other words, $\mathit{PFE}_{10} = E(\delta_{10})$.

As already stated above, $X \perp\!\!\!\perp_P U \mid Z$ holds as well, where $Z := \mathit{sex}$. Because $C_X = U$ and Z is U-measurable, Theorem 7.11 implies that the conditional expectation $E(Y\,|\,X,Z)$ is unbiased with respect to total effects. Furthermore, because $P(X{=}0,Z{=}z), P(X{=}1,Z{=}z) > 0$ for both values z of Z, this implies τ_x-unbiasedness of all conditional expectation values $E(Y\,|\,X{=}x,Z{=}z)$, and this in turn implies that the prima facie effects

$$\mathit{PFE}_{10;Z=m} \;=\; E(Y\,|\,X{=}1,Z{=}m) - E(Y\,|\,X{=}0,Z{=}m) \;\approx\; 92.50 - 83.00 \;\approx\; 9.50$$

and

$$\mathit{PFE}_{10;Z=f} \;=\; E(Y\,|\,X{=}1,Z{=}f) - E(Y\,|\,X{=}0,Z{=}f) \;=\; 122 - 111 \;=\; 11$$

(see Theorem 7.11 and Corollary 7.14) are δ_{10}-unbiased.

Also note that

$$\mathit{PFE}_{10} \;=\; E(\mathit{PFE}_{10;Z}) \;=\; 9.50 \cdot \frac{4}{6} + 11 \cdot \frac{2}{6} \;=\; 10.$$

This means that the prima facie effect is the expectation of the corresponding conditional prima facie effects. This property, which will be called *expectation stability*, is no coincidence. Instead it follows from $X \perp\!\!\!\perp_P C_X$ (see PR-Th. 15.14 and Rem. 15.17). ◁

Example 7.21 (Conditionally Independent Cause) In Table 6.3 (see p. 130) we already treated an example in which $X \underset{P}{\perp\!\!\!\perp} C_X \mid Z$ holds, whereas $X \underset{P}{\perp\!\!\!\perp} C_X$ does not. Again, $C_X = U$. As is easily seen in this table, the individual treatment probabilities are the same for all units *within each of the two subsets of males and females*, i. e.:

$$P(X{=}1 \mid U{=}u, Z{=}m) \;=\; P(X{=}1 \mid Z{=}m) \;=\; \frac{3}{4} \qquad \text{for each male unit } u$$

and

$$P(X{=}1 \mid U{=}u, Z{=}f) \;=\; P(X{=}1 \mid Z{=}f) \;=\; \frac{1}{4} \qquad \text{for each female unit } u.$$

Hence, the individual treatment probabilities are 3/4 for males and 1/4 for females. These individual treatment probabilities differ for different values of the covariate, but they are invariant *given a value of the covariate* Z. Note that the conditional treatment probability $P(X{=}1|Z{=}m) = 3/4$ is also the individual treatment probability $P(X{=}1 \mid U{=}u)$ for the male units u_1 to u_4, and the conditional treatment probability $P(X{=}1 \mid Z{=}f) = 1/4$ is also the individual treatment probability $P(X{=}1 \mid U{=}u)$ for the female units u_5 and u_6. This follows from

$$\begin{aligned} P(X{=}1 \mid U) \;&=\; E[P(X{=}1 \mid U, Z) \mid U] && \text{[PR-Box 10.2 (v)]} \\ &=\; E[P(X{=}1 \mid Z) \mid U] && [P(X{=}1 \mid U, Z) = P(X{=}1 \mid Z)] \\ &=\; P(X{=}1 \mid Z). && \text{[PR-Box 10.2 (vii)]} \end{aligned}$$

In Table 6.3 (p. 130) there is only *conditional* independence of units and treatments given Z, i. e., $X \underset{P}{\perp\!\!\!\perp} U \mid Z$. Therefore, in this table, the unconditional prima facie effect

$$PFE_{10} \;=\; E(Y|X{=}1) - E(Y|X{=}0) \;\approx\; 96.715 - 99.800 \;\approx\; -3.085$$

is biased, because the average total effect in this example is $E(\delta_{10}) = 10$ (see Exercise 7-2). However, the $(Z{=}z)$*-conditional* prima facie effects are unbiased. In fact, they are the same as in Table 6.2 (see p. 129 and Example 7.20). Hence, we can use the conditional prima facie effects to compute the average total effect. Remember, if the conditional prima facie effects are unbiased, i. e., if they are equal to the conditional total effects, then the expectation of the conditional prima facie effects is equal to the average total effect. In our example, this expectation is:

$$\begin{aligned} E(\delta_{10}) \;=\; E(PFE_{xx';Z}) \;&=\; PFE_{10;Z=m} \cdot \frac{4}{6} + PFE_{10;Z=f} \cdot \frac{2}{6} \\ &=\; 9.50 \cdot \frac{4}{6} + 11 \cdot \frac{2}{6} \;=\; 10.00. \end{aligned}$$

◁

Example 7.22 (Joe and Ann With Randomized Assignment – continued) In Table 4.5 we presented an example with an intermediate variable M and in Example 4.24 we showed how to compute the t-direct true-outcome variables $\tau_{0,t}$ and

$\tau_{1,t}$ as well as their difference $\delta_{10,t}$, the atomic t-direct-effect variable. Now we illustrate how to compute various total and direct causal effects, including the average t-direct effect and some conditional t-direct effects using the results of section 7.3.

First of all, the average total treatment effect is

$$E(\delta_{10}) = .10 \cdot .50 + .00 \cdot .50 \;=\; .05.$$

Because we assume $C_X = U$ and in this example $X \perp\!\!\!\perp_P U$, according to Theorem 7.6, this effect can also be computed by

$$E(\delta_{10}) \;=\; PFE_{10} \;=\; E(Y|X{=}1) - E(Y|X{=}0) \;=\; .50 - .45 \;=\; .05.$$

Next, we compute the U-conditional total-effect function $E(\delta_{10}|U)$ for the covariate U. The values of the conditional expectation $E(\delta_{10}\,|\,U)$ are the individual total treatment effects $E(\delta_{10}\,|\,U{=}\mathit{Joe}) = .10$ and $E(\delta_{10}\,|\,U{=}\mathit{Ann}) = .00$. Because $X \perp\!\!\!\perp_P C_X\,|U$, according to Corollary 7.14, these effects can also be computed by

$$PFE_{10;U} \;=\; E^{X=1}(Y|U) - E^{X=0}(Y|U),$$

the values of which are $PFE_{10;U=\mathit{Joe}} = E^{X=1}(Y\,|\,U{=}\mathit{Joe}) - E^{X=0}(Y\,|\,U{=}\mathit{Joe}) = .10$ for Joe and $PFE_{10;U=\mathit{Ann}} = E^{X=1}(Y\,|\,U{=}\mathit{Ann}) - E^{X=0}(Y\,|\,U{=}\mathit{Ann}) = .00$ for Ann.

Next, we consider the atomic t_M-direct-effect function δ_{10,t_M}, whose values are $\delta_{10,t_M}(\omega) = .058$ if $\omega \in \{M{=}0, U{=}\mathit{Joe}\}$ and $\delta_{10,t_M}(\omega) = .091$ if $\omega \in \{M{=}1, U{=}\mathit{Joe}\}$, and $\delta_{10,t_M}(\omega) = .003$ if $\omega \in \{M{=}0, U{=}\mathit{Ann}\}$ and $\delta_{10,t_M}(\omega) = -.023$ if $\omega \in \{M{=}1, U{=}\mathit{Ann}\}$ (see Table 4.5). Because $X \perp\!\!\!\perp_P C_{X,t_M}\,|M,U$ and $\delta_{10,t_M} = E(\delta_{10,t_M}\,|\,M,U)$, Corollary 7.18 implies that these effects can also be computed by

$$PFE_{10;M,U} \;=\; E^{X=1}(Y|M,U) - E^{X=0}(Y|M,U),$$

whose values are $E^{X=1}(Y|M{=}0,U{=}\mathit{Joe}) - E^{X=0}(Y|M{=}0,U{=}\mathit{Joe}) = .667 - .609 = .058$ and $E^{X=1}(Y|M{=}1,U{=}\mathit{Joe}) - E^{X=0}(Y|M{=}1,U{=}\mathit{Joe}) = .818 - .727 = .091$ for Joe, as well as $E^{X=1}(Y|M{=}0,U{=}\mathit{Ann}) - E^{X=0}(Y|M{=}0,U{=}\mathit{Ann},) = .179 - .176 = .003$ and, finally, $E^{X=1}(Y|M{=}1,U{=}\mathit{Ann}) - E^{X=0}(Y|M{=}1,U{=}\mathit{Ann},) = .227 - .250 = -.023$ for Ann.

Last but not least, the average t_M-direct treatment effect is

$$\begin{aligned} E(\delta_{10,t_M}) &= E(PFE_{10;M,U}) \\ &= .058 \cdot (.027 + .042 + .008 + .016) \;+\; .091 \cdot (.063 + .168 + .032 + .144) \;+ \\ &\quad\; .003 \cdot (.168 + .036 + .092 + .020) \;-\; .023 \cdot (.072 + .024 + .068 + .020) \\ &= .039147. \end{aligned}$$

Note that this effect can *not* be computed from the conditional expectation $E(Y|X,M)$ even though this is a randomized experiment such that $X \perp\!\!\!\perp_P C_X$ holds. In this example, the observational-unit variable U is the only covariate for which the prima facie effect function $E^{X=1}(Y|M,U) - E^{X=0}(Y|M,U)$ is unbiased, and, in this example, $\delta_{10,t_M} = E^{X=1}(Y|M,U) - E^{X=0}(Y|M,U)$. ◁

Table 7.1. Mean-Independent Outcome Condition

Unit u	Sampling probability $P(U=u)$	Covariate sex Z	True outcome variable under control τ_0	True outcome variable under treatment τ_1	True treatment probability $P(X=1\,\vert\,U)$	True total effect variable $\delta_{10}=\tau_1-\tau_0$	$P(U=u\vert X=0)$	$P(U=u\vert X=1)$
u_1	1/6	male	90	95	6/8	5	2/19	6/19
u_2	1/6	male	90	95	7/8	5	1/19	7/19
u_3	1/6	male	90	95	6/8	5	2/19	6/19
u_4	1/6	male	90	95	7/8	5	1/19	7/19
u_5	1/6	female	100	110	1/8	10	7/19	1/19
u_6	1/6	female	100	110	2/8	10	6/19	2/19

$E(\tau_0)$	$E(\tau_1)$	$P(X=1)$	$E(\delta_{10})$
93.333	100.000	29/48	6.667

$E(Y\vert X=0)$	$E(Y\vert X=1)$	PFE_{10}
96.842	96.552	−0.290

	male	female
Conditional causal effects	5.000	10.000
Conditional *PFE*s	5.000	10.000

7.4.2 Examples for the Mean-Independent Outcome Conditions

Now we illustrate the mean-independent outcome condition by a numerical example.

Example 7.23 Table 7.1 (p. 175) shows an example in which the mean-independent outcome condition $Y \vdash C_X \,|\, X, Z$ with respect to total effects holds. In this example, $C_X = U$, which implies

$$E(Y \,|\, X, C_X) \underset{P}{=} E(Y \,|\, X, U).$$

Hence, in this example, the true outcomes are the individual conditional expectation values $E(Y \,|\, X{=}x, U{=}u)$. As is easily seen in the table, the true outcomes under control and the atomic total effects of treatment 1 vs. 0 are the same for all individuals *within males* and *within females*. For all males, the true outcomes *under control* are

$$E(Y|X{=}0,Z{=}m,U{=}u_i) \;=\; E(Y|X{=}0,Z{=}m) \;=\; 90, \quad i=1,\ldots,4,$$

and for all females they are

$$E(Y|X{=}0,Z{=}f,U{=}u_i) \;=\; E(Y|X{=}0,Z{=}f) \;=\; 100, \quad i=5,6.$$

Under treatment, the true outcomes for all males are

$$E(Y|X{=}1,Z{=}m,U{=}u_i) \;=\; E(Y|X{=}1,Z{=}m) \;=\; 95, \quad i=1,\ldots,4,$$

and for all females they are

$$E(Y|X{=}1,Z{=}f,U{=}u_i) \;=\; E(Y|X{=}1,Z{=}f) \;=\; 110, \quad i=5,6.$$

Obviously, in both treatment groups, the expectations of the outcomes only depend on Z, i. e., $E^{X=x}(Y|U,Z) = E^{X=x}(Y|U) = E^{X=x}(Y|Z)$. Hence, we conclude that in this example

$$E(Y|X,C_X) \;\overline{=}_P\; E(Y|X,U) \;\overline{=}_P\; E(Y|X,Z).$$

Because $C_X = U$, this is the mean-independent outcome condition $Y \vdash C_X|X,Z$.

According to Theorem 7.11, $Y \vdash C_X|X,Z$ implies that $E(Y|X,Z)$ is unbiased with respect to total effects, provided that $P(X{=}1|Z) > 0$. Because, $P(X{=}x,Z{=}z) > 0$ for all values of X and Z, according to Corollary 7.14, this in turn implies that the conditional prima facie effects

$$PFE_{10;Z=m} \;=\; E(Y|X{=}1,Z{=}m) - E(Y|X{=}0,Z{=}m) \;=\; 95-90 \;=\; 5$$

and

$$PFE_{10;Z=f} \;=\; E(Y|X{=}1,Z{=}f) - E(Y|X{=}0,Z{=}f) \;=\; 110-100 \;=\; 10$$

are unbiased as well, i. e., $PFE_{10;Z=m} = E(\delta_{10}|Z{=}m)$ and $PFE_{10;Z=f} = E(\delta_{10}|Z{=}f)$. In other words, the $(Z{=}z)$-conditional prima facie effects are equal to the $(Z{=}z)$-conditional total effects.

Another conclusion is that the average of the conditional total effects is equal to the average total effect (see Remark 6.7), i. e.,

$$\begin{aligned} E(\delta_{10}) \;=\; E(PFE_{xx';Z}) \;&=\; PFE_{10;Z=m}\cdot 4/6 + PFE_{10;Z=f}\cdot 2/6 \\ &=\; 5\cdot 4/6 + 10\cdot 2/6 \;\approx\; 6.667. \end{aligned}$$

Finally, note that if $Y \vdash C_X|X,Z$ holds, the individual treatment probabilities $P(X{=}1\,|\,U{=}u)$ that played a crucial role in our examples for independence and Z-conditional independence of X and C_X do not matter any more in the following sense: Even though, in this example, the individual treatment probabilities $P(X{=}1\,|\,U{=}u)$ are different within the male and within the female subpopulations, the Z-conditional prima facie effects are unbiased. Just as the independent cause conditions, the mean-independent outcome conditions are sufficient for unbiasedness. ◁

7.5 Methodological Implications

Now we discuss the implications of the independent cause and the mean-independent outcome conditions for the design of experiments and quasi-experiments. Theorem 7.6 — specifically its proposition about the independent cause condition $X \perp\!\!\!\perp_P C_X$ — is the theoretical foundation of the design technique of *randomization* and of the analysis of total treatment effects in experiments by comparing (unadjusted) means between treatment conditions. Theorem 7.11 may also be used in randomized experiments when it comes to analyzing Z-conditional and average total treatment effects, and Theorem 7.16 can be used in randomized experiments as well as in (nonrandomized) quasi-experiments whenever we wish to analyze total, direct, and indirect treatment effects. In the analysis of direct and indirect effects, the independent cause condition $X \perp\!\!\!\perp_P C_{X,t} | Z$ and the mean-independent outcome condition $Y \vdash C_{X,t} | X,Z$ may be used for covariate selection. Finally, the causality conditions $X \perp\!\!\!\perp_P C_X | Z$ and $Y \vdash C_X | X,Z$ can also be used for *covariate selection* in quasi-experiments aiming only at the analysis of total effects.

Remark 7.24 (Randomization) It is well-known at least since (Fisher, 1925/1946) that *randomization* plays a crucial role in ruling out biases in comparisons of means. In a randomized experiment, there is always an observational-unit variable U whose values u denote the observational units. If the filtration $(\mathfrak{F}_t, t \in T)$ is appropriately specified, then U will be prior to X. Therefore, the definition of a global covariate C_X [see Def. **??** (b)] guarantees that U is measurable with respect to C_X. Hence, if X represents the treatment variable and C_X a global covariate, then $X \perp\!\!\!\perp_P C_X$, and this if $P(X{=}x) > 0$, then this implies

$$P(X{=}x | C_X) \;=\; P(X{=}x | U) \;=\; P(X{=}x).$$

According to this equation, each unit u has the same probability $P(X{=}x | U{=}u) = P(X{=}x)$ to be assigned to treatment x.

Remember that we are talking about a *single-unit trial*. A simple example of such a single-unit trial consists of sampling a single unit from a set of units, assessing a number of covariates, assigning the unit (or observing its assignment) to one of the treatment conditions and observing the outcome variable (see ch. 2 for more details and other kinds of single-unit trials).

Note that $P(X{=}x | U) = P(X{=}x)$ does *not* imply that the probabilities $P(X{=}x)$ are the same for all treatment conditions x. If, e. g., there are two treatment conditions, say 0 and 1, then $P(X{=}1)$ might be $1/4$ and $P(X{=}0) = 3/4$. However, $X \perp\!\!\!\perp_P C_X$ implies $P(X{=}1 | U) = P(X{=}1)$, provided, of course, that we consider a random experiment in which U is C_X-measurable. In other words, $X \perp\!\!\!\perp_P C_X$ implies that the individual treatment probabilities $P(X{=}1 | U{=}u)$ are identical between units, and they are equal to the unconditional treatment probability $P(X{=}1) = 1/4$. Hence, in a perfect randomized experiment we ensure $X \perp\!\!\!\perp_P C_X$, and this condition implies that the prima facie effect $E(Y | X{=}x) - E(Y | X{=}x')$ is identical to the average total effect $E(\delta_{xx'})$ (see Cor. 7.9).

◁

Remark 7.25 (Conditional Randomization) What has just been said about randomization also applies to to *conditional randomization* based on a covariate Z. Within the single-unit trial of an experiment or quasi-experiment, in which a unit u is sampled and a value z of the covariate Z is assessed prior to treatment, *conditional randomization* refers to randomized assignment of the unit u to treatment condition x with probability $P(X{=}x\,|\,Z{=}z)$, where $x = 0, 1, \ldots, J$. In this procedure we have to make sure that treatment assignment only depends on the values of Z, but not on any other attribute of the units or any other covariate. In more formal terms, we create $P(X{=}x\,|\,C_X) = P(X{=}x\,|\,Z)$.

We distinguish two cases. *First*, if Z is measured without measurement error, each value z of Z will represent subset of observational units (see, e. g., Rubin, 1984). Table 6.3 (p. 130) contains an example with independence of units and treatments within each of the two subsets of males and females.

Second, if the covariate Z is measured *with* error, i. e., $Z{=}f(U){+}\varepsilon$, then a value z of Z does *not* represent a subset of units, because z is not only determined by u but also by other factors and/or chance. Nevertheless, treatment can be assigned such that the treatment probability depends on the *observed score* z, not on the attribute $f(U)$ of the unit itself. In both cases, $Z{=}f(U)$ and $Z{=}f(U){+}\varepsilon$, conditional random assignment of a unit u to one of the treatment conditions given Z secures that a global covariate C_X satisfies $X \underset{P}{\perp\!\!\!\perp} C_X\,|\,Z$, i. e., Z-conditionally independence of X and C_X.

Conditional random assignment allows that units with different values z_1 and z_2 of Z have different probabilities to be assigned to treatment x. Thus it is possible to assign a unit for which we observe a covariate value z_1 (e. g., high need) to treatment x with a higher probability than a unit with covariate value z_2 (e. g., low need). In this way we may respect ethical and/or other requirements without compromising the validity of causal inference in such an experiment. Remember, unconditional random assignment means that each unit has the *same* probability of being assigned to a given treatment, regardless of his or her need (and any other attribute of the unit).

For simplicity, suppose there are just two treatment conditions, $X{=}0$ (control) and $X{=}1$ (treatment). Conditional randomization consists of:

(a) fixing the conditional treatment probabilities $P(X{=}1|Z)$ as a function of the covariate Z,
(b) sampling a unit u and assessing the value z of the covariate, and
(c) randomized assignment of the unit with probability $P(X{=}1\,|\,Z{=}z)$ to the treatment.

If, e. g., the covariate has three values, say *high*, *medium*, and *low need*, we may toss a dice and assign a unit with *high need* to treatment if the dice shows less than six dots, and assign it to control otherwise. A unit with *medium need* might be assigned to treatment if the dice shows less than four dots, and to control oth-

erwise. Finally, a unit with *low need* might be assigned to treatment if the dice shows one dot, and to control otherwise.

If the covariate Z is discrete, then the conditional treatment probabilities $P(X{=}1\,|\,Z{=}z)$ may be fixed in a table by assigning a treatment probability to each value z that seems appropriate with respect to ethical and other requirements. In the example above, these were the values 5/6 for *high need*, 3/6 for *medium need* and 1/6 for *low need*. If the covariate is univariate continuous, we may also use a function, such as

$$P(X{=}1\,|\,Z) \;=\; \frac{\exp(\lambda_0+\lambda_1\cdot Z)}{1+\exp(\lambda_0+\lambda_1\cdot Z)}$$

with coefficients λ_0 and λ_1 that seem appropriate for the experiment considered.

If $P(Z{=}z)>0$, then conditional randomization implies that the $(Z{=}z)$-conditional prima facie effects $PFE_{xx';Z=z} := E(Y\,|\,X{=}x,Z{=}z) - E(Y\,|\,X{=}x',Z{=}z)$ are the conditional total effects given a value z of the covariate Z. Hence, studying conditional prima facie effects in a perfect conditionally randomized experiment is tantamount to studying conditional total effects. As has been shown in chapter 5 [see, e. g., Eq. (5.4)], once we know the conditional total effects given the values z of a covariate Z, we can also compute the average total effect.

◁

Remark 7.26 (Mean-Independent Outcome Condition) Using $Y \vdash C_{X,t}\,|\,X,Z$ instead of $X \underset{P}{\perp\!\!\!\perp} C_{X,t}\,|\,Z$ in covariate selection might be easier. In many cases, a pre-test Z_1 of the outcome variable Y and a pre-test Z_2 of the intermediate variable M may already suffice to satisfy the mean-independent outcome condition $Y \vdash C_{X,t}\,|\,X,Z$ with $Z := (M,Z_1,Z_2)$. As has been shown in the example presented in section 1.3.2, such a pre-test of an intermediate variable is of utmost importance, whenever this pre-test Z_2 and the intermediate variable M are correlated, Z_2 and the pre-test Z_1 of the outcome are correlated, Z_1 and outcome Y correlate, and the treatment variable X affects the intermediate variable M. Note that these conditions are usually satisfied whenever there are systematic and relatively stable inter-individual differences in the outcome variable and in the intermediate variable, and X affects the intermediate variable M. A typical example is $Y :=$ *achievement* and $M :=$ *achievement motivation* in an educational experiment. ◁

Remark 7.27 (Covariate Selection in Randomized Experiments) Selecting the t-covariates in the vector $Z := (Z_1,\ldots,Z_m)$ for which we can hope that $X \underset{P}{\perp\!\!\!\perp} C_{X,t}\,|\,Z$ or $Y \vdash C_{X,t}\,|\,X,Z$ hold, are useful design techniques in the analysis of direct treatment effects. Even if there is independence of X and C_X — e. g., created by randomization — ignoring covariates and just analyzing $E(Y\,|\,X,M)$ will usually not yield the average and conditional direct treatment effects (see Example 7.22 and the example in section 1.3.2). ◁

Remark 7.28 (Covariate Selection in Quasi-Experiments) Covariate selection is also the first step in a number of techniques for the analysis of conditional and average total effects in *quasi-experiments*, in which the experimenter cannot fix the true treatment probabilities himself. The steps to follow distinguish different

techniques of analysis (see ch. 10). By definition, there is no randomization in a quasi-experiment. However, even under initial randomization, systematic attrition of subjects may invalidate the independent cause condition $X \perp\!\!\!\perp_P C_X$ (see, e. g., Abraham & Russell, 2004; Fichman & Cummings, 2003; Graham & Donaldson, 1993; Shadish et al., 2002). In this case, we will say that randomization failed and the initially randomized experiment turned into a quasi-experiment.

In quasi-experiments, selecting the covariates in the covariate vector $Z := (Z_1, \ldots, Z_m)$ for which we can hope that $X \perp\!\!\!\perp_P C_X \,|\, Z$ or $Y \vdash C_X \,|\, X, Z$ holds, is a useful strategy in the analysis of conditional and average total treatment effects. Again, compared to the independent cause condition $X \perp\!\!\!\perp_P C_X \,|\, Z$, it might be easier to find covariates such that the mean-independent outcome condition $Y \vdash C_X \,|\, X, Z$ holds. If, e. g., X is a treatment variable, in many cases a pre-test of the outcome variable Y will already suffice. In contrast, there might be many covariates determining the treatment probabilities. For instance, the *severity of the disorder*, *knowing about the treatment*, and *availability of the treatment* are candidates for such covariates. Often only $Z :=$ *pre-test severity of the disorder* may be important as a covariate to satisfy the mean-independent outcome condition $Y \vdash C_X \,|\, X, Z$ if the outcome variable Y is *post-test severity of the disorder*. However, there is no guarantee that the pre-test is sufficient as a covariate for the mean-independent outcome condition to hold. Whether or not it holds depends on the application considered. ◁

Remark 7.29 (Falsifiability) As already mentioned before, in contrast to some of the other causality conditions treated in the subsequent chapters, both the independent cause and the mean-independent outcome conditions can be tested in empirical applications, at least in the sense that some consequences of these assumptions can be checked. Falsifiability is important, because otherwise we would not have any criterion for covariate selection, i. e., for deciding whether or not a specific covariate should be included in the vector $Z := (Z_1, \ldots, Z_m)$ of covariates with respect to which we hope that either the independent cause condition or the corresponding mean-independent outcome condition holds.

Let us summarize how we can falsify the independent cause conditions. If $X \perp\!\!\!\perp_P C_X$, then

$$P(X{=}x \,|\, W) \underset{P}{=} P(X{=}x) \tag{7.42}$$

holds for each W that is measurable with respect to C_X. Examples for such variables W are sex, educational status, but also a fallible pre-test. Similarly, $X \perp\!\!\!\perp_P C_X \,|\, Z$ implies:

$$P(X{=}x \,|\, Z, W) \underset{P}{=} P(X{=}x \,|\, Z), \tag{7.43}$$

for each W that is measurable with respect to C_X. Furthermore, $X \perp\!\!\!\perp_P C_{X,t} \,|\, Z$ implies:

$$P(X{=}x \,|\, Z, W) \underset{P}{=} P(X{=}x \,|\, Z), \tag{7.44}$$

for each W that is measurable with respect to $C_{X,t}$. All these equations are easily tested using standard procedures for conditional probabilities such as logistic

(see, e. g., Agresti, 2007; Bonney, 1987; Green, 2003; Hosmer & Lemeshow, 2000) or probit regression (see, e. g., McCullagh & Nelder, 1989; Borooah, 2001; Liao, 1994).

Similarly, if W is measurable with respect to C_X, it is easy to test the mean-independent outcome conditions. For example, in regression analysis, it is well-known how to test

$$E(Y|X,W) \underset{P}{=} E(Y|X), \tag{7.45}$$

which follows from $Y \vdash C_X|X$. The same applies to

$$E(Y|X,Z,W) \underset{P}{=} E(Y|X,Z), \tag{7.46}$$

which follows from $Y \vdash C_X|X,Z$ and to

$$E(Y|X,Z,W) \underset{P}{=} E(Y|X,Z), \tag{7.47}$$

following from $Y \vdash C_{X,t}|X,W$, where Z and W are measurable with respect to $C_{X,t}$. For such tests we can use the well-known techniques of regression analysis (see, e. g., Aiken & West, 1996; Allen, 1997; Cohen et al., 2003; Draper & Smith, 1998; Gelman & Hill, 2007; Györfi, Kohler, Krzyzak, & Walk, 2002; von Eye & Schuster, 1998; West & Aiken, 2005). ◁

Remark 7.30 (Covariate Selection for Identifying Total and Direct Effects) Note that $Y \vdash C_X|X,Z$ does not imply $X \underset{P}{\perp\!\!\!\perp} C_X|Z$ and vice versa, $X \underset{P}{\perp\!\!\!\perp} C_X|Z$ does not imply $Y \vdash C_X|X,Z$. Furthermore, $Y \vdash C_{X,t}|X,Z$ does not imply $Y \vdash C_X|X,Z$ and $X \underset{P}{\perp\!\!\!\perp} C_{X,t}|Z$ does not imply $X \underset{P}{\perp\!\!\!\perp} C_X|Z$. Hence, covariate selection for the identification of total effects is a different process than selecting covariates for the identification of direct effects. ◁

Outlook

In the next chapters we will treat further sufficient conditions for unbiasedness that can replace the independent cause or the mean-independent outcome conditions and that will be the foundations of additional design techniques and methods of data analysis for the analysis of causal effects. In chapter 10, we will then show how to identify the various total, direct, and indirect effects from conditional expectations that can be estimated from samples of treatment, covariates, intermediate variables, and outcome variables, provided that one of the causality conditions holds.

7.6 Proofs

Proof of Theorem 7.6

(i) $X \underset{P}{\perp\!\!\!\perp} C_X \Rightarrow E(Y|X{=}x) = E(\tau_x)$:

Box 7.1 Glossary of New Concepts

We assume that $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ is a causality space, that Y is numerical and nonnegative or with finite expectation, and that Z denotes a covariate of X.

Conditions Implying Unbiasedness With Respect to Total Effects

$X \underset{P}{\perp\!\!\!\perp} C_X$ — *Independence of X and C_X*. This condition can be created by randomized assignment of units to treatments x. If $P(X{=}x) > 0$, then it implies $P(X{=}x \,|\, C_X) \underset{P}{>} 0$ and τ_x-unbiasedness of $E(Y \,|\, X{=}x)$.

$X \underset{P}{\perp\!\!\!\perp} C_X \,|\, Z$ — *Conditional independence of X and C_X given Z*. This condition can be created by conditional randomized assignment of units to treatment conditions based on the values z of Z. If $P(X{=}x \,|\, C_X) \underset{P}{>} 0$, then it implies that $E^{X=x}(Y \,|\, Z)$ is τ_x-unbiased. Furthermore, we can try to select the (possibly multivariate) covariate $Z = (Z_1, \ldots, Z_m)$ such that $X \underset{P}{\perp\!\!\!\perp} C_X \,|\, Z$ is satisfied. If $X \underset{P}{\perp\!\!\!\perp} C_X \,|\, Z$, then $1_{X=x} \underset{P}{\perp\!\!\!\perp} C_X \,|\, \pi_x$ holds for the Z-conditional propensity $\pi_x := P(X{=}x \,|\, Z)$, a fact that is used in adjustment techniques based on propensity scores.

$Y \vdash C_X \,|\, X$ — *Mean-independence of Y from C_X given X*. This condition is defined by $E(Y \,|\, X, C_X) \underset{P}{=} E(Y \,|\, X)$. It may or may not hold depending on the empirical phenomenon considered. If $P(X{=}x \,|\, C_X) \underset{P}{>} 0$, then $Y \vdash C_X \,|\, X$ implies that $E(Y \,|\, X{=}x)$ is τ_x-unbiased.

$Y \vdash C_X \,|\, X, Z$ — *Mean-independence of Y from C_X given X and Z*. It is defined by $E(Y \,|\, X, C_X) \underset{P}{=} E(Y \,|\, X, Z)$. We can try to select the (possibly multivariate) variable $Z = (Z_1, \ldots, Z_m)$ such that $Y \vdash C_X \,|\, X, Z$ holds. If $P(X{=}x \,|\, C_X) \underset{P}{>} 0$, then $Y \vdash C_X \,|\, X, Z$ implies that $E^{X=x}(Y \,|\, Z)$ is τ_x-unbiased.

Conditions Implying Unbiasedness With Respect to Direct Effects

Now we assume $t \in T$ with $t_X \le t < t_Y$ and that Z is a t-covariate of X.

$X \underset{P}{\perp\!\!\!\perp} C_{X,t} \,|\, Z$ — *Conditional independence of X and $C_{X,t}$ given Z*. If $t > t_X$, then this condition can *not* be created by any design technique. However, we may try to select the (possibly multivariate) t-covariate $Z = (Z_1, \ldots, Z_m)$ of X such that $X \underset{P}{\perp\!\!\!\perp} C_{X,t} \,|\, Z$ is satisfied. In this case, $1_{X=x} \underset{P}{\perp\!\!\!\perp} C_{X,t} \,|\, \pi_x$ holds for $\pi_x = P(X{=}x \,|\, Z)$, a fact that is used in adjustment techniques based on propensity scores. If $P(X{=}x \,|\, C_{X,t}) \underset{P}{>} 0$, then $X \underset{P}{\perp\!\!\!\perp} C_{X,t} \,|\, Z$ implies that $E^{X=x}(Y \,|\, Z)$ is $\tau_{x,t}$-unbiased.

$Y \vdash C_{X,t} \,|\, X, Z$ — *Mean-independence of Y from $C_{X,t}$ given X and Z*. It is defined by $E(Y \,|\, X, C_{X,t}) \underset{P}{=} E(Y \,|\, X, Z)$. We can try to select a (possibly multivariate) covariate $Z = (Z_1, \ldots, Z_m)$ such that $Y \vdash C_{X,t} \,|\, X, Z$ holds. If $P(X{=}x \,|\, C_{X,t}) \underset{P}{>} 0$, then $Y \vdash C_{X,t} \,|\, X, Z$ implies that $E^{X=x}(Y \,|\, Z)$ is $\tau_{x,t}$-unbiased.

$$\begin{aligned} E(Y|X{=}x) &= E^{X=x}(Y) && \text{[PR-(9.11)]} \\ &= E^{X=x}\big[E^{X=x}(Y|C_X)\big] && \text{[PR-Box 10.2 (iv)]} \\ &= E\big[E^{X=x}(Y|C_X)\big] && \big[X \perp\!\!\!\perp_P C_X, \text{PR-(16.40)}\big] \\ &= E(\tau_x). && \text{[(4.8)]} \end{aligned}$$

$Y \vdash C_X|X \Rightarrow E(Y|X{=}x) = E(\tau_x)$: Note that $P(X{=}x\,|\,C_X) \gtrsim_P 0$ implies that $E^{X=x}(Y|C_X)$ is P-unique [see Cor. PR-14.48 (a) and (c)]. Furthermore, according to (7.15), $E^{X=x}(Y|Z)$ is a version of $E^{X=x}(Y|C_X)$. Therefore, P-uniquenessof $E^{X=x}(Y|C_X)$ implies that $E^{X=x}(Y|Z)$ is P-unique as well. If Y is nonnegative, then there is a version $E^{X=x}(Y|C_X)$ that is nonnegative [see PR-(10.17)], and if $E(Y) < \infty$, then $E[E^{X=x}(Y|C_X)] < \infty$. Therefore, and because $E^{X=x}(Y|Z)$ is Z-measurable, we can apply PR-Box 10.2 (vii). Hence,

$$\begin{aligned} E(Y|X{=}x) &= E^{X=x}(Y) && \text{[PR-(9.11)]} \\ &= E\big[E^{X=x}(Y)\big] && \text{[PR-Box 6.1 (i)]} \\ &= E\big[E^{X=x}(Y|C_X)\big] && \text{[(7.14), } P\text{-uniqueness of } E^{X=x}(Y|C_X)\text{]} \\ &= E(\tau_x). && \text{[(4.8)]} \end{aligned}$$

(ii) If $P(X{=}x, Z{=}z) > 0$, then $P(X{=}x) > 0$ and $P(Z{=}z) > 0$. Furthermore, according to Remark PR-14.37, $P(X{=}x, Z{=}z) > 0$ implies that $E^{X=x}(Y|Z{=}z)$ is identical to $E(Y|X{=}x, Z{=}z)$. Therefore, Equation (7.32) follows from (7.31). This proves (ii).

Proof of Theorem 7.11

First of all note that $P(X{=}x\,|\,C_X) \gtrsim_P 0$ implies $P(X{=}x) > 0$, because

$$\begin{aligned} E[P(X{=}x\,|\,C_X)] &= E[E(\mathbb{1}_{X=x}\,|\,C_X)] && \text{[PR-(10.4)]} \\ &= E(\mathbb{1}_{X=x}) && \text{[PR-Box 10.2 (iv)]} \\ &= P(X{=}x). && \text{[PR-(6.5)]} \end{aligned}$$

Applying Theorem 3.52 (ii) then yields the proposition.

(i) $X \perp\!\!\!\perp_P C_X|Z \Rightarrow E^{X=x}(Y|Z) \eqsim_P E(\tau_x|Z)$:

$$\begin{aligned} E^{X=x}(Y|Z) &\underset{P^{X=x}}{=} E^{X=x}\big[E^{X=x}(Y|C_X)\,\big|\,Z\big] && \text{[PR-Box 10.2 (v)]} \\ &\underset{P^{X=x}}{=} E\big[E^{X=x}(Y|C_X)\,\big|\,Z\big] && \big[X \perp\!\!\!\perp_P C_X|Z, \text{PR-(16.39)}\big] \\ &\underset{P^{X=x}}{=} E(\tau_x|Z) && \text{[(4.8)]} \\ &\eqsim_P E(\tau_x|Z). && [P(X{=}x\,|\,C_X) \gtrsim_P 0, \text{ Cor. PR-14.48 (a), (c)}] \end{aligned}$$

$Y \vdash C_X|X, Z \Rightarrow E^{X=x}(Y|Z) \eqsim_P E(\tau_x|Z)$: Note that $P(X{=}x\,|\,C_X) \gtrsim_P 0$ implies that $E^{X=x}(Y|C_X)$ is P-unique [see Cor. PR-14.48 (a) and (c)]. Furthermore, according to (7.15), $E^{X=x}(Y|Z)$ is a version of $E^{X=x}(Y|C_X)$. Therefore, P-uniqueness of $E^{X=x}(Y|C_X)$ implies that $E^{X=x}(Y|Z)$ is P-unique as well. If Y is nonnegative, then there is a version $E^{X=x}(Y|C_X)$ that is nonnegative [see PR-(10.17)], and if $E(Y) < \infty$, then $E[E^{X=x}(Y|C_X)] < \infty$. Therefore, and because $E^{X=x}(Y|Z)$ is Z-measurable, we can apply PR-Box 10.2 (vii). Hence,

$$\begin{aligned} E^{X=x}(Y|Z) &\underset{P}{=} E\big[E^{X=x}(Y|Z)\,\big|\,Z\big] && \big[\text{PR-Box 10.2 (vii)}\big] \\ &\underset{P}{=} E\big[E^{X=x}(Y|C_X)\,\big|\,Z\big] && [(7.15),\ P\text{-uniqueness of } E^{X=x}(Y|C_X)] \\ &\underset{P}{=} E(\tau_x|Z). && [(4.8)] \end{aligned}$$

(ii) If $P(X{=}x, Z{=}z) > 0$, then $P(X{=}x) > 0$ and $P(Z{=}z) > 0$. Furthermore, according to Remark PR-14.37, $P(X{=}x, Z{=}z) > 0$ implies that $E^{X=x}(Y|Z{=}z)$ is identical to $E(Y|X{=}x, Z{=}z)$. Therefore, Equation (7.32) follows from (7.31). This proves (ii).

Proof of Corollary 7.14

(i) This proposition immediately follows from Theorem 7.11, because

$$\begin{aligned} PFE_{xx';Z} &:= E^{X=x}(Y|Z) - E^{X=x'}(Y|Z) \\ &\underset{P}{=} E(\tau_x|Z) - E(\tau_{x'}|Z) && [\text{Th. 7.11 (i)}] \\ &\underset{P}{=} E(\delta_{xx'}|Z). \end{aligned}$$

(ii) If $P(X{=}x, Z{=}z) > 0$ and $P(X{=}x', Z{=}z) > 0$, then proposition (ii) of Theorem 7.11 implies $E^{X=x}(Y|Z{=}z) = E(\tau_x|Z{=}z)$ and $E^{X=x'}(Y|Z{=}z) = E(\tau_{x'}|Z{=}z)$. Hence,

$$E^{X=x}(Y|Z{=}z) - E^{X=x'}(Y|Z{=}z) \;=\; E(\tau_x|Z{=}z) - E(\tau_{x'}|Z{=}z) \;=\; E(\delta_{xx'}|Z{=}z).$$

Proof of Corollary 7.15

Corollary 7.15 immediately follows from Theorem 7.11. Also note that Corollary 7.10 is a special case of Corollary 7.15 for Z being a constant.

Proof of Theorem 7.16

First of all note that $P(X{=}x|C_{X,t}) \underset{P}{>} 0$ implies $P(X{=}x) > 0$. This can be shown as follows:

$$\begin{aligned} E[P(X{=}x|C_{X,t})] &= E[E(1_{X=x}|C_{X,t})] && [(10.4)] \\ &= E(1_{X=x}) && [\text{PR-Box 10.2 (iv)}] \\ &= P(X{=}x). && [\text{PR-(6.5)}] \end{aligned}$$

Applying Theorem 3.52 (ii) then yields the proposition.

(i) $X \underset{P}{\perp\!\!\!\perp} C_{X,t}|Z \Rightarrow E^{X=x}(Y|Z) \underset{P}{=} E(\tau_{x,t}|Z)$:

$$\begin{aligned} E^{X=x}(Y|Z) &\underset{P^{X=x}}{=} E^{X=x}\big[E^{X=x}(Y|C_{X,t})\,\big|\,Z\big] && [\text{PR-Box 10.2 (v)}] \\ &\underset{P^{X=x}}{=} E\big[E^{X=x}(Y|C_{X,t})\,\big|\,Z\big] && \big[X \underset{P}{\perp\!\!\!\perp} C_{X,t}|Z, \text{PR-(16.39)}\big] \\ &\underset{P^{X=x}}{=} E(\tau_{x,t}|Z) && [(5.14)] \\ &\underset{P}{=} E(\tau_{x,t}|Z). && [P(X{=}x|C_{X,t}) \underset{P}{>} 0, \text{ Cor. PR-14.48 (a), (c)}] \end{aligned}$$

$Y \vdash C_{X,t} | X,Z \;\Rightarrow\; E^{X=x}(Y|Z) \underset{P}{=} E(\tau_{x,t}|Z)$: Note that $P(X{=}x \,|\, C_{X,t}) \underset{P}{>} 0$ implies that $E^{X=x}(Y|C_{X,t})$ is P-unique [see Cor. PR-14.48 (a) and (c)]. Furthermore, according to (7.20), $E^{X=x}(Y|Z)$ is a version of $E^{X=x}(Y|C_{X,t})$. Therefore, P-uniquenessof $E^{X=x}(Y|C_{X,t})$ implies that $E^{X=x}(Y|Z)$ is P-unique as well. If Y is nonnegative, then there is a version $E^{X=x}(Y|C_{X,t})$ that is nonnegative [see PR-(10.17)], and if $E(Y) < \infty$, then $E[E^{X=x}(Y|C_{X,t})] < \infty$. Therefore, and because $E^{X=x}(Y|Z)$ is Z-measurable, we can apply PR-Box 10.2 (vii). Hence,

$$\begin{aligned} E^{X=x}(Y|Z) &\underset{P}{=} E\big[\,E^{X=x}(Y|Z)\,\big|\,Z\big] && \big[\text{PR-Box 10.2 (vii)}\big] \\ &\underset{P}{=} E\big[\,E^{X=x}(Y|C_{X,t})\,\big|\,Z\big] && [(7.15),\ P\text{-uniqueness of } E^{X=x}(Y|C_{X,t})] \\ &\underset{P}{=} E(\tau_{x,t}|Z). && [(5.14)] \end{aligned}$$

(ii) If $P(X{=}x, Z{=}z) > 0$, then $P(X{=}x) > 0$ and $P(Z{=}z) > 0$. Furthermore, according to Remark PR-14.37, $P(X{=}x, Z{=}z) > 0$ implies that $E^{X=x}(Y|Z{=}z)$ is identical to $E(Y|X{=}x, Z{=}z)$. Therefore, Equation (7.34) follows from (7.33). This proves (ii).

Proof of Corollary 7.18

The proof is analogous to the proof of Corollary 7.14.

(i) This proposition immediately follows from Theorem 7.16, because

$$\begin{aligned} PFE_{xx';Z} &:= E^{X=x}(Y|Z) - E^{X=x'}(Y|Z) \\ &\underset{P}{=} E(\tau_{x,t}|Z) - E(\tau_{x',t}|Z) && [\text{Th. 7.16 (i)}] \\ &\underset{P}{=} E(\delta_{xx',t}|Z). \end{aligned}$$

(ii) If $P(X{=}x, Z{=}z) > 0$ and $P(X{=}x', Z{=}z) > 0$, then proposition (ii) of Theorem 7.16 implies $E^{X=x}(Y|Z{=}z) = E(\tau_{x,t}|Z{=}z)$ and $E^{X=x'}(Y|Z{=}z) = E(\tau_{x',t}|Z{=}z)$. Hence,

$$E^{X=x}(Y|Z{=}z) - E^{X=x'}(Y|Z{=}z) \;=\; E(\tau_{x,t}|Z{=}z) - E(\tau_{x',t}|Z{=}z) \;=\; E(\delta_{xx',t}|Z{=}z).$$

7.7 Exercises

▷ **Exercise 7-1** Proof that Equation (7.4) and C_X-measurability of Z imply Equation (7.6).

▷ **Exercise 7-2** Compute the conditional expectation values $E(Y|X{=}0)$ and $E(\tau_0)$ in the example presented in Table 6.3 (p. 130) and compare these numbers to each other.

▷ **Exercise 7-3** Assume that X is discrete with a finite number of values $x = 0, 1, \ldots, J$, and that $P(X{=}x \,|\, C_{X,t}) \underset{P}{>} 0$ for all $x = 0, 1, \ldots, J$. Which unbiasedness condition follows from $X \perp\!\!\!\perp_P C_X$ and which one from $X \perp\!\!\!\perp_P C_X | Z$?

▷ **Exercise 7-4** Assume that X is discrete with a finite number of values $x = 0, 1, \ldots, J$, and that $P(X{=}x \,|\, C_{X,t}) \underset{P}{>} 0$ for all $x = 0, 1, \ldots, J$. Which unbiasedness condition follows from $Y \vdash C_X | X$ and which one from $Y \vdash C_X | X, Z$?

▷ **Exercise 7-5** Describe randomized assignment of a unit to one of two treatment conditions!

▷ **Exercise 7-6** Describe *conditional* randomized assignment of a unit to one of two treatment conditions given a covariate!

▷ **Exercise 7-7** Show that $E[P(X{=}x\,|\,U)] = P(X{=}x)$.

▷ **Exercise 7-8** Show that $X \underset{P}{\perp\!\!\!\perp} U$ implies $X \underset{P}{\perp\!\!\!\perp} U\,|\,Z$, provided that Z is measurable with respect to U. Assume that X is discrete with a finite number of values $x = 0,1,\ldots,J$ and that $P(X{=}x) > 0$ for all values x of X.

▷ **Exercise 7-9** Show that $X \underset{P}{\perp\!\!\!\perp} C_X$ implies:

$$P(X{=}x\,|\,W) \underset{P}{=} P(X{=}x) \qquad \text{for each } x = 0,1,\ldots,J,$$

for each random variable W that is measurable with respect to C_X. Assume that X is discrete with a finite number of values $x = 0,1,\ldots,J$ and that $P(X{=}x) > 0$ for all values x of X.

▷ **Exercise 7-10** Show that $Y \vdash C_X\,|\,X,Z$ implies $E(Y\,|\,X,Z,W) \underset{P}{=} E(Y\,|\,X,Z)$ if both Z and W are measurable with respect to the global covariate C_X.

Solutions

▷ **Solution 7-1** For $x = 0,1,\ldots,J$,

$$\begin{aligned}
P(X{=}x\,|\,\pi_x) &\underset{P}{=} E(1_{X=x}\,|\,\pi_x) && \text{[PR-(10.4)]}\\
&\underset{P}{=} E[P(X{=}x\,|\,C_X)\,|\,\pi_x] && \text{[PR-Box 10.2 (v)]}\\
&\underset{P}{=} E[P(X{=}x\,|\,Z)\,|\,\pi_x] && \text{[(7.4)]}\\
&\underset{P}{=} E(\pi_x\,|\,\pi_x) && \text{[(7.5)]}\\
&\underset{P}{=} \pi_x && \text{[PR-Box 10.2 (vii)]}\\
&\underset{P}{=} P(X{=}x\,|\,Z) && \text{[(7.5)].}
\end{aligned}$$

Inserting $P(X{=}x\,|\,\pi_x)$ for $P(X{=}x\,|\,Z)$ in Equation (7.4) completes the proof.

▷ **Solution 7-2** The conditional expectation value $E(Y\,|\,X{=}0)$ can be computed via

$$E(Y\,|\,X{=}x) = \sum_u E(Y\,|\,X{=}x, U{=}u)\cdot P(U{=}u\,|X{=}x)$$

[see Rule (iv) in PR-Box 9.2]. In this example, this equation yields

$$\begin{aligned}
E(Y\,|\,X{=}0) &= \sum_u E(Y\,|\,X{=}0, U{=}u)\cdot P(U{=}u\,|X{=}0)\\
&= (68+78+88+98)\cdot\frac{1}{10} + (106+116)\cdot\frac{3}{10} = 99.8.
\end{aligned}$$

In contrast, the expectation $E(\tau_0)$ can be computed via

$$E(\tau_0) = \sum_u E(Y|X{=}x, U{=}u)\cdot P(U{=}u)$$
$$= (68+78+88+98+106+116)\cdot\frac{1}{6} = 92.3333.$$

Comparing $E(Y|X{=}0) = 99.8$ to $E(\tau_0) = 92.3333$ shows that $E(Y|X{=}0)$ is strongly biased.

▷ **Solution 7-3** The independent cause condition $X \perp\!\!\!\perp_P C_X$ implies that $E(Y|X)$ is unbiased with respect to total effects, whereas $X \perp\!\!\!\perp_P C_X|Z$ implies that $E(Y|X,Z)$ is unbiased with respect to total effects.

▷ **Solution 7-4** The mean-independent outcome condition $Y \vdash C_X|X$ implies that $E(Y|X)$ is unbiased with respect to total effects, whereas $Y \vdash C_X|X,Z$ implies that $E(Y|X,Z)$ is unbiased with respect to total effects.

▷ **Solution 7-5** In a random experiment, in which a unit u is sampled and assigned to one of two treatment conditions, we may assign the unit by coin toss, for instance. This makes sure that $P(X{=}1|C_X) \underset{P}{=} P(X{=}1)$, i. e., that the treatment probabilities do not depend on any covariate. If the global covariate C_X is constructed such that U is measurable with respect to C_X, then $P(X{=}1|C_X) \underset{P}{=} P(X{=}1)$ implies that each unit has the same probability $P(X{=}1|U{=}u) = P(X{=}1)$ to be in treatment 1, and this implies that each unit has the same probability $P(X{=}0)$ to be assigned to treatment 0 as well.

▷ **Solution 7-6** In a random experiment, in which a unit u is sampled and a value z of the covariate Z is assessed before the unit is assigned to one of the treatment conditions, *conditional random assignment* of a unit to one of the two treatment conditions refers to assigning the unit u to treatment condition 1 with probability $P^{Z=z}(X{=}1|C_X) \underset{P}{=} P^{Z=z}(X{=}1)$. Among other things, this makes sure that all units with the same value z of the covariate Z have the same probability to be assigned to treatment 1. This assignment procedure allows for different treatment probabilities for units for which we observe different values z of the covariate Z.

▷ **Solution 7-7** By definition of the expectation,

$$E[P(X{=}x|U)] = \sum_u P(X{=}x|U{=}u)\cdot P(U{=}u).$$

Because, according to the theorem of total probability

$$P(X{=}x) = \sum_u P(X{=}x|U{=}u)\cdot P(U{=}u)$$

(see Th. PR-4.25), this shows that $E[P(X{=}x|U)] = P(X{=}x)$. Note that $E[P(X{=}x|U)] = P(X{=}x)$ is also a special case of Rule (iv) in PR-Box 10.2, if we use $P(X{=}x|U) := E(\mathbb{1}_{X=x}|U)$, where $\mathbb{1}_{X=x}$ denotes the indicator variable indicating with values 1 and 0 whether or not X takes on the value x.

▷ **Solution 7-8** If $P(X{=}x) > 0$ for all values x of X, then the condition $U \perp\!\!\!\perp_P X$ is equivalent to: $P(X{=}x|U) \underset{P}{=} P(X{=}x)$, for each $x = 0,1,\ldots,J$. If Z is measurable with respect to U, then this equation implies

$$P(X{=}x|U,Z) \underset{P}{=} P(X{=}x|U) \underset{P}{=} P(X{=}x).$$

Using this equation as well as $P(X{=}x|Z) := E(\mathbb{1}_{X=x}|Z)$ and $E(\mathbb{1}_{X=x}) = P(X{=}x)$:

$$\begin{aligned}
E(1_{X=x}\,|\,Z) &\underset{P}{=} E\big[E(1_{X=x}\,|\,U,Z)\,\big|\,Z\big] && \text{[PR-Box 10.2, (v)]}\\
&\underset{P}{=} E\big[E(1_{X=x}\,|\,U)\,\big|\,Z\big] && \text{[Z is measurable w.r.t. U]}\\
&\underset{P}{=} E\big[E(1_{X=x})\,\big|\,Z\big] && \text{[$U \perp\!\!\!\perp_P X$, PR-Box 10.2, (vi)]}\\
&\underset{P}{=} E(1_{X=x}) && \text{[PR-Box 10.2, (i)]}\\
&= P(X{=}x),
\end{aligned}$$

for each $x = 0,1,\ldots,J$. Hence, $P(X{=}x|U,Z) \underset{P}{=} P(X{=}x|Z)$, for each $x = 0,1,\ldots,J$. This equation is equivalent to $U \perp\!\!\!\perp_P X\,|\,Z$ if X is discrete with a finite number of values $x = 0,1,\ldots,J$.

▷ **Solution 7-9** Note that $P(X{=}x) = E(1_{X=x})$ and $P(X{=}x\,|\,W) := E(1_{X=x}\,|\,W)$. Hence,

$$\begin{aligned}
E(1_{X=x}\,|\,W) &\underset{P}{=} E\big[E(1_{X=x}\,|\,C_X)\,\big|\,W\big] && \text{[PR-Box 10.2, (v)]}\\
&\underset{P}{=} E\big[P(X{=}x\,|\,C_X)\,\big|\,W\big] && \text{[(10.3)]}\\
&\underset{P}{=} E\,[P(X{=}x)\,|\,W\,] && [X \perp\!\!\!\perp_P C_X]\\
&\underset{P}{=} P(X{=}x) && \text{[PR-Box 10.2, (i)]}\\
&= E(1_{X=x}),
\end{aligned}$$

for each $x = 0,1,\ldots,J$. Because $P(X{=}x) = E(1_{X=x})$ and $P(X{=}x\,|\,W) := E(1_{X=x}\,|\,W)$, this proves the proposition.

▷ **Solution 7-10**

$$\begin{aligned}
E(Y\,|\,X,Z,W) &\underset{P}{=} E\big[E(Y\,|\,X,C_X)\,\big|\,X,Z,W\big] && \text{[PR-Box 10.2, (v)]}\\
&\underset{P}{=} E\big[\,E(Y\,|\,X,Z)\,\big|\,X,Z,W\big] && [Y \vdash C_X\,|\,X,Z]\\
&\underset{P}{=} E(Y\,|\,X,Z) && \text{[Box PR-10.2, (vii)]}.
\end{aligned}$$

Chapter 8
Unconfoundedness

In chapter 7 we treated the independent cause conditions and the regressively independent outcome conditions. Both kinds of conditions imply C_X-unbiasedness of the regressions $E(Y|X)$ and $E(Y|X,Z)$ if Z is a covariate, and (C_X, M)-unbiasedness of $E(Y|X,W)$, if W is measurable with respect to (C_X, M) and M is an intermediate variable of X and Y. In this chapter we study another kind of causality conditions implying unbiasedness. These new causality conditions will be called *unconfoundedness conditions*. These conditions deserve special attention, because they are — to our knowledge — the weakest conditions implying unbiasedness that are still empirically testable. 'Empirically testable' means that, in empirical applications, the hypothesis of unconfoundedness of the regressions $E(Y|X)$, $E(Y|X,Z)$, or $E(Y|X,W)$ can be tested by statistical procedures.

In this and some other respects, the unconfoundedness conditions are similar to the independent cause conditions and the regressively independent outcome conditions. If, e. g., unconfoundedness of the regression $E(Y|X)$ does not hold in an empirical application, then there is serious doubt that $E(Y|X)$ is unbiased. Just like the independent cause condition $X \underset{P}{\perp\!\!\!\perp} C_X$, unconfoundedness of $E(Y|X)$ with respect to C_X holds in a perfect randomized experiment in which X is the treatment variable and Y the outcome variable. Similarly, just like the independent cause condition $X \underset{P}{\perp\!\!\!\perp} C_X|Z$, unconfoundedness of $E(Y|X,Z)$ with respect to C_X holds in a perfect Z-conditionally randomized experiment. Furthermore, just as the independent cause conditions and the regressively independent outcome conditions, the unconfoundedness conditions can be used as criteria in selecting the relevant covariates, i. e., those variables that would induce bias if not included in the vector $Z := (Z_1, \ldots, Z_m)$ of covariates in the regressions $E(Y|X,Z)$ and $E(Y|X,W)$.

Overview

We start with a formal definition of the unconfoundedness conditions, show that they are implied by the independent cause conditions and the regressively independent outcome conditions. Then we illustrate them by an example. Next, we show that the unconfoundedness conditions imply the corresponding unbiasedness conditions. Finally, we summarize and discuss the implication structure between the different causality conditions treated so far.

8.1 Unconfoundedness of Regressions

In this section, we introduce *unconfoundedness of regressions*, a causality condition that has already been used by Steyer (1984, 1992) who called it "weak causality". To our knowledge, the idea has also been used by Cartwright (1979), who used it to define "effective strategies". Note that unconfoundedness has a number of different colloquial and technical meanings. The term as used here has been introduced by Steyer, Gabler, von Davier, and Nachtigall (2000).

8.1.1 Unconfoundedness With Respect to Total Effects

The basic idea of unconfoundedness of the simple regression $E(Y|X)$ is to postulate that, for each covariate W, each conditional expected value $E(Y|X{=}x)$ is the expectation of the conditional expected values $E(Y|X{=}x, W{=}w)$ over the distribution of W. Obviously, under this condition Simpson's paradox cannot occur (see ch. 1).

Remark 8.1 (Unconfoundedness of a Simple Regression) In more formal terms, let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a numerical causality space and suppose that X is discrete with values $x = 0, 1, \ldots, J$ and $P(X{=}x) > 0$, for all $x = 0, 1, \ldots, J$. Furthermore, let $\mathscr{W}_X$ denote the set of all random variables on the probability space $(\Omega, \mathscr{A}, P)$ that are measurable with respect to C_X, and let $W \in \mathscr{W}_X$. Now, we define *unconfoundedness of the regression* $E(Y|X)$ by postulating

$$\begin{gathered} E(Y|X{=}x) \;=\; E[E^{X=x}(Y|W)] \\ \text{for all } x = 0, 1, \ldots, J \text{ and all } W \in \mathscr{W}_X. \end{gathered} \tag{8.1}$$

As mentioned above, this equation presumes that X is discrete with a finite number of values $x = 0, 1, \ldots, J$ and $P(X{=}x) > 0$ for all $x = 0, 1, \ldots, J$. In contrast, W can be also be continuous. The general concept of unconfoundedness, which also applies if $X\colon (\Omega, \mathscr{A}, P) \to (\Omega'_X, \mathscr{A}'_X)$ is continuous, will be defined below by

$$\begin{gathered} E(Y|X{=}x) \;=\; \int E(Y|X{=}x, W{=}w)\; P_W\,(dw) \\ \text{for } P_X\text{-a.a. } x \in \Omega'_X \text{ and all } W \in \mathscr{W}_X. \end{gathered} \tag{8.2}$$

Note that Propositions (8.1) and (8.2) are equivalent if X is discrete with a finite number of values $x = 0, 1, \ldots, J$ and $P(X{=}x) > 0$ for all $x = 0, 1, \ldots, J$.

Proposition (8.2) explicitly shows that the expectation is taken *with respect to the marginal distribution P_W of W*. Hence, for a given value x of X with $P(X{=}x) > 0$ and a discrete covariate W, Proposition (8.2) implies

$$E(Y|X{=}x) \;=\; \sum_w E(Y|X{=}x, W{=}w) \cdot P(W{=}w). \tag{8.3}$$

$\lhd$

Definition 8.2 (Unconfoundedness of $E(Y|X)$)
Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a numerical causality space and let $\mathscr{W}_X$ denote the set of all random variables on $(\Omega, \mathscr{A}, P)$ that are measurable with respect to C_X. Then $E(Y|X)$ is called C_X-unconfounded [abbr.: $E(Y|X) \perp\!\!\!\perp C_X$] if Proposition (8.2) *holds.*

Remark 8.3 (Falsifiability) The definition of unconfoundedness already shows that it can easily be tested empirically, at least in the sense of falsifiability. In order to test $E(Y|X) \perp\!\!\!\perp C_X$, we simply have to select a covariate $W \in \mathscr{W}_X$ and see if the equation in Proposition (8.2) — or, if X and W is discrete, Equation (8.3) — holds. Obviously, all terms in these equations are empirically estimable. If there is a $W \in \mathscr{W}_X$ for which the equation in (8.2) does not hold, then $E(Y|X)$ is not unconfounded with respect to C_X. ◁

Remark 8.4 (Unconfoundedness, Conditioning on a Covariate) The basic idea outlined for the simple regression $E(Y|X)$ can be extended to the case in which we also condition on a covariate $Z\colon (\Omega, \mathscr{A}, P) \to (\Omega'_Z, \mathscr{A}'_Z)$, $Z \in \mathscr{W}_X$. The analog to Proposition (8.2) is

$$\begin{gathered} E(Y|X{=}x, Z{=}z) = \int E(Y|X{=}x, Z{=}z, W{=}w)\, P_{W|Z=z}(dw) \\ \text{for } P_{X,Z}\text{-a.a. } (x,z) \in \Omega'_X \times \Omega'_Z \text{ and all } W \in \mathscr{W}_X. \end{gathered} \tag{8.4}$$

Obviously, now we take the conditional expected value of the conditional expected values $E(Y|X{=}x, Z{=}z, W{=}w)$ with respect to the $(Z{=}z)$-conditional distribution $P_{W|Z=z}$ of W. Note that Proposition (8.4) is *not* always true. Instead it defines C_X-*unconfoundedness* of the regression $E(Y|X,Z)$. If we consider $E(Y|X,Z)$ we condition on a given (possibly multi-dimensional) covariate Z and 'ignore' all other covariates $W \in \mathscr{W}_X$. ◁

Definition 8.5 (Unconfoundedness of $E(Y|X,Z)$)
Let the assumption of Definition 8.2 be satisfied and let $Z \in \mathscr{W}_X$. Then $E(Y|X,Z)$ is called C_X-unconfounded (abbr.: $E(Y|X,Z) \perp\!\!\!\perp C_X$) if Proposition (8.4) *holds.*

Remark 8.6 (Causal Dependencies Beyond Causal Effects) Let us emphasize that the unconfoundedness conditions are well-defined even if X and Z are continuous and, therefore, the true outcome variables τ_x do not exist. As we will see, unconfoundedness implies unbiasedness (see Th. 8.15). ◁

Definition 8.7 (Unconfoundedness of $E^{X=x}(Y|Z)$)
Let the assumption of Definition 8.2 be satisfied and let $Z \in \mathscr{W}_X$. Furthermore,

let $x \in \Omega'_X$ with $\{x\} \in \mathscr{A}'_X$ and $P(X{=}x|Z) \gneq_P 0$. Then $E^{X=x}(Y|Z)$ is called C_X-unconfounded if

$$E^{X=x}(Y|Z) \underset{P}{=} E[E^{X=x}(Y|Z,W)\,|\,Z] \qquad \text{for all } W \in \mathscr{W}_X. \tag{8.5}$$

Remark 8.8 (An Equivalent Condition) If $x \in \Omega'_X$ with $\{x\} \in \mathscr{A}'_X$ and $P(X{=}x|Z) \gneq_P 0$, then Equation (8.5) is equivalent to

$$\begin{gathered} E(Y|X{=}x, Z{=}z) = \int E(Y|X{=}x, Z{=}z, W{=}w)\; P_{W|Z=z}\,(dw) \\ \text{for } P_Z\text{-a.a. } z \in \Omega'_Z \text{ and all } W \in \mathscr{W}_X \end{gathered} \tag{8.6}$$

(see Exercise ???). ◁

Remark 8.9 (Später?) If $x \in \Omega'_X$ with $\{x\} \in \mathscr{A}'_X$ and $P(X{=}x|Z) \gneq_P 0$, then C_X-unconfoundedness of $E(Y|X,Z)$ implies that $E^{X=x}(Y|Z)$ is C_X-unbiased, i. e.,

$$E(Y|X,Z) \perp\!\!\!\perp C_X \;\Rightarrow\; E^{X=x}(Y|Z) \perp C_X. \tag{8.7}$$

◁

8.1.2 Unconfoundedness With Respect to Direct Effects

Now we also include an intermediate variable $M\colon (\Omega, \mathscr{A}, P) \to (\Omega'_M, \mathscr{A}'_M)$ (see Def. 4.10) and consider unconfoundedness *with respect to direct effects*. Reading the following definition, remember that M is called an *intermediate variable of X and Y*, if X is pre-ordered to M and M is pre-ordered to Y with respect to the filtration $(\mathscr{F}_t, t \in T)$ (see Def. 4.10).

Definition 8.10 (Unconfoundedness of Regressions)
Let the assumption of Def. 8.5 be satisfied, let $\mathscr{M}_{X,Y}$ denote the set of all intermediate variables of X and Y, and let $M \in \mathscr{M}_{X,Y}$. Then $E(Y|X,M)$ is called $(C_X; M)$-unconfounded [abbr.: $E(Y|X,M) \perp\!\!\!\perp (C_X; M)$] if

$$\begin{gathered} E(Y|X{=}x, M{=}m) = \int E(Y|X{=}x, M{=}m, W{=}w)\; P_{W|M=m}\,(dw) \\ \text{for } P_{X,M}\text{-a.a. } (x,m) \in \Omega'_X \times \Omega'_M \text{ and all } W \in \mathscr{W}_X. \end{gathered} \tag{8.8}$$

Using the semicolon in $E(Y|X,M) \perp\!\!\!\perp (C_X; M)$ indicates that M is the intermediate variable considered. If we condition not only an intermediate variable M but also on a covariate Z, then unconfoundedness of the regression $E(Y|X,M,Z)$ with respect to direct effects is defined as follows:

Definition 8.11 (Unconfoundedness of Regressions)
Let the assumption of Def. 8.10 be satisfied, let $Z \in \mathcal{W}_X$, and $M \in \mathcal{M}_{X,Y}$. Then $E(Y|X,M,Z)$ is called $(C_X;M)$-unconfounded $[E(Y|X,M,Z) \Uparrow\!\!\!\Uparrow (C_X;M)]$ if

$$E(Y|X{=}x,Z{=}z,M{=}m) = \int E(Y|X{=}x,Z{=}z,M{=}m,W{=}w)\, P_{W|Z=z,M=m}(dw)$$
$$\text{for } P_{X,Z,M}\text{-a.a. } (x,z,m) \in \Omega'_X \times \Omega'_Z \times \Omega'_M \text{ and all } W \in \mathcal{W}_X. \tag{8.9}$$

Again, the semicolon in $E(Y|X,M,Z) \Uparrow\!\!\!\Uparrow (C_X;M)$ indicates that M is the (possibly multivariate) intermediate variable. Using X as the subscript of C_X indicates that X is the cause. Hence, Z is the (possibly multivariate) covariate that is controlled.

8.2 Conditions Implying Unconfoundedness

The independent cause conditions and the regressively independent outcome conditions both imply the corresponding unconfoundedness conditions. This is most easily seen for the regression $E(Y|X)$ if, for $W \in \mathcal{W}_X$, we consider the equation

$$E(Y|X{=}x) = \int E(Y|X{=}x,W{=}w)\, P_{W|X=x}(dw) \quad \text{for } P_X\text{-a.a. } x \in \Omega'_X \tag{8.10}$$

which is always true, provided that $\{x\} \in \mathcal{A}'_X$ for all $x \in \Omega'_X$ (see Cor. PR-17.60). Because W is measurable with respect to C_X, the independent cause condition $X \underset{P}{\perp\!\!\!\perp} C_X$ implies

$$P_{W|X=x} = P_W \qquad \text{for } P_X\text{-a.a. } x \in \Omega'_X,$$

(see Cor. PR-**??**). Hence, Equation (8.2) immediately follows from $X \underset{P}{\perp\!\!\!\perp} C_X$.

Similarly, if $W \in \mathcal{W}_X$, $\{x\} \in \mathcal{A}'_X$ for all $x \in \Omega'_X$, and we assume that the regressively independent outcome condition $Y \vdash C_X|X$ holds, then $E(Y|X,W) \underset{P}{=} E(Y|X)$. However, according to Definition PR-10.33 (i) and Equation PR-(10.26), this is equivalent to

$$E(Y|X{=}x,W{=}w) = E(Y|X{=}x) \qquad \text{for } P_{X,W}\text{-a.a. } (x,w) \in \Omega'_X \times \Omega'_W,$$

which in turn implies

$$\begin{aligned} E(Y|X{=}x) &= E(Y|X{=}x) \cdot 1 \\ &= E(Y|X{=}x) \cdot \int P_W(dw) \\ &= \int E(Y|X{=}x) \cdot P_W(dw) \\ &= \int E(Y|X{=}x,W{=}w) \cdot P_W(dw) \qquad \text{for } P_X\text{-a.a. } x \in \Omega'_X. \end{aligned}$$

Of course, the corresponding implications also hold if we include a covariate $Z \in \mathscr{W}_X$ and/or an intermediate variable $M \in \mathscr{M}_{X,Y}$.

Theorem 8.12 (Sufficient Conditions of Unconfoundedness)
Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a numerical causality space.

(i) Then $E(Y|X) \perp\!\!\!\perp C_X$ if

$$X \underset{P}{\perp\!\!\!\perp} C_X \quad \text{or} \quad Y \vdash C_X | X .$$

(ii) Let $Z \in \mathscr{W}_X$. Then $E(Y|X,Z) \perp\!\!\!\perp C_X$ if

$$X \underset{P}{\perp\!\!\!\perp} C_X | Z \quad \text{or} \quad Y \vdash C_X | X, Z .$$

(iii) Let $M \in \mathscr{M}_{X,Y}$. Then $E(Y|X,M) \perp\!\!\!\perp (C_X; M)$ if

$$X \underset{P}{\perp\!\!\!\perp} C_X | M \quad \text{or} \quad Y \vdash C_X | X, M .$$

(iv) Finally, let $Z \in \mathscr{W}_X$ and $M \in \mathscr{M}_{X,Y}$. Then $E(Y|X,M,Z) \perp\!\!\!\perp (C_X; M)$ if

$$X \underset{P}{\perp\!\!\!\perp} C_X | M, Z \quad \text{or} \quad Y \vdash C_X | X, M, Z .$$

(Proof p. 207)

Note that the unconfoundedness conditions do not imply the disjunction of the corresponding independent cause condition and the regressively independent outcome condition. Hence, unconfoundedness is less restrictive than this disjunction and less restrictive than each of these conditions alone. However, according to Theorem 8.12, we can conclude, e. g., that unconfoundedness holds in a perfect randomized experiment, because, in this case, $X \underset{P}{\perp\!\!\!\perp} C_X$ is satisfied (see section 7.5). Similarly, we can conclude, that $E(Y|X,Z)$ is Z-conditionally unconfounded in a perfect Z-conditionally randomized experiment, because, in this case, $X \underset{P}{\perp\!\!\!\perp} C_X | Z$ is satisfied.

8.3 Numerical Example

Before we turn to the implications of unconfoundedness on unbiasedness, let us illustrate unconfoundedness by a numerical example. In this example, we again assume $C_X = U$, which implies

$$E(Y|X,C_X) \underset{P}{=} E(Y|X,U) \tag{8.11}$$

for the regression $E(Y|X,C_X)$, and

$$P(X{=}1|C_X) \underset{P}{=} P(X{=}1|U) \tag{8.12}$$

Table 8.1. Covariate-Treatment Regression That is C_X-Unconfounded

				Fundamental parameters							Derived parameters				
Unit u	Sampling probability $P(U=u)$	Covariate *sex* Z	Potential confounder *marital status* W	True outcome under control $\tau_0 = E^{X=0}(Y\|U)$	True outcome under treatment 1 $\tau_1 = E^{X=1}(Y\|U)$	True outcome under treatment 2 $\tau_2 = E_{X=2}(Y\|U)$	True treatment probability $P(X=0\|U)$	True treatment probability $P(X=1\|U)$	True treatment probability $P(X=2\|U)$	True total effect $\delta_{12} = \tau_1 - \tau_0$	True total effect $\delta_{20} = \tau_2 - \tau_0$	$P(U=u\|X=0)$	$P(U=u\|X=1)$	$P(U=u\|X=2)$
u_1	1/6	male (m)	married (m)	82	100	110	2/6	3/6	1/6	18	28	4/18	3/19	2/16
u_2	1/6	male (m)	not married (n)	89	100	110	2/6	1/6	3/6	11	21	4/18	1/19	6/16
u_3	1/6	male (m)	not married (n)	95	100	110	2/6	3/6	1/6	5	15	4/18	3/19	2/16
u_4	1/6	female (f)	married (m)	102	113	120	2/12	8/12	2/12	11	18	2/18	4/19	2/16
u_5	1/6	female (f)	not married (n)	102	117	120	1/12	8/12	3/12	15	18	1/18	4/19	3/16
u_6	1/6	female (f)	divorced (d)	102	125	120	3/12	8/12	1/12	23	18	3/18	4/19	1/16
	Expectations of true outcomes etc.			95.333	109.167	115.00	.250	.528	.222	13.833	19.667			
	Conditional expectations $E(Y\|X=x)$			93.111	111.579	113.750								

	1 vs. 0	2 vs. 0
Prima facie effect (*PFE*)	18.468	20.639
Conditional *PFEs* for males ($Z=m$)	11.333	21.333
Conditional *PFEs* for females ($Z=f$)	16.333	18.000

	1 vs. 0	2 vs. 0
Average total effect	13.833	19.667
Conditional total effects for males ($Z=m$)	11.333	21.333
Conditional total effects for females ($Z=f$)	16.333	18.000

for the true treatment probabilities. If the values u of U represent the observational units *at the onset of treatment*, these assumptions are realistic also in empirical applications if a simple experiment described in section 2.1 is considered.

Example 8.13 (Unconfoundedness of the Treatment Regression) In Table 8.1, *unconfoundedness of the treatment regression* $E(Y|X)$ *does not hold*, because, in this example, it is *not true* that Equation (8.3) is satisfied for each of the three values x of X. Let us consider the covariate := *sex*. In this case, Z takes the role of W in the equation occurring in (8.3). In order to see that Equation (8.3) does not hold we compute both sides of this equation and see if they are identical for all three values $x = 0, 1, 2$ of the treatment variable X.

The conditional expected value $E(Y|X{=}x)$ of the outcome variable Y in treatment x can be computed applying

$$E(Y|X{=}x) \;=\; \sum_z E(Y|X{=}x, Z{=}z) \cdot P(Z{=}z|X{=}x), \tag{8.13}$$

which is always true if Z is discrete [see Eq. (ii) in Box 9.2]. For $X{=}0$, this yields

$$E(Y|X{=}0) \;=\; \sum_z E(Y|X{=}0, Z{=}z) \cdot P(Z{=}z|X{=}0) \;=\; 88.\overline{6} \cdot \frac{2}{3} + 102 \cdot \frac{1}{3} \;\approx\; 93.111$$

(see Exercises 8-3, 8-4, and 8-5). In contrast, computing the right-hand side of Equation (8.3) for $X{=}0$ yields

$$\sum_z E(Y|X{=}0, Z{=}z) \cdot P(Z{=}z) \;=\; 88.\overline{6} \cdot \frac{1}{2} + 102 \cdot \frac{1}{2} \;\approx\; 95.333.$$

Because $93.111 \neq 95.333$, Equation (8.3) does not hold in this example. Therefore, we can stop checking unconfoundedness of $E(Y|X)$ and conclude that, in this example, the treatment regression $E(Y|X)$ *is confounded* with respect to C_X. ◁

Example 8.14 (Unconfoundedness of the Covariate - Treatment Regression) Now we illustrate C_X-unconfoundedness of the regression $E(Y|X,Z)$ for the covariate Z := *sex* and the covariate W := *marital status*. Because W is discrete, the equation occurring in (8.4) can also be written

$$E(Y|X{=}x, Z{=}z) \;=\; \sum_w E(Y|X{=}x, Z{=}z, W{=}w) \cdot P(W{=}w|Z{=}z). \tag{8.14}$$

Again, note that this equation does not necessarily hold. It does hold, however, if the regression $E(Y|X,Z)$ is C_X-unconfounded and W is a covariate. Hence, again we just have to compute both sides of this equation, this time for all combinations of values x and z and check if both sides are identical.

The conditional expected value $E(Y|X{=}x, Z{=}z)$ can be computed applying

$$E(Y|X{=}x, Z{=}z) \;=\; \sum_w E(Y|X{=}x, Z{=}z, W{=}w) \cdot P(W{=}w|X{=}x, Z{=}z), \tag{8.15}$$

which is always true if $W{=}w$ is discrete [see Eq. (ii) in Box 9.2]. For $X{=}0$ and $Z{=}m$ (males), this equation yields

$$E(Y|X=0,Z=m) = \sum_w E(Y|X=0,Z=m,W=w)\cdot P(W=w|X=0,Z=m)$$
$$= 82\cdot\frac{1}{3} + \frac{89+95}{2}\cdot\frac{2}{3} \approx 88.667.$$

(see Exercises 8-6, 8-7, and 8-8), and computing the right-hand side of Equation (8.14) for $X=0$ and $Z=m$ yields

$$\sum_w E(Y|X=0,Z=m,W=w)\cdot P(W=w|Z=m) = 82\cdot\frac{1}{3} + \frac{89+95}{2}\cdot\frac{2}{3} \approx 88.667$$

(see Exercise 8-9). Obviously, the two numbers are identical and Equation (8.14) holds in this example for $X=0$ and $Z=m$. Checking this equation for the other five combinations of values of X and Z yields the same result, namely that Equation (8.14) holds for the other five pairs (x,z) of values of X and Z as well.

Hence, checking unconfoundedness of the regression $E(Y|X,Z)$ with respect to Z-conditional total effects for covariate $Z :=$ *sex* and covariate $W :=$ *marital status* does not give any reason for rejecting this hypothesis. In other words, we failed to falsify the hypothesis that the covariate-treatment regression $E(Y|X,Z)$ is unconfounded with respect to Z-conditional total effects. Note that this does not prove that $E(Y|X,Z)$ is unconfounded. It is still possible that there is another covariate such that the equation occurring in (8.4) does not hold. For instance, in our example, $W = U$ or the indicator $I_{W=n}$ would be other candidates of covariates. ◁

8.4 Implications of Unconfoundedness on Unbiasedness

Now let us study the implications of the unconfoundedness conditions on unbiasedness of various conditional expectations, regressions and prima facie effects, and prima facie effect functions.

8.4.1 Unbiasedness With Respect to Total Effects

We start considering the implications of unconfoundedness of the simple regression $E(Y|X)$ on unbiasedness of conditional expectations, regressions and prima facie effects *with respect to total effects*. Reading the following definition, remember that $\tau_x := E^{X=x}(Y|C_X)$ denotes the true-outcome variable with respect to total effects (see Def. 4.12), provided that the regression $E^{X=x}(Y|C_X)$ (of Y given the global covariate C_X) with respect to the $(X=x)$-conditional-probability measure $P^{X=x}$ is P-unique.

Theorem 8.15 (Implications on Unbiasedness of Regressions)
Let $\langle(\Omega,\mathscr{A},P),(\mathscr{F}_t,t\in T),X,Y\rangle$ be a numerical causality space.

(i) If $E(Y|X) \Perp C_X$, then $E(Y|X) \perp C_X$.

(ii) If $E(Y|X) \Uparrow C_X$ *and* $P(X{=}x) > 0$*, then* $E^{X=x}(Y|C_X)$ *is* P*-unique and*

$$E(Y|X{=}x) \;=\; E(\tau_x). \tag{8.16}$$

(Proof p. 208)

Hence, the simple regression $E(Y|X)$ is τ_x-unbiased if it is τ_x-unconfounded. Of course, both, unbiasedness and unconfoundedness refer to average total effects. Note that no additional assumption is necessary for this proposition to hold. In contrast, we have to assume $P(X{=}x) > 0$ in order to conclude from unconfoundedness of $E(Y|X)$ that the conditional expectation $E(Y|X{=}x)$ is unbiased with respect to average total effects.

Of course, this theorem has immediate implications on unbiasedness of the prima facie effect $E(Y|X{=}x) - E(Y|X{=}x')$. This is obvious if we consider Equation (8.16) and remember that $E(\tau_x) - E(\tau_{x'}) \;=\; E(\delta_{xx'})$ is the average total effect of x vs. x', where $\delta_{xx'} := \tau_x - \tau_{x'}$ denotes the true total effect variable (see Def. 4.12).

Corollary 8.16 (Implications on Unbiasedness of Prima Facie Effects)
Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ *be a numerical causality space and let* $\delta_{xx'} := \tau_x - \tau_{x'}$*. If* $E(Y|X) \Uparrow C_X$ *and* $P(X{=}x)$*,* $P(X{=}x') > 0$*, then* $E^{X=x}(Y|C_X)$ *and* $E^{X=x'}(Y|C_X)$ *are* P*-unique, and*

$$PFE_{xx'} \;=\; E(\tau_x) - E(\tau_{x'}) \;=\; E(\delta_{xx'}). \tag{8.17}$$

(Proof p. 208)

Hence, if the simple regression $E(Y|X)$ is unconfounded with respect to average total effects and we can assume $P(X{=}x) > 0$ and $P(X{=}x') > 0$, then we can conclude that the prima facie effect $E(Y|X{=}x) - E(Y|X{=}x')$ is unbiased with respect to the average total effect $E(\delta_{xx'})$ of x vs. x'. Of course, the additional requirement $P(X{=}x) > 0$ and $P(X{=}x') > 0$ is trivial if X represents a discrete (treatment) variable. However, it does not hold if X is a continuous random variable. In this case, the regression $E(Y|X)$ can be unbiased, even though no average total effect $E(\delta_{xx'})$ of x vs. x' can be defined.

Now we will study the implications of unconfoundedness on unbiasedness, additionally considering a covariate $Z \in \mathscr{W}_X$.

Theorem 8.17 (Implications on Unbiasedness of Regressions)
Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ *be a numerical causality space and* $Z \in \mathscr{W}_X$*.*

(i) If $E(Y|X,Z) \Uparrow C_X$*, then* $E(Y|X,Z) \perp\!\!\!\perp C_X$*.*
(ii) Let $x \in \Omega'_X$ *with* $\{x\} \in \mathscr{A}'_X$*. If* $E(Y|X,Z) \Uparrow C_X$ *and* $P(X{=}x\,|\,Z) \gneq_P 0$*, then* $E^{X=x}(Y|Z) \perp\!\!\!\perp C_X$*.*

(Proof p. 208)

Because $C_X \in \mathscr{W}_X$, Proposition (8.4) implies

$$E(Y|X{=}x, Z{=}z) \;= \int E(Y|X{=}x, Z{=}z, C_X{=}c_X)\; P_{C_X|Z=z}\,(dc_X)$$
$$\text{for } P_{X,Z}\text{-a.a. } (x,z) \in \Omega'_X \times \Omega'_Z,$$

Measurability of Z with respect to C_X implies $\sigma(X,Z,C_X) = \sigma(X,C_X)$. Therefore, a factorization of $E(Y|X,Z,C_X)$ is also a factorization of $E(Y|X,C_X)$. Because $E(Y|X{=}x, Z{=}z, C_X{=}c_X)$ and $E(Y|X{=}x, C_X{=}c_X)$ denote values of a factorization of a version V of $E(Y|X,Z,C_X)$ and $E(Y|X,C_X)$, respectively, we can conclude: $E(Y|X{=}x, Z{=}z, C_X{=}c_X) = E(Y|X{=}x, C_X{=}c_X)$. Therefore, Proposition (8.4) also implies

$$\begin{aligned} E(Y|X{=}x, Z{=}z) \;&= \int E(Y|X{=}x, C_X{=}c_X)\; P_{C_X|Z=z}\,(dc_X) \\ &\text{for } P_{X,Z}\text{-a.a. } (x,z) \in \Omega'_X \times \Omega'_Z. \end{aligned} \tag{8.18}$$

However, this equation defines $E(Y|X,Z) \perp\!\!\!\perp C_X$, which proves (i).

(ii) Now we additionally assume $x \in \Omega'_X$ with $\{x\} \in \mathscr{A}'_X$ and $P(X{=}x\,|\,Z) \gneq_P 0$, which implies $P(X{=}x) > 0$. According to Theorem PR-**??** and Remark PR-14.36,

$$E(Y|X{=}x, Z{=}z) \;=\; E^{X=x}(Y|Z{=}z) \qquad P_Z^{X=x}\text{-a.a. } z \in \Omega'_Z. \tag{8.19}$$

If $g(x,z)$ and $f(x,z)$ denote the left- and right-hand sides of Equation (8.18),

$$A'_0 := \{(x,z) \in \Omega'_X \times \Omega'_Z \colon g(x,z) \neq f(x,z)\},$$

and, for a fixed $x \in \Omega'_X$,

$$C'_{x0} := \{z \in \Omega'_Z \colon g(x,z) \neq f(x,z)\},$$

then $Z^{-1}(C'_{x0}) = (X,Z)^{-1}(A'_0) \cap X^{-1}(\{x\})$. Now

$$\begin{aligned} P_Z^{X=x}(C'_{x0}) \;&=\; P^{X=x}[Z^{-1}(C'_{x0})] \\ &=\; P^{X=x}[(X,Z)^{-1}(A'_0) \cap X^{-1}(\{x\})] \\ &=\; \frac{P[(X,Z)^{-1}(A'_0) \cap \{X{=}x\}]}{P(X{=}x)}. \end{aligned}$$

The last equation shows that $P_{X,Z}(A'_0) = P[(X,Z)^{-1}(A'_0)] = 0$ implies $P_Z^{X=x}(C'_{x0}) = 0$. Therefore, Equations (8.19) and (8.18) imply

$$\begin{aligned} E^{X=x}(Y|Z{=}z) \;&= \int E(Y|X{=}x, C_X{=}c_X)\; P_{C_X|Z=z}\,(dc_X) \\ &\text{for } P_Z^{X=x}\text{-a.a. } z \in \Omega'_Z. \end{aligned} \tag{8.20}$$

Jetzt der nächste Schritt, der erlaubt, $E(Y|X{=}x, C_X{=}c_X)$ durch $E^{X=x}(Y|C_X{=}c_X)$ zu ersetzen. Das Folgende daz stimmt noch nicht.

Now we have to show that $x \in \Omega'_X$ with $\{x\} \in \mathscr{A}'_X$ and $P(X{=}x \,|\, Z) \gneq_P 0$ implies

$$E(Y \,|\, X{=}x, C_X{=}c_X) \;=\; E^{X=x}(Y \,|\, C_X{=}c_X) \qquad \text{for } P_{C_X|Z=z}\text{-a.a. } c_x \in \Omega'_{C_X}. \tag{8.21}$$

Again, according to Theorem PR-**??** and Remark PR-14.36,and

$$E(Y \,|\, X{=}x, C_X{=}c_X) \;=\; E^{X=x}(Y \,|\, C_X{=}c_X) \qquad \text{for } P^{X=x}_{C_X}\text{-a.a. } c_x \in \Omega'_{C_X}. \tag{8.22}$$

If $g_x(c_x)$ and $f_x(c_x)$ denote the left- and right-hand sides of Equation (8.22), respectively, and

$$B'_0 \;=\; \{c_x \in \Omega'_{C_X} \colon g_x(c_x) \neq f_x(c_x)\}$$

then $P^{X=x}_{C_X}(B'_0) = 0$, implies $P^{X=x}_Z(B'_0) = 0$. This can be seen as follows: Consider

$$P^{X=x}_{C_X}(B'_0) \;=\; P^{X=x}[C_X^{-1}(B'_0)] \;=\; \frac{P[C_X^{-1}(B'_0) \cap \{X{=}x\}]}{P(X{=}x)} \;=\; 0$$

Because $C_X^{-1}(B'_0) \cap \{X{=}x\}$ is not empty and $P(X{=}x) > 0$, we can conclude $P[C_X^{-1}(B'_0)] = 0$. Because

$$P_{C_X|Z=z}(B'_0) \;=\; P[C_X^{-1}(B'_0) \,|\, Z{=}z],$$

$P[C_X^{-1}(B'_0)] = 0$ implies $P_{C_X|Z=z}(B'_0) = 0$. Hence,

$$E(Y \,|\, X{=}x, C_X{=}c_X) \;=\; E^{X=x}(Y \,|\, C_X{=}c_X) \qquad \text{for } P_{C_X|Z=z}\text{-a.a. } c_x \in \Omega'_{C_X}.$$

Therefore, Equation (8.20) yields

$$E^{X=x}(Y \,|\, Z{=}z) \;= \int E^{X=x}(Y \,|\, C_X{=}c_X)\; P_{C_X|Z=z}\,(dc_X)$$

$$\text{for } P^{X=x}_Z\text{-a.a. } z \in \Omega'_Z.$$

???

??? Wo verwende ich $P(X{=}x \,|\, Z) > 0$?

???

Because $\tau_x := E^{X=x}(Y \,|\, C_X)$ is P-unique, this implies

$$E^{X=x}(Y \,|\, Z) \;\underset{P}{=}\; E(\tau_x \,|\, Z).$$

END OF PROOF

Hence, the regression $E(Y \,|\, X, Z)$ is C_X-unbiased if it is C_X-unconfounded. Again, note that this proposition holds without any additional assumption. In contrast, we have to assume $x \in \Omega'_X$ with $\{x\} \in \mathscr{A}'_X$ and $P(X{=}x \,|\, Z) \gneq_P 0$ in order to conclude from C_X-unconfoundedness of $E(Y \,|\, X, Z)$ that the $(X{=}x)$-conditional regression $E^{X=x}(Y \,|\, Z)$ of Y on Z is C_X-unbiased. In Remark 7.12 we already commented on the 'common support' requirement $P(X{=}x|Z) \gneq_P 0$ for the covariate Z. This condition requires that $P(X{=}x \,|\, Z{=}z) = 0$ only for a subset A'_Z of values z of Z with $P_Z(A'_Z) = 0$, which implies $P(X{=}x \,|\, Z{=}z) > 0$ for all values z of Z that have a positive probability $P(Z{=}z) > 0$.

Theorem 8.17 has immediate implications on unbiasedness of the prima-facie-effect function $PFE_{xx';Z}$. This is obvious if we consider Equation

$$E^{X=x}(Y|Z) \underset{P}{=} E(\tau_x|Z), \tag{8.23}$$

which defines C_X-unbiasedness of $E^{X=x}(Y|Z)$, and remember that $E(\tau_x|Z) - E(\tau_{x'}|Z) = E(\delta_{xx'}|Z)$ is the Z-conditional total effect function of x vs. x' (see Def. 5.6).

Corollary 8.18 (Implications on Unbiasedness of Prima Facie Effects)
Let $\langle(\Omega,\mathscr{A},P),(\mathscr{F}_t,t\in T),X,Y\rangle$ be a numerical causality space, let $Z\in\mathscr{W}_X$, and let $\delta_{xx'}:=\tau_x-\tau_{x'}$. If $x,x'\in\Omega'_X$ with $\{x\},\{x'\}\in\mathscr{A}'_X$, and if $P(X{=}x|Z)$, $P(X{=}x'|Z)\underset{P}{>}0$, then $E(Y|X,Z)\perp\!\!\!\perp C_X$ implies that both $E^{X=x}(Y|C_X)$ and $E^{X=x'}(Y|C_X)$ are P-unique, and

$$PFE_{xx';Z} \underset{P}{=} E(\tau_x|Z) - E(\tau_{x'}|Z) \underset{P}{=} E(\delta_{xx'}|Z). \tag{8.24}$$

(Proof p. 208)

Hence, if the regression $E(Y|X,Z)$ is C_X-unconfounded and we can assume $x,x'\in\Omega'_X$ with $\{x\},\{x'\}\in\mathscr{A}'_X$ and $P(X{=}x|Z)$, $P(X{=}x'|Z)\underset{P}{>}0$, then we can conclude that the prima-facie-effect function $PFE_{xx';Z}$ is C_X-unbiased. Note that the additional common support requirements $P(X{=}x|Z)\underset{P}{>}0$ and $P(X{=}x'|Z)\underset{P}{>}0$ are not trivial even if X represents a discrete (treatment) variable. If, e. g., X is a treatment variable, then $P(X{=}x|Z{=}z)>0$ means that units obtaining a covariate value z must have a positive probability to be in treatment x, provided that z has a nonzero probability $P(Z{=}z)$.

8.4.2 Unbiasedness With Respect to Direct Effects

Now we will study the implications of unconfoundedness on unbiasedness *with respect to direct effects*, where 'direct' is relative to a (possibly multivariate) intermediate variable M. Again we will start with the simplest case in which no covariate is considered and then turn to the most general case in which we also condition of a (possibly multivariate) covariate Z. Reading the following definition, remember that $\tau_{x,t_M}:=E^{X=x}(Y|M,C_X)$ denotes the true outcome variable with respect to direct effects, provided that $E^{X=x}(Y|M,C_X)$ is P-unique.

Theorem 8.19 (Implications on Unbiasedness of Regressions)
Let $\langle(\Omega,\mathscr{A},P),(\mathscr{F}_t,t\in T),X,Y\rangle$ be a numerical causality space and let $M\in\mathscr{M}_{X,Y}$ be an intermediate variable.

(i) If $E(Y|X,M)\perp\!\!\!\perp(C_X;M)$, then $E(Y|X,M)\perp(C_X;M)$.
(ii) If $x\in\Omega'_X$ with $\{x\}\in\mathscr{A}'_X$ and $P(X{=}x|M)\underset{P}{>}0$, then $E(Y|X,M)\perp\!\!\!\perp(C_X;M)$ implies $E^{X=x}(Y|M)\perp C_X$.

(Proof p. 208)

In other words, if $E(Y|X,M)$ is $(C_X;M)$-unconfounded, then it is $(C_X;M)$-unbiased. Furthermore, if $x \in \Omega'_X$ with $\{x\} \in \mathscr{A}'_X$ and $P(X{=}x|M) \gneq_P 0$, then (C_X,M)-unbiasedness of the regression $E(Y|X,M)$ implies that $E^{X=x}(Y|M,C_X)$ is P-unique and

$$E^{X=x}(Y|M) \eqP E(\tau_{x,t_M}|M). \tag{8.25}$$

Theorem 8.19 has immediate implications on unbiasedness of the prima-facie-effect function $PFE_{xx';M} = E^{X=x}(Y|M) - E^{X=x'}(Y|M)$. This is obvious if we consider Equation (8.25) and remember that $E(\tau_{x,t_M}|M) - E(\tau_{x',t_M}|M) = E(\delta_{xx',t_M}|M)$ is the M-conditional direct-effect function of x vs. x' with respect to the intermediate variable M (see section 5.3).

Corollary 8.20 (Implications on Unbiasedness of Prima Facie Effects)
Let $\langle (\Omega,\mathscr{A},P),(\mathscr{F}_t,t\in T),X,Y\rangle$ be a numerical causality space, let M be an intermediate variable, and let $\delta_{xx',t_M} := \tau_{x,t_M} - \tau_{x',t_M}$. If $x,x' \in \Omega'_X$ with $\{x\},\{x'\} \in \mathscr{A}'_X$, and $P(X{=}x|M)$, $P(X{=}x'|M) \gneq_P 0$, then $E(Y|X,M) \perp\!\!\!\perp (C_X;M)$ implies that $E^{X=x}(Y|M,C_X)$ and $E^{X=x'}(Y|M,C_X)$ are P-unique and

$$PFE_{xx';M} \eqP E(\tau_{x,t_M}|M) - E(\tau_{x',t_M}|M) \eqP E(\delta_{xx',t_M}|M).$$

(Proof p. 208)

Hence, if the regression $E(Y|X,M)$ is unconfounded with respect to M-conditional direct effects and we can assume $x,x' \in \Omega'_X$ with $\{x\},\{x'\} \in \mathscr{A}'_X$ and $P(X{=}x|M)$, $P(X{=}x'|M) \gneq_P 0$, then the prima-facie-effect function $PFE_{xx';M}$ is $C_X;M)$-unbiased. Again, note that the additional common support requirements $P(X{=}x|M) \gneq_P 0$ and $P(X{=}x'|M) \gneq_P 0$ are *not trivial* even if X represents a discrete (treatment) variable. If, e. g., X is a treatment variable, then $P(X{=}x|M{=}m) > 0$ means that units obtaining a value m of the intermediate must have a positive probability to be in treatment x, provided that m has a nonzero probability $P(M{=}m)$.

Let us now also consider a covariate Z. Because, both Z and M can be multivariate, this is the most general case.

Theorem 8.21 (Implications on Unbiasedness of Regressions)
Let $\langle (\Omega,\mathscr{A},P),(\mathscr{F}_t,t\in T),X,Y\rangle$ be a numerical causality space, let $Z \in \mathscr{W}_X$, and $M \in \mathscr{M}_{X,Y}$.

(i) If $E(Y|X,M,Z) \perp\!\!\!\perp (C_X;M)$, then $E(Y|X,M,Z) \perp\!\!\!\perp (C_X;M)$.
(ii) If $E(Y|X,M,Z) \perp\!\!\!\perp (C_X;M)$, then $E^{X=x}(Y|M,Z) \perp\!\!\!\perp (C_X;M)$, provided that $x \in \Omega'_X$ with $\{x\} \in \mathscr{A}'_X$ and $P(X{=}x|Z,M) \gneq_P 0$.

(Proof p. 208)

Hence, the regression $E(Y|X,M,Z)$ is $(C_X;M)$-unbiased if it is $(C_X;M)$-unconfounded. Again, note that this proposition holds without any additional assumption. In contrast, we need the additional assumption $P(X{=}x\,|\,Z,M) \gneq_P 0$ in order to conclude from $(C_X;M)$-unconfoundedness of $E(Y|X,M,Z)$ that the $(X{=}x)$-conditional regression $E^{X=x}(Y|Z,M)$ of Y on Z and M is $(C_X;M)$-unbiased. In other words, if $x \in \Omega'_X$ with $\{x\} \in \mathscr{A}'_X$ and $P(X{=}x\,|\,Z,M) \gneq_P 0$, then $(C_X;M)$-unconfoundedness of $E^{X=x}(Y|Z,M)$ implies that $\tau_{x,t_M} := E^{X=x}(Y|M,C_X)$ is P-unique and

$$E^{X=x}(Y|Z,M) \underset{P}{=} E(\tau_{x,t_M}|Z,M). \tag{8.26}$$

Theorem 8.21 has immediate implications on unbiasedness of the prima-facie-effect function $PFE_{xx';M,Z}$. This is obvious if we consider Equation (8.26) and remember that $E(\tau_{x,t_M}|Z,M) - E(\tau_{x',t_M}|Z,M) = E(\delta_{xx',t_M}|Z,M)$ is the (Z,M)-conditional direct-effect function of x vs. x' with respect to the intermediate variable M (see section 5.3).

Corollary 8.22 (Implications on Unbiasedness of Prima Facie Effects)
Let $\langle (\Omega,\mathscr{A},P), (\mathscr{F}_t, t\in T), X, Y\rangle$ be a numerical causality space and let $\delta_{xx',t_M} = \tau_{x,t_M} - \tau_{x',t_M}$. Furthermore, let $x,x' \in \Omega'_X$ with $\{x\},\{x'\} \in \mathscr{A}'_X$ and assume $P(X{=}x|Z,M)$, $P(X{=}x'\,|\,Z,M) \gneq_P 0$. Then (C_X,M)-unconfoundedness of $E(Y|X,M,Z)$ implies that $E^{X=x}(Y|M,C_X)$ and $E^{X=x'}(Y|M,C_X)$ are P-unique, and

$$PFE_{xx';M,Z} \underset{P}{=} E(\tau_{x,t_M}|Z,M) - E(\tau_{x',t_M}|Z,M) \underset{P}{=} E(\delta_{xx',t_M}|Z,M). \tag{8.27}$$

(Proof p. 209)

Hence, if the regression $E(Y|X,M,Z)$ is (C_X,M)-unconfounded and we can assume $P(X{=}x\,|\,Z,M) \gneq_P 0$ and $P(X{=}x'\,|\,Z,M) \gneq_P 0$, then the prima-facie-effect function $PFE_{xx';M,Z}$ is unbiased with respect to the (Z,M)-conditional direct effects. Again, note that the requirements $P(X{=}x|Z,M) \gneq_P 0$ and $P(X{=}x'|Z,M) \gneq_P 0$ are *not trivial* even if X represents a discrete (treatment) variable. If, e. g., X is a treatment variable, then $P(X{=}x\,|\,Z{=}z,M{=}m) > 0$ implies that units obtaining a pair (z,m) of values of Z and M must have a positive probability to be in treatment x, provided that (z,m) has a nonzero probability $P(Z{=}z,M{=}m)$.

8.5 Implications Between Causality Conditions

In Theorem 8.12 we showed that the independent cause conditions and the complete conditions imply the corresponding unconfoundedness conditions, and in Theorem 8.21 we proved that unconfoundedness implies unbiasedness. Table 8.2 summarizes these and other implication relations between the various causality conditions, some of which will only be proved in chapter ???

From these implications relations we can conclude, e. g., that a property that holds under the unconfoundedness $E(Y|X) \perp\!\!\!\perp C_X$ also holds for the independent

cause condition $X \perp\!\!\!\perp_P C_X$ and the regressively independent outcome condition $Y \vdash C_X | X$. For example, it would have been sufficient to prove that $E(Y|X) \Uparrow C_X$ implies that the simple regression $E(Y|X)$ is C_X-unbiased. Because, $X \perp\!\!\!\perp_P C_X$ and $Y \vdash C_X | X$ both imply $E(Y|X) \Uparrow C_X$ we could conclude that both, $X \perp\!\!\!\perp_P C_X$ and $Y \vdash C_X | X$ also imply C_X-unbiasedness of $E(Y|X)$.

If $Z \in \mathscr{W}_X$, the independent cause condition $X \perp\!\!\!\perp_P C_X$ implies $X \perp\!\!\!\perp_P C_X | Z$. Hence, we can conclude that $X \perp\!\!\!\perp_P C_X$ also implies C_X-unbiasedness of $E(Y|X,Z)$. Similarly, because $Y \vdash C_X | X$ implies $Y \vdash C_X | X, Z$, we can conclude that $X \perp\!\!\!\perp_P C_X$ also implies C_X-unbiasedness of the regression $E(Y|X,Z)$.

Instead of commenting on all the implications represented in Table 8.2 let us contend ourselves with hinting at the Generalizability Theorem, which will be treated in chapter ??? **??**. This theorem is the key to many of the implications represented in Table 8.2. Aside from that, this theorem also has important methodological and substantive implications.

8.6 Summary and Conclusions

In this chapter we studied another class of sufficient conditions for unbiasedness, called *unconfoundedness conditions*. Like the independent cause conditions and the regressively independent outcome conditions, the unconfoundedness conditions are well-defined even in cases in which the true outcome variables are not defined. Together with the additional structural components of a causality space, each of these causality conditions make the difference between *ordinary regressive dependencies* (i. e., dependencies that can be described by regressions) and *causal regressive dependencies*.

Unconfoundedness and the Randomized Experiment

The unconfoundedness conditions are implied by their corresponding independent cause conditions and regressively independent outcome conditions that have been treated in chapter 7. Hence, C_X-unconfoundedness of $E(Y|X)$ holds in randomized experiments, presuming, of course, that X is the treatment variable and Y the outcome variable. Similarly, C_X-unconfoundedness of $E(Y|X,Z)$ holds in Z-conditionally randomized experiments (see Remark 7.25). Table 8.2 displays the implication structure of the causality conditions discussed so far.

Covariate Selection in the Analysis of Total Effects

To our knowledge, unconfoundedness is the *weakest condition implying unbiasedness that is still empirically falsifiable*. If unconfoundedness does not hold in a particular empirical application, then there is serious doubt that unbiasedness holds in this application. Hence, unconfoundedness may serve as a criterion for selecting those covariates in the vector $Z := (Z_1, \ldots, Z_Q)$ for which the regression

Table 8.2. Implications Between Causality Conditions

		(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)	(x)	(xi)	(xii)	(xiii)	(xiv)	(xv)	(xvi)
(i)	$X \underset{P}{\perp\!\!\!\perp} C_X$	⇒	⇒							⇒	⇒			⇒	⇒		
(ii)	$X \underset{P}{\perp\!\!\!\perp} C_X \mid Z$		⇒								⇒				⇒		
(iii)	$X \underset{P}{\perp\!\!\!\perp} C_X \mid M$			⇒	⇒							⇒	⇒			⇒	⇒
(iv)	$X \underset{P}{\perp\!\!\!\perp} C_X \mid M, Z$				⇒								⇒				⇒
(v)	$Y \vdash C_X \mid X$					⇒	⇒			⇒	⇒			⇒	⇒		
(vi)	$Y \vdash C_X \mid X, Z$						⇒				⇒				⇒		
(vii)	$Y \vdash C_X \mid X, M$							⇒	⇒			⇒	⇒			⇒	⇒
(viii)	$Y \vdash C_X \mid X, M, Z$								⇒				⇒				⇒
(ix)	$E(Y \mid X) \Uparrow C_X$									⇒	⇒			⇒	⇒		
(x)	$E(Y \mid X, Z) \Uparrow C_X$										⇒				⇒		
(xi)	$E(Y \mid X, M) \Uparrow (C_X; M)$											⇒	⇒			⇒	⇒
(xii)	$E(Y \mid X, M, Z) \Uparrow (C_X; M)$												⇒				⇒
(xiii)	$E(Y \mid X) \uparrow\!\!\!\perp C_X$													⇒			
(xiv)	$E(Y \mid X, Z) \uparrow\!\!\!\perp C_X$														⇒		
(xv)	$E(Y \mid X, M) \uparrow\!\!\!\perp (C_X; M)$															⇒	
(xvi)	$E(Y \mid X, M, Z) \uparrow\!\!\!\perp (C_X; M)$																⇒

Note: We assume that there is a numerical causality space $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$, that $Z \in \mathscr{W}_X$ and $M \in \mathscr{M}_{X,Y}$. Then '⇒' indicates that the condition in the row implies the condition in the column.

Box 8.1 Glossary of New Concepts

The unconfoundedness conditions summarized below presume that there is a numerical causality space $\langle (\Omega, \mathcal{A}, P), (\mathcal{F}_t, t \in T), X, Y \rangle$, a covariate $Z \in \mathcal{W}_X$, and an intermediate variable $M \in \mathcal{M}_{X,Y}$.

$E(Y\|X) \Uparrow C_X$	*C_X-unconfoundedness of $E(Y\|X)$*. This causality condition can be created by random assignment of units to treatment conditions. It is implied by $X \underset{P}{\perp\!\!\!\perp} C_X$ and by $Y \vdash C_X\|X$.
$E(Y\|X,Z) \Uparrow C_X$	*C_X-unconfoundedness of $E(Y\|X,Z)$*. This causality condition can be created by conditional random assignment of units to treatment conditions based on the values z of Z. Units with different values z can be assigned to treatment conditions with different $(Z{=}z)$-conditional probabilities. Furthermore, the covariate $Z = (Z_1,\ldots,Z_m)$ can be selected such that $E(Y\|X,Z) \Uparrow C_X$ might hold. It is implied by $X \underset{P}{\perp\!\!\!\perp} C_X$, by $Y \vdash C_X\|X$, by $X \underset{P}{\perp\!\!\!\perp} C_X\|Z$, and by $Y \vdash C_X\|X,Z$.
$E(Y\|X,M) \Uparrow\!(C_X;M)$	*(C_X,M)-unconfoundedness of $E(Y\|X,M)$*. This causality condition can *not* be created by random assignment of units to treatment conditions. It can also *not* be created by conditional random assignment of units to treatment conditions based on the values m of M, because M is an intermediate variable whose values are not known before treatment. It can also *not* be created by covariate selection. It is implied by $X \underset{P}{\perp\!\!\!\perp} C_X\|M$ and by $Y \vdash C_X\|X,M$.
$E(Y\|X,M,Z) \Uparrow\!(C_X;M)$	*(C_X,M)-unconfoundedness of $E(Y\|X,M,Z)$*. This causality condition can *not* be created by random assignment to treatment conditions. However, the covariate $Z = (Z_1,\ldots,Z_m)$ can be selected such that $E(Y\|X,M,Z) \Uparrow\!(C_X;M)$ might hold. It is implied by $E(Y\|X,M) \Uparrow\!(C_X;M)$, by $X \underset{P}{\perp\!\!\!\perp} C_X\|M$, by $X \underset{P}{\perp\!\!\!\perp} C_X\|M,Z$, by $Y \vdash C_X\|X,M$, and by $Y \vdash C_X\|X,M,Z$.

$E(Y|X,Z)$ can hoped for being C_X-unconfoundedness. Similarly, in the analysis of direct and nondirect effects, unconfoundedness may serve as a criterion for selecting those covariates in the vector $Z := (Z_1,\ldots,Z_m)$ for which the regression $E(Y|X,M,Z)$ can hoped for being (C_X,M)-unconfounded.

Covariate Selection in the Analysis of Direct and nondirect Effects

Neither randomization nor Z-conditional randomization guarantee (C_X,M)-unbiasedness of $E(Y|X,M,Z)$. Hence, the only available design technique in the analysis of direct and nondirect effects is covariate selection. Hence, using (C_X,M)-unconfoundedness of $E(Y|X,M,Z)$ as a criterion in covariate selection is the most viable option, because both, the corresponding independent cause condition and the corresponding regressively independent outcome conditions are more restrictive.

Outlook

In chapter **??** ??? we will go into more details of testing unconfoundedness, which will be important if unconfoundedness is used as a criterion in covariate selection. We will also prove additional theorems on *generalizability* and *expectation stability*, and extend the unconfoundedness conditions to conditional distributions.

8.7 Proofs

Proof of Theorem 8.12

Propositions (i), (ii), and (iii) are particular cases of (iv) for (Z,M), Z, and M being constants, respectively. Hence, we only have to prove (iv). Consider the equation

$$E(Y|X{=}x,M{=}m,Z{=}z) = \int E(Y|X{=}x,Z{=}z,M{=}m,W{=}w)\cdot P_{W|X=x,Z=z,M=m}(dw)$$
$$\text{for } P_{X,Z,M}\text{-a.a. } (x,z,m)\in\Omega'_X\times\Omega'_Z\times\Omega'_M,$$

which is always true (see Cor. 17.60). Because W is measurable with respect to $\mathscr{C}_X$, the independent cause condition $X \underset{P}{\perp\!\!\!\perp} C_X|M,Z$ implies

$$P_{W|X=x,Z=z,M=m} = P_{W|Z=z,M=m}$$
$$\text{for } P_{X,Z,M}\text{-a.a. } (x,z,m)\in\Omega'_X\times\Omega'_Z\times\Omega'_M.$$

Hence, Equation (8.9) immediately follows from $X \underset{P}{\perp\!\!\!\perp} C_X|M,Z$. Similarly, if the complete cause condition $Y \vdash C_X|X,M,Z$ holds, then

$$E(Y|X{=}x,Z{=}z,M{=}m,W{=}w) = E(Y|X{=}x,M{=}m,Z{=}z)$$
$$\text{for } P_{X,Z,M,W}\text{-a.a. } (x,z,m,w)\in\Omega'_X\times\Omega'_Z\times\Omega'_M\times\Omega'_W.$$

will hold for a covariate W, which in turn implies

$$\begin{aligned} E(Y|X{=}x, M{=}m, Z{=}z) &= E(Y|X{=}x, M{=}m, Z{=}z) \cdot 1 \\ &= E(Y|X{=}x, M{=}m, Z{=}z) \cdot \int P_{W|Z=z,M=m}(dw) \\ &= \int E(Y|X{=}x, M{=}m, Z{=}z) \cdot P_{W|Z=z,M=m}(dw) \\ &= \int E(Y|X{=}x, Z{=}z, M{=}m, W{=}w) \cdot P_{W|Z=z,M=m}(dw) \end{aligned}$$

for $P_{X,Z,M}$-a.a. $(x,z,m) \in \Omega'_X \times \Omega'_Z \times \Omega'_M$.

However, this proves proposition (iv).

Proof of Theorem 8.15

This theorem is a particular case of Theorem 8.17 .

Proof of Corollary 8.16

This corollary is a particular case of Corollary 8.18.

Proof of Theorem 8.17

Proof of Corollary 8.18

According to Theorem 8.17, both $E^{X=x}(Y|C_X)$ and $E^{X=x'}(Y|C_X)$ are P-unique. Furthermore, according to this theorem,

$$E^{X=x}(Y|Z) \underset{P}{=} E(\tau_x|Z)$$

as well as

$$E^{X=x'}(Y|Z) \underset{P}{=} E(\tau_{x'}|Z).$$

Hence, taking the difference between these two equations yields Equation (8.24).

Proof of Theorem 8.19

This theorem is a particular case of Theorem 8.21.

Proof of Corollary 8.20

This corollary is a particular case of Corollary 8.22.

Proof of Theorem 8.21

The proof is analog to the proof of Theorem 8.17 (see Exercise 8-1).

Proof of Corollary 8.22

The proof is analog to the proof of Corollary 8.18 (see Exercise 8-2).

8.8 Exercises

▷ **Exercise 8-1** Prove Theorem 8.21

▷ **Exercise 8-2** Prove Corollary 8.22.

▷ **Exercise 8-3** Compute the probabilities $P(U=u\,|X=0,Z=m)$ and $P(U=u\,|X=0,Z=f)$ in the example presented in Table 8.1.

▷ **Exercise 8-4** Compute the conditional expectations $E(Y|X=0,Z=z)$ in the example presented in Table 8.1.

▷ **Exercise 8-5** Compute the conditional probabilities $P(Z=z\,|\,X=0)$ in the example presented in Table 8.1 for the covariate $Z :=$ *sex*.

▷ **Exercise 8-6** Compute the conditional probabilities $P(U=u\,|X=0,Z=m,W=m)$ and $P(U=u\,|X=0,Z=m,W=n)$ in the example presented in Table 8.1.

▷ **Exercise 8-7** Compute the conditional expectations $E(Y\,|\,X=0,Z=m,W=w)$ in the example presented in Table 8.1.

▷ **Exercise 8-8** Compute the conditional probabilities $P(W=w\,|\,X=0,Z=m)$ in the example presented in Table 8.1.

▷ **Exercise 8-9** Compute the conditional probabilities $P(W=w|Z=m)$ in the example presented in Table 8.1.

▷ **Exercise 8-10** Define a mapping h such that the variable $Z :=$ *sex* in the example of Table 8.1 is the composition $h(U)$ of h and U. [An alternative notation for $h(U)$ is $h\circ U$.]

▷ **Exercise 8-11** Using the example of Table 8.1, give five examples of a variable $W=f(U)$ and write down the associated five partitions of the total set of observational units $\Omega_U := \{u_1,u_2,\ldots,u_6\}$. (A *partition* of the set Ω_U is a set of subsets of Ω_U such that the union of these subsets is Ω_U, the subsets are disjunct, i. e. they do not intersect, and the empty set is not an element in the partition.)

▷ **Exercise 8-12** How many different partitions of the set of units $\Omega_U := \{u_1,u_2,\ldots,u_6\}$ — excluding the partition $\{\Omega_U\}$ — exist in the example of Table 8.1?

▷ **Exercise 8-13** Show that *Z-conditional independence* of X and U does not hold in the example of Table 8.1.

▷ **Exercise 8-14** Show that $Y \vdash C_X\,|\,X,Z$ with $Z :=$ *sex* does not hold in the example of Table 8.1.

Solutions

▷ **Solution 8-1** (i) The unconfoundedness condition $E(Y|X,M,Z) \perp\!\!\!\perp (C_X; M)$ implies (8.9). Because the global covariate C_X is a potential confounder and Z is measurable with respect to C_X, (8.9) implies

$$\begin{aligned} E(Y|X{=}x, M{=}m, Z{=}z) &= \int E(Y|X{=}x, Z{=}z, M{=}m, C_X{=}\omega)\, P_{C_X|Z=z,\,M=m}(d\omega) \\ &= \int E(Y|X{=}x, M{=}m, C_X{=}\omega)\, P_{C_X|Z=z,\,M=m}(d\omega) \end{aligned} \tag{8.28}$$

$$\text{for } P_{X,Z,M}\text{-a.a. } (x,z,m) \in \Omega'_X \times \Omega'_Z \times \Omega'_M,$$

which defines unbiasedness of $E(Y|X,M,Z)$ w.r.t. (Z,M)-conditional M-direct effects.

(ii) $P(X{=}x\,|\,Z,M) \underset{P}{>} 0$ implies $P(X{=}x) > 0$. Hence, according to Remark **??**, Equation (8.28) implies

$$\begin{aligned} E^{X=x}(Y|Z{=}z, M{=}m) &= \int E^{X=x}(Y|M{=}m, C_X{=}\omega)\, P_{C_X|Z=z,\,M=m}(d\omega) \\ &\text{for } P^{X=x}_{Z,M}\text{-a.a. } (z,m) \in \Omega'_Z \times \Omega'_M. \end{aligned} \tag{8.29}$$

However, according to Theorem **??**, $P(X{=}x\,|\,Z,M) \underset{P}{>} 0$ and Equation (8.29) imply that $E^{X=x}(Y|M,C_X)$ is P-unique and $E^{X=x}(Y|Z,M) \underset{P}{=} E(\tau_{x,t_M}|Z,M)$.

▷ **Solution 8-2** According to Theorem 8.21, both $E^{X=x}(Y|M,C_X)$ and $E^{X=x'}(Y|M,C_X)$ are P-unique. Furthermore, according to this theorem,

$$E^{X=x}(Y|Z,M) \underset{P}{=} E(\tau_{x,t_M}|Z,M) \quad \text{and} \quad E^{X=x'}(Y|Z,M) \underset{P}{=} E(\tau_{x',t_M}|Z,M).$$

Hence, taking the difference between these two equations yields Equation (8.27).

▷ **Solution 8-3** For $Z{=}m$, these probabilities can be computed as follows:

$$\begin{aligned} P(U{=}u\,|X{=}0, Z{=}m) &= \frac{P(X{=}0\,|\,Z{=}m, U{=}u)\cdot P(Z{=}m, U{=}u)}{P(X{=}0\,|\,Z{=}m)\cdot P(Z{=}m)} \\ &= \frac{2/6\cdot 1/6}{2/6\cdot 1/2} = \frac{1}{3}, \end{aligned}$$

for units u_1, u_2 and u_3 and $P(U{=}u\,|X{=}0, Z{=}m) = 0$ for units u_4, u_5 and u_6.

For $Z{=}f$, these probabilities are:

$$\begin{aligned} P(U{=}u_4\,|X{=}0, Z{=}f) &= \frac{P(X{=}0\,|\,Z{=}f, U{=}u_4)\cdot P(Z{=}f, U{=}u_4)}{P(X{=}0\,|\,Z{=}f)\cdot P(Z{=}f)} \\ &= \frac{2/12\cdot 1/6}{2/12\cdot 1/2} = \frac{2}{6}, \\ P(U{=}u_5\,|X{=}0, Z{=}f) &= \frac{P(X{=}0\,|\,Z{=}f, U{=}u_5)\cdot P(Z{=}f, U{=}u_5)}{P(X{=}0\,|\,Z{=}f)\cdot P(Z{=}f)} \\ &= \frac{1/12\cdot 1/6}{2/12\cdot 1/2} = \frac{1}{6}, \end{aligned}$$

$$P(U{=}u_6\,|\,X{=}0, Z{=}f) \;=\; \frac{P(X{=}0\,|\,Z{=}f, U{=}u_6)\cdot P(Z{=}f, U{=}u_6)}{P(X{=}0\,|\,Z{=}f)\cdot P(Z{=}f)}$$
$$= \frac{3/12\cdot 1/6}{2/12\cdot 1/2} \;=\; \frac{3}{6},$$

and $P(U{=}u\,|X{=}0, Z{=}f) = 0$ for units u_1, u_2, and u_3.

▷ **Solution 8-4** According to Eq. (**??**) in Box 9.2,

$$E(Y\,|\,X{=}0, Z{=}z) \;=\; \sum_u E(Y\,|\,X{=}0, U{=}u, Z{=}z)\cdot P(U{=}u\,|X{=}0, Z{=}z)$$

For $Z{=}m$ (males) this yields

$$E(Y\,|\,X{=}0, Z{=}m) \;=\; \sum_u E(Y\,|\,X{=}0, U{=}u, Z{=}m)\cdot P(U{=}u\,|X{=}0, Z{=}m)$$
$$= 82\cdot\frac{1}{3}+89\cdot\frac{1}{3}+95\cdot\frac{1}{3} \;=\; 88.\overline{6},$$

using the results of Exercise **??**. For $Z{=}f$ (females) we receive

$$E(Y\,|\,X{=}0, Z{=}f) \;=\; \sum_u E(Y\,|\,X{=}0, U{=}u, Z{=}f)\cdot P(U{=}u\,|X{=}0, Z{=}f)$$
$$= 102\cdot\frac{2}{6}+102\cdot\frac{1}{6}+102\cdot\frac{3}{6} \;=\; 102,$$

again, using the results of Exercise **??**.

▷ **Solution 8-5** Applying the theorem of total probability to the conditional probability $P(Z{=}z\,|\,X{=}0)$ yields

$$P(Z{=}z\,|\,X{=}0) \;=\; \sum_u P(Z{=}z, U{=}u\,|X{=}0).$$

Hence, for $Z{=}m$ (males) we receive

$$P(Z{=}m\,|\,X{=}0) \;=\; \sum_u P(Z{=}m, U{=}u\,|X{=}0) \;=\; \frac{4}{18}+\frac{4}{18}+\frac{4}{18} \;=\; \frac{2}{3},$$

whereas for $Z{=}f$ (females) we receive

$$P(Z{=}f\,|\,X{=}0) \;=\; \sum_u P(Z{=}f, U{=}u\,|X{=}0) \;=\; \frac{2}{18}+\frac{1}{18}+\frac{3}{18} \;=\; \frac{1}{3}.$$

▷ **Solution 8-6** The conditional probabilities $P(U{=}u\,|X{=}0, Z{=}m, W{=}m)$ can be computed by

$$P(U{=}u\,|X{=}0, Z{=}m, W{=}m) \;=\; \frac{P(X{=}0\,|\,Z{=}m, W{=}m, U{=}u)\cdot P(Z{=}m, W{=}m, U{=}u)}{P(X{=}0\,|\,Z{=}m, W{=}m)\cdot P(Z{=}m, W{=}m)}.$$

For $U{=}u_1$, this equation yields

$$P(U{=}u_1\,|X{=}0, Z{=}m, W{=}m) \;=\; \frac{P(X{=}0\,|\,Z{=}m, W{=}m, U{=}u_1)\cdot P(Z{=}m, W{=}m, U{=}u_1)}{P(X{=}0\,|\,Z{=}m, W{=}m)\cdot P(Z{=}m, W{=}m)}$$
$$= \frac{2/6\cdot 1/6}{2/6\cdot 1/6} = 1,$$

whereas it yields $P(U{=}u\,|X{=}0, Z{=}m, W{=}m) = 0$ for all other units.

For $X{=}0$, $Z{=}m$, $W{=}n$ and $U{=}u_2$, these probabilities are:

$$P(U{=}u_2 \,|\, X{=}0, Z{=}m, W{=}n) = \frac{P(X{=}0 \,|\, Z{=}m, W{=}n, U{=}u_2) \cdot P(Z{=}m, W{=}n, U{=}u_2)}{P(X{=}0 \,|\, Z{=}m, W{=}n) \cdot P(Z{=}m, W{=}n)}$$
$$= \frac{2/6 \cdot 1/6}{2/6 \cdot 2/6} = 1/2.$$

The same result, $P(U{=}u_3 \,|\, X{=}0, Z{=}m, W{=}n) = 1/2$, is obtained for unit u_3, whereas the conditional probabilities $P(U{=}u \,|\, X{=}0, Z{=}m, W{=}n)$ are 0 for the other four units.

▷ **Solution 8-7** According to Eq. (**??**) in Box 9.2,

$$E(Y \,|\, X{=}x, Z{=}z, W{=}w) = \sum_u E(Y \,|\, X{=}x, Z{=}z, W{=}w, U{=}u) \cdot P(U{=}u \,|\, X{=}x, Z{=}z, W{=}w)$$

For $X{=}0$, $Z{=}m$ (males), and $W{=}m$ (married), this equation yields

$$E(Y \,|\, X{=}0, Z{=}m, W{=}m) = \sum_u E(Y \,|\, X{=}0, Z{=}m, W{=}m, U{=}u) \cdot P(U{=}u \,|\, X{=}0, Z{=}m, W{=}m)$$
$$= 82 \cdot 1 = 82,$$

using the results of Exercise 8-6. In contrast, for $X{=}0$, $Z{=}m$ (males), and $W{=}n$ (not married), it yields

$$E(Y \,|\, X{=}0, Z{=}m, W{=}n) = \sum_u E(Y \,|\, X{=}0, Z{=}m, W{=}n, U{=}u) \cdot P(U{=}u \,|\, X{=}0, Z{=}m, W{=}n)$$
$$= 89 \cdot \frac{1}{2} + 95 \cdot \frac{1}{2} = 92,$$

again, using the results of Exercise 8-6.

▷ **Solution 8-8** Applying the theorem of total probability to the conditional probability $P(W{=}w \,|\, X{=}0, Z{=}m)$ yields

$$P(W{=}w \,|\, X{=}0, Z{=}m) = \sum_u P(W{=}w, U{=}u \,|\, X{=}0, Z{=}m).$$

Hence, for $W{=}m$ (married) we receive

$$P(W{=}m \,|\, X{=}0, Z{=}m) = \sum_u P(W{=}m, U{=}u \,|\, X{=}0, Z{=}m) = \frac{1}{3},$$

whereas for $W{=}n$ we receive

$$P(W{=}n \,|\, X{=}0, Z{=}m) = \sum_u P(W{=}n, U{=}u \,|\, X{=}0, Z{=}m) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}.$$

▷ **Solution 8-9** Applying the theorem of total probability to the conditional probability $P(W{=}w \,|\, Z{=}m)$ yields

$$P(W{=}w \,|\, Z{=}m) = \sum_u P(W{=}w, U{=}u \,|\, Z{=}m).$$

Hence, for $W{=}m$ (married) we receive

$$P(W{=}m \,|\, Z{=}m) = \sum_u P(W{=}m, U{=}u \,|\, Z{=}m) = \frac{1}{3},$$

whereas for $W{=}n$ we receive

$$P(W{=}n \,|\, Z{=}m) = \sum_u P(W{=}n, U{=}u \,|\, Z{=}m) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}.$$

▷ **Solution 8-10** In the example of Table 8.1, the random variable $U : \Omega \to \Omega_U$, where $\Omega_U := \{u_1, \ldots, u_6\}$, is defined by $U(\omega) = u_i$, where $\omega = (u_i, \omega_X, \omega_Y)$ is one of the possible outcomes of the single-unit trial, which indicates that unit u_i is sampled, then assigned (or self-assigned) to a treatment condition ω_X, and outcome ω_Y is observed. Therefore we can now define the mapping $h : \Omega_U \to \{m, f\}$ with $h(u_i) = m$, if u_i is a male unit and $h(u_i) = f$, if the unit sampled is female. The variable $Z :=$ *sex* with values m and f is the composition of this mapping h with U, i. e.: $Z := h(U)$. Both, U and Z, are mappings on the set Ω of possible outcomes of the single-unit trial, i. e., $U : \Omega \to \Omega_U$ and $Z : \Omega \to \{m, f\}$.

▷ **Solution 8-11** The first example is $W_1 := U$, the second is $W_2 := Z :=$ *sex*, and the third is $W_3 := W :=$ *marital status*. A fourth example can be defined by $W_4(\omega) = 1$ if $\tau_1(\omega) = 100$ and $W_4(\omega) = 0$, otherwise. A fifth example is $W_5(\omega) = 1$, if $\tau_0(\omega) < 90$, $W_5(\omega) = 2$, if $90 \leq \tau_0(\omega) < 100$, and $W_5(\omega) = 3$, if $\tau_0(\omega) \geq 100$.

The partition of Ω_U associated with U is the following set of subsets of Ω_U: $S_1 = \{\{u_1\}, \{u_2\}, \ldots, \{u_6\}\}$. The partition of Ω_U associated with Z is: $S_2 = \{\{u_1, u_2, u_3\}, \{u_4, u_5, u_6\}\}$. The partition of Ω_U associated with W is $S_3 = \{\{u_1, u_4\}, \{u_2, u_3, u_5\}, \{u_6\}\}$. The partition of Ω_U associated with W_4 is again S_2. Finally, $S_5 = \{\{u_1, u_2\}, \{u_3\}, \{u_4, u_5, u_6\}\}$ is the partition of Ω_U associated with W_5.

▷ **Solution 8-12** In total there are 122 possible partitions of the total population Ω_U into subsets. It is tedious to obtain this number by writing out all possible partitions. However it can be obtained by the formula for the *Stirling numbers of the second kind*, which gives the number of possible partitions of a set with n elements into k distinct, non-empty and exhaustive sets. Stirling numbers of the second kind can be calculated by:

$$S(n,k) = \frac{1}{k!} \sum_{i=0}^{k} (-1)^i \binom{k}{i} (k-i)^n,$$

where $k! := k \cdot (k-1) \cdot \ldots \cdot 1$ and

$$\binom{k}{i} := \frac{k!}{i! \cdot (k-i)!}.$$

The sum $\sum_{j=2}^{6} S(6, j)$ gives the number of different partitions of $\Omega_U := \{u_1, u_2, \ldots, u_6\}$ into two or more subsets, i. e., excluding the partition $\{\Omega_U\}$. Hence, in our case this sum is $31 + 10 + 65 + 15 + 1 = 122$.

▷ **Solution 8-13** Conditional independence of X and U given $Z :=$ *sex* does not hold in this example, because in treatment condition 0, the individual treatment probabilities $P(X{=}0 \,|\, Z{=}f, U{=}u)$ are not the same for each unit u in the female subpopulation, i. e., $P(X{=}0 \,|\, Z{=}f, U{=}u) \neq P(X{=}0 \,|\, Z{=}f)$.

▷ **Solution 8-14** The complete cause condition $Y \vdash C_X \,|\, X, Z$ does not hold in this example, because, for example, in treatment 0, the male units u_1, u_2, and u_3 do not have the same true outcomes $E^{X=0}(Y \,|\, U{=}u)$. Instead, these true outcomes are 82, 89, and 95, respectively.

Chapter 9
Other Causality Conditions

In chapter 6 we have seen that the expectations of the outcomes in the treatment conditions and their differences, the prima facie effects, can be biased and we have defined what it means that they are unbiased. Furthermore, we also introduced the general concepts of unbiasedness of the regressions $E(Y|X)$ and $E(Y|X,Z)$, which imply unbiasedness of the corresponding prima facie effects $E(Y|X{=}x) - E(Y|X{=}x')$ and $E(Y|X{=}x, Z{=}z) - E(Y|X{=}x', Z{=}z)$, respectively. Finally, the concept of unbiasedness has been extended to conditional distributions.

???

RESTE

9.0.1 Strong Ignorability With Respect to Total Effects

???

Now we consider independence and conditional independence of the cause X and true-outcome variables τ_x. Hence, let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a numerical causality space, let $x \in X(\Omega)$ be a particular value of X with $P(X{=}x) > 0$, and let $1_{X=x}$ denote the indicator of $\{X{=}x\} := \{\omega \in \Omega\colon X(\omega) = x\}$. Then we abbreviate independence of $1_{X=x}$ and a true-outcome variable τ_x by $1_{X=x} \perp\!\!\!\perp \tau_x$. If Z is a random variable on $(\Omega, \mathscr{A}, P)$, we will abbreviate Z-conditional independence of $1_{X=x}$ and a true-outcome variable τ_x by $1_{X=x} \perp\!\!\!\perp \tau_x | Z$. In this case, the σ-algebras $\mathscr{D}$ and $\mathscr{E}$ used in Definition **??** are the σ-algebras generated by $1_{X=x}$ and by τ_x, respectively.

Similarly, if the σ-algebras $\mathscr{D}$ and $\mathscr{E}$ used in Definition **??** are the σ-algebras generated by X and by τ_x, respectively, this defines independence of the cause X and the true-outcome variable τ_x, which will be abbreviated by $X \perp\!\!\!\perp \tau_x$. Correspondingly, we will use $X \perp\!\!\!\perp \tau_x | Z$ to denote Z-conditional independence of X and τ_x. If Z is measurable with respect to (X, C_X), then these conditions will be referred to as the *strong ignorability conditions with respect to total effects*.[1]

[1] In conjunction with $0 < P(X{=}x | Z) < 1$, the condition $X \perp\!\!\!\perp \tau_x | Z$ corresponds to the Rosenbaum-Rubin *strong ignorability* condition (see, e. g., Rosenbaum & Rubin, 1983b). The true-outcome variables τ_x with respect to total effects take the role of Rubin's 'potential outcome variables' Y_x (see chapter 5). We will make additional requirements only when needed for a specific proposition.

9.0.2 Strong Ignorability With Respect to Direct Effects

Strong ignorability can be extended if we additionally include an intermediate variable M. In this case, we consider the true-outcome variables τ_{x,t_M} with respect to M-direct effects. If W is measurable with respect to (X, M, C_X), then *W-conditional strong ignorability with respect to M-direct effects* is defined by $X \perp\!\!\!\perp \tau_{x,t_M} | M, W$.

Again, sometimes we will consider independence and conditional independence of an indicator $1_{X=x}$ of a value x of X and a true-outcome variable τ_{x,t_M} abbreviated by $1_{X=x} \perp\!\!\!\perp \tau_{x,t_M}$ and $1_{X=x} \perp\!\!\!\perp \tau_{x,t_M} | W$, respectively.

Remark 9.1 (Independent Cause Conditions Imply Strong Ignorability) The true-outcome variable *with respect to total effects* has been defined by $\tau_x := E^{X=x}(Y|C_X)$. Therefore $X \perp\!\!\!\perp \tau_x$ follows from $X \perp\!\!\!\perp C_X$. Of course, the same applies to $X \perp\!\!\!\perp \tau_x | Z$ that follows from the independent cause condition $X \perp\!\!\!\perp C_X | Z$.

Similarly, the true-outcome variable *with respect to M-direct effects* has been defined by $\tau_{x,t_M} := E^{X=x}(Y|M, C_X)$. Hence, $X \underset{P}{\perp\!\!\!\perp} C_X | M, W$ implies $X \perp\!\!\!\perp \tau_{x,t_M} | M, W$, because, according to Rule (vii) in Box 16.2 of Steyer & Nagel, in press,

$$X \underset{P}{\perp\!\!\!\perp} C_X | M, W \quad \Leftrightarrow \quad X \perp\!\!\!\perp M, C_X | M, W . \tag{9.1}$$

◁

9.0.3 Weak Ignorability With Respect to Total Effects

The strong ignorability condition $1_{X=x} \perp\!\!\!\perp \tau_x | Z$ requires Z-conditional stochastic independence of $1_{X=x}$ and τ_x. However, as will be shown in section 7.3, unbiasedness of $E(Y|X=x)$ already follows from $E(\tau_x | 1_{X=x}) = E(\tau_x)$. Therefore, we will now consider a weaker kind of ignorability conditions assuming only regressive independence or conditional regressive independence.

Hence, let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a numerical causality space, let $x \in X(\Omega)$ be a particular value of X with $P(X=x) > 0$, and let $1_{X=x}$ denote the indicator of $\{X=x\} := \{\omega \in \Omega\colon X(\omega) = x\}$. Then we abbreviate

$$E(\tau_x | 1_{X=x}) \;=\; E(\tau_x) \tag{9.2}$$

by $\tau_x \vdash 1_{X=x}$. Similarly, let X be discrete with values $x = 0, 1, \ldots J$ such that $P(X=x) > 0$ for all values $x = 0, 1, \ldots J$ and let $\tau := (\tau_0, \tau_1, \ldots, \tau_J)$ denote the vector of true-outcome variables with respect to total effects. Then we abbreviate

$$E(\tau_x | X) \;=\; E(\tau_x) \quad \text{for each value } x = 0, 1, \ldots J \tag{9.3}$$

by

$$\tau \vdash X.$$

Furthermore, we abbreviate

$$E(\tau_x | 1_{X=x}, Z) \underset{P}{=} E(\tau_x | Z) \tag{9.4}$$

by

$$\tau_x \vdash 1_{X=x} \mid Z,$$

and

$$E(\tau_x \mid X, Z) \underset{P}{=} E(\tau_x \mid Z) \quad \text{for each value } x = 0, 1, \ldots J \tag{9.5}$$

by

$$\tau \vdash X \mid Z.$$

If Z is measurable with respect to (X, C_X), then these conditions will be referred to as the *weak ignorability conditions with respect to total effects.*

9.0.4 Weak Ignorability With Respect to Direct Effects

If we also include an intermediate variable M and consider the true-outcome variables with respect to M-direct effects, which are defined by $\tau_{x,t_M} := E^{X=x}(Y \mid M, C_X)$, and still consider a particular value x of X with $P(X{=}x) > 0$, then we abbreviate

$$E(\tau_{x,t_M} \mid 1_{X=x}, M) \underset{P}{=} E(\tau_{x,t_M} \mid M) \tag{9.6}$$

by

$$\tau_{x,t_M} \vdash 1_{X=x} \mid M.$$

Similarly, if X is discrete with values $x = 0, 1, \ldots J$ such that $P(X{=}x) > 0$ for all values $x = 0, 1, \ldots J$ and $\tau_M := (\tau_{0,t_M}, \tau_{1,t_M}, \ldots, \tau_{J,t_M})$, then we abbreviate

$$E(\tau_{x,t_M} \mid X, M) \underset{P}{=} E(\tau_{x,t_M} \mid M) \quad \text{for each value } x = 0, 1, \ldots J \tag{9.7}$$

by

$$\tau_M \vdash X \mid M.$$

Finally, we abbreviate

$$E(\tau_{x,t_M} \mid 1_{X=x}, M, W) \underset{P}{=} E(\tau_{x,t_M} \mid M, W) \tag{9.8}$$

by

$$\tau_{x,t_M} \vdash 1_{X=x} \mid M, W,$$

and

$$E(\tau_{x,t_M} \mid X, M, W) \underset{P}{=} E(\tau_{x,t_M} \mid M, W) \quad \text{for each value } x = 0, 1, \ldots J \tag{9.9}$$

by

$$\tau_M \vdash X \mid M, W.$$

If W is measurable with respect to (X, M, C_X), then these conditions will be referred to as the *weak ignorability conditions with respect to direct effects.*

???

HEIR Original weiter

???

The specific goals of this chapter are to learn more about:

(a) other important causality conditions,
(b) the implication structure between these conditions, and
(c) the consequences of these conditions for the design of experiments and quasi-experiments as well as for data analysis in quasi-experiments and beyond.

Overview

The following sufficient conditions for unbiasedness will be treated in this chapter: *independence of X and true outcomes* — also known as *strong ignorability* — and *regressive independence of true outcomes from X*. Then we study the implication structure between these conditions and briefly discuss their practical consequences. A more detailed discussion will be given in chapter 14.

9.1 Independence of X and True Outcomes

In this section we will show that independence of X and the true outcomes is also a sufficient condition for unbiasedness, which is somewhat weaker than independence of X and covariates. We start with some useful implications of the independence of X and the true outcomes. Then we show that independence of X and the true outcomes implies unbiasedness. Next we illustrate the theory by a simple numerical example in which the conjectured cause is again a treatment variable. Finally, we discuss independence of X and the true outcomes from a substantive point of view.

9.1.1 Theory

Definition 9.2 (Independence of Random Variables)
Let $X : (\Omega, \mathscr{A}) \to (\Omega'_X, \mathscr{A}'_X)$ and $Y : (\Omega, \mathscr{A}) \to (\Omega'_Y, \mathscr{A}'_Y)$ be a random variables on a probability space $(\Omega, \mathscr{A}, P)$. Then:

(i) X and Y are called $independent$ (abbr.: $X \perp\!\!\!\perp Y$) if

$$\begin{aligned} P(A \cap C) &= P(A) \cdot P(C) \\ \text{for all } (A, C) &\in X^{-1}(\mathscr{A}'_X) \times Y^{-1}(\mathscr{A}'_Y). \end{aligned} \tag{9.10}$$

(ii) Let Z also be a random variable on $(\Omega, \mathscr{A}, P)$, and let $P(A|Z) := E(1_A|Z)$ denote the regression of the indicator 1_A on Z. Then X and Y are called Z-$conditionally\ independent$ (abbr.: $X \perp\!\!\!\perp Y \mid Z$) if

$$\begin{aligned} P(A \cap C|Z) &\underset{P}{=} P(A|Z) \cdot P(C|Z) \\ \text{for all } (A, C) &\in X^{-1}(\mathscr{A}'_X) \times Y^{-1}(\mathscr{A}'_Y). \end{aligned} \tag{9.11}$$

Remark 9.3 Note that $X \perp\!\!\!\perp Y$ is a particular case of $X \perp\!\!\!\perp Y \mid Z$ in which Z is a constant, i. e., in which Z is a random variable taking the same value for all $\omega \in \Omega$. In this chapter, we will use these concepts of independence and conditional independence for the treatment variable X and true outcome variables τ_x. Hence, their independence can be abbreviated by $X \perp\!\!\!\perp \tau_x$ and their conditional independent given a covariate Z by $X \perp\!\!\!\perp \tau_x | Z$. We will also consider independence and conditional independence of an indicator of a value x of X and a true outcome variable τ_x (abbr.: $1_{X=x} \perp\!\!\!\perp \tau_x$ and $1_{X=x} \perp\!\!\!\perp \tau_x | Z$, respectively). ◁

Remark 9.4 Remember, that the true outcome variable has been defined by $\tau_x := E^{X=x}(Y|\mathscr{C}_X)$. Hence, $X \perp\!\!\!\perp \tau_x$ and $1_{X=x} \perp\!\!\!\perp \tau_x$ follow from $X \perp\!\!\!\perp \mathscr{C}_X$ and $1_{X=x} \perp\!\!\!\perp \mathscr{C}_X$, respectively. Of course, the same applies to $X \perp\!\!\!\perp \tau_x | Z$ and $1_{X=x} \perp\!\!\!\perp \tau_x | Z$ that follow from $X \perp\!\!\!\perp \mathscr{C}_X | Z$ and $1_{X=x} \perp\!\!\!\perp \mathscr{C}_X | Z$, respectively. (The independent cause conditions $X \perp\!\!\!\perp \mathscr{C}_X$ and $X \perp\!\!\!\perp \mathscr{C}_X | Z$ have been treated in chapters 7 and **??**.) ◁

Theorem 9.5 (Independence of X and True Outcomes)
Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a numerical causality space, let $x \in X(\Omega)$ be a value of X with $P(X{=}x) > 0$, and let $1_{X=x}$ denote the indicator of the event $\{X{=}x\} := \{\omega \in \Omega \colon X(\omega) = x\}$. Then:

(i) $1_{X=x} \perp\!\!\!\perp \tau_x$ is equivalent to

$$P(X{=}x \mid \tau_x) \underset{P}{=} P(X{=}x) \tag{9.12}$$

(ii) If Z a random variable on $(\Omega, \mathscr{A}, P)$, then $1_{X=x} \perp\!\!\!\perp \tau_x | Z$ is equivalent to

$$P(X{=}x \mid Z, \tau_x) \underset{P}{=} P(X{=}x \mid Z) \tag{9.13}$$

Furthermore, let X be discrete with a finite number of values $0, 1, \ldots, J$ such that $P(X{=}x) > 0$ for each $x = 0, 1, \ldots, J$ and let $\tau := (\tau_0, \tau_1, \ldots, \tau_J)$ denote the vector of the associated true outcome variables. Then:

(iii) $X \perp\!\!\!\perp \tau$ is equivalent to

$$P(X{=}x \mid \tau) \underset{P}{=} P(X{=}x) \quad \text{for each } x = 0, 1, \ldots, J. \tag{9.14}$$

(iv) If Z a random variable on $(\Omega, \mathscr{A}, P)$, then $X \perp\!\!\!\perp \tau \mid Z$ is equivalent to

$$P(X{=}x \mid Z, \tau) \underset{P}{=} P(X{=}x \mid Z) \quad \text{for each } x = 0, 1, \ldots, J. \tag{9.15}$$

(Proof p. 233)

Remark 9.6 (Symmetry and Implications of Independence) Note that $\tau_x \perp\!\!\!\perp I_{X=x}$ is equivalent to $I_{X=x} \perp\!\!\!\perp \tau_x$ (see Exercise **??**). Similarly, $\tau \perp\!\!\!\perp X \mid Z$ is equivalent to $X \perp\!\!\!\perp \tau \mid Z$. Furthermore, $\tau \perp\!\!\!\perp X \mid Z$ implies that X and each τ_x, $x = 0, 1, \ldots, J$, are independent given Z. Conditional independence of X and true outcomes, i. e.,

$\tau \perp\!\!\!\perp X \mid Z$, also implies that X and each subset of components of the vector τ are independent given Z.

◁

Remark 9.7 (Strong Ignorability) Independence of X and the true outcomes corresponds to Rubin's *strong ignorability* condition. The true outcome variables τ_x replace the potential outcome variables Y_x in Rubin's terminology (see chapter 5). Rosenbaum and Rubin (1983b) additionally require $0 < P(X{=}x \mid Z) < 1$. In the following theorem we make a slightly weaker requirement postulating $P(X{=}x \mid Z) \gneq_P 0$, which is equivalent to $0 < P(X{=}x \mid Z) < 1$ if Z is discrete with $P(Z{=}z) > 0$ for all values $z \in \Omega'_Z$ of Z. ◁

Theorem 9.8 (Unbiasedness)
Let the assumptions of Theorem 9.5 hold true. Then:

(i) $1_{X=x} \perp\!\!\!\perp \tau_x$ *implies that* $E(Y \mid X{=}x)$ *is unbiased.*

(ii) If Z *is a covariate and* $P(X{=}x \mid Z) \gneq_P 0$*, then:* $1_{X=x} \perp\!\!\!\perp \tau_x \mid Z$ *implies that* $E^{X=x}(Y \mid Z)$ *is unbiased w.r.t. total effects.*

(Proof p. 233)

Now, let $\tau := (\tau_0, \tau_1, \ldots, \tau_J)$. Since $\tau \perp\!\!\!\perp X \mid Z$ implies $\tau_x \perp\!\!\!\perp I_{X=x} \mid Z$ for each value x of X and Z-conditional unbiasedness of $E(Y \mid X, Z)$ is defined by unbiasedness of $E^{X=x}(Y \mid Z)$ for each value x of X, Theorem 9.8 implies the following corollary.

Corollary 9.9 *Let* $\tau := (\tau_0, \tau_1, \ldots, \tau_J)$*. Then:*

(i) $\tau \perp\!\!\!\perp X$ *implies that* $E(Y \mid X)$ *is unbiased.*
(ii) $\tau \perp\!\!\!\perp X \mid Z$ *implies that* $E(Y \mid X, Z)$ *is* Z*-conditionally unbiased.*

(Proof p. 234)

9.1.2 Substantive Meaning

According to Theorem 9.8 and Corollary 9.9, independence of X and true outcomes implies unbiasedness of the regression $E(Y \mid X)$, and conditional independence of X and true outcomes given a covariate Z implies Z-conditional unbiasedness of the regression $E(Y \mid X, Z)$.

Remark 9.10 (Relationship to Randomization) Let us assume that X represents a treatment variable. Random assignment of the unit to one of the treatment conditions (randomization) will not only imply that the observational-unit variable U and the treatment variable X are independent but also that the true outcome variables and X are stochastically independent. This follows from the fact that the true outcome variable τ is measurable w.r.t. the covariate projection c_X and that randomization creates independence of X and c_X. ◁

Remark 9.11 (Testability) Independence of X and true outcomes as well as Z-conditional independence of X and true outcomes do not imply anything that could be tested empirically. In contrast, independence of c_X and X (see section **??**) and homogeneity of the regression $E(Y|X)$ (see section **??**) *do* have empirically testable consequences. Hence, although independence of X and true outcomes is theoretically important, it is useless in empirical applications, because, in contrast to independence of c_X and X as well as homogeneity of $E(Y|X)$, it cannot be used in covariate selection. Note that, in chapter **??**, we will treat another, weaker sufficient condition for unbiasedness, which *does* have implications that are empirically testable, and which can also be used as a criterion in covariate selection. ◁

9.1.3 Numerical Example

Example 9.12 We again assume that assumption (**??**), i. e.,

$$E(Y|X,c_X) = E(Y|X,U) \quad a.s. \tag{9.16}$$

holds for the true outcome variables τ_x and that

$$P(X{=}1|c_X) = P(X{=}1|U) \quad a.s. \tag{9.17}$$

holds for the treatment probabilities. As mentioned before, if U represents the observational units at the onset of treatment, these assumptions are realistic also in empirical applications.

Table 9.1 then displays the true outcomes under treatment and under control as well as the probabilities for each observational unit to be assigned to treatment condition 1. These are the *fundamental parameters*; all other parameters, such as the associated individual causal effects, the conditional probabilities $P(U{=}u|X{=}x)$, etc. can be computed from these fundamental parameters. Note again that the table does not contain any sample data. It displays the parameters describing the laws of the single-unit trial of sampling a unit, assigning (or observe assigment) to one of the treatment conditions and registering the outcome Y (for more details, see section 2.1).

Verifying Conditional Independence of X and True Outcomes. In the random experiment presented in Table 9.1, the individual treatment probabilities and true outcomes are such that conditional independence of X — the treatment variable — and true outcomes given $Z := sex$ holds. This means that we have to check if Equation (9.13) holds.

We start showing that this equation holds for $X{=}1$, i. e.:

$$P(X{=}1|Z,\tau_1) = P(X{=}1|Z) \quad a.s. \tag{9.18}$$

Since, in our example, Z and τ_1 are discrete, this equation is equivalent to

$$\begin{gathered} P(X{=}1|Z{=}z,\tau_1{=}t) = P(X{=}1|Z{=}z) \\ \text{for all pairs } (z,t) \text{ with } P(Z{=}z,\tau_1 = t) > 0. \end{gathered} \tag{9.19}$$

Table 9.1. Conditional independence of X and true outcomes

Unit u	Sampling probability $P(U=u)$	Covariate sex Z	True outcome variable under control τ_0	True outcome variable under treatment τ_1	True treatment probability $P(X=1\,\vert\,U)$	True causal effect variable $\delta_{10}=\tau_1-\tau_0$	$P(U=u\,\vert\,X=0)$	$P(U=u\,\vert\,X=1)$
u_1	1/6	male	72	83	7/8	11	1/20	7/28
u_2	1/6	male	72	83	5/8	11	3/20	5/28
u_3	1/6	male	95	100	7/8	5	1/20	7/28
u_4	1/6	male	95	100	5/8	5	3/20	5/28
u_5	1/6	female	106	114	2/8	8	6/20	2/28
u_6	1/6	female	116	130	2/8	14	6/20	2/28

$E(\tau_0)$	$E(\tau_1)$	$P(X=1)$	$E(\delta_{10})$
92.667	101.667	7/12	9.000

$E(Y\,\vert\,X=0)$	$E(Y\,\vert\,X=1)$	PFE_{10}
100.000	95.857	−4.143

	male	female
Conditional causal effects	8.000	11.000
Conditional *PFE*s	8.000	11.000

In the next steps, we will use

$$P(X=x\,|\,V=v) \;=\; \sum_u P(X=x\,|\,V=v,U=u)\cdot P(U=u\,|V=v), \tag{9.20}$$

which is always true if U is discrete. Using this equation with Z taking the role of V and considering the case $X=1$ and $Z=m$ yields:

$$\begin{aligned} P(X=1\,|\,Z=m) \;&=\; \sum_u P(X=1\,|\,Z=m,U=u)\cdot P(U=u|Z=m) \\ &=\; 7/8\cdot 1/4+5/8\cdot 1/4+7/8\cdot 1/4+5/8\cdot 1/4 \;=\; 6/8. \end{aligned}$$

(In this case we only have to sum over the first four units displayed in Table 9.1, because, for the two female units, the probabilities $P(U=u|Z=m)$ are zero.) Applying (9.20) to $V=(Z,\tau_1)$ yields the treatment probability

$$P(X{=}1\,|\,Z{=}m,\tau_1{=}83) \;=\; 7/8\cdot 1/2+5/8\cdot 1/2 \;=\; 6/8$$

for $Z{=}m$ and $\tau_1{=}83$ (see the first two rows of Table 9.1). The same result is obtained for $Z{=}m$ and $\tau_1{=}100$, i. e., for rows 3 and 4 of Table 9.1:

$$P(X{=}1\,|\,Z{=}m,\tau_1{=}100) \;=\; 7/8\cdot 1/2+5/8\cdot 1/2 \;=\; 6/8.$$

Hence,

$$\begin{gathered}P(X{=}1\,|\,Z{=}m,\tau_1{=}t) \;=\; P(X{=}1\,|\,Z{=}m)\\ \text{for all values } t \text{ of } \tau_1, \text{ for which } P(Z{=}m,\tau_1{=}t)>0.\end{gathered}$$

The corresponding equation also holds for $Z{=}f$, because in the female subpopulation, the treatment probabilities are constant.

Now we show that Equation (9.13) holds for $X{=}0$, i. e.:

$$P(X{=}0\,|\,Z,\tau_0) \;=\; P(X{=}0\,|\,Z) \quad a.\,s. \tag{9.21}$$

Since Z and τ_0 are discrete, this equation is equivalent to

$$\begin{gathered}P(X{=}0\,|\,Z{=}z,\tau_0{=}t) \;=\; P(X{=}0\,|\,Z{=}z)\\ \text{for all pairs } (z,t) \text{ with } P(Z{=}z,\tau_0{=}t)>0.\end{gathered} \tag{9.22}$$

For $X{=}0$ and $Z{=}m$, Equation (9.20) yields

$$\begin{aligned}P(X{=}0\,|\,Z{=}m) \;&=\; \sum_u P(X{=}0\,|\,Z{=}m,U{=}u)\cdot P(U{=}u|Z{=}m)\\ &=\; 1/8\cdot 1/4+3/8\cdot 1/4+1/8\cdot 1/4+3/8\cdot 1/4 \;=\; 1/4.\end{aligned}$$

Applying (9.20) to $V=(Z,\tau_0)$ yields the treatment probability

$$P(X{=}0\,|\,Z{=}m,\tau_0{=}72) \;=\; 1/8\cdot 1/2+3/8\cdot 1/2 \;=\; 1/4$$

for $Z{=}m$ and $\tau_0{=}72$ (see the first two rows of Table 9.1). The same result is obtained for $Z{=}m$ and $\tau_0{=}95$, i. e., for rows 3 and 4 of Table 9.1:

$$P(X{=}0\,|\,Z{=}m,\tau_0{=}95) \;=\; 1/8\cdot 1/2+3/8\cdot 1/2 \;=\; 1/4.$$

Hence,

$$\begin{gathered}P(X{=}0\,|\,Z{=}m,\tau_0{=}t) \;=\; P(X{=}0\,|\,Z{=}m)\\ \text{for all values } t \text{ of } \tau_0, \text{ for which } P(Z{=}m,\tau_0{=}t)>0.\end{gathered}$$

The corresponding equation also holds for $Z{=}f$, because in the female subpopulation, the treatment probabilities are constant. This proves that Equation (9.21) holds as well.

Since, according to Theorem 9.8, conditional independence of τ_x and $I_{X=x}$ given the covariate Z implies unbiasedness of the regression $E^{X=x}(Y\,|\,Z)$, the conditional prima facie effects are unbiased as well, i. e., $E^{X=1}(Y\,|\,Z)-E^{X=0}(Y\,|\,Z)=averagetotaleffect_{10;Z}$. Therefore, we can also compute the conditional and unconditional average causal effects from the conditional prima facie effects, just in the same way as demonstrated in Example 7.4.

◁

9.2 Regressive Independence

The last condition treated in this chapter, regressive independence, is also the weakest sufficient condition for both biases, the baseline bias and the effect bias, to be zero.[2] 'Weakest' condition means that all other conditions treated in this chapter imply this last one.

9.2.1 Theory

Again, we start with some definitions. Then we show that regressive independence of the true outcomes from X implies unbiasedness. We then illustrate the theory by a simple numerical example in which the conjectured cause is again a treatment variable. Finally, we discuss regressive independence from a substantive point of view.

Definition 9.13 (Regressive Independence of True Outcomes From X)
Let $\langle(\Omega,\mathscr{A},P),(\mathscr{F}_t,t\in T),X,Y\rangle$ be a numerical causality space, let the true outcome variable τ_x exist, and let $I_{X=x}$ denote the indicator of $\{X{=}x\}:=\{\omega\in\Omega\colon X(\omega)=x\}$. Then:

(i) The true outcomes given $X{=}x$ are called regressively independent from $1_{X=x}$, if

$$E(\tau_x|1_{X=x}) \;=\; E(\tau_x). \tag{9.23}$$

We will abbreviate this equation by $\tau_x \vdash 1_{X=x}$.

Let X be discrete with values $x=0,1,\dots J$, let the true outcome variables τ_x, $x=0,1,\dots J$, exist, and let $\tau:=(\tau_0,\tau_1,\dots,\tau_J)$.

(ii) The true outcomes are called regressively independent from X, if

$$E(\tau_x|X) \;=\; E(\tau_x) \qquad \text{for each value } x \text{ of } X. \tag{9.24}$$

We will abbreviate this equation by $\tau \vdash X$.

Remark 9.14 Note that Equation (9.23) is equivalent to

$$E(\tau_x|X{=}x) \;=\; E(\tau_x) \tag{9.25}$$

and that Equation (9.25) implies

$$E(\tau_x|X{\neq}x) \;:=\; E(\tau_x|1_{X=x}{\neq}0) \;=\; E(\tau_x) \tag{9.26}$$

(see Exercise ???). ◁

[2] See, e. g., Equations 11 to 14 of Morgan and Harding (2006) for a formulation of (conditional) regressive independence in terms of Rubin's potential outcomes.

Remark 9.15 The proposition that the true outcome variable τ_x is regressively independent from the indicator variable $1_{X=x}$ for the event $X{=}x$ may sound strange at first sight. Remember, however, that τ_x does not refer to data but to a random experiment to be conducted *in the future*. If, e. g., X represents a treatment variable and we assume

$$E(Y|X,c_X) = E(Y|X,U) \quad a.s.,$$

this implies $\tau_x = E^{X=x}(Y|U)$ for each treatment x. Hence, the values of τ_x represent an *unknown attribute of the units*, namely their true outcome if they *would* be in treatment x. Hence, the expectation of this unknown attribute will not depend on $1_{X=x}$, if the individual probabilities to be in treatment x do not depend on whether or not the unit has a high score on the true outcome under treatment x. In other words, Equation (9.25) will hold if assignment of the units to treatment x do not depend on their values on the true outcome variable τ_x. Of course, formulating it this way — treatment assignment should not depend on an attribute of the unit — is the more natural way of thinking. However, looking for the logically weakest condition for both biases to vanish, we have to consider Equation (9.25). The same arguments apply to regressive independence of true outcomes given all values x of X from the treatment variable X [see Eq. (9.24)]. ◁

Using the abbreviations introduced in Definition 9.13, we can now formulate the following theorem:

Theorem 9.16 (Unbiasedness)

(i) $\tau_x \vdash 1_{X=x}$ *implies that* $E(Y|X{=}x)$ *is unbiased*
(ii) $\tau_x \vdash X$ *for each value* x *of* X *implies that the following propositions hold as well:*

(a) The treatment regression $E(Y|X)$ *is unbiased.*
(b) Both biases, the baseline bias and the effect bias, are zero for all PFEs .
(c) All prima facie effects are unbiased.

(Proof p. 234)

We now consider the generalization of regressive independence and the associated theorem to conditioning on a covariate.

Definition 9.17 (Conditional Regressive Independence) *Let* $\langle(\Omega,\mathscr{A},P),(\mathscr{F}_t,t\in T),X,Y\rangle$ *be a numerical causality space, let the true outcome variable* τ_x *exist, let* $I_{X=x}$ *denote the indicator of* $\{X{=}x\} := \{\omega \in \Omega\colon X(\omega) = x\}$*, and let* Z *be a covariate, i. e., let* Z *be measurable w.r.t.* c_X*. Then:*

(i) The true outcomes given $X{=}x$ *are called* Z*-conditionally regressively independent from* $1_{X=x}$*, if*

$$E(\tau_x | 1_{X=x}, Z) = E(\tau_x | Z) \quad a.s. \tag{9.27}$$

This equation will be abbreviated by $\tau_x \vdash 1_{X=x} | Z$.

Let X *be discrete with values* $x = 0, 1, \ldots, J$, *let the true outcome variables* τ_x, $x = 0, 1, \ldots, J$, *exist, and let* $\tau := (\tau_0, \tau_1, \ldots, \tau_J)$.

(ii) The true outcomes are called Z*-conditionally regressively independent from* X, *if*

$$E(\tau_x | X, Z) = E(\tau_x | Z) \quad a.s. \qquad \text{for each value } x \text{ of } X. \tag{9.28}$$

We will abbreviate this equation by $\tau \vdash X | Z$.

Remark 9.18 Equation (9.27) is equivalent to

$$E^{X=x}(\tau_x | Z) = E(\tau_x | Z) \quad a.s. \tag{9.29}$$

(see Exercise ???). Equation (9.28) is equivalent to

$$E^{X=x'}(\tau_x | Z) = E(\tau_x | Z) \quad a.s. \qquad \text{for all pairs } x, x' = 0, 1, \ldots, J. \tag{9.30}$$

◁

Theorem 9.19 (Conditional Unbiasedness) *Let the presumptions of Definition 9.17 hold true. Then:*

(i) $\tau_x \vdash 1_{X=x} | Z$ *implies that* $E^{X=x}(Y | Z)$ *is unbiased.*

Let X *be discrete with values* $x = 0, 1, \ldots, J$, *let the true outcome variables* τ_x, $x = 0, 1, \ldots, J$, *exist, let* $\tau := (\tau_0, \tau_1, \ldots, \tau_J)$, *and let* $E^{X=x}(Y | Z)$ *denote the extension of the conditional expectation* $E^{X=x}(Y | Z)$.

(ii) $\tau \vdash X | Z$ *implies:*
(a) The regression $E(Y | X, Z)$ *is* Z*-conditionally unbiased.*
(b) Both biases of the Z*-conditional PFE-functions* $PFE_{xx';Z}$ *are zero. Hence, the* Z*-conditional PFE-functions are unbiased.*
(c) For each value x *of* X*:*

$$E\big[E^{X=x}(Y | Z) \,\big|\, X \big] = E(\tau_x | X). \tag{9.31}$$

(d) For each pair $x, x' = 0, 1, \ldots, J$, $x \neq x'$ *:*

$$E(PFE_{xx';Z} | X) = E(\delta_{xx'} | X) = E(deltajkatom | X). \tag{9.32}$$

(Proof p. 235)

Remark 9.20 Propositions *(ii c)* and *(ii d)* are new. Equation (9.32) shows how to identify the X-conditional causal-effect functions from the Z-conditional prima facie effect functions $PFE_{xx';Z}$ (see section 5.2.3). Equation (9.32) implies that the conditional causal effect of $X{=}x$ compared to $X{=}x'$ given $X{=}x^*$ is equal to the $(X{=}x^*)$-conditional expectation $E(PFE_{xx';Z} \,|\, X{=}x^*)$ of the *PFE*-function $PFE_{xx';Z}$, i. e., for each pair (x,x'), $x \neq x'$, $x,x' = 0,1,\ldots,J$:

$$\begin{aligned} E(PFE_{xx';Z} \,|\, X{=}x^*) \;&=\; E\Big[E^{X=x}(Y|Z) - E^{X=x'}(Y|Z)\,\Big|\, X{=}x^*\Big] \\ &=\; E(\delta_{xx'} \,|\, X{=}x^*) \;=\; CTE_{xx';X=x^*}. \end{aligned} \tag{9.33}$$

If, for example, X represents a treatment variable and there are two treatment conditions, $X{=}0$ and $X{=}1$, then this equation can be used to compute the conditional causal effect $CTE_{10;X=1}$ given $X{=}1$, but also the conditional causal effect $CTE_{10;X=0}$ given $X{=}0$ (see section 5.2.3 for more details and the last paragraph of section 9.2.3 for an example). ◁

9.2.2 Substantive Meaning

Theorems 9.16 and 9.19 are important for two reasons: *First*, (conditional) regressive independence of the true outcomes from X is the logically weakest sufficient condition for both biases to vanish among the four conditions treated in this chapter. As will be shown in section 9.3, all other sufficient conditions treated in this chapter imply (conditional) regressive independence. *Second*, if conditional regressive independence holds in an application for a given (possibly multivariate) covariate Z, we can estimate the conditional expectations $E(Y|X{=}x, Z{=}z)$ and their differences, the conditional prima facie effects $PFE_{xx';Z=z} = E(Y|X{=}x, Z{=}z) - E(Y|X{=}x', Z{=}z)$, and then use these statistics to estimate all kinds of conditional and unconditional causal effects introduced in chapter 5: the (unconditional) average causal effects $E(\delta_{xx'})$, and both kinds of conditional causal effects, i. e., the conditional causal effects $E(\delta_{xx'} \,|\, Z{=}z)$ given a value z of a covariate, and the conditional causal effects $CTE_{10;X=x^*}$ given $X{=}x^*$. This will be exemplified in example 9.2.3.

9.2.3 Numerical Example

Example 9.21 We again assume that

$$E(Y|X,c_X) \;=\; E(Y|X,U) \quad a.s. \tag{9.34}$$

and

$$P(X{=}x|c_X) \;\underset{P}{=}\; P(X{=}x|U) \qquad \text{for each value } x \text{ of } X \tag{9.35}$$

hold for the true outcome variables τ_x and the true treatment probabilities, respectively. As mentioned before, if U represents the observational units at the onset of treatment, these assumptions are realistic also in empirical applications. ◁

Table 9.2. Conditional regressive independence

Unit u	Sampling probability $P(U=u)$	Covariate sex Z	True outcome variable under control τ_0	True outcome variable under treatment τ_1	True treatment probability $P(X=1\,\|\,U)$	True causal effect variable $\delta_{10}=\tau_1-\tau_0$	$P(U=u\,\|X=0)$	$P(U=u\,\|X=1)$
u_1	1/6	male	75	87	16/20	12	4/50	16/70
u_2	1/6	male	70	80	14/20	10	6/50	14/70
u_3	1/6	male	93	97	16/20	4	4/50	16/70
u_4	1/6	male	98	104	14/20	6	6/50	14/70
u_5	1/6	female	106	114	5/20	8	15/50	5/70
u_6	1/6	female	116	130	5/20	14	15/50	5/70

$E(\tau_0)$	$E(\tau_1)$	$P(X=1)$	$E(\delta_{10})$
93.000	102.000	7/12	9.000

$E(Y\,\|X=0)$	$E(Y\,\|X=1)$	PFE_{10}
100.200	96.286	−3.914

	male	female
Conditional causal effects	8.000	11.000
Conditional *PFE*s	8.000	11.000

Table 9.2 presents an example for *conditional regressive independence* given a covariate Z. Under the assumptions (9.34) and (9.35), this table displays the true outcomes under treatment and under control as well as the probabilities for each observational unit to be assigned to treatment condition $X=1$. These are the *fundamental parameters*; all other parameters, such as associated individual causal effects, the conditional probabilities $P(U=u\,|\,X=x)$, etc. can be computed from these fundamental parameters. The table also displays the values of the covariate $Z :=$ *sex*. Note again that the table does not contain any sample data. It displays the parameters describing the laws of the single-unit trial.

Verifying Conditional Regressive Independence. In this random experiment, the true treatment probabilities and true outcomes are such that conditional regressive independence ($\tau \vdash X\,|\,Z$) holds, i. e.:

$$E(\tau_0\,|X,Z) \;=\; E(\tau_0\,|Z) \qquad \text{and} \qquad E(\tau_1\,|X,Z) \;=\; E(\tau_1\,|Z). \tag{9.36}$$

Table 9.3. Conditional probabilities $P(U=u|X=x,Z=z)$ for Table 9.2

Unit	$X=0, Z=m$	$X=1, Z=m$	$X=0, Z=f$	$X=1, Z=f$
u_1	2/10	8/30	0	0
u_2	3/10	7/30	0	0
u_3	2/10	8/30	0	0
u_4	3/10	7/30	0	0
u_5	0	0	1/2	1/2
u_6	0	0	1/2	1/2

Note that, in this example, *none of the other sufficient conditions for unbiasedness* treated in this chapter *holds*.

In order to verify that $\tau \vdash X \,|\, Z$ holds, consider

$$E(\tau_x|X=x^*,Z=z) \;=\; \sum_u E(Y|X=x,U=u)\cdot P(U=u|X=x^*,Z=z), \tag{9.37}$$

which is true if Assumption (9.34) holds and $P(X=x^*,Z=z)>0$ (see Exercise ???). Note that the indices i and x may be different in this equation, although both are values of the treatment variable X. If we want to apply this equation to the example in Table 9.2, we have to use the probabilities $P(U=u|X=x^*,Z=z)$ displayed in Table 9.3 (see Exercise 6-10) ???.

Using Equation (9.37), we can now compute

$$\begin{aligned} E(\tau_0|X=1,Z=m) \;&=\; \sum_u E(Y|X=0,U=u)\cdot P(U=u|X=1,Z=m) \\ &=\; 75\cdot 8/30+70\cdot 7/30+93\cdot 8/30+98\cdot 7/30 \;=\; 84. \end{aligned}$$

This is exactly the same result as obtained for

$$E(\tau_0|Z=m) \;=\; \sum_u E(Y|X=0,U=u)\cdot P(U=u|Z=m) \;=\; (75+70+93+98)\cdot 1/4 \;=\; 84.$$

Hence, $E(\tau_0|X=0,Z=m) = E(\tau_0|Z=m)$, and the same can be shown for all other combinations of values of X and Z. Since the corresponding property also holds for τ_1 [see Eq. (9.36)], the true outcomes in Table 9.2 are conditionally regressively independent from the treatment variable X given Z.

Computing the Average Causal Effects Given the Value of the Covariate. Since, in this example, conditional regressive independence [Eq. (9.28)] holds, Theorem 9.19 implies that the *conditional PFE*s are identical to the *conditional* causal effects *given a value of the covariate*, although, in this example, the *unconditional PFE*s and *ATE*s are not identical. Hence, once the conditional expectations $E(Y|X=x,Z=z)$ are known for all four combinations of values of X and Z, we can

also compute the conditional causal effects given the values of the covariate, i. e., the conditional causal effects within the sex subpopulations.

In empirical applications, the conditional expectations $E(Y|X{=}x, Z{=}z)$ are easy to estimate provided that we have some observations of the outcome variable Y within each pair (x,z) of values of X and Z. Computing them from the true outcomes and the true treatment probabilities displayed in Table 9.2 is also possible. Although the true outcomes are not available in empirical applications, let us illustrate the formula by our example. According to the equation

$$E(Y|X{=}x, Z{=}z) \;=\; \sum_u E(Y|X{=}x, U{=}u) \cdot P(U{=}u|X{=}x, Z{=}z) \tag{9.38}$$

already treated in chapter 6 [see Eq. (iii) in Box 6.1], we need the conditional probabilities $P(U{=}u|X{=}x, Z{=}z)$ for all units and all combinations of values of X and Z displayed in Table 9.3.

For $X{=}0$ and $Z{=}m$, Equation (9.38) yields:

$$\begin{aligned} E(Y|X{=}0, Z{=}m) \;&=\; \sum_u E(Y|X{=}0, U{=}u) \cdot P(U{=}u|X{=}0, Z{=}m) \\ &=\; (75+70+93+98) \cdot 1/4 \;=\; 84. \end{aligned}$$

For the other combinations of values of X and Z we receive $E(Y|X{=}1, Z{=}m) = 92$, $E(Y|X{=}0, Z{=}f) = 111$, and $E(Y|X{=}1, Z{=}f) = 122$.

Taking differences between these conditional expectations yields the conditional prima facie effects given the value of a covariate which are, according to Theorem 9.16, also the average causal effects in the two subpopulations:

$$\begin{aligned} CTE_{10;Z=m} \;&=\; PFE_{10;Z=m} \\ &=\; E(Y|X{=}1, Z{=}m) - E(Y|X{=}0, Z{=}m) \\ &=\; 92-84 \;=\; 8 \end{aligned}$$

and

$$\begin{aligned} CTE_{10;Z=f} \;&=\; PFE_{10;Z=f} \\ &=\; E(Y|X{=}1, Z{=}f) - E(Y|X{=}0, Z{=}f) \\ &=\; 122-111 \;=\; 11. \end{aligned}$$

Average Causal Effect. Since Z-conditional regressive independence holds in this example, which implies that $E(Y|X,Z)$ is Z-conditionally unbiased, we can also use Equation (**??**) in Box **??** to compute the average causal effect from the conditional causal effects in the two subpopulations:

$$averagetotaleffect_{10} \;=\; E(CTE_{10;Z}) \;=\; 8 \cdot 4/6 + 11 \cdot 2/6 \;=\; 9.$$

Box 9.1 A Summary of the Causality Conditions

We treated the following causality conditions for the conditional expectation $E(Y|X)$, referring to the numerical causality space $\langle(\Omega,\mathscr{A},P),(\mathscr{F}_t,t\in T),X,Y\rangle$.

(1) Independence of X and all covariates (short: $X \perp\!\!\!\perp c_X$), which implies, and if X is discrete, is equivalent to

$$P(X{=}x|c_X) = P(X{=}x) \quad a.s. \qquad \text{for each value } x \text{ of } X. \tag{i}$$

(2) Homogeneity of $E(Y|X)$ (short: $Y \vdash c_X|X$), defined by

$$E(Y|X,c_X) = E(Y|X) \quad a.s. \tag{ii}$$

Both causality conditions do not need any further assumptions. However, if we make the *additional assumptions* (a) that X is discrete with values $x = 0,1,\ldots,J$, (b) the true outcome variables τ_x exist, and (c) $\tau := (\tau_0,\tau_1,\ldots,\tau_J)$, then (1), (2), and each of the following two conditions imply unbiasedness of the conditional expectation $E(Y|X)$.

(3) Independence of X and true outcome variables ($X \perp\!\!\!\perp \tau$), which implies, and if X is discrete, is equivalent to

$$P(X{=}x|\tau) = P(X{=}x) \quad a.s. \qquad \text{for each value } x \text{ of } X. \tag{iii}$$

(4) Regressive independence of true outcome variables from X ($\tau \vdash X$), defined by

$$E(\tau_x|X) = E(\tau_x) \quad a.s. \qquad \text{for each value } x \text{ of } X. \tag{iv}$$

Let Z denote a covariate, i. e., let Z be measurable w.r.t. c_X. We treated the following causality conditions for the conditional expectation $E(Y|X,Z)$:

(5) Independence of X and covariates given Z (short: $X \perp\!\!\!\perp c_X|Z$), which implies, and if X is discrete, is equivalent to

$$P(X{=}x|c_X) = P(X{=}x|Z) \quad a.s. \qquad \text{for each value } x \text{ of } X. \tag{v}$$

(6) Z-conditional homogeneity of the regression $E(Y|X,Z)$ (short: $Y \vdash c_X|X,Z$), defined by

$$E(Y|X,c_X) = E(Y|X,Z) \quad a.s. \tag{vi}$$

Again, these two causality conditions need no further assumptions. Under the additional assumptions (a) to (c) mentioned above, (5), (6), and each of the following conditions imply Z-conditional unbiasedness of the regression $E(Y|X,Z)$:

(7) Independence of X and true outcome variables given Z (short: $X \perp\!\!\!\perp \tau|Z$), which implies, and if X is discrete, is equivalent to

$$P(X{=}x|Z,\tau) = P(X{=}x|Z) \quad a.s. \qquad \text{for each value } x \text{ of } X. \tag{vii}$$

(8) Conditional regressive independence of true outcome variables from X given Z (short: $\tau \vdash X|Z$), which is defined by

$$E(\tau_x|X,Z) = E(\tau_x|Z) \quad a.s. \qquad \text{for each value } x \text{ of } X. \tag{viii}$$

Table 9.4. Implication structure between causality conditions

		(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
(i)	$U \perp\!\!\!\perp X$			$\Rightarrow$	$\Rightarrow$				
(ii)	$Y \vdash U \mid X$			$\Rightarrow$	$\Rightarrow$		$\Rightarrow$		
(iii)	$\tau \perp\!\!\!\perp X$				$\Rightarrow$				
(iv)	$\tau \vdash X$								
(v)	$U \perp\!\!\!\perp X \mid Z$							$\Rightarrow$	$\Rightarrow$
(vi)	$Y \vdash U \mid X, Z$							$\Rightarrow$	$\Rightarrow$
(vii)	$\tau \perp\!\!\!\perp X \mid Z$								$\Rightarrow$
(viii)	$\tau \vdash X \mid Z$								

Note: The symbol $\Rightarrow$ indicates that the condition in the row implies the condition in the column. Conditions (i) to (iv) imply unbiasedness of $E(Y \mid X)$, and Conditions (v) to (viii) imply Z-conditional unbiasedness of $E(Y \mid X, Z)$. For definitions of conditions (i) to (viii) see Box 9.1.

9.3 Implications Between Causality Conditions

Table 9.4 displays the implications between the causality conditions treated in this chapter (see Box 9.1 for a summary and the abbreviations used). The proofs for the propositions summarized in Table 9.4 are given in Exercise 9-16. Hence, starting with the unconditional case [i. e., conditions (i) to (iv)], we can see that stochastic independence of X and covariates ($X \perp\!\!\!\perp c_X$) is the strongest condition: It implies stochastic independence of X and the true outcomes ($X \perp\!\!\!\perp \tau$) as well as regressive independence of true outcomes from X ($\tau \vdash X$). However, independence of X and covariates does not imply homogeneity ($Y \vdash c_X \mid X$), which itself implies stochastic independence of X and true outcomes ($X \perp\!\!\!\perp \tau$), regressive independence of true outcomes from X ($\tau \vdash X$), and Z-conditional homogeneity ($Y \vdash c_X \mid X, Z$). Last but not least, stochastic independence of X and true outcomes ($X \perp\!\!\!\perp \tau$) implies regressive independence of true outcomes from X ($\tau \vdash X$). The implication structure in the conditional case is similar (see Table 9.4).

Theorem 9.22

(i) Each of the conditions (1) to (4) in Box 9.1 implies Propositions (a) to (c) in Theorem 9.16.

(ii) Each of the conditions (5) to (8) in Box 9.1 implies Propositions (a) to (d) in Theorem 9.19.

(Proof p. 236)

9.4 Summary and Conclusions

In the previous chapter we showed that unbiasedness of the conditional or unconditional prima facie effects (*PFE*s) is crucial for estimating conditional and unconditional causal effects. In this chapter, we studied four causality conditions that imply unbiasedness of the conditional and unconditional *PFE*s , both in their unconditional and their conditional versions. The first two of these causality conditions, the *independent cause* and the *sufficient cause* conditions, are defined even in those cases in which the true outcome variables do not exist such that unbiasedness and even causal effects are not defined.

In experiments, the independent cause condition can be created by the experimenter via random assignment of the observational unit to one of the treatment conditions represented by X. Outside the randomized experiment, and even in cases in which X does not denote a treatment variable, conditional independence of X and covariates given Z may also hold if the covariate Z is carefully selected. In this case, however, there is no *guarantee* that conditional independence of X and covariates will hold.

In contrast to the conditional independent cause condition, the conditional sufficient cause condition cannot be created by experimental design techniques such as randomization, but — just like Z-conditional independence — it might be true for a carefully chosen covariate. The third condition, *(Z-conditional) independence of X and true outcomes* ("strong ignorability" in Rubin's terms) , can also be created by (conditional) randomization. It is implied by the first condition and it implies itself the fourth condition, *(Z-conditional) regressive independence* of true outcomes from X. This condition is of specific theoretical interest, because all other conditions treated in this chapter imply this last one. Furthermore, this last condition implies that both biases, the baseline bias and the effect bias, are zero. Hence, we can conclude that the first three conditions also imply that both biases are zero.

9.5 Proofs

Proof of Theorem 9.5

Propositions (i) to (iv) are immediate consequences of Theorem 14.64.

Proof of Theorem 9.8

Again, we only have to prove (ii), because (i) is a particular case in which Z is a constant. Hence, we have to prove

$$E^{X=x}(Y|Z) \;=\; E(\tau_x|Z)$$

However, this equation directly follows from proposition (ii) of Theorem **??** .

Proof of Corollary 9.9

Since $\tau \perp\!\!\!\perp X \,|\, Z$ implies $\tau_x \perp\!\!\!\perp I_{X=x} \,|\, Z$ for each value x of X and Z-conditional unbiasedness of $E(Y|X,Z)$ is equivalent to unbiasedness of $E^{X=x}(Y|Z)$ for each value x of X, Theorem 9.8 implies this corollary.

Proof of Theorem 9.16

For proving (i) consider

$$\begin{aligned}
E(Y|X{=}x) &= E^{X=x}(Y) \qquad \text{[Lemma \textbf{??}]} \\
&= E^{X=x}\big[E^{X=x}(Y\,|\,c_X)\big] \qquad \text{[Rule (iv) in Box \textbf{??}]} \\
&= E^{X=x}\big[E^{X=x}(Y\,|\,c_X)\big] \qquad \text{[Remark \textbf{??}]} \\
&= E^{X=x}(\tau_x) \qquad \text{[def. of } \tau_x\text{]} \\
&= E(\tau_x|X{=}x) \qquad \text{[Lemma \textbf{??}]} \\
&= E(\tau_x) \qquad \text{[Eq. (9.25)]}.
\end{aligned}$$

We now prove the propositions summarized in (ii). Proposition (ii a) immediately follows from Proposition (i), because Equation (9.24) implies Equation (9.25) for each value x of X and unbiasedness of the regression $E(Y|X)$ is equivalent to unbiasedness of the expectation $E(Y|X{=}x)$ for each value x of X.

Equation (9.24) implies that the baseline-bias

$$\textit{baseline bias}_{xx'} \;:=\; E(\tau_{x'}|X{=}x) - E(\tau_{x'}|X{=}x') \tag{9.39}$$

is zero, because Equation (9.24) is equivalent to

$$E(\tau_{x'}|X{=}x) \;=\; E(\tau_{x'}) \qquad \text{for each pair } x, x' = 0, 1, \ldots J. \tag{9.40}$$

This equation implies that both terms on the right-hand side of Equation (9.39) are identical.

Similarly, Equation (9.24) also implies that the effect-bias

$$\textit{effect bias}_{xx'} \;:=\; E(\delta_{xx'}|X{=}x) - E(\delta_{xx'}) \tag{9.41}$$

is zero, because, if the regression of the true outcome variables τ_x and $\tau_{x'}$ do not depend on X, then this will also be true for their differences $\delta_{xx'} = \tau_x - \tau_{x'}$, i. e.:

$$E(\delta_{xx'}\,|\,X) \;=\; E(\delta_{xx'}) \qquad \text{for all pairs } x, x' = 0, 1, \ldots J. \tag{9.42}$$

Since $E(\delta_{xx'}) := E(\delta_{xx'})$ and (9.42) is equivalent to

$$E(\delta_{xx'}\,|\,X{=}x^*) \;=\; E(\delta_{xx'}) \qquad \text{for all triples } x^*, x, x' = 0, 1, \ldots J,$$

both terms on the right-hand side of Equation (9.41) are identical, which proves Proposition (ii b). Now Proposition (ii c) immediately follows from Proposition (i) of Theorem 6.27, the Bias Theorem.

Proof of Theorem 9.19

According to Remark 9.18, Equation (9.27) — and therefore $\tau_x \vdash 1_{X=x} \,|\, Z$ — is equivalent to Equation (9.29). Furthermore, we assumed that the extension $E^{X=x}(Y\,|\,\mathscr{F}_{t_X}))$ exists, which implies that $E^{X=x}(Y\,|\,Z)$ exists as well. Hence,

$$\begin{aligned} E^{X=x}(Y\,|\,Z) &= E^{X=x}\big[\,E^{X=x}(Y\,|\,\mathscr{F}_{t_X}))\,\big|\,Z\big] \qquad \text{[Rule (??) in Box ??]}. \\ &= E^{X=x}(\tau_x\,|Z) \qquad \text{[def. of } \tau_x\text{]} \\ &= E(\tau_x\,|Z) \quad \textit{a.s.} \qquad \text{[Eq. (9.29)]}. \end{aligned}$$

This proves Proposition (i).

Equation (9.28) implies that the baseline bias function

$$\textit{baseline bias}_{xx';Z} := E^{X=x}(\tau_{x'}\,|Z) - E^{X=x'}(\tau_{x'}\,|Z) \quad \textit{a.s.} \tag{9.43}$$

is zero, because Equation (9.28) is equivalent to (9.30) and because this equation implies that both terms on the right-hand side of Equation (9.43) are identical.

Similarly, Equation (9.28) also implies that the effect bias function

$$\textit{effect bias}_{xx';Z} := E^{X=x}(\delta_{xx'}\,|\,Z) - E(\delta_{xx'}\,|\,Z) \quad \textit{a.s.} \tag{9.44}$$

is zero, because Equation (9.28) implies Equation (9.30) (see Remark 9.18), which itself implies

$$E^{X=x^*}(\delta_{xx'}\,|\,Z) \;=\; E(\delta_{xx'}\,|\,Z) \quad \textit{a.s.} \qquad \text{for all triples } x^*, x, x' = 0, 1, \ldots, J. \tag{9.45}$$

This equation and $E(\delta_{xx'}\,|\,Z) := E(\delta_{xx'}\,|\,Z)$ that both terms on the right-hand side of Equation (9.44) are identical, which proves Proposition (ii b).

For proving Proposition (ii c), consider

$$\begin{aligned} E\big[\,E^{X=x}(Y\,|\,Z)\,\big|\,X\big] &= E\big[E(\tau_x\,|Z)\,\big|\,X\big] \qquad \text{[(i)]} \\ &= E\big[E(\tau_x\,|X,Z)\,\big|\,X\big] \qquad \text{[Eq. (9.28)]} \\ &= E(\tau_x\,|X) \quad \textit{a.s.} \qquad \text{[Rule (??) in Box ??]}. \end{aligned}$$

Finally, for proving Proposition (ii d), consider

$$\begin{aligned} E(PFE_{xx';Z}\,|\,X) &= E\big[E(\delta_{xx'}\,|\,Z)\,\big|\,X\big] \qquad \text{[(ii b)]} \\ &= E\big[E(\tau_x - \tau_{x'}\,|Z)\,\big|\,X\big] \qquad \text{[def. of } \delta_{xx'}\text{]} \\ &= E\big[E(\tau_x\,|Z) - E(\tau_{x'}\,|Z)\,\big|\,X\big] \qquad \text{[Rule (??) in Box ??]} \\ &= E\big[E(\tau_x\,|X,Z) - E(\tau_{x'}\,|X,Z)\,\big|\,X\big] \qquad \text{[Eq. (9.28)]} \\ &= E\big[E(\tau_x\,|X,Z)\,\big|\,X\big] - E(\tau_{x'}\,|X,Z)\,\big|\,X\big] \qquad \text{[Rule (??) in Box ??]} \\ &= E(\tau_x\,|X) - E(\tau_{x'}\,|X) \qquad \text{[Rule (??) in Box ??]} \\ &= E(\delta_{xx'}\,|X) \qquad \text{[def. of } \delta_{xx'}\text{]} \\ &= E(deltajkatom\,|X) \quad \textit{a.s.} \qquad \text{[def. of } E(deltajkatom\,|X)\text{]}. \end{aligned}$$

Proof of Theorem 9.22

This theorem follows from Theorems 9.16 and 9.19 and the propositions summarized Table 9.4, the proofs of which are found in Exercise 9-16.

9.6 Exercises

▷ **Exercise 9-1** Which of the sufficient conditions for unbiasedness of the prima facie effects treated in this chapter can be created by random assignment of units to treatment conditions?

▷ **Exercise 9-2** Describe random assignment of a unit to one of two treatment conditions!

▷ **Exercise 9-3** Describe *conditional* random assignment of a unit to one of two treatment conditions given a covariate!

▷ **Exercise 9-4** Which of the sufficient conditions for unbiasedness of the conditional prima facie effects treated in this chapter can be created by conditional random assignment of units to treatment conditions given a covariate?

▷ **Exercise 9-5** Shown that ξ occurring in Equation (**??**) is a linear function of the *logit* variable

$$logit := ln\left(\frac{P(Y_i=1\,|\,U)}{1-P(Y_i=1\,|\,U)}\right).$$

▷ **Exercise 9-6** Which are the sufficient conditions for unbiasedness of the unconditional and the conditional prima facie effects treated in this chapter? Describe them in words and define them mathematically!

▷ **Exercise 9-7** Which is the probability $P(X{=}0\,|\,U{=}u_1)$ in Table 9.2?

▷ **Exercise 9-8** Show that $E\big[P(X{=}x\,|\,U)\big] = P(X{=}x)$.

▷ **Exercise 9-9** Compute $E\big[P(X{=}1\,|\,U)\big] = P(X{=}1)$ for the example in Table 9.2.

▷ **Exercise 9-10** Compute the probabilities $P(U{=}u\,|X{=}1, Z{=}m)$ displayed in Table 9.3.

▷ **Exercise 9-11** Show that Equation (**??**) implies $E(Y\,|\,X,Z,W) = E(Y\,|\,X,Z)$ if both Z and W are covariates, i. e., if both Z and W are measurable w.r.t. the covariate projection c_X.

▷ **Exercise 9-12** Which of the sufficient conditions for unbiasedness is logically implied by which other condition?

▷ **Exercise 9-13** Show that $U \perp\!\!\!\perp X$ implies $U \perp\!\!\!\perp X\,|\,Z$, if Z is a measurable function of U.

▷ **Exercise 9-14** Compute the probability $P(Z{=}m\,|\,X{=}0)$ for the example in Table 9.2.

▷ **Exercise 9-15** Show that independence of X and covariates implies:

$$P(X{=}x\,|\,W) \;=\; P(X{=}x) \qquad \text{for each } x = 0,1,\ldots,J,$$

for each covariate W.

▷ **Exercise 9-16** Proof the propositions on the implication structure between the conditions for unbiasedness displayed in Table 9.4.

Solutions

▷ **Solution 9-1** Random assignment of units to treatment conditions assures: (a) independence of X and covariates ($X \perp\!\!\!\perp c_X$), (b) independence of X and true outcomes ($X \perp\!\!\!\perp \tau$), and (c) regressive independence of true outcomes from X ($\tau \vdash X$).

▷ **Solution 9-2** The unit is assigned to one of the treatment conditions by coin toss, for instance. This makes sure that $P(X{=}1 \,|\, c_X) = P(X{=}1)$, i. e., that the true treatment probabilities are constant and do not depend on any covariate. On the level of the observational units, $P(X{=}1 \,|\, c_X) = P(X{=}1)$ implies that each unit has the same probability $P(X{=}1 \,|\, U{=}u) = P(X{=}1)$ to be in treatment 1, which implies that each unit has the same probability $P(X{=}0)$ to be assigned to treatment 0 as well.

▷ **Solution 9-3** In a single-unit trial, in which a unit u is sampled and a value z of the covariate Z is assessed before the unit is assigned to one of the treatment conditions (see section **??**), *conditional random assignment* of a unit to one of the two treatment conditions refers to assigning the unit u to treatment condition 1 with probability $P_{Z=z}(X{=}1 \,|\, c_X) = P_{Z=z}(X{=}1) = P(X{=}1 \,|\, Z{=}z)$. Among other things, this makes sure that all units with the same value z of the covariate Z have the same probability to be assigned to treatment 1. This assignment procedure allows for different treatment probabilities for units with different values z of the covariate Z.

▷ **Solution 9-4** Conditional random assignment of units to treatment conditions given a covariate Z creates the following conditions: (a) conditional independence of X and covariates, i. e., $X \perp\!\!\!\perp c_X \,|\, Z$, (b) conditional independence of X and true outcomes, i. e., $X \perp\!\!\!\perp \tau \,|\, Z$, and (c) conditional regressive independence of true outcomes from X, i. e., $\tau \vdash X \,|\, Z$.

▷ **Solution 9-5** Fehlt noch !!!

▷ **Solution 9-6** The answers are summarized in Box 9.1.

▷ **Solution 9-7** This probability is: $P(X{=}0 \,|\, U{=}u_1) = 1 - P(X{=}1 \,|\, U{=}u_1) = 1 - 16/20 = 4/20$. Note that only $P(X{=}1 \,|\, U{=}u_1)$ is displayed in the table.

▷ **Solution 9-8** By definition of the expectation,

$$E\big[P(X{=}x \,|\, U)\big] \;=\; \sum_u P(X{=}x \,|\, U{=}u) \cdot P(U{=}u).$$

Since, according to the theorem of total probability

$$P(X{=}x) \;=\; \sum_u P(X{=}x \,|\, U{=}u) \cdot P(U{=}u)$$

(see Box **??**), this shows that $E\big[P(X{=}x|U)\big] = P(X{=}x)$. Note that $E\big[P(X{=}x|U)\big] = P(X{=}x)$ is also a particular case of Rule (iv) in appendix **??**, if we use $P(X{=}x \,|\, U) = E(1_{X=x} \,|\, U)$, where $1_{X=x}$ denotes the indicator variable indicating with values 1 and 0 whether or not X takes on the value x.

▷ **Solution 9-9**

$$\begin{aligned} P(X{=}1) \;&=\; E\big[P(X{=}1 \,|\, U)\big] \;=\; \sum_u P(X{=}1 \,|\, U{=}u) \cdot P(U{=}u) \\ &=\; (16/20 + 14/20 + 16/20 + 14/20 + 5/20 + 5/20) \cdot 1/6 \;=\; 7/12. \end{aligned}$$

▷ **Solution 9-10** The general formula needed for these computations is

$$P(U{=}u\,|X{=}x,Z{=}z) \;=\; \frac{P(X{=}x\,|\,U{=}u,Z{=}z)\cdot P(U{=}u,Z{=}z)}{P(X{=}x,Z{=}z)}, \tag{9.46}$$

where

$$P(X{=}x,Z{=}z) \;=\; \sum_u P(X{=}x\,|\,U{=}u,Z{=}z)\cdot P(U{=}u,Z{=}z) \tag{9.47}$$

(see Exercise 6-16.)

Let us start with $Z{=}m$ and $X{=}1$. For this case, Equation (9.47) yields

$$\begin{aligned} P(X{=}1,Z{=}m) \;&=\; \sum_u P(X{=}1\,|\,U{=}u,Z{=}m)\cdot P(U{=}u,Z{=}m) \\ &=\; (16/20+14/20+16/20+14/20)\cdot 1/6 \;=\; 1/2, \end{aligned}$$

and for u_1 we obtain from Equation (9.46):

$$\begin{aligned} P(U{=}u_1\,|\,X{=}1,Z{=}m) \;&=\; \frac{P(X{=}1\,|\,U{=}u_1,Z{=}m)\cdot P(U{=}u_1,Z{=}m)}{P(X{=}1,Z{=}m)} \\ &=\; \frac{16/20\cdot 1/6}{1/2} \;=\; 8/30. \end{aligned}$$

Computing the corresponding probabilities for the other units, we receive 7/30, 8/30, and 7/30 for u_2, u_3, and u_4, respectively. The corresponding probabilities for u_5 and u_6 are 0.

▷ **Solution 9-11**

$$\begin{aligned} E(Y\,|\,X,Z,W) \;&=\; E\big[E(Y\,|\,X,Z,c_X)\,\big|\,X,Z,W\big] && \text{[Rule (??) in Box ??]} \\ &=\; E\big[E(Y\,|\,X,c_X)\,\big|\,X,Z,W\big] && \text{[measurability of } Z \text{ w. r. to } c_X\text{]} \\ &=\; E\big[E(Y\,|\,X,Z)\,\big|\,X,Z,W\big] && \text{[Eq. (??)]} \\ &=\; E(Y\,|\,X,Z) && \text{[Rule (??) in Box ??]}. \end{aligned}$$

▷ **Solution 9-12** The answers are summarized in Table 9.4.

▷ **Solution 9-13** $U \perp\!\!\!\perp X$ is equivalent to: $P(X{=}x\,|\,U) = P(X{=}x)$, for each $x = 0,1,\ldots,J$. If $Z{=}f(U)$, this equation implies $P(X{=}x\,|\,U,Z) = P(X{=}x\,|\,U) = P(X{=}x)$. Using this equation as well as $E(\mathbb{1}_{X=x}\,|\,Z) = P(X{=}x\,|\,Z)$ and $E(\mathbb{1}_{X=x}) = P(X{=}x)$:

$$\begin{aligned} E(\mathbb{1}_{X=x}\,|\,Z) \;&=\; E\big[E(\mathbb{1}_{X=x}\,|\,U,Z)\,\big|\,Z\big] && \text{[Rule (??) in Box ??]} \\ &=\; E\big[E(\mathbb{1}_{X=x}\,|\,U)\,\big|\,Z\big] && \text{[}Z \text{ is measurable w. r. to } U\text{]} \\ &=\; E\big[E(\mathbb{1}_{X=x})\,\big|\,Z\big] && \text{[}U \perp\!\!\!\perp X\text{]} \\ &=\; E(\mathbb{1}_{X=x}) && \text{[Rule (??) in Box ??]} \\ &=\; P(X{=}x) && \text{for each } x = 0,1,\ldots,J. \end{aligned}$$

Hence, $P(X{=}x\,|\,U,Z) = P(X{=}x\,|\,Z)$, for each $x = 0,1,\ldots,J$.

▷ **Solution 9-14** Table 9.2 displays the probabilities $P(X{=}1\,|\,U{=}u)$ from which we can compute the probabilities $P(X{=}0\,|\,U{=}u) = 1 - P(X{=}1\,|\,U{=}u)$. Since $Z = f(U)$, the probabilities $P(X{=}0\,|\,U{=}u)$ also yield the probabilities $P(X{=}0\,|\,U{=}u,Z{=}m)$, which can be used to compute

$$\begin{aligned} P(X{=}0\,|\,Z{=}m) \;&=\; \sum_u P(X{=}0\,|\,U{=}u,Z{=}m)\cdot P(U{=}u\,|Z{=}m) \\ &=\; (4/20+6/20+4/20+6/20)\cdot 1/4 \;=\; 1/4. \end{aligned}$$

Using the Bayes Theorem, yields

$$\begin{aligned} P(Z{=}m\,|\,X{=}0) &= P(X{=}0\,|\,Z{=}m)\cdot P(Z{=}m)/P(X{=}0) \\ &= \frac{1/4\cdot 2/3}{5/12} = 2/5. \end{aligned}$$

▷ **Solution 9-15** First, note that $P(X{=}x) = E(1_{X=x})$ and $P(X{=}x\,|\,W) := E(1_{X=x}\,|\,W)$. Hence,

$$\begin{aligned} E(1_{X=x}\,|\,W) &= E\big[E(1_{X=x}\,|\,c_X)\,\big|\,W\big] \qquad \text{[Rule (??) in Box ??]} \\ &= E\big[P(X{=}x\,|\,c_X)\,\big|\,W\big] \\ &= E\big[P(X{=}x)\,|\,W\big] \qquad \text{[Eq. (??)]} \\ &= P(X{=}x) \qquad \text{[Rule (??) in Box ??]} \\ &= E(1_{X=x}), \qquad \text{for each } x = 0,1,\ldots,J. \end{aligned}$$

Since $P(X{=}x) = E(1_{X=x})$ and $P(X{=}x\,|\,W) = E(1_{X=x}\,|\,W)$, this proves the proposition.

▷ **Solution 9-16** We start with (ii) ⇒ (vi), i. e., with $Y \vdash c_X\,|\,X \Rightarrow Y \vdash c_X\,|\,X,Z$ (see Table 9.4). According to Box 9.1, $Y \vdash c_X\,|\,X$ means that

$$E(Y\,|\,X,c_X) = E(Y\,|\,X) \quad a.s. \tag{9.48}$$

holds, and we have to show that this equation implies

$$E(Y\,|\,X,c_X,Z) = E(Y\,|\,X,Z) \quad a.s.$$

This can be shown as follows:

$$\begin{aligned} E(Y\,|\,X,Z) &= E\big[E(Y\,|\,X,c_X,Z)\,\big|\,X,Z\big] \qquad \text{[Rule (??) in Box ??]} \\ &= E\big[E(Y\,|\,X,c_X)\,\big|\,X,Z\big] \qquad \text{[Z is measurable w. r. to c_X]} \\ &= E\big[E(Y\,|\,X)\,\big|\,X,Z\big] \qquad \text{[Eq. (9.48)]} \\ &= E(Y\,|\,X) \qquad \text{[Rule (??) in Box ??]} \\ &= E(Y\,|\,X,c_X) \qquad \text{[Eq. (9.48)]} \\ &= E(Y\,|\,X,c_X,Z) \quad a.s. \qquad \text{[Z is measurable w. r. to c_X]} \end{aligned}$$

which had to be shown.

Proving the other implications in Table 9.4, we only have to consider the conditional case, because the unconditional case is a particular case with Z being a constant. The first two implications, $X \perp\!\!\!\perp c_X\,|\,Z \Rightarrow \tau \perp\!\!\!\perp X\,|\,Z$ and $X \perp\!\!\!\perp c_X\,|\,Z \Rightarrow \tau \vdash X\,|\,Z$ are trivial, because conditional stochastic independence of c_X and X implies that also X and all functions that are measurable w.r.t. c_X are conditionally independent, and τ is such a function that is measurable w.r.t. X. Furthermore conditional stochastic independence implies conditional regressive independence.

Now we show that $Y \vdash c_X\,|\,X,Z$ implies $X \perp\!\!\!\perp \tau\,|\,Z$. First of all, note that $Y \vdash c_X\,|\,X,Z$ is equivalent to

$$E^{X=x}(Y\,|\,c_X,Z) = E^{X=x}(Y\,|\,Z) \quad a.s. \qquad \text{for each } x = 0,1,\ldots,J. \tag{9.49}$$

Since Z is measurable w.r.t. c_X,

$$\begin{aligned} E^{X=x}(Y\,|\,\mathscr{F}_{t_X}),Z) &= E^{X=x}(Y\,|\,\mathscr{F}_{t_X})) \\ &= \tau_x \quad P^{X=x}\text{-}a.s. \qquad \text{[def. of } \tau_x\text{]} \qquad \text{for each } x = 0,1,\ldots,J. \end{aligned}$$

Assuming that the extension exists, this implies that this equation also holds for the extensions

$$\begin{aligned} E^{X=x}(Y|\mathscr{F}_{t_X}),Z) &= E^{X=x}(Y|\mathscr{F}_{t_X})) \\ &= \tau_x \quad a.s. \qquad [\text{def. of } \tau_x] \qquad \text{for each } x=0,1,\ldots,J. \end{aligned} \tag{9.50}$$

Hence, comparing Equation (9.49) to Equation (9.50) shows that under $Y \vdash c_X \,|\, X,Z$, each τ_x is a measurable function of Z. But this implies:

$$P(X{=}x\,|\,\tau,Z) = P(X{=}x\,|\,Z) \quad a.s. \qquad \text{for each } x=0,1,\ldots,J. \tag{9.51}$$

This proves that $Y \vdash c_X \,|\, X,Z$ implies $X \perp\!\!\!\perp \tau \,|\, Z$.

The next implication is $Y \vdash c_X \,|\, X,Z \Rightarrow \tau \vdash X \,|\, Z$. Proving this proposition, we again will use Equation (9.49). For each $x=0,1,\ldots,J$,

$$\begin{aligned} E(\tau_x\,|X,Z) &= E\big[\, E^{X=x}(Y|\mathscr{F}_{t_X}))\,\big|\,X,Z\big] \qquad [\text{def. of } \tau_x] \\ &= E\big[\, E^{X=x}(Y|\mathscr{F}_{t_X}),Z)\,\big|\,X,Z\big] \qquad [Z \text{ is measurable w. r. to } c_X] \\ &= E\big[\, E^{X=x}(Y|Z)\,\big|\,X,Z\big] \qquad [\text{ Eq. (9.49)}] \\ &= E^{X=x}(Y|Z) \quad a.s. \qquad [\text{Rule (}\mathbf{??}\text{) in Box }\mathbf{??}]. \end{aligned} \tag{9.52}$$

Similarly, for each $x=0,1,\ldots,J$,

$$\begin{aligned} E(\tau_x\,|Z) &= E\big[\, E^{X=x}(Y|\mathscr{F}_{t_X}))\,\big|\,Z\big] \qquad [\text{def. of } \tau_x] \\ &= E\big[\, E^{X=x}(Y|\mathscr{F}_{t_X}),Z)\,\big|\,Z\big] \qquad [Z \text{ is measurable w. r. to } c_X] \\ &= E\big[\, E^{X=x}(Y|Z)\,\big|\,Z\big] \qquad [\text{ Eq. (9.49)}] \\ &= E^{X=x}(Y|Z) \quad a.s. \qquad [\text{Rule (}\mathbf{??}\text{) in Box }\mathbf{??}]. \end{aligned} \tag{9.53}$$

Equations (9.52) and (9.53) imply

$$E(\tau_x\,|X,Z) = E(\tau_x\,|Z) \qquad \text{for each } x=0,1,\ldots,J.$$

This equation is equivalent to $E(\tau\,|\,X,Z) = E(\tau\,|\,Z)$ *a.s.*, which had to be shown.

Finally, we have to show that $\tau \perp\!\!\!\perp X \,|\, Z$ implies $\tau \vdash X \,|\, Z$. This is again trivial, because conditional regressive independence follows from conditional stochastic independence (see Theorem 14.64).

Chapter 10
Identification of Causal Effects and Effect Functions

In chapter 6, we have seen that the conditional expectation values such as $E(Y \mid X{=}x)$, $E(Y \mid X{=}x, Z{=}z)$, $E(Y \mid X{=}x, M{=}m)$, and $E(Y \mid X{=}x, M{=}m, Z{=}z)$, but also the regressions $E(Y \mid X)$, $E(Y \mid X, Z)$, $E(Y \mid X, M)$, and $E(Y \mid X, M, Z)$, as well as the conditional and unconditional prima facie effects are usually biased. Furthermore, in chapters 7 to 9, we studied sufficient conditions for unbiasedness.

In this chapter, we will show how to adjust all these terms for bias, both for bias with respect to total effects as well as for bias with respect to direct effects. The specific goal is to provide the theoretical foundation for a number of data analysis procedures such as analysis of covariance and its generalizations, as well as stratifying and matching on covariates and on propensity scores. Adjustment will also be treated in chapter 11 dealing with propensity scores and their role in various adjustment techniques. Chapter 13 will then focus on procedures of data analysis utilizing the results of chapters 10 and 11.

Overview. We start with the adjustment theorems showing how to identify the expectations of the true outcome variables and the average total effects if we can assume that the regression $E(Y \mid X, Z)$ is Z-conditionally unbiased with respect to total effects. Then we turn to the identification of average total effects. ???

10.1 Identification of Total Effects

In this section, we present adjustments that aim at identifying average and conditional total effects. We show how the expectations $E(\tau_x)$ of the true outcome variables with respect to total effects can be identified (computed) from the regressions $E^{X=x}(Y \mid Z)$ of the outcome variable Y on the covariate Z with respect to the conditional-probability measure $P^{X=x}$, provided that the regressions $E^{X=x}(Y \mid Z)$ are τ_x-unbiased. This is important, because the difference $E(\tau_x) - E(\tau_{x'})$ is equal to the average total effect $E(\delta_{xx'})$. Similarly, the regressions $E(\tau_x \mid V)$ and $E(\delta_{xx'} \mid V)$ can be computed from the regressions $E^{X=x}(Y \mid Z)$, where V denotes a random variable that is measurable with respect to the covariate Z. Note that this also holds if the regression $E^{X=x}(Y \mid V)$ is *not* τ_x*-unbiased*. Furthermore, if we assume that the mean independent true-outcome (RIT) condition $\tau_x \vdash X \mid Z$ holds, then we can also identify the conditional expectation values $E(\tau_x \mid X{=}x^*)$ of the true

outcome variables τ_x and the conditional total effects $E(\delta_{xx'}|X=x^*)$ given a value x^* of X.

10.1.1 Theory

The following theorem is the theoretical foundation of several adjustment procedures aiming at identifying average and conditional total effects. Among these procedures are traditional analysis of covariance and its generalizations, matching on covariates, stratifying on covariates, matching on propensity scores, and stratifying on propensity scores.

Remark 10.1 (τ_x-Unbiasedness) In the following theorem $E^{X=x}(Y|Z)$ denotes the conditional expectation of Y on Z with respect to the conditional-probability measure $P^{X=x}$. In this theorem, we will assume that the $E^{X=x}(Y|Z)$ is τ_x-unbiased. Remember that *τ_x-unbiasedness of* $E^{X=x}(Y|Z)$ has been defined by P-uniqueness of $\tau_x = E^{X=x}(Y|C_X)$ and

$$E^{X=x}(Y|Z) \underset{P}{=} E(\tau_x|Z) \tag{10.1}$$

(see section 6.1). Also remember that the causality conditions (referring to total effects) treated in chapters 7 to 9 imply that $E^{X=x}(Y|Z)$ is τ_x-unbiased. ◁

Theorem 10.2 (Identification of Expectations and Regressions I)
Let $\langle (\Omega,\mathscr{A},P), (\mathscr{F}_t, t\in T), X, Y\rangle$ be a causality space, let Y be numerical and nonnegative or with finite expectation, let Z denote a covariate, let x be a value of X with $P(X=x)>0$, and let $E^{X=x}(Y|Z)$ be τ_x-unbiased.

(i) Then

$$E(\tau_x) = E[E^{X=x}(Y|Z)]. \tag{10.2}$$

(ii) Let $V\colon (\Omega,\mathscr{A},P)\to(\Omega'_V,\mathscr{A}'_V)$ be a random variable that is measurable w.r.t. Z. Then

$$E(\tau_x|V) \underset{P}{=} E\left[E^{X=x}(Y|Z)\,\middle|\,V\right]. \tag{10.3}$$

(iii) If $v\in\Omega'_V$ is a value of V with $P(V=v)>0$, then

$$E(\tau_x|V=v) = E\left[E^{X=x}(Y|Z)\,\middle|\,V=v\right]. \tag{10.4}$$

(Proof p. 258)

This theorem tells us how to identify the expectation of the true-outcome variable τ_x, i. e., how to compute $E(\tau_x)$ from an estimable quantity, the conditional expectation $E^{X=x}(Y|Z)$. It also shows how to compute the conditional expectation of the true-outcome variable on any Z-measurable random variable V. For both purposes we have to assume that $E^{X=x}(Y|Z)$ is τ_x-unbiased. Under this assumption, all we need for identifying $E(\tau_x|V)$ is the $(X=x)$-conditional regression of Y on Z and its conditional expectation on V, both of which can be estimated in empirical applications (see Examples 10.11 and 10.14).

Remark 10.3 (RIT Condition) Instead of τ_x-unbiasedness of $E^{X=x}(Y|Z)$, let us assume that the *RIT condition* $\tau_x \vdash X \mid Z$ holds, which is defined by P-uniqueness of $\tau_x := E^{X=x}(Y|C_X)$ and

$$E(\tau_x|X,Z) \underset{P}{=} E(\tau_x|Z) \tag{10.5}$$

(see section 9.0.3). This assumption implies τ_x-unbiasedness of $E^{X=x}(Y|Z)$ [see Eq. (10.1)], but is not equivalent to τ_x-unbiasedness. Therefore the RIT condition also implies somewhat stronger results using a random variable V that is measurable with respect to (X,Z) and not only with respect to Z. This implies that the random variable V may be a function $f(X,Z)$ of Z *and* X, and not only a function $f(Z)$ of Z alone. Note that $\tau_x \vdash X \mid Z$ follows from some of the causality conditions treated in chapters 7 and ?? (see Table ??). ◁

Theorem 10.4 (Identification of Expectations and Regressions II)
Let the assumptions of Theorem 10.2 be true. If $\tau_x \vdash X \mid Z$, then propositions (ii) and (iii) of Theorem 10.2 also hold for V being measurable w.r.t. (X,Z).

(Proof p. 258)

Before commenting on these theorems in more detail, let us consider their implications on total effects and total effect functions comparing two values x and x' of X to each other.

Theorem 10.5 (Adjusting Effects and Effect Functions I)
Let the assumptions of Theorem 10.2 be true, let x, x' be two values of X such that $P(X{=}x)$, $P(X{=}x') > 0$, let $\delta_{xx'}$ denote the true total effect variable, let $PFE_{xx';Z} := E^{X=x}(Y|Z) - E^{X=x'}(Y|Z)$, and let $E^{X=x}(Y|Z)$ and $E^{X=x'}(Y|Z)$ be τ_x-unbiased.

(i) Then

$$E(\delta_{xx'}) = E(PFE_{xx';Z}) \tag{10.6}$$

and

$$E(\delta_{xx'}|Z) \underset{P}{=} PFE_{xx';Z}. \tag{10.7}$$

(ii) If z is a value of Z with $P(Z{=}z) > 0$, then

$$E(\delta_{xx'}|Z{=}z) = PFE_{xx';Z=z}. \tag{10.8}$$

(iii) If V is measurable w.r.t. Z, then

$$E(\delta_{xx'}|V) \underset{P}{=} E(PFE_{xx';Z}|V). \tag{10.9}$$

(iv) If v is a value of V with $P(V{=}v) > 0$, then

$$E(\delta_{xx'}|V{=}v) = E(PFE_{xx';Z}|V{=}v). \tag{10.10}$$

(Proof p. 258)

Remark 10.6 (RIT Condition) Instead of assuming only τ_x-unbiasedness of the two $(X{=}x)$-conditional regressions $E^{X=x}(Y|Z)$ and $E^{X=x'}(Y|Z)$, let us now assume that the RIT conditions $\tau_x \vdash X \mid Z$ and $\tau_{x'} \vdash X \mid Z$ hold [see Eq. (10.5)]. Again, this allows us to condition on a random variable V that is measurable with respect to (X,Z) and not only with respect to Z. Note again that $\tau_{x'} \vdash X|Z$ follows from some of the causality conditions treated in chapters 7 and ?? (see Table ??). ◁

Theorem 10.7 (Identification of Effects and Effect Functions II)
Let the assumptions of Theorem 10.5 be true. If $\tau_x \vdash X|Z$ and $\tau_{x'} \vdash X|Z$, then propositions (iii) and (iv) of Theorem 10.5 still hold for V being measurable w.r.t. (X,Z).

(Proof p. 259)

Remark 10.8 (Average Total Effect Given $X{=}x^*$) The additional implications of the weak ignorability condition $\tau_x \vdash X|Z$ are useful if we are not only interested in the average total effects and in the Z-conditional total effects given a covariate Z, but also in the *V-conditional total effects* $E(\delta_{xx'}|V{=}v)$, where V is any measurable function of (X,Z). For example, X is a measurable function of (X,Z) and may therefore take the role of V. Hence, under $\tau_x \vdash X \mid Z$, we can use equations

$$E(\delta_{xx'} \mid X) \underset{P}{=} E(PFE_{xx';Z} \mid X) \tag{10.11}$$

and

$$E(\delta_{xx'} \mid X{=}x^*) = E(PFE_{xx';Z} \mid X{=}x^*) \tag{10.12}$$

for the $(X{=}x^*)$-conditional total effects. For a given pair (x,x') of values of X, $E(\delta_{xx'}|X{=}x^*)$ may differ from $E(\delta_{xx'}|X{=}x^\star)$ if $x^\star$ denotes a value of X that is different from x^*. In section 5.2.3 we introduced the conditional expectation values $E(\delta_{xx'}|X{=}x^*)$ and discussed their substantive meaning in some detail. ◁

10.1.2 Methodological Implications

The theorems presented in section 10.1.1 are the theoretical foundation of several procedures for identifying conditional and average total effects and total-effect functions. These procedures will be outlined in some detail in chapter 13. In this section we content ourselves with some remarks on the methodological implications of these theorems.

Remark 10.9 (No Functional Form Assumption) Note that Equations (10.2) to (10.4) and Equations (10.6) to (10.10) can be used whenever the conditional expectations $E^{X=x}(Y|Z)$ and $E^{X=x'}(Y|Z)$ are τ_x-unbiased. No assumptions about the functional form of $E^{X=x}(Y|Z)$ and $E^{X=x'}(Y|Z)$ are made. However, in empirical applications, assuming an incorrect functional form of these conditional expectations may lead to incorrect estimates of the expectations $E(\tau_x)$ and $E(\tau_{x'})$,

the average total effect $E(\delta_{xx'})$, and the $(V=v)$-conditional total effects $E(\delta_{xx'} \mid V=v)$. ◁

Remark 10.10 (Adjusted Means) Equation (10.2) shows how to compute the expectations $E(\tau_x)$ of the true outcomes, if the conditional expectation $E^{X=x}(Y|Z)$ is τ_x-unbiased. In empirical applications, it is these expectations $E(\tau_x)$ that are of interest, and not the (usually biased) conditional expectation values $E(Y|X=x)$ of the outcome variable Y given $X=x$. Estimates of the expectations $E(\tau_x)$ are also called *adjusted means* of the outcome variable Y given $X=x$. Note, however, that the adjusted means are guaranteed to be τ_x-unbiased estimators of the expectations $E(\tau_x)$ only if (a) $E^{X=x}(Y|Z)$ is τ_x-unbiased, (b) no incorrect assumption about the functional form of $E^{X=x}(Y|Z)$ is made, and (c) the assumptions of the statistical model are correct. (Remember, statistical models and their assumptions are not treated in this book.) ◁

Example 10.11 (Linear Regressions) If X represents a treatment variable, then $E^{X=x}(Y|Z)$ is the conditional expectation of Y on Z in treatment x. If we assume that $E^{X=x}(Y|Z)$ is a linear function of Z, i. e.,

$$E^{X=x}(Y|Z) \underset{P^{X=x}}{=} \beta_0^{(x)} + \beta_1^{(x)} Z, \tag{10.13}$$

and that $E^{X=x}(Y|Z)$ is τ_x-unbiased, then the expectation of $E^{X=x}(Y|Z)$ is

$$\begin{aligned} E[E^{X=x}(Y|Z)] &\underset{P}{=} E(\beta_0^{(x)} + \beta_1^{(x)} Z) \\ &\underset{P}{=} \beta_0^{(x)} + \beta_1^{(x)} E(Z). \end{aligned} \tag{10.14}$$

Note that $E(Z)$ denotes the expectation with respect to the measure P, not with respect to the conditional-probability measure $P^{X=x}$. Hence, if X represents a treatment variable, the sample estimate of $E(Z)$ is the mean in the *total sample*, not the mean in the subsample of treatment condition x. ◁

Remark 10.12 (Average Total Effects) According to Equation (10.6), once the *conditional* prima facie effects are identified, we can also identify the average total effects by taking the expectation $E(PFE_{xx';Z})$ of the Z-conditional prima-facie-effect function

$$PFE_{xx';Z} := E^{X=x}(Y|Z) - E^{X=x'}(Y|Z)$$

over the distribution of the (possibly multivariate) covariate Z, *provided that the conditional expectations $E^{X=x}(Y|Z)$ and $E^{X=x'}(Y|Z)$ are τ_x-unbiased.* ◁

Remark 10.13 (Conditional Total Effect Functions) If the conditional expectations $E^{X=x}(Y|Z)$ and $E^{X=x'}(Y|Z)$ are both τ_x-unbiased, then the conditional prima facie effect function $PFE_{xx';Z}$ is equal to the conditional total-effect function $E(\delta_{xx'}|Z)$ [see Eq. (10.7)]. If, e. g., Z:=*sex* and $Z=m$ represents the male subpopulation, then this implies that the conditional prima facie effect

$$PFE_{xx';Z=m} := E(Y|X=x, Z=m) - E(Y|X=x', Z=m)$$

in the male subpopulation is also the conditional total effect $E(\delta_{xx'}|Z=m)$ in this subpopulation [see Eq. (10.8)]. ◁

If we are not only interested in the average total effects and in the Z-conditional total effects given a covariate Z, but also in the $(V{=}v)$-*conditional total effects* $E(\delta_{xx'} \mid V{=}v)$, where V is any measurable function of Z, then Equations (10.9) and (10.10) show how these $(V{=}v)$-conditional effects can be identified. For example, if $Z = (V, W)$, then V is such a measurable function of Z. Hence, we can use

$$E(\delta_{xx'} \mid V) \underset{P}{=} E(PFE_{xx';Z} \mid V) \tag{10.15}$$

[see Eq. (10.9)] and

$$E(\delta_{xx'} \mid V{=}v) = E(PFE_{xx';Z} \mid V{=}v) \tag{10.16}$$

[see Eq. (10.10)], provided that $E^{X=x}(Y \mid Z)$ and $E^{X=x'}(Y \mid Z)$ are τ_x-unbiased.

Example 10.14 (Multiple Linear Regressions) If X represents a treatment variable, then $E^{X=x}(Y \mid Z)$ is the conditional expectation of Y on Z in treatment x. If we assume that $E^{X=x}(Y \mid Z)$ is P-unique and that it is a linear function of $Z = (V, W)$, i. e.,

$$E^{X=x}(Y \mid Z) \underset{P}{=} \beta_0^{(x)} + \beta_1^{(x)} V + \beta_2^{(x)} W, \tag{10.17}$$

then the conditional expectation of $E^{X=x}(Y \mid Z)$ on V is

$$\begin{aligned} E[E^{X=x}(Y \mid Z) \mid V] &\underset{P}{=} E(\beta_0^{(x)} + \beta_1^{(x)} V + \beta_2^{(x)} W \mid V) \quad \text{[Eq. (10.17)]} \\ &\underset{P}{=} \beta_0^{(x)} + \beta_1^{(x)} V + \beta_2^{(x)} E(W \mid V) \quad \text{[Box PR-10.2 (v), (iii)]}. \end{aligned} \tag{10.18}$$

If we additionally assume

$$E(W \mid V) \underset{P}{=} \alpha_0 + \alpha_1 V, \tag{10.19}$$

then

$$\begin{aligned} E[E^{X=x}(Y \mid Z) \mid V] &\underset{P}{=} \beta_0^{(x)} + \beta_1^{(x)} V + \beta_2^{(x)} E(W \mid V) \qquad \text{[Eq. (10.18)]} \\ &\underset{P}{=} \beta_0^{(x)} + \beta_1^{(x)} V + \beta_2^{(x)} (\alpha_0 + \alpha_1 V) \qquad \text{[Eq. (10.19)]} \\ &\underset{P}{=} (\beta_0^{(x)} + \beta_2^{(x)} \alpha_0) + (\beta_1^{(x)} + \beta_2^{(x)} \alpha_1) V. \end{aligned}$$

Hence, in this case the conditional expectation of $E^{X=x}(Y \mid Z)$ on V is a linear function of V with intercept $\beta_0^{(x)} + \beta_2^{(x)} \alpha_0$ and slope $\beta_1^{(x)} + \beta_2^{(x)} \alpha_1$. ◁

Remark 10.15 (Propensity as a Covariate) Note that the conditional probability $\pi_x := P(X{=}x \mid Z)$ is measurable with respect to Z. This function, which is called the *propensity of* x given Z, can also play the role of the covariate Z in Theorems 10.2 and 10.4. Hence, all equations in these two theorems involving Z also hold for π_x taking the role of Z, provided that $E^{X=x}(Y \mid \pi_x)$ is τ_x-unbiased.

Similarly, the vector $\pi := (\pi_1, \ldots, \pi_J)$, with $\pi_x := P(X{=}x \mid Z)$, $x = 1, \ldots, J$, is measurable with respect to Z. This propensity vector π can also play the role of Z in the theorems presented in section 10.1.1. Hence, all equations in these theorems involving Z also hold for π taking the role of Z. (More details about propensities will be treated in chapter 11.) ◁

Table 10.1. Strong Ignorability

	Fundamental parameters						Derived parameters			
Unit u	Sampling probability $P(U=u)$	Covariate sex V	Covariate college W	True outcomes under control τ_0	True outcomes under treatment τ_1	True treatment probability $P(X=1\,\vert\,U)$	True causal effect δ_{10}	$P(X=1\,\vert\,V,W)$	$P(U=u\,\vert X=0)$	$P(U=u\,\vert X=1)$
u_1	1/6	m	n	72	83	7/8	11	6/8	1/22	7/26
u_2	1/6	m	n	72	83	5/8	11	6/8	3/22	5/26
u_3	1/6	m	y	95	100	5/8	5	5/8	3/22	5/26
u_4	1/6	m	y	100	105	5/8	5	5/8	3/22	5/26
u_5	1/6	f	y	106	114	2/8	8	2/8	6/22	2/26
u_6	1/6	f	y	116	130	2/8	14	2/8	6/22	2/26
			$E(\tau_x)$	93.50	102.50					
			$E(Y\,\vert X=x)$	100.227	96.50					

Average causal effect (average total effect) 9.000
prima facie effect (*PFE*) −3.727

	male, no college	male, college	female, college
Conditional total effects	11	5	11
Conditional *PFE*s	11	5	11

10.1.3 Numerical Example

In chapter 6 we already treated some examples, in which the conditional expectation $E(Y\,\vert X,Z)$ is τ_x-unbiased. In these examples, we assumed

$$E(Y\,\vert X,C_X) \;=\; E(Y\,\vert X,U) \tag{10.20}$$

for the true-outcome variables τ_x and

$$P(X{=}1\,\vert\mathscr{C}_X) \underset{P}{=} P(X{=}1\,\vert U) \tag{10.21}$$

for the true treatment probability function. There, we already noted that these assumptions are realistic also in empirical applications if U represents the observational units *at the onset of treatment*.

Example 10.16 In the example presented in Table 10.1, the IT condition $X \perp\!\!\!\perp \tau\,\vert Z$ (see section 9.1) holds, because

Table 10.2. Conditional Expectations of Y Given Treatment and Covariates

	Covariates		
Treatment	male, no college	male, college	female, college
$X=0$	72.00	97.50	111.00
$X=1$	83.00	102.50	122.00

$$P(X=1\,|\,Z,\tau) \;=\; P(X=1\,|\,Z)$$

(see Exercise 11-5). Because $X \perp\!\!\!\perp \tau_x|Z$ implies $\tau_x \vdash X|Z$ and in this example there are two covariates, V and W, we will use it to illustrate the theorems treated in section 10.1.1.

In this example, we can compute the expectations of the true outcomes in the control and treatment conditions as follows:

$$\begin{aligned} E(\tau_0) \;&= \sum_u E(Y\,|\,X=0, U=u)\cdot P(U=u) \\ &= \;72\cdot\frac{1}{6}+95\cdot\frac{1}{6}+\ldots+116\cdot\frac{1}{6} \;=\; 93.50, \end{aligned} \tag{10.22}$$

$$\begin{aligned} E(\tau_1) \;&= \sum_u E(Y\,|\,X=1, U=u)\cdot P(U=u) \\ &= \;83\cdot\frac{1}{6}+100\cdot\frac{1}{6}+\ldots+130\cdot\frac{1}{6} \;=\; 102.50 \end{aligned} \tag{10.23}$$

[see Eq. (ii) in Box 6.1]. Using $E(\delta_{10}) = E(\tau_1) - E(\tau_0)$, the difference $E(\tau_1) - E(\tau_0) = 102.50 - 93.50 = 9.00$ yields the average total effect. However, in empirical applications, the true outcomes used in these two equations are unknown and not all are estimable. Hence, we need alternative ways to compute the expectations of the true outcomes and the average total effect.

According to Equation (10.2), using the two-dimensional covariate $Z := (V, W)$ = (*sex*, *college*), these expectations can also be computed as follows:

$$\begin{aligned} E(\tau_0) \;=\; E\big[E^{X=0}(Y\,|\,Z)\big] \;&= \sum_z E(Y\,|\,X=0, Z=z)\cdot P(Z=z) \\ &= 72.00\cdot\frac{1}{3}+97.50\cdot\frac{1}{3}+111.00\cdot\frac{1}{3} \\ &= 93.50 \end{aligned} \tag{10.24}$$

and

$$\begin{aligned} E(\tau_1) \;=\; E\big[E^{X=1}(Y\,|\,Z)\big] \;&= \sum_z E(Y\,|\,X=1, Z=z)\cdot P(Z=z) \\ &= 83.00\cdot\frac{1}{3}+102.50\cdot\frac{1}{3}+122.00\cdot\frac{1}{3} \\ &= 102.5. \end{aligned} \tag{10.25}$$

In this example, the covariate $Z := (V,W)$ has three values: $z_1 = (m,n)$, $z_2 = (m,y)$, and $z_3 = (f,y)$. In the equations above, we used the conditional expectation values of Y displayed in Table 10.2, which can be computed as follows:

$$\begin{aligned} E(Y|X{=}x,Z{=}z) &= \sum_u E(Y|X{=}x,Z{=}z,U{=}u)\cdot P(U{=}u|X{=}x,Z{=}z) \\ &= \sum_u E(Y|X{=}x,U{=}u)\cdot P(U{=}u|X{=}x,Z{=}z). \end{aligned} \tag{10.26}$$

The first of these two equation is always true and the second one holds, because, in this example, $Z := (V,W)$ is a function of the observational-unit variable U. Note that, in contrast to Equations (10.22) and (10.23), all terms needed in Equations (10.24) and (10.25) *are* estimable in empirical applications.

Similarly, the expectation of the true total effects can be computed via

$$E(\delta_{10}) = 11\cdot\frac{2}{6}+5\cdot\frac{2}{6}+8\cdot\frac{1}{6}+14\cdot\frac{1}{6} = 9$$

[see Table 10.1]. However, in empirical applications the true total effects, i. e., the values of δ_{10}, are neither available nor are they estimable. What can be done, however, is to take the difference between $E(\tau_1)$ and $E(\tau_0)$ — obtained from Equations (10.25) and from Equation (10.24), respectively — or alternatively, use Equation (10.6):

$$\begin{aligned} E(\delta_{10}) &= E(PFE_{10;Z}) = \sum_z PFE_{10;Z=z}\cdot P(Z{=}z) \\ &= 11\cdot\frac{2}{6}+5\cdot\frac{2}{6}+11\cdot\frac{2}{6} = 9, \end{aligned}$$

which yields the average total effect from the prima facie effects 11, 5 and 11 in the three subpopulations. Note that all terms appearing on the right-hand side of this equation *are* empirically estimable. Also note that $E(\delta_{10}) = E(PFE_{10;Z})$ only holds, because, in the example presented in Table 10.1, the covariate-treatment conditional expectation $E(Y|X,Z)$ is τ_x-unbiased for the covariate $Z := (V,W)$.

Hence, Z-conditional τ_x-unbiasedness of the conditional expectation $E(Y|X,Z)$ — implying τ_x-unbiased conditional prima facie effects — is crucial for estimating the expectations of the true outcomes, the average total effects, and the conditional total effects. This underlines the importance of the chapters in which we treated sufficient conditions for unbiasedness.

If we are interested in the conditional total effects of treatment 1 vs. 0 given $W :=$ *sex*, we can use Equation (10.4), which can be written

$$\begin{aligned} E(\tau_x|V{=}v) &= E(E^{X=x}(Y|Z)|V{=}v) \\ &= \sum_z E(Y|X{=}x,Z{=}z)\cdot P(Z{=}z|V{=}v). \end{aligned} \tag{10.27}$$

For $X{=}0$ and $V{=}m$ (males), we receive

$$\begin{aligned} E(\tau_0|V{=}m) &= \sum_z E(Y|X{=}0,Z{=}z)\cdot P(Z{=}z|V{=}m) \\ &= 72\cdot\frac{1}{2}+97.5\cdot\frac{1}{2}+111\cdot 0 \approx 84.75, \end{aligned} \tag{10.28}$$

and for $X=1$ and $V=m$,

$$\begin{aligned} E(\tau_1 \mid V=m) &= \sum_z E(Y \mid X=1, Z=z) \cdot P(Z=z \mid V=m) \\ &= 83 \cdot \frac{1}{2} + 102.5 \cdot \frac{1}{2} + 122 \cdot 0 \approx 92.75, \end{aligned} \tag{10.29}$$

Hence, the conditional total effect of treatment 1 vs. 0 for males is

$$E(\delta_{10} \mid V=m) = E(\tau_1 \mid V=m) - E(\tau_0 \mid V=m) = 92.75 - 84.75 = 8.$$

This is also the $V=m$-conditional expected value of $PFE_{10;Z}$:

$$\begin{aligned} E(PFE_{10;Z} \mid V=m) &= \sum_z PFE_{10;Z=z} \cdot P(Z=z \mid V=m) \\ &= 11 \cdot \frac{1}{2} + 5 \cdot \frac{1}{2} + 11 \cdot 0 = 8 \end{aligned}$$

[see Table 10.1 and Eq. (10.10)].

The $V=m$-conditional prima facie effect is $PFE_{10;V=m} \approx 91.86 - 87.3 \approx 4.56$ is biased in this example. It can be computed by

$$PFE_{10;V=m} := E(Y \mid X=1, V=m) - E(Y \mid X=0, V=m)$$

where, in this example, in which V is a function of U,

$$\begin{aligned} E(Y \mid X=1, V=m) &= \sum_u E(Y \mid X=1, V=v, U=u) \cdot P(U=u \mid X=1, V=m) \\ &= \sum_u E(Y \mid X=1, U=u) \cdot P(U=u \mid X=1, V=m) \qquad [W = f(U)] \\ &= 83 \cdot \frac{7}{22} + 83 \cdot \frac{5}{22} + 100 \cdot \frac{5}{22} + 105 \cdot \frac{5}{22} \approx 91.86 \end{aligned}$$

and

$$\begin{aligned} E(Y \mid X=0, V=m) &= \sum_u E(Y \mid X=0, V=v, U=u) \cdot P(U=u \mid X=0, V=m) \\ &= \sum_u E(Y \mid X=0, U=u) \cdot P(U=u \mid X=0, V=m) \qquad [V = f(U)] \\ &= 72 \cdot \frac{1}{10} + 72 \cdot \frac{3}{10} + 95 \cdot \frac{3}{10} + 100 \cdot \frac{3}{10} = 87.3 \end{aligned}$$

(see Exercise ??). ◁

10.2 Identification of Direct Effects

Now we generalize the adjustment theorems presented in section 10.1 considering the *identification of average and conditional direct effects* with respect to an intermediate variable M.

10.2.1 Theory

Conditioning will now refer to a random variable W that is measurable with respect to (M,Z), where Z denotes a covariate. Hence (M,Z) itself may play the role of W. Another example of W is the conditional probability $P(X{=}x\,|\,M,Z)$. We will also consider a random variable V that is measurable with respect to W. Examples of V are $V=M$, $V=Z$, as well as $P(X{=}x\,|\,M)$ and $P(X{=}x\,|\,Z)$.

Remark 10.17 (τ_{x,t_M}-Unbiasedness) Remember that the conditional expectation $E^{X=x}(Y|W)$ is called τ_{x,t_M}-*unbiased*, if $\tau_{x,t_M} := E^{X=x}(Y|C_X,M)$ is P-unique and

$$E^{X=x}(Y|W) \underset{P}{=} E(\tau_{x,t_M}|W) \tag{10.30}$$

(see section 6.4). ◁

Theorem 10.18 (Adjustments Identifying Expectations and Regressions I) *Let $\langle(\Omega,\mathscr{A},P),(\mathscr{F}_t,t\in T),X,Y\rangle$ be a numerical causality space, let M denote an intermediate variable w.r.t. X and Y, let Z denote a covariate, let x be a value of X with $P(X{=}x)>0$, let W be measurable w.r.t. (M,Z), and let $E^{X=x}(Y|W)$ be τ_{x,t_M}-unbiased.*

(i) Then

$$E(\tau_{x,t_M}) = E[E^{X=x}(Y|W)]. \tag{10.31}$$

(ii) If V is measurable w.r.t. W, then

$$E(\tau_{x,t_M}|V) \underset{P}{=} E[E^{X=x}(Y|W)\,|\,V]. \tag{10.32}$$

(iii) If v is a value of V with $P(V{=}v)>0$, then

$$E(\tau_{x,t_M}|V{=}v) = E[E^{X=x}(Y|W)\,|\,V{=}v]. \tag{10.33}$$

(Proof p. 260)

This theorem tells us how to compute the expectation of τ_{x,t_M}, the true-outcome variable with respect to M-direct effects. It also tells us how to compute the conditional expectation of τ_{x,t_M} on any W-measurable function V. In both cases we have to assume τ_{x,t_M}-unbiasedness of the conditional expectation $E^{X=x}(Y|W)$ with respect to M-direct effects. Under this unbiasedness assumption, all we need is $E^{X=x}(Y|W)$, the $(X{=}x)$-conditional conditional expectation of Y on W, its expectation, and its conditional expectation on V.

Remark 10.19 (RIT Condition for Direct Effects) In the following theorem we replace τ_{x,t_M}-unbiasedness by the RIT condition $\tau_{x,t_M} \vdash X\,|\,W$ defined by P-uniqueness of $E^{X=x}(Y|C_X,M)$ and

$$E(\tau_{x,t_M}|X,W) \underset{P}{=} E(\tau_{x,t_M}|W). \tag{10.34}$$

This allows us to consider a random variable V that is measurable with respect to (X,W) and not only with respect to W. ◁

Theorem 10.20 (Identification of Expectations and Regressions II)
Let the assumptions of Theorem 10.18 be true. If $\tau_{x,t_M} \vdash X \mid W$, then propositions (ii) and (iii) of Theorem 10.18 also hold for V being measurable w.r.t. (X,W).

(Proof p. 260)

Now we consider the implications of these theorems for identifying average direct effects and conditional direct-effect functions considering two different values x and x' of X and the true total effect variable δ_{xx',t_M}.

Theorem 10.21 (Adjustments Identifying Effects and Effect Functions I)
Let the assumptions of Theorem 10.18 be true, let x, x' be two values of X such that $P(X{=}x)$, $P(X{=}x') > 0$, let $\delta_{xx',t_M} := \tau_{x,t_M} - \tau_{x',t_M}$ denote the true M-direct effect variable, let W be measurable w.r.t. (M,Z), let $PFE_{xx';W} := E^{X=x}(Y|W) - E^{X=x'}(Y|W)$, and let $E^{X=x}(Y|W)$ and $E^{X=x'}(Y|W)$ both be τ_{x,t_M}-unbiased.

(i) Then

$$E(\delta_{xx',t_M}) = E(PFE_{xx';W}) \tag{10.35}$$

and

$$E(\delta_{xx',t_M} \mid W) \underset{P}{=} PFE_{xx';W}. \tag{10.36}$$

(ii) If v is a value of W with $P(W{=}w) > 0$, then

$$E(\delta_{xx',t_M} \mid W{=}w) = PFE_{xx';W=w}. \tag{10.37}$$

(iii) If V is measurable w.r.t. W, then

$$E(\delta_{xx',t_M} \mid V) \underset{P}{=} E(PFE_{xx';W} \mid V), \tag{10.38}$$

(iv) If v is a value of V with $P(V{=}v) > 0$, then

$$E(\delta_{xx',t_M} \mid V{=}v) = E(PFE_{xx';W} \mid V{=}v). \tag{10.39}$$

(Proof p. 260)

Remark 10.22 (Intermediate and Covariate) The random variable W occurring in Theorems 10.18 and 10.21 may be the vector consisting of the intermediate variable M and the covariate Z, i. e., $W := (M,Z)$. In this case all equations occurring in Theorem 10.21 can be written replacing W by (M,Z). For instance, Equation (10.35) can then be written

$$E(\delta_{xx',t_M}) = E(PFE_{xx';M,Z}), \tag{10.40}$$

and it will hold if $E^{X=x}(Y \mid M,Z)$ and $E^{X=x'}(Y \mid M,Z)$ are τ_{x,t_M}-unbiased. Hence, $E(PFE_{xx';M,Z})$ is identical with the average direct effect $E(\delta_{xx',t_M})$, if $E^{X=x}(Y|M,Z)$ and $E^{X=x'}(Y|M,Z)$ are τ_{x,t_M}-unbiased.

Similarly, if $W := (M, Z)$, then Equation (10.36) can be written

$$E(\delta_{xx',t_M} \mid M, Z) \underset{P}{=} PFE_{xx';M,Z}, \tag{10.41}$$

and we can conclude that the prima facie effect function $PFE_{xx';M,Z}$ is identical with the M-direct-effect function $E(\delta_{xx',t_M} \mid M, Z)$, provided that $E^{X=x}(Y \mid M, Z)$ and $E^{X=x'}(Y \mid M, Z)$ are both τ_{x,t_M}-unbiased.

Furthermore, if $W := (M, Z)$ and $V := Z$, then Equation (10.38) can be written

$$E(\delta_{xx',t_M} \mid Z) \underset{P}{=} E(PFE_{xx';M,Z} \mid Z). \tag{10.42}$$

Hence, $E(PFE_{xx';M,Z} \mid Z)$ is identical with the Z-conditional M-direct-effect function $E(\delta_{xx',t_M} \mid Z)$, provided that $E^{X=x}(Y \mid M, Z)$ and $E^{X=x'}(Y \mid M, Z)$ are both τ_{x,t_M}-unbiased.

Finally, if $W := (M, Z)$ and $V := M$, then Equation (10.38) can be written

$$E(\delta_{xx',t_M} \mid M) \underset{P}{=} E(PFE_{xx';M,Z} \mid M). \tag{10.43}$$

Hence, $E(PFE_{xx';M,Z} \mid M)$ is identical with the M-conditional M-direct-effect function $E(\delta_{xx',t_M} \mid M)$, provided that $E^{X=x}(Y \mid M, Z)$ and $E^{X=x'}(Y \mid M, Z)$ are both τ_{x,t_M}-unbiased. ◁

In the following theorem we replace τ_{x,t_M}-unbiasedness [see Eq. (10.30)] by the RIT condition $\tau_{x,t_M} \vdash X \mid W$ (see Remark 10.19). This allows us to consider a random variable V that is measurable with respect to (X, W) and not only with respect to W.

Theorem 10.23 (Adjustments Identifying Effects and Effect Functions II) *Let the assumptions of Theorem 10.21 be true. If $\tau_{x,t_M} \vdash X \mid W$ and $\tau_{x',t_M} \vdash X \mid W$, then propositions (iii) and (iv) of Theorem 10.21 still hold for V being measurable with respect to (X, W).*

(Proof p. 261)

Remark 10.24 (Propensities) In chapter 11, we will show that W can also be the conditional treatment probability $P(X{=}x \mid M, C_X)$ or the conditional treatment probability $P(X{=}x \mid M, Z)$. In this case, Theorems 10.18 to 10.23 provide the theoretical foundation of several procedures for identifying conditional and average direct effects using propensities. Some of these procedures for identifying conditional and average direct effects will be outlined in more detail in chapter 13. ◁

10.2.2 Methodological Implications

The theorems presented in section 10.2.1 are the theoretical foundation of several procedures for identifying conditional and average direct effects and direct-effect

functions. These procedures will be outlined in some detail in chapter 13. In this section we content ourselves with some remarks on the methodological implications of these theorems. What has been noted in section 10.1.2 still applies with minor changes. We slightly have to adapt our remarks, because now we consider *direct* effects.

Remark 10.25 (No Functional Form Assumption) What has been noted in Remark 10.9 still applies with minor changes. Equations (10.31) to (10.33) and Equations (10.35) to (10.39) can be used whenever the conditional expectations $E^{X=x}(Y|W)$ and $E^{X=x'}(Y|W)$ are τ_{x,t_M}-unbiased. Again, no assumptions about the functional form of $E^{X=x}(Y|W)$ and $E^{X=x'}(Y|W)$ are made. However, in empirical applications, assuming an incorrect functional form of these conditional expectations may yield biased estimates of the expectations $E(\tau_{x,t_M})$ and $E(\tau_{x',t_M})$, the average direct effect $E(\delta_{xx',t_M})$, and the $(W=w)$-conditional direct effects $E(\delta_{xx',t_M}|W=w)$. ◁

Remark 10.26 (Expectations of True Outcome Variables w.r.t. Direct Effects) Equation (10.31) shows how to compute the expectations $E(\tau_{x,t_M})$ of the true outcomes, if the conditional expectation $E^{X=x}(Y|W)$ is τ_{x,t_M}-unbiased. It is these expectations $E(\tau_{x,t_M})$ that are of interest, because the differences $E(\tau_{x,t_M}) - E(\tau_{x',t_M})$ are the average M-direct effects of x vs. x'. ◁

Example 10.27 (Multiple Linear Regression I) If X represents a treatment variable, then $E^{X=x}(Y|M,Z)$ is the conditional expectation of Y on Z and M in treatment x. If we assume that $E^{X=x}(Y|M,Z)$ is P-unique and that it is a linear function of Z and M, i. e.,

$$E^{X=x}(Y|M,Z) \underset{P}{=} \beta_0^{(x)} + \beta_1^{(x)} M + \beta_2^{(x)} Z, \tag{10.44}$$

then the expectation of $E^{X=x}(Y|M,Z)$ is

$$\begin{aligned} E[E^{X=x}(Y|M,Z)] &\underset{P}{=} E\left(\beta_0^{(x)} + \beta_1^{(x)} M + \beta_2^{(x)} Z\right) \\ &\underset{P}{=} \beta_0^{(x)} + \beta_1^{(x)} E(M) + \beta_2^{(x)} E(Z). \end{aligned} \tag{10.45}$$

Again, $E(M)$ and $E(Z)$ are the expectations with respect to the measure P, not with respect to the conditional-probability measure $P^{X=x}$. Hence, if X represents a treatment variable, the sample estimates of $E(M)$ and $E(Z)$ are the means in the *total sample*, not the means in the subsample of treatment condition x. ◁

Remark 10.28 (Average Direct Effects) According to Equation (10.35), once the *conditional* prima facie effects are identified, we can also identify the average direct effects by taking the expectation $E(PFE_{xx';W})$ of the W-conditional prima-facie-effect function

$$PFE_{xx';W} := E^{X=x}(Y|W) - E^{X=x'}(Y|W)$$

over the distribution of W, *provided, of course, that the conditional expectations* $E^{X=x}(Y|W)$ *and* $E^{X=x'}(Y|W)$ *are* τ_{x,t_M}*-unbiased.* ◁

Remark 10.29 (Conditional Direct-Effect Functions) If the conditional conditional expectations $E^{X=x}(Y|W)$ and $E^{X=x'}(Y|W)$ are both τ_{x,t_M}-unbiased, then the conditional prima facie effect function $PFE_{xx';W}$ is equal to the conditional direct-effect function $E(\delta_{xx',t_M}|W)$ [see Eq. (10.36)]. This implies that the conditional prima facie effect

$$PFE_{xx';W=w} := E(Y|X=x,W=w) - E(Y|X=x',W=w)$$

is also the conditional M-direct effect $E(\delta_{xx',t_M}|W=w)$ [see Eq. (10.37)]. ◁

If we are not only interested in the average direct effects and in the W-conditional direct effects given the variable W, but also in the $(V=v)$-*conditional direct effects* $E(\delta_{xx',t_M}|V=v)$, where V is any measurable function of W, then Equations (10.38) and (10.39) show how these $(V=v)$-conditional direct effects can be identified. For example, if $W=(M,Z)$, then $V=Z$ is such a measurable function of W. Hence, we can use

$$E(\delta_{xx',t_M}|Z) \underset{P}{=} E(PFE_{xx';M,Z}|Z) \tag{10.46}$$

[see Eq. (10.38)] and

$$E(\delta_{xx',t_M}|Z=z) = E(PFE_{xx';M,Z}|Z=z) \tag{10.47}$$

[see Eq. (10.39)], provided that $E^{X=x}(Y|M,Z)$ and $E^{X=x'}(Y|M,Z)$ are τ_{x,t_M}-unbiased.

Similarly, if $W=(M,Z)$ and $V=M$, we can use

$$E(\delta_{xx',t_M}|M) \underset{P}{=} E(PFE_{xx';M,Z}|M) \tag{10.48}$$

[see Eq. (10.38)] and

$$E(\delta_{xx',t_M}|M=m) = E(PFE_{xx';M,Z}|M=m) \tag{10.49}$$

[see Eq. (10.39)], provided that $E^{X=x}(Y|M,Z)$ and $E^{X=x'}(Y|M,Z)$ are τ_{x,t_M}-unbiased.

Example 10.30 (Multiple Linear Regression II) If X represents a treatment variable, then $E^{X=x}(Y|M,Z)$ is the conditional expectation of Y on M and Z in treatment x. If we assume that $E^{X=x}(Y|M,Z)$ is P-unique and that it is a linear function of $W=(M,Z)$, i. e.,

$$E^{X=x}(Y|M,Z) \underset{P}{=} \beta_0^{(x)} + \beta_1^{(x)} M + \beta_2^{(x)} Z, \tag{10.50}$$

then the conditional expectation of $E^{X=x}(Y|M,Z)$ on Z is

$$\begin{aligned} E[E^{X=x}(Y|M,Z)|Z] &\underset{P}{=} E\left(\beta_0^{(x)} + \beta_1^{(x)} M + \beta_2^{(x)} Z \,\middle|\, Z\right) && \text{[Eq. (10.50)]} \\ &\underset{P}{=} \beta_0^{(x)} + \beta_1^{(x)} E(M|Z) + \beta_2^{(x)} Z && \text{[Box PR-10.2 (v), (iii)].} \end{aligned}$$

If we additionally assume

$$E(M|Z) \underset{P}{=} \alpha_0 + \alpha_1 Z, \tag{10.51}$$

then

$$\begin{aligned} E[E^{X=x}(Y|M,Z)|Z] &\underset{P}{=} \beta_0^{(x)} + \beta_1^{(x)} E(M|Z) + \beta_2^{(x)} Z \\ &\underset{P}{=} \beta_0^{(x)} + \beta_1^{(x)} (\alpha_0 + \alpha_1 Z) + \beta_2^{(x)} Z \qquad \text{[Eq. (10.51)]} \\ &\underset{P}{=} (\beta_0^{(x)} + \beta_1^{(x)} \alpha_0) + (\beta_2^{(x)} + \beta_1^{(x)} \alpha_1) Z . \end{aligned}$$

Hence, in this case the conditional expectation of $E^{X=x}(Y|M,Z)$ on Z is a linear function of Z with intercept $\beta_0^{(x)} + \beta_1^{(x)} \alpha_0$ and slope $\beta_2^{(x)} + \beta_1^{(x)} \alpha_1$. ◁

Remark 10.31 (Average Direct Effect Given $X{=}x^*$) The additional implications of the RIT condition $\tau_{x,t_M} \vdash X \mid W$ are useful if we are not only interested in the average M-direct effects and in the conditional M-direct effects given a function V of intermediate variable M and covariate Z, but also in the *V-conditional direct effects* $E(\delta_{xx',t_M}|V{=}v)$, where V is any measurable function of (X,W). For example, if W is measurable with respect to (M,Z), then X is measurable with respect to (X,W) and may therefore take the role of V. Hence, under $\tau_{x,t_M} \vdash X|W$, where W is measurable with respect to (M,Z), we can use equations

$$E(\delta_{xx',t_M}|X) \underset{P}{=} E(PFE_{xx';W}|X) \tag{10.52}$$

and

$$E(\delta_{xx',t_M}|X{=}x^*) = E(PFE_{xx';W}|X{=}x^*) \tag{10.53}$$

for the $(X{=}x^*)$-conditional total effects. For a given pair (x,x') of values of X, the conditional expectation values $E(\delta_{xx',t_M}|X{=}x^*)$ may differ from $E(\delta_{xx',t_M}|X{=}x^\star)$ if $x^\star$ denotes a value of X that is different from x^*. ◁

Remark 10.32 (Propensity as a Covariate) Note that the conditional probability $\pi_x := P(X{=}x|M,Z)$ is measurable with respect to (M,Z). This function, which is called the *propensity of x* given (M,Z), can also play the role of the random variable W in Theorems 10.18 and 10.20. Hence, all equations in these two theorems involving W also hold for π_x taking the role of W, provided that $E^{X=x}(Y|\pi_x)$ is τ_{x,t_M}-unbiased.

Similarly, the vector $\pi := (\pi_1,\ldots,\pi_J)$, with $\pi_x := P(X{=}x|M,Z),\ x = 1,\ldots,J,$ is measurable with respect to (M,Z). This propensity vector π can also play the role of W in the theorems presented in section 10.2.2. Hence, all equations in these theorems involving W also hold for π taking the role of W. (For more details on propensities see chapter 11.) ◁

10.3 Summary and Conclusions

In this chapter, we treated several ways to adjust for the bias, some of which are based on the *adjustment theorems*. According to these theorems, τ_x-unbiasedness of the regression $E(Y|X,Z)$ does not only allow to identify conditional total effects, but also yields the average total effects by taking the expectation of the conditional *PFE*s over the distribution of the covariate. According to these theorems,

Box 10.1 Adjusting Strategies

(1) **Adjusting expectations and effects w.r.t. total effects**
Choose a (possibly multivariate) covariate Z pertaining to X such that the conditional conditional expectations $E^{X=x}(Y|Z)$ and $E^{X=x'}(Y|Z)$ are τ_x-unbiased. Then:

(a) the expectation of the conditional expectation $E^{X=x}(Y|Z)$ is the expectation of the true-outcome variable τ_x, i. e.,

$$E[E^{X=x}(Y|Z)] = E(\tau_x).$$

(b) the expectation of the difference $PFE_{xx';Z} := E^{X=x}(Y|Z) - E^{X=x'}(Y|Z)$ is the average total effect, i. e.,

$$E(PFE_{xx';Z}) = E(\delta_{xx'}).$$

(2) **Adjusting conditional expectations and effect functions w.r.t. total effects**
If $E^{X=x}(Y|Z)$ and $E^{X=x'}(Y|Z)$ are τ_x-unbiased and V is measurable w.r.t. Z, then the conditional expectation of the prima facie effect function $PFE_{xx';Z}$ on V is also the total-effect function $E(\delta_{xx'}|V)$, i. e.,

$$E(PFE_{xx';Z}|V) = E(\delta_{xx'}|V).$$

(3) **Adjusting expectations and effects w.r.t. direct effects**
Choose a (possibly multivariate) intermediate variable M w.r.t. X and Y and a (possibly multivariate) covariate Z pertaining to X such that the conditional conditional expectations $E^{X=x}(Y|M,Z)$ and $E^{X=x'}(Y|M,Z)$ are τ_{x,t_M}-unbiased. Then:

(a) the expectation of the conditional expectation $E^{X=x}(Y|M,Z)$ is the expectation of the true-outcome variable τ_{x,t_M}, i. e.,

$$E[E^{X=x}(Y|M,Z)] = E(\tau_{x,t_M}).$$

(b) the expectation of $PFE_{xx';M,Z} := E^{X=x}(Y|M,Z) - E^{X=x'}(Y|M,Z)$ is the average direct effect, i. e.,

$$E(PFE_{xx';M,Z}) = E(\delta_{xx',t_M}).$$

(4) **Adjusting conditional expectations and effect functions w.r.t. direct effects**
If $E^{X=x}(Y|M,Z)$ and $E^{X=x'}(Y|M,Z)$ are τ_{x,t_M}-unbiased and V is measurable w.r.t. (M,Z), then the conditional expectation of the prima facie effect function $PFE_{xx';M,Z}$ on V is also the M-direct-effect function $E(\delta_{xx',t_M}|V)$, i. e.,

$$E(PFE_{xx';M,Z}|V) = E(\delta_{xx'}|V).$$

assuming conditional regressive independence of the true outcomes from X, also allows identification of the conditional total effects $E(\delta_{xx'} \mid X{=}x^*)$ given a treatment condition x^* (see section 5.2.3 for details and substantive meaning of these concepts).

There are several possibilities for choosing an appropriate covariate. The first possibility is to choose a measured covariate, say Z, assuming that the conditional expectation $E(Y \mid X, Z)$ is τ_x-unbiased. In chapter 13 we will show that traditional analysis of covariance (ANCOVA) and generalized ANCOVA are particular cases of this strategy.

10.4 Proofs

Proof of Theorem 10.2

Assuming that $E^{X=x}(Y \mid Z)$ is τ_x-unbiased includes the assumption that $\tau_x = E^{X=x}(Y \mid C_X)$ is P-unique, which implies that the expectation $E(\tau_x)$ and the conditional expectation $E(\tau_x \mid Z)$ are well-defined. Hence, Equation (10.2) can be derived as follows:

$$\begin{aligned} E[E^{X=x}(Y \mid Z)] &= E[E(\tau_x \mid Z)] \qquad \text{[Eq. (10.1)]} \\ &= E(\tau_x). \qquad \text{[Box PR-10.2 (iv)]} \end{aligned}$$

(ii) Similarly,

$$\begin{aligned} E[E^{X=x}(Y \mid Z) \mid V] &\underset{P}{=} E[E(\tau_x \mid Z) \mid V] \qquad \text{[Eq. (10.1)]} \\ &\underset{P}{=} E(\tau_x \mid V). \qquad \text{[Box PR-10.2 (v)]} \end{aligned}$$

(iii) (10.4) directly follows from Equation (10.3) and the assumption $P(V{=}v) > 0$.

Proof of Theorem 10.4

If V is measurable with respect to (X, Z) and we assume $\tau_x \vdash X \mid Z$, then Equation (10.5) implies

$$\begin{aligned} E[E^{X=x}(Y \mid Z) \mid V] &\underset{P}{=} E[E(\tau_x \mid Z) \mid V] \qquad [\tau_x \vdash X \mid Z \Rightarrow \text{Eq. (10.1)}] \\ &\underset{P}{=} E[E(\tau_x \mid X, Z) \mid V] \qquad \text{[Eq. (10.5)]} \\ &\underset{P}{=} E(\tau_x \mid V) \qquad \text{[Box PR-10.2 (v)]}. \end{aligned}$$

If v is a value of V with $P(V{=}v) > 0$, this equation immediately implies

$$E(\tau_x \mid V{=}v) = E[E^{X=x}(Y \mid Z) \mid V{=}v].$$

Proof of Theorem 10.5

Assuming that $E^{X=x}(Y \mid Z)$ and $E^{X=x'}(Y \mid Z)$ are τ_x-unbiased also means that we assume that $\tau_x := E^{X=x}(Y \mid C_X)$ and $\tau_{x'} := E^{X=x'}(Y \mid C_X)$ are P-unique, implying that the expectations $E(\tau_x)$ and $E(\tau_{x'})$ as well as the conditional expectations $E(\tau_x \mid V)$, $E(\tau_{x'} \mid V)$, and $E(\delta_{xx'} \mid V)$ are well-defined.

(i) Equation (10.6) follows from Equation (10.2), because $\delta_{xx'} \underset{P}{=} \tau_x - \tau_{x'}$, which implies

$$E(\delta_{xx'}) \;=\; E(\tau_x) - E(\tau_{x'}),$$

and $PFE_{xx';Z} := E^{X=x}(Y \mid Z) - E^{X=x'}(Y \mid Z)$, which implies

$$\begin{aligned} E(PFE_{xx';Z}) \;&=\; E[E^{X=x}(Y \mid Z)] - E[E^{X=x'}(Y \mid Z)] \\ &=\; E(\tau_x) - E(\tau_{x'}) \qquad \text{[Eq. (10.2)]}. \end{aligned}$$

(ii) Equations (10.7) and (10.8) follow from Equations (10.9) and (10.10), respectively, because for $V = Z$, Equation (10.9) implies

$$\begin{aligned} E(\delta_{xx'} \mid Z) \;&\underset{P}{=}\; E(PFE_{xx';Z} \mid Z) \\ &\underset{P}{=}\; PFE_{xx';Z} \qquad \text{[Box PR-10.2 (vii)]}. \end{aligned}$$

(iii) Equation (10.9) follows from Equation (10.3), because

$$\begin{aligned} E(PFE_{xx';Z} \mid V) \;&\underset{P}{=}\; E[E^{X=x}(Y \mid Z) - E^{X=x'}(Y \mid Z) \mid V] \qquad \text{[def. of } PFE_{xx';Z}] \\ &\underset{P}{=}\; E[E^{X=x}(Y \mid Z) \mid V] - E[E^{X=x'}(Y \mid Z) \mid V] \qquad \text{[Box PR-10.2 (xvi)]} \\ &\underset{P}{=}\; E(\tau_x \mid V) - E(\tau_{x'} \mid V) \qquad \text{[Eq. (10.3)]} \\ &\underset{P}{=}\; E(\tau_x - \tau_{x'} \mid V) \qquad \text{[Box PR-10.2 (xvi)]} \\ &\underset{P}{=}\; E(\delta_{xx'} \mid V) \qquad \text{[def. of } \delta_{xx'}]. \end{aligned}$$

(iv) Equation (10.10) directly follows from (10.9) and the assumption $P(V = v) > 0$.

Proof of Theorem 10.7

If V is measurable with respect to (X, Z), then Equation (10.9) follows from $\tau_x \vdash X \mid Z$ and $\tau_{x'} \vdash X \mid Z$, because

$$\begin{aligned} E(PFE_{xx';Z} \mid V) \;&\underset{P}{=}\; E[E^{X=x}(Y \mid Z) - E^{X=x'}(Y \mid Z) \mid V] \qquad \text{[def. of } PFE_{xx';Z}] \\ &\underset{P}{=}\; E[E^{X=x}(Y \mid Z) \mid V] - E[E^{X=x'}(Y \mid Z) \mid V] \qquad \text{[Box PR-10.2 (xvi)]} \\ &\underset{P}{=}\; E[E(\tau_x \mid Z) \mid V] - E[E(\tau_{x'} \mid Z) \mid V] \qquad [\tau_x \vdash X \mid Z \Rightarrow \text{Eq. (10.1)}] \\ &\underset{P}{=}\; E[E(\tau_x \mid X, Z) \mid V] - E[E(\tau_{x'} \mid X, Z) \mid V] \qquad [\tau_x, \tau_{x'} \vdash X \mid Z] \\ &\underset{P}{=}\; E(\tau_x \mid V) - E(\tau_{x'} \mid V) \qquad \text{[Box PR-10.2 (v)]} \\ &\underset{P}{=}\; E(\tau_x - \tau_{x'} \mid V) \qquad \text{[Box PR-10.2 (xvi)]} \\ &\underset{P}{=}\; E(\delta_{xx'} \mid V) \qquad \text{[def. of } \delta_{xx'}]. \end{aligned}$$

Equation (10.10) directly follows from (10.9) and the assumption $P(V = v) > 0$.

Proof of Theorem 10.18

(i) Equation (10.31) can be derived as follows:

$$\begin{aligned} E[E^{X=x}(Y|W)] &= E[E(\tau_{x,t_M}|W)] \qquad \text{[Eq. (10.30)]} \\ &= E(\tau_{x,t_M}) \qquad \text{[Box PR-10.2 (iv)]}. \end{aligned}$$

(ii) Similarly,

$$\begin{aligned} E[E^{X=x}(Y|W)|V] &\underset{P}{=} E[E(\tau_{x,t_M}|W)|V] \qquad \text{[Eq. (10.30)]} \\ &\underset{P}{=} E(\tau_{x,t_M}|V) \qquad \text{[Box PR-10.2 (v)]}. \end{aligned}$$

(iii) (10.33) immediately follows from Equation (10.32) and $P(V=v)>0$.

Proof of Theorem 10.20

If V is measurable with respect to (X,W) and we assume $\tau_{x,t_M} \vdash X \mid W$, then

$$\begin{aligned} E[E^{X=x}(Y|W)|V] &\underset{P}{=} E[E(\tau_{x,t_M}|W)|V] \qquad [\tau_{x,t_M} \vdash X \mid W \Rightarrow \text{Eq. (10.30)}] \\ &\underset{P}{=} E[E(\tau_{x,t_M}|X,W)|V] \qquad \text{[Eq. (10.34)]} \\ &\underset{P}{=} E(\tau_{x,t_M}|V) \qquad \text{[(v) in Box PR-10.2]}. \end{aligned}$$

If v is a value of V with $P(V=v)>0$, the equation t $E(\tau_{x,t_M}|V) \underset{P}{=} E[E^{X=x}(Y|W)|V]$ immediately implies

$$E(\tau_{x,t_M}|V=v) = E[E^{X=x}(Y|W)|V=v].$$

Proof of Theorem 10.21

Equation (10.35) follows from Equation (10.31), because $\delta_{xx',t_M} := \tau_{x,t_M} - \tau_{x',t_M}$, which implies

$$E(\delta_{xx',t_M}) = E(\tau_{x,t_M}) - E(\tau_{x',t_M}).$$

Furthermore, $PFE_{xx';W} := E^{X=x}(Y|W) - E^{X=x'}(Y|W)$, which implies

$$\begin{aligned} E(PFE_{xx';W}) &= E[E^{X=x}(Y|W)] - E[E^{X=x'}(Y|W)] \\ &= E(\tau_{x,t_M}) - E(\tau_{x',t_M}) \qquad \text{[Eq. (10.31)]}. \end{aligned}$$

Equations (10.36) and (10.37) follow from Equation (10.38) and (10.39), respectively, because for $V=W$, Equation (10.38) yields

$$\begin{aligned} E(\delta_{xx',t_M}|W) &\underset{P}{=} E(PFE_{xx';V}|W) \\ &\underset{P}{=} PFE_{xx';W} \qquad \text{[Box PR-10.2 (vii)]}. \end{aligned}$$

Equation (10.38) follows from Equation (10.32), because

$$
\begin{aligned}
E(PFE_{xx';W} \mid V) &\underset{P}{=} E[E^{X=x}(Y \mid W) - E^{X=x'}(Y \mid W) \mid V] && \text{[def. of } PFE_{xx';W}\text{]} \\
&\underset{P}{=} E[E^{X=x}(Y \mid W) \mid V] - E[E^{X=x'}(Y \mid W) \mid V] && \text{[Box PR-10.2 (xvi)]} \\
&\underset{P}{=} E(\tau_{x,t_M} \mid V) - E(\tau_{x',t_M} \mid V) && \text{[Eq. (10.32)]} \\
&\underset{P}{=} E(\tau_{x,t_M} - \tau_{x',t_M} \mid V) && \text{[(Box PR-10.2 xvi)]} \\
&\underset{P}{=} E(\delta_{xx',t_M} \mid V) && \text{[def. of } \delta_{xx',t_M}\text{]}.
\end{aligned}
$$

Equation (10.39) directly follows from (10.38) and the assumption $P(V=v) > 0$.

Proof of Theorem 10.23

If V is measurable with respect to (X, W), then Equation (10.38) follows from $\tau_{x,t_M} \vdash X \mid W$, because

$$
\begin{aligned}
E(PFE_{xx';W} \mid V) &\underset{P}{=} E[E^{X=x}(Y \mid W) - E^{X=x'}(Y \mid W) \mid V] && \text{[def. of } PFE_{xx';W}\text{]} \\
&\underset{P}{=} E[(\tau_{x,t_M} \mid W) - E(\tau_{x',t_M} \mid W) \mid V] && [\tau_{x,t_M} \vdash X \mid W \Rightarrow \text{Eq. (10.30)}] \\
&\underset{P}{=} E[E(\tau_{x,t_M} \mid X, W) - E(\tau_{x',t_M} \mid X, W) \mid V] && \text{[Eq. (10.34)]} \\
&\underset{P}{=} E[E(\tau_{x,t_M} \mid X, W) \mid V] - E[E(\tau_{x',t_M} \mid X, W) \mid V] && \text{[Box PR-10.2 (xvi)]} \\
&\underset{P}{=} E(\tau_{x,t_M} \mid V) - E(\tau_{x',t_M} \mid V) && \text{[Box PR-10.2 (v)]} \\
&\underset{P}{=} E(\tau_{x,t_M} - \tau_{x',t_M} \mid V) && \text{[Box PR-10.2 (xvi)]} \\
&\underset{P}{=} E(\delta_{xx',t_M} \mid V) && \text{[def. of } \delta_{xx',t_M}\text{]}.
\end{aligned}
$$

Equation (10.39) directly follows from (10.38) and the assumption $P(V=v) > 0$.

10.5 Exercises

▷ **Exercise 10-1** Which is the content of Theorem 10.2?

▷ **Exercise 10-2** Why is it not possible to use Equation (10.6) to compute the average causal effect in the total population from the prima facie effects in the two subpopulations for the random experiment presented in Table 6.1?

▷ **Exercise 10-3** Show that $X \perp\!\!\!\perp \tau \mid Z$, where $\tau := (\tau_1, \ldots, \tau_J)$ and $\tau_x := E^{X=x}(Y \mid U)$, $x = 1, \ldots, J$, holds in the example presented in Table 10.1.

▷ **Exercise 10-4** Use Equation (10.6) to compute the average total effect in the total population from the average total effects in the two subpopulations for the random experiment presented in Table 10.1.

Solutions

▷ **Solution 10-1** According to Theorem 10.2, we cannot only compute the conditional, but also the unconditional total effects and the expectations $E(\tau_x)$ of the true-outcome variables τ_x from the conditional expectation $E(Y \mid X,Z)$ *provided that it is* τ_x*-unbiased*. Hence, even if the conditional expectation $E(Y \mid X)$ is biased, there is the possibility for valid causal inference.

▷ **Solution 10-2** It is not possible to use Equation (10.6) for the random experiment presented in Table 6.1 because, in this example, the conditional expectation $E(Y \mid X,Z)$ and, therefore, the conditional prima facie effects given Z, are biased.

▷ **Solution 10-3** If X is discrete with a finite number of values, each of which has a positive probability, the $X \perp\!\!\!\perp \tau \mid Z$ is equivalent to

$$P(X{=}1 \mid Z,\tau) \;=\; P(X{=}1 \mid Z).$$

Looking at Table 10.1 shows that the conditional treatment probability $P(X{=}1 \mid U)$ is identical with $P(X{=}1 \mid Z)$. Because τ is a function of U this implies $P(X{=}1 \mid Z,\tau) \;=\; P(X{=}1 \mid Z)$.

▷ **Solution 10-4** According to Equation (10.6) and because the two-dimensional covariate Z is discrete, we can use

$$E(\delta_{xx'}) \;=\; E(PFE_{xx';Z}) \;=\; \sum_{z} PFE_{xx';Z=z} \cdot P(Z{=}z).$$

For the example presented in Table 10.1 this yields:

$$\begin{aligned} E(\delta_{10}) \;&=\; PFE_{12;Z=(m,n)} \cdot 1/3 + PFE_{12;Z=(m,y)} \cdot 1/3 + PFE_{12;Z=(f,y)} \cdot 1/3 \\ &=\; 11.00 \cdot 1/3 + 5.00 \cdot 1/3 + 11.00 \cdot 1/3 \;=\; 9. \end{aligned}$$

Chapter 11
Propensities

In chapter 6, we have seen that the conditional expectation values $E(Y|X{=}x)$ and $E(Y|X{=}x,Z{=}z)$, the conditional expectation $E(Y|X)$ and $E(Y|X,Z)$ as well as the conditional and unconditional prima facie effects are biased, unless one of the sufficient conditions for unbiasedness holds (see chapters 8 and 9). In this chapter, we show how to adjust for bias. The crucial assumption is that the conditional expectation $E(Y|X,Z)$ is Z-conditionally unbiased, which implies that the *conditional* prima facie effects given a covariate Z are unbiased as well.

The specific goal of this chapter is to provide the theoretical foundation for a number of data analysis procedures such as the analysis of covariance and its generalizations, stratifying on covariates and on propensity scores, and weighting the outcome variable by propensity scores. These procedures will be treated in chapter 13.

Overview. We start with the adjustment theorems showing how to identify the expectations of the true outcome variables and the average causal effects if we can assume that the conditional expectation $E(Y|X,Z)$ is Z-conditionally unbiased. Then we turn to the true propensities which be can used as covariates, provided that they are known. Next, we turn to the Z-conditional propensities which can also be used as covariates in the adjustment theorems, provided that X and true outcomes are independent given the original covariate Z. In the last section we treat *weighting the outcome variable* by a function of the Z-conditional propensities. This is an alternative to *adjusting means* of the outcome variable which does not involve the conditional expectation $E(Y|X,Z)$.

11.1 True Propensities for Total Effects

In this section, we show that the probability functions $P(X{=}x|\mathscr{C}_X)$, which are called *true propensities of x*, may replace the covariates in the adjustment theorems. First, this can be useful in the conditionally randomized experiment, in which, by definition, the true propensity of x is under the control of the experimenter. Second, it can also be useful if a covariate Z is selected such that

$$P(X{=}x|\mathscr{C}_X) \underset{P}{=} P(X{=}x|Z). \tag{11.1}$$

Furthermore, in section 11.4 we show that the true propensities can also be used in weighting the outcome variable.

11.1.1 Theory

For simplicity, we use φ_x to denote a version of the conditional probability $P(X{=}x|\mathscr{C}_X)$, i. e.,

$$\varphi_x := P(X{=}x|\mathscr{C}_X). \tag{11.2}$$

We start showing that the independent cause condition $1_{X=x} \perp\!\!\!\perp_P \mathscr{C}_X|\varphi_x$ always holds provided that $P(X{=}x) > 0$. In this case $1_{X=x} \perp\!\!\!\perp_P \mathscr{C}_X|\varphi_x$ is equivalent to

$$P(X{=}x|\mathscr{C}_X) \underset{P}{=} P(X{=}x|\varphi_x) \tag{11.3}$$

(see PR-Th. 16.26 and PR-Rem. 10.4). Furthermore, for $\varphi := (\varphi_1,\ldots,\varphi_J)$, we show that the independent cause condition $X \perp\!\!\!\perp_P \mathscr{C}_X|\varphi$ always holds if $P(X{=}x) > 0$ for all $x = 0,1,\ldots,J$. Note that in this case $X \perp\!\!\!\perp_P \mathscr{C}_X|\varphi$ is equivalent to

$$P(X{=}x|\mathscr{C}_X) \underset{P}{=} P(X{=}x|\varphi), \quad \text{for all } x = 0,1,\ldots,J \tag{11.4}$$

(see again PR-Th. 16.26 and PR-Rem. 10.4).

Theorem 11.1 (Conditional Independence of X and $\mathscr{C}_X$)
Let $\langle(\Omega,\mathscr{A},P),(\mathscr{F}_t,t\in T),X,Y\rangle$ be a causality space, let Y be numerical and nonnegative or with finite expectation, let $x\in\Omega'_X$ be a value of X with $P(X{=}x)>0$, and let $\varphi_x := P(X{=}x|\mathscr{C}_X)$. Then:

(i) $1_{X=x} \perp\!\!\!\perp_P \mathscr{C}_X|\varphi_x$.
(ii) If X is discrete with a finite number of different values $x = 0,1,\ldots,J$ such that $P(X{=}x) > 0$ for all $x = 0,1,\ldots,J$, and if $\varphi := (\varphi_1,\ldots,\varphi_J)$ denotes the vector of true propensities, then $X \perp\!\!\!\perp_P \mathscr{C}_X|\varphi$.

(Proof p. 291)

According to this theorem, conditional independence of $1_{X=x}$ and $\mathscr{C}_X$ given φ_x, i. e., $1_{X=x} \perp\!\!\!\perp_P \mathscr{C}_X|\varphi_x$, as well as conditional independence of X and $\mathscr{C}_X$ given φ, i. e., $X \perp\!\!\!\perp_P \mathscr{C}_X|\varphi$, *always hold*, provided that X is discrete with a finite number of values $x = 0,1,\ldots,J$ such that $P(X{=}x) > 0$ for all $x = 0,1,\ldots,J$. This result is remarkable, because it only rests on assuming that X has a finite number of different values each of which has a positive probability. According to Theorem 7.11, we can conclude that the conditional expectations $E(Y|X,\varphi_x)$ and $E(Y|X,\varphi)$ are always unbiased.

Corollary 11.2 (Unbiasedness I)
Let the assumptions of Theorem 11.1 be true. Then:

(i) $E^{X=x}(Y \mid \varphi_x)$ *is unbiased.*
(ii) If X *is discrete with a finite number of values* $x = 0, 1, \ldots, J$ *such that* $P(X{=}x) > 0$ *for all* $x = 0, 1, \ldots, J$*, then* $E(Y \mid X, \varphi)$ *is unbiased.*

(Proof p. 291)

Before discussing the methodological implications of this corollary, let us study some further implications of Theorem 11.1.

Remark 11.3 In the next corollary we still assume that X is discrete with a finite number of different values $x = 0, 1, \ldots, J$ such that $P(X{=}x) > 0$ for all $x = 0, 1, \ldots, J$ and we consider the vector $\tau := (\tau_0, \tau_1, \ldots, \tau_J)$ of all $J+1$ true outcome variables τ_x with respect to total effects. In this case, $X \perp\!\!\!\perp_P \tau \mid \varphi$ is equivalent to

$$P(X{=}x \mid \tau, \varphi) \underset{P}{=} P(X{=}x \mid \varphi) \quad \text{for all } x = 0, 1, \ldots, J. \tag{11.5}$$

◁

Corollary 11.4 (Conditional Independence of X and τ_x)
Let the assumptions of Theorem 11.1 be true and let τ_x *denote the true outcome variable of* x *w.r.t. total effects. Then:*

(i) $1_{X=x} \perp\!\!\!\perp_P \tau_x \mid \varphi_x$.
(ii) If X *is discrete with a finite number of values* $x = 0, 1, \ldots, J$ *such that* $P(X{=}x) > 0$ *for all* $x = 0, 1, \ldots, J$*, and* $\tau := (\tau_0, \tau_1, \ldots, \tau_J)$*, then* $X \perp\!\!\!\perp_P \tau \mid \varphi$.

(Proof p. 291)

Hence, according to this corollary, $1_{X=x}$ and τ_x are *conditionally independent* given the true propensity φ_x. Furthermore, X and the vector τ are *conditionally independent* given the vector φ of the true propensities. According to Theorem PR-16.34 and Remark 16.35, this also implies that the true outcome variable τ_x is *mean independent* from $1_{X=x}$ given φ_x and that each τ_x is mean independent from X given φ, provided that X is discrete with positive probabilities for each of its values.

Corollary 11.5 (Conditional Mean Independence of τ_x from X)
Let the assumptions of Theorem 11.1 be true.

(i) Then $\tau_x \vdash 1_{X=x} \mid \varphi_x$*, i. e.,*

$$E(\tau_x \mid 1_{X=x}, \varphi_x) \underset{P}{=} E(\tau_x \mid \varphi_x). \tag{11.6}$$

(ii) If X is discrete with a finite number of values $x = 0,1,\dots,J$ such that $P(X{=}x) > 0$ for all $x = 0,1,\dots,J$, and if $\tau := (\tau_0,\tau_1,\dots,\tau_J)$, then $\tau \vdash X\,|\,\varphi$, i.e.,

$$E(\tau_x\,|\,X,\varphi) \underset{P}{=} E(\tau_x\,|\,\varphi) \quad \text{for each } x = 0,1,\dots,J. \tag{11.7}$$

According to Theorem 7.11, $\tau_x \vdash 1_{X=x}\,|\,\varphi_x$ implies unbiasedness of $E^{X=x}(Y\,|\,\varphi_x)$ and $\tau \vdash X\,|\,\varphi$ implies unbiasedness of each $E^{X=x}(Y\,|\,\varphi)$, for each $x = 0,1,\dots,J$.

Corollary 11.6 (Unbiasedness II)
Let the assumptions of Theorem 11.1 be true.

(i) Then $E^{X=x}(Y\,|\,\varphi_x)$ is unbiased, i.e.,

$$E^{X=x}(Y\,|\,\varphi_x) \underset{P}{=} E(\tau_x\,|\,\varphi_x). \tag{11.8}$$

(ii) If X is discrete with a finite number of values $x = 0,1,\dots,J$ such that $P(X{=}x) > 0$ for all $x = 0,1,\dots,J$, then each conditional expectation $E^{X=x}(Y\,|\,\varphi)$ is unbiased, i.e.,

$$E^{X=x}(Y\,|\,\varphi) \underset{P}{=} E(\tau_x\,|\,\varphi) \quad \text{for each } x = 0,1,\dots,J. \tag{11.9}$$

According to Theorem 10.2, unbiasedness of $E^{X=x}(Y\,|\,\varphi_x)$ implies that we can use this conditional expectation for identifying the expectation of the true outcome variable τ_x and for identifying the conditional expectation $E(\tau_x|V_x)$ of τ_x given any measurable mapping $V_x = f(X,\varphi_x)$ of X and φ_x. Similarly, unbiasedness of the conditional expectation $E^{X=x}(Y\,|\,\varphi)$ implies that we can use it for identifying the expectation of τ_x and for identifying the conditional expectation $E(\tau_x|V)$ of τ_x given any measurable mapping $V = f(X,\varphi)$ of X and φ.

Corollary 11.7 (Adjusting)
Let the assumptions of Theorem 11.1 be true.

(i) Then

$$E\big[E^{X=x}(Y\,|\,\varphi_x)\big] = E(\tau_x). \tag{11.10}$$

(ii) If $V_x = f(X,\varphi_x)$ is a measurable mapping of (X,φ_x), then

$$E\big[E^{X=x}(Y\,|\,\varphi_x)\,\big|\,V_x\big] \underset{P}{=} E(\tau_x\,|\,V_x). \tag{11.11}$$

(iii) If X is discrete with a finite number of values $x = 0,1,\dots,J$ such that $P(X{=}x) > 0$ for all $x = 0,1,\dots,J$, then

$$E\big[E^{X=x}(Y\,|\,\varphi)\big] = E(\tau_x) \quad \text{for each } x = 0,1,\dots,J. \tag{11.12}$$

(iv) If $V = f(X,\varphi)$ is a measurable mapping of (X,φ), then

$$E\big[E^{X=x}(Y|\,\varphi)\,\big|\,V\big] \underset{P}{=} E(\tau_x|\,V) \quad \textit{for each } x = 0,1,\ldots,J. \tag{11.13}$$

11.1.2 Methodological Implications

If the true propensities $\varphi_x := P(X{=}x|\mathscr{C}_X)$ are known, we may use them in order to identify the average total effects. If, e. g., there are only two values of X, say $X{=}0$ and $X{=}1$, then the conditional expectation $E(Y|X,\varphi_1)$, with $\varphi_1 := P(X{=}1|\mathscr{C}_X)$, will be unbiased and, therefore, Theorem 10.2 will also apply to φ_1 taking the role of the covariate. *No restrictive assumption is necessary*, other than $0 < P(X{=}1) < 1$. In contrast, using ordinary covariates, there is no guarantee at all for unbiasedness of the conditional expectation $E(Y|\,X,Z)$ even if Z is a vector consisting of many covariates.

Remark 11.8 (Identifying Average and Conditional Total Effects) According to Corollary 11.7, we can use the conditional expectations $E^{X=x}(Y|\,\varphi_x)$ to compute the expectations $E(\tau_x)$ of the true outcomes [see Eq.(11.10)] and also to compute the conditional expectations $E(\tau_x|\,V)$ of the true outcomes given any measurable mapping V of X and φ_x [see (11.11)]. Because $E(\delta_{xx'}) = E(\tau_x) - E(\tau_{x'})$ and $E(\delta_{xx'}|\,V) = E(\tau_x|\,V) - E(\tau_{x'}|\,V)$, the conditional expectations $E^{X=x}(Y|\varphi_x)$ can be also be used for identifying the average and conditional total effects. ◁

Remark 11.9 (Determining True Propensities) A prerequisite for using Theorem 11.1 and its corollaries is that the true propensities φ_x are known. This will be the case if they are under the control of the experimenter, i. e., if it is the experimenter who determines the true treatment probabilities. Note that these treatment probabilities do not have to be the same for each observational unit. Instead, the experimenter may determine them according to his diagnostic criteria of severity of symptoms, his rated probability of success for the specific patient, or other substantive or statistical optimality criteria (see, e. g., Pukelsheim, 2006; Berger & Wong, 2005). ◁

Remark 11.10 (True Propensities vs. Z-Conditional Propensities) In section 11.2, we will extend Theorem 11.1 and its corollaries to the case in which we use the Z-conditional propensities $P(X{=}x\,|\,Z)$ instead of the true propensities $P(X{=}x\,|\,\mathscr{C}_X)$, where Z denotes a (possibly multivariate) covariate. These Z-conditional propensities $P(X{=}x\,|\,Z)$ can be estimated in quasi-experiments and can be used in identifying the average and conditional total effects, provided that we can assume Z-conditional independence of X and the true outcome variables τ_x. Also note that the true propensities can only be utilized for identifying *total* effects. For the identification of *direct* effects see section 11.2. ◁

11.1.3 Numerical Example

Example 11.11 (continued) Let us use the example displayed in Table 10.1 to illustrate adjustment based on the true propensities. This example is constructed in such a way that

$$P(X{=}1\,|\,\mathscr{C}_X) \;=\; P(X{=}1\,|\,U). \tag{11.14}$$

In this specific example, there are three different individual probabilities (true propensities) for treatment 1: The first unit has probability $P(X{=}1\,|\,U{=}u_1) = 7/8$ for treatment 1, the next three units have the probability 5/8, and the last two units have the probability 2/8 for treatment 1. These probabilities are the values of the true propensity $\varphi_1 = P(X{=}1\,|\,C_X) = P(X{=}1\,|\,U)$, which can be used as a covariate. Because X is dichotomous with values 0 and 1, we only need one single function $P(X{=}1\,|\,U)$ of true propensities, i. e., $\varphi_1 = \varphi$.

According to Corollary 11.4, the conditional expectations $E^{X=x}(Y\,|\,\varphi)$ of the outcomes given the true propensity φ within treatment conditions are always unbiased. Because, in this example, φ has only three different values, these conditional expectations can always be can be parameterized as

$$E^{X=x}(Y\,|\,\varphi) \;=\; \gamma_0^{(x)} + \gamma_1^{(x)} \cdot I_{\varphi=5/8} + \gamma_2^{(x)} \cdot I_{\varphi=2/8}, \tag{11.15}$$

where $I_{\varphi=5/8}$ and $I_{\varphi=2/8}$ denote the indicator variables for $\varphi{=}5/8$ and $\varphi{=}2/8$, respectively. In treatment 0, this conditional expectation is

$$\begin{aligned} E^{X=0}(Y\,|\,\varphi) \;&=\; \gamma_0^{(0)} + \gamma_1^{(0)} \cdot I_{\varphi=5/8} + \gamma_2^{(0)} \cdot I_{\varphi=2/8} \\ &=\; 72 + 17 \cdot I_{\varphi=5/8} + 39 \cdot I_{\varphi=2/8}, \end{aligned} \tag{11.16}$$

and in treatment 1 it is

$$\begin{aligned} E^{X=1}(Y\,|\,\varphi) \;&=\; \gamma_0^{(1)} + \gamma_1^{(1)} \cdot I_{\varphi=5/8} + \gamma_2^{(1)} \cdot I_{\varphi=2/8} \\ &=\; 83 + 13 \cdot I_{\varphi=5/8} + 39 \cdot I_{\varphi=2/8}. \end{aligned} \tag{11.17}$$

The coefficients of these equations can be computed from the conditional expectations displayed in Table 11.1 as follows:

$$\begin{aligned} E(Y\,|\,X{=}x, \varphi{=}p) \;&=\; \sum_u E(Y\,|\,X{=}x, \varphi{=}p, U{=}u) \cdot P(U{=}u\,|\,X{=}x, \varphi{=}p) \\ &=\; \sum_u E(Y\,|\,X{=}x, U{=}u) \cdot P(U{=}u\,|\,X{=}x, \varphi{=}p) \end{aligned}$$

(see Exercise 11-4).

Because the conditional expectations $E^{X=x}(Y|\varphi)$ are unbiased, the prima facie effect function

$$\begin{aligned} PFE_{12;\varphi} \;&=\; E^{X=1}(Y\,|\,\varphi) - E^{X=0}(Y\,|\,\varphi) \\ &=\; 83 + 13 \cdot I_{\varphi=5/8} + 39 \cdot I_{\varphi=2/8} - (72 + 17 \cdot I_{\varphi=5/8} + 39 \cdot I_{\varphi=2/8}) \\ &=\; 11 - 4 \cdot I_{\varphi=5/8} \end{aligned} \tag{11.18}$$

Table 11.1. Expectations of the Outcome Variable Given Treatment and True Propensity in the Example of Table 10.1

	$\varphi := P(X{=}1 \mid c_X) = P(X{=}1 \mid U)$		
Treatment	7/8	5/8	2/8
$X{=}0$	72.00	89.00	111.00
$X{=}1$	83.00	96.00	122.00

is also the φ-conditional total effect function $E(\delta_{10} \mid \varphi)$. Hence, the conditional causal effect $E(\delta_{10} \mid \varphi{=}7/8)$ of treatment 1 as compared to the control given $\varphi = 7/8$ is

$$E(\delta_{10} \mid \varphi{=}7/8) = 11 - 4 \cdot 0 = 11,$$

the conditional causal effect $E(\delta_{10} \mid \varphi{=}5/8)$ of treatment 1 as compared to the control given $\varphi = 5/8$ is

$$E(\delta_{10} \mid \varphi{=}5/8) = 11 - 4 \cdot 1 = 7,$$

and

$$E(\delta_{10} \mid \varphi{=}2/8) = 11 - 4 \cdot 0 = 11.$$

According to Theorem 10.2, the average causal effect is

$$E(\delta_{10}) = E(PFE_{12;\varphi}) = E(11 - 4 \cdot I_{\varphi=5/8}) = 11 - 4 \cdot E(I_{\varphi=5/8}) = 11 - 4 \cdot 3/6 = 9.$$

This is exactly the same result as already obtained in example 10.16.

To summarize, this example shows how we can use the true propensity φ as a covariate for identifying average and conditional total effects. The values of the effect function $E(\delta_{10} \mid \varphi)$ are not only the conditional total effects given $\varphi(u) := P(X{=}1 \mid U{=}u)$, but we can also apply Theorem 10.2 in order to compute the average causal effect. As mentioned before, the conditional expectation $E(Y \mid X, \varphi)$ is *always* unbiased. Hence, in empirical applications, we only have to estimate (the values of) this conditional expectation. The drawback is, however, that the true propensities cannot be estimated in empirical applications. They must either be fixed by the experimenter or assumed to be identical to the Z-conditional propensities, which *can* be estimated. ◁

11.2 Conditional Propensities for Total Effects

True propensities can be used as covariates in adjusting for bias with respect to total effects, provided that the true propensities are known. This is the case only in the true experiment, in which, by definition of the true experiment, these probabilities are under full control of the experimenter. Now we show that we can use

the conditional propensities $P(X{=}x\,|\,Z)$ given the covariate Z instead of the true propensities $P(X{=}x\,|\,\mathscr{C}_X)$, provided that X and true outcome variables τ_x are Z-conditionally independent (see section 9.1).

11.2.1 Theory

The conditional probability $P(X{=}x\,|\,Z{=}z)$ is called the *propensity score* of $X{=}x$ given $Z{=}z$ (cf. Rosenbaum & Rubin, 1983a, 1984), and the function

$$\pi_x \;:=\; P(X{=}x\,|\,Z) \tag{11.19}$$

the (Z-conditional) propensity of x. Note that we only need J such propensities if there are $J+1$ different values of X. If X represents a treatment variable and there are only two treatment conditions, then $P(X{=}0\,|\,Z)$ can be computed from $P(X{=}1\,|\,Z)$, because $P(X{=}0\,|\,Z) = 1 - P(X{=}1\,|\,Z)$. Hence, we can gather the propensities in the J-dimensional vector $\pi := (\pi_1,\dots,\pi_J)$, keeping in mind that $\pi_0 = 1-(\pi_1+\dots+\pi_J)$. Also note that the vector $\pi := (\pi_1,\dots,\pi_J)$ of the propensities $\pi_x := P(X{=}x\,|\,Z)$, $x = 1,\dots,J$, as well as each π_x fulfill the requirements of covariates, i. e., each π_x and π are measurable w.r.t. $\mathscr{C}_X$, provided that Z itself is a covariate and is therefore measurable w.r.t. $\mathscr{C}_X$.

Remark 11.12 (Conditional Independence of X and True Outcome Variables)
In this section we assume that X is discrete with values $x = 0,1,\dots,J$. In this case, $1_{X=x} \perp\!\!\!\perp_P \tau_x\,|\,Z$ is equivalent to

$$P(X{=}x\,|\,Z,\tau_x) \;=_P\; P(X{=}x\,|\,Z). \tag{11.20}$$

If we consider the vector $\tau := (\tau_0,\tau_1,\dots,\tau_J)$ of all $J+1$ true outcome variables with respect to total effects τ_x, then $X \perp\!\!\!\perp_P \tau\,|\,Z$ is equivalent to

$$P(X{=}x\,|\,Z,\tau) \;=_P\; P(X{=}x\,|\,Z) \quad \text{for each } x = 0,1,\dots,J. \tag{11.21}$$

◁

Theorem 11.13
Let $\langle\,(\Omega,\mathscr{A},P),(\mathscr{F}_t,t\in T),X,Y\,\rangle$ be a causality space, let Y be numerical and nonnegative or with finite expectation, let X be discrete with a finite number of values $x = 0,1,\dots,J$ such that $P(X{=}x) > 0$ for all $x = 0,1,\dots,J$, let $\pi := (\pi_1,\dots,\pi_J)$ denote the vector of the propensities $\pi_x := P(X{=}x\,|\,Z)$, and let $\tau := (\tau_0,\tau_1,\dots,\tau_J)$ denote the vector of the true outcome variables with respect to total effects.

(i) If $1_{X=x} \perp\!\!\!\perp_P \tau_x\,|\,Z$, then

$$P(X{=}x\,|\,\tau_x,\pi_x) \;=_P\; P(X{=}x\,|\,\pi_x) \;=_P\; \pi_x. \tag{11.22}$$

(ii) If $X \perp\!\!\!\perp_P \tau \,|\, Z$, *then*

$$P(X{=}x \,|\, \tau, \pi) \underset{P}{=} P(X{=}x \,|\, \pi) \underset{P}{=} \pi_x \quad \forall x = 0, 1, \ldots, J. \tag{11.23}$$

(Proof p. 291)

Remark 11.14 (Conditional Independence of X and True Outcome Variables)
Hence, according to this theorem,

$$1_{X=x} \perp\!\!\!\perp_P \tau_x \,|\, Z \quad \text{implies} \quad 1_{X=x} \perp\!\!\!\perp_P \tau_x \,|\, \pi_x$$

and

$$X \perp\!\!\!\perp_P \tau \,|\, Z \quad \text{implies} \quad X \perp\!\!\!\perp_P \tau \,|\, \pi.$$

Furthermore, each of $1_{X=x} \perp\!\!\!\perp_P \tau_x \,|\, Z$ and $X \perp\!\!\!\perp_P \tau \,|\, Z$ implies

$$\pi_x := P(X{=}x \,|\, Z) \underset{P}{=} P(X{=}x \,|\, \pi) \underset{P}{=} P(X{=}x \,|\, \pi_x) \quad \forall x = 0, 1, \ldots, J. \tag{11.24}$$

Also note that $X \perp\!\!\!\perp_P \tau \,|\, Z$ implies $1_{X=x} \perp\!\!\!\perp_P \tau_x | Z$.

Of course, $\tau_x \perp\!\!\!\perp_P 1_{X=x} \,|\, Z$ also implies that the true outcome variable τ_x is *mean independent* from $1_{X=x}$ given π_x, i. e.,

$$E(\tau_x \,|\, 1_{X=x}, \pi_x) \underset{P}{=} E(\tau_x \,|\, \pi_x). \tag{11.25}$$

This equation is abbreviated by $\tau_x \vdash 1_{X=x} \,|\, \pi_x$. Furthermore, if X is discrete, then $\tau \perp\!\!\!\perp_P X \,|\, Z$ implies that each τ_x is *mean independent* from X given π, i. e.,

$$E(\tau_x \,|\, X, \pi) \underset{P}{=} E(\tau_x \,|\, \pi) \quad \forall x = 0, 1, \ldots, J. \tag{11.26}$$

◁

Corollary 11.15 (Conditional Regressive Independence of τ_x from X)
Let the assumptions of Theorem 11.13 be true.

(i) If $\tau_x \perp\!\!\!\perp_P 1_{X=x} \,|\, Z$, *then* $\tau_x \vdash 1_{X=x} \,|\, \pi_x$.
(ii) Let X *be discrete with a finite number of values* $x = 0, 1, \ldots, J$ *such that* $P(X{=}x) > 0$ *for all* $x = 0, 1, \ldots, J$. *Then* $\tau \perp\!\!\!\perp_P X \,|\, Z$ *implies* $\tau \vdash X \,|\, \pi$.

(Proof p. 292)

According to Theorem 7.11, $\tau_x \vdash 1_{X=x} \,|\, \pi_x$ implies unbiasedness of $E^{X=x}(Y \,|\, \pi_x)$ and $\tau \vdash X \,|\, \pi$ implies unbiasedness of each $E^{X=x}(Y \,|\, \pi)$, $x = 0, 1, \ldots, J$.

Corollary 11.16 (Unbiasedness II)
Let the assumptions of Theorem 11.13 be true.

(i) If $\tau_x \perp\!\!\!\perp_P 1_{X=x} \,|\, Z$, *then*

$$E^{X=x}(Y \,|\, \pi_x) \underset{P}{=} E(\tau_x \,|\, \pi_x). \tag{11.27}$$

(ii) Let X *be discrete with a finite number of values* $x = 0, 1, \ldots, J$ *such that* $P(X{=}x) > 0$ *for all* $x = 0, 1, \ldots, J$. *Then* $\tau \perp\!\!\!\perp_P X \,|\, Z$ *implies*

$$E^{X=x}(Y \,|\, \pi) \underset{P}{=} E(\tau_x \,|\, \pi) \quad \forall x = 0, 1, \ldots, J. \tag{11.28}$$

According to Theorem 10.2, unbiasedness of $E^{X=x}(Y \,|\, \pi_x)$ implies that we can use this conditional expectation for identifying the expectation of the true outcome variable τ_x and for identifying the conditional expectation $E(\tau_x|V_x)$ of τ_x given any measurable mapping $V_x = f(1_{X=x}, \pi_x)$ of $1_{X=x}$ and π_x. Similarly, unbiasedness of the conditional expectation $E^{X=x}(Y \,|\, \pi)$ implies that we can use it for identifying the expectation of τ_x and for identifying the conditional expectation $E(\tau_x|V)$ of τ_x given any measurable mapping $V = f(X, \pi)$ of X and π.

Corollary 11.17 (Adjusting)
Let the assumptions of Theorem 11.13 be true.

(i) If $\tau_x \perp\!\!\!\perp_P 1_{X=x} \,|\, Z$, *then*

$$E\big[E^{X=x}(Y \,|\, \pi_x)\big] = E(\tau_x). \tag{11.29}$$

(ii) Let $V_x = f(1_{X=x}, \pi_x)$ *be a measurable mapping of* $(1_{X=x}, \pi_x)$. *Then* $\tau_x \perp\!\!\!\perp_P 1_{X=x} \,|\, Z$ *implies*

$$E\big[E^{X=x}(Y \,|\, \pi_x) \,\big|\, V_x\big] \underset{P}{=} E(\tau_x \,|\, V_x). \tag{11.30}$$

(iii) Let X *be discrete with a finite number of values* $x = 0, 1, \ldots, J$ *such that* $P(X{=}x) > 0$ *for all* $x = 0, 1, \ldots, J$. *If* $\tau \perp\!\!\!\perp_P X \,|\, Z$, *then*

$$E\big[E^{X=x}(Y \,|\, \pi)\big] = E(\tau_x) \quad \textit{for each } x = 0, 1, \ldots, J. \tag{11.31}$$

(iv) Let $V = f(X, \pi)$ *be a measurable mapping of* (X, π). *If* $\tau \perp\!\!\!\perp_P X \,|\, Z$, *then*

$$E\big[E^{X=x}(Y \,|\, \pi) \,\big|\, V\big] \underset{P}{=} E(\tau_x \,|\, V) \quad \forall x = 0, 1, \ldots, J. \tag{11.32}$$

(Proof p. 292)

Because $E(\delta_{xx'}) = E(\tau_x) - E(\tau_{x'})$ and $E(\delta_{xx'} \,|\, V) \underset{P}{=} E(\tau_x \,|\, V) - E(\tau_{x'} \,|\, V)$, the conditional expectations $E^{X=x}(Y \,|\, \pi_x)$ can be also be used for identifying the average and conditional total effects.

11.2.2 Methodological Implications

Remark 11.18 (Propensities as Covariates in Mean Adjustment) If $1_{X=x}$ and the true outcome variable τ_x are conditionally independent given the covariate Z (i. e., if $\tau_x \perp\!\!\!\perp_P 1_{X=x} \mid Z$), we may use the propensity π_x instead of Z itself and still have conditional independence of τ_x and $1_{X=x}$ given π_x. Similarly, if X and the vector τ of true outcome variables are conditionally independent given the covariate Z (i. e., if $\tau \perp\!\!\!\perp_P X \mid Z$), we may use the vector π of propensities instead of Z itself and still have conditional independence of τ and X given π (see Th. 11.13).

If $\tau_x \perp\!\!\!\perp_P 1_{X=x} \mid Z$, then there is also mean independence of the true outcome variable τ_x from the indicator $1_{X=x}$ given π_x, and if $\tau \perp\!\!\!\perp_P X \mid Z$, then there is also mean independence of all true outcome variables τ_x from X given π (see Cor. 11.15). Furthermore, $\tau_x \perp\!\!\!\perp_P 1_{X=x} \mid Z$ implies that the conditional expectation $E^{X=x}(Y \mid \pi_x)$ is unbiased and $\tau \perp\!\!\!\perp_P X \mid Z$ implies that the conditional expectations $E^{X=x}(Y \mid \pi)$ are unbiased for all $x = 0, 1, \ldots, J$ (see Cor. 11.16). ◁

Remark 11.19 (Reduction of Dimensionality) Using the vector $\pi := (\pi_1, \ldots, \pi_J)$ of propensities instead of the original vector $Z := (Z_1, \ldots, Z_Q)$ of covariates can be advantageous. For a small number $J+1$ of values of X and a large number of covariates $Z_1, \ldots, Z_Q$ this may simplify the conditional expectations $E^{X=x}(Y \mid \pi_x)$ considerably, if compared to the conditional expectations $E^{X=x}(Y \mid Z)$. If, for instance, X represents a treatment variable and there are $J+1 = 2$ treatment conditions and five covariates $Z_1, \ldots, Z_5$, then conditional independence of τ and X given the covariate $Z := (Z_1, \ldots, Z_5)$ implies that τ and X are also conditionally independent given $\pi_1 := P(X{=}1 \mid Z)$. This fact allows us to compute the average total effect $E(\delta_{10})$ not only from the conditional expectations $E^{X=x}(Y \mid Z_1, \ldots, Z_5)$, $x = 0, 1$, but also from the conditional expectations $E^{X=x}(Y \mid \pi_1)$ and from the π_1-conditional prima facie effect function $PFE_{10;\pi_1}$ (see Theorems 10.2 and 10.5). Furthermore, the π_1-conditional total effect function $E(\delta_{10} \mid \pi_1)$ is identical to the π_1-conditional prima facie effect function $PFE_{10;\pi_1}$.

Note, however, that the conditional total effect function $E(\delta_{10} \mid Z)$ given the covariate Z may be more informative than the conditional total effect function $E(\delta_{10} \mid \pi_1)$ given π_1, because $\pi_1 := P(X{=}1 \mid Z)$ may not be a one-to-one function. Furthermore, the interpretation of π_1 may not be as easy as the interpretation of the covariate Z. If, e. g., Z is a pre-test, then we exactly know the meaning of the conditional treatment effect of a person scoring z on the pre-test. These conditional effects immediately answer questions such as 'Is the treatment effect on the outcome variable Y higher for those having a low score on the pre-test as compared to those having a high pre-test score?' In contrast, the meaning of the conditional treatment effect of a person scoring p on the propensity π_1 is less clear.

Finally, if there are many values of X and few covariates it will be more convenient and more informative to use the conditional expectations $E^{X=x}(Y \mid Z)$ rather than the conditional expectations $E^{X=x}(Y \mid \pi_x)$ in adjustment procedures for total effects. For example, if there are six treatment conditions and only one

single covariate Z for which $\tau \underset{P}{\perp\!\!\!\perp} X|Z$ holds, we would need five different propensities $\pi_x := P(X{=}x\,|\,Z)$ and six conditional expectations $E^{X=x}(Y|\pi_x)$ instead of just six conditional expectations $E^{X=x}(Y|Z)$. Because, in empirical applications, the propensities $\pi_x := P(X{=}x\,|\,Z)$ often have to estimated, using propensities instead of covariates is much more error-prone in such a case with many values of X and only few covariates. ◁

Remark 11.20 (Experiment) In empirical applications, we distinguish two different cases. In the *true experiment*, the true propensities are fixed by the experimenter, for example, as a function of a measured covariate, such as the rated *neediness* of a person for a therapy. The experimenter may decide that a person with a *high* need receives the treatment with probability 5/6, a person with a *medium* need receives the treatment with probability 3/6, and person with a *low* need receives the treatment with probability 1/6. If a person is categorized to have a *high* need, the experimenter tosses a dice and treats the subject unless he tosses a six, and he proceeds in the same way with persons belonging to the other two categories, except for the different treatment probabilities and associated decision rules: If a person is categorized to have a *medium* need, the experimenter treats the subject unless he tosses a four, five or six. If a person is rated having a low need, the experimenter treats the person if he tosses a one. In the true experiment, the true propensities $\varphi_x := P(X{=}x|\mathscr{C}_X)$ and the Z-conditional propensities $\pi_x := P(X{=}x\,|\,Z)$ are identical if Z denotes *neediness* with values *high*, *medium*, and *low*. ◁

Remark 11.21 (Quasi-Experiment) In a *quasi-experiment*, the propensities have to be estimated. In this case, there are two sources of errors. The *first* one is in the selection of the covariates. The critical question is: 'Did we select those covariates for which treatments and true outcomes are independent given these covariates?' The *second* source of error is in the specification of the function $P(X{=}x|Z)$. While a correct specification is simple if Z has only a few values such as *neediness* in the example above, it is difficult if Z has many values and/or is multivariate, i. e., $Z := (Z_1, \ldots, Z_Q)$. If, e. g., we assume

$$P(X{=}x\,|\,Z) \underset{P}{=} \frac{exp(\alpha_0 + \alpha_1 \cdot Z_1 + \ldots + \alpha_Q \cdot Z_Q)}{1 + exp(\alpha_0 + \alpha_1 \cdot Z_1 + \ldots + \alpha_Q \cdot Z_Q)}, \tag{11.33}$$

the critical question is, whether or not this assumption is correct, i. e.: 'Did we correctly specify the function $P(X{=}x\,|\,Z)$?' Of course, there are many other types of parametric functions for $P(X{=}x\,|\,Z)$. These functions may include quadratic and higher order terms as well as product terms representing interactions between different covariates. Furthermore, it is also possible that there is no simple parametric function at all that describes $P(X{=}x\,|\,Z)$. In fact, there are many alternatives to logistic regression for estimating $P(X{=}x\,|\,Z)$, such as neuronal networks (e. g., King & Zeng, 2001), boosted regressions (e. g., McCaffrey, Ridgeway, & Morral, 2004), classification trees (e. g., B. K. Lee, Lessler, & Stuart, 2010), and smoothing splines generalized additive models (e. g., Keele, 2008; Woo, Reiter, & Karr, 2008). ◁

Remark 11.22 (Balancing) In the next theorem we will show that π_x-conditional independence of the indicator $1_{X=x}$ and $\mathscr{C}_X$ implies π_x-conditional independence of the indicator $1_{X=x}$ and Z, which implies that the π_x-conditional distributions of all covariates are identical given $\{X{=}x\}$ and $\{X{\neq}x\}$. Of course, this has implications on the identity of conditional expectation values and other moments of the conditional distributions of the covariates. This fact is often used in propensity score adjustments in order to test the causality condition (strict ignorability) $1_{X=x} \perp\!\!\!\perp_P \mathscr{C}_X \,|\, \pi_x$. ◁

Theorem 11.23
Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a causality space, let Y be numerical and nonnegative or with finite expectation, let x denote a value of X with $P(X{=}x) > 0$, let $\pi_x := P(X{=}x \,|\, Z)$, and let $\varphi_x := P(X{=}x \,|\, \mathscr{C}_X)$. If Z is measurable w.r.t. $\mathscr{C}_X$ and $1_{X=x} \perp\!\!\!\perp_P \mathscr{C}_X \,|\, \pi_x$, then $1_{X=x} \perp\!\!\!\perp_P Z \,|\, \pi_x$ and

$$\pi_x \underset{P}{=} \varphi_x \,. \tag{11.34}$$

The fact that $1_{X=x} \perp\!\!\!\perp_P \mathscr{C}_X \,|\, \pi_x$ implies $1_{X=x} \perp\!\!\!\perp_P Z \,|\, \pi_x$ immediately follows from measurability of Z with respect to $\mathscr{C}_X$. Furthermore,

$$\begin{aligned}
\pi_x &:= P(X{=}x \,|\, Z) && [\text{def. of } \pi_x] \\
&\underset{P}{=} E[P(X{=}x \,|\, Z) \,|\, \pi_x] && [\pi_x := P(X{=}x \,|\, Z), \text{ PR-Box 10.2 (vii)}] \\
&\underset{P}{=} P(X{=}x \,|\, \pi_x) && [\pi_x := P(X{=}x \,|\, Z), \text{ PR-Box 10.2 (v)}] \\
&\underset{P}{=} P(X{=}x \,|\, \pi_x, \mathscr{C}_X) && [1_{X=x} \perp\!\!\!\perp_P \mathscr{C}_X \,|\, \pi_x, \text{ PR-(16.36)}] \\
&\underset{P}{=} P(X{=}x \,|\, \mathscr{C}_X) && [\pi_x \text{ is } \mathscr{C}_X\text{-measurable, PR-Rem. 10.47}] \\
&\underset{P}{=} \varphi_x \,. && [(11.2)]
\end{aligned}$$

11.2.3 Numerical Examples

We again use the example displayed in Table 10.1 to illustrate adjustment based on the Z-conditional propensities. Remember, this example is constructed such that

$$E^{X=x}(Y \,|\, \mathscr{C}_X) \underset{P}{=} E^{X=x}(Y \,|\, U) \tag{11.35}$$

and

$$P(X{=}1 \,|\, \mathscr{C}_X) \underset{P}{=} P(X{=}1 \,|\, U). \tag{11.36}$$

Table 10.1 gives an example for conditional independence of treatments and true outcomes given the covariate $Z := (Z_1, Z_2)$, where $Z_1 :=$ *sex* and $Z_2 :=$ *college*. In this table, the covariate Z is *two-dimensional*. Note that there is also conditional independence of X (here: the treatment variable) and the true outcomes given the *one-dimensional* propensity $\pi := P(X{=}1 \,|\, Z)$.

Table 11.2. Conditional expectations of Y given treatment and Z-conditional propensity scores in the example of Table 10.1

	$\pi := P(X{=}1 \mid Z)$		
Treatment	6/8	5/8	2/8
$X{=}0$	72.00	97.50	111.00
$X{=}1$	83.00	102.50	122.00

Example 11.24 (continued) Let us check conditional independence of X and true outcomes given the two-dimensional covariate. In this example, the covariate $Z := (Z_1, Z_2)$ has only three values, and the associated Z-conditional propensities (treatment probabilities) are: $P(X{=}1 \mid Z_1{=}m, Z_2{=}n) = 6/8$, $P(X{=}1 \mid Z_1{=}m, Z_2{=}y) = 5/8$, and $P(X{=}1 \mid Z_1{=}f, Z_2{=}y) = 2/8$.

Given $Z = (m,n)$, both true outcome variables τ_0 and τ_1 are constant, i. e.,

$$E^{X=0}(Y \mid U{=}u_1) \;=\; E^{X=0}(Y \mid U{=}u_2) = 72$$

and

$$E^{X=1}(Y \mid U{=}u_1) \;=\; E^{X=1}(Y \mid U{=}u_2) = 83.$$

Because, for $Z = (m,n)$, the vector $\tau := (\tau_0, \tau_1)$ of the true outcome variables is constant, the $Z = (m,n)$-conditional propensity does not depend on τ, i. e.,

$$P_{Z=(m,n)}(X{=}1 \mid \tau) \;=\; P_{Z=(m,n)}(X{=}1) \;=\; P[X{=}1 \mid Z{=}(m,n)]$$

[see Eq. (11.23)].

Given $Z = (m,y)$, the two individual treatment probabilities $P(X{=}1 \mid U{=}u_3) = P(X{=}1 \mid U{=}u_4) = 5/8$ are constant. Hence,

$$P_{Z=(m,y)}(X{=}1 \mid \tau) \;=\; P_{Z=(m,y)}(X{=}1) \;=\; P[X{=}1 \mid Z{=}(m,y)]$$

[see Eq. (11.23)]. Similarly, given $Z = (f,y)$, the conditional treatment probabilities $P_{Z=(f,y)}(X{=}1 \mid \tau) = P_{Z=(f,y)}(X{=}1) = P\big(X{=}1 \mid Z{=}(f,y)\big) = 1/4$ are constant. Hence, conditional independence of X and true outcomes given the covariates Z_1 and Z_2 holds. Exactly the same result is obtained checking conditional independence of X and true outcomes given the propensity π. ◁

Example 11.25 (continued) Because there is conditional independence of X and true outcomes given the two covariates Z_1 and Z_2, the conditional expectation $E(Y \mid X, \pi)$ is unbiased (see Cor. 11.16).

In this example, there are three different Z-conditional propensity scores for treatment 1: The first two units have the Z-conditional propensity score $P(X{=}1 \mid Z_1{=}m, Z_2{=}n) = 6/8$, whereas the next two units have the propensity score

$P(X{=}1 \mid Z_1{=}m, Z_2{=}y) = 5/8$, and the last two units have the propensity score $P(X{=}1 \mid Z_1{=}f, Z_2{=}y) = 2/8$ for treatment 1 (see Table 10.1). These propensities are the values of the propensity $\pi_1 = P(X{=}1 \mid Z)$, which, if known, can be used as a covariate, because there is independence of X and true outcomes given the two covariates Z_1 and Z_2. Because X is dichotomous with values 0 and 1, we only need one a single propensity $\pi_1 = \pi = P(X{=}1 \mid Z_1, Z_2)$.

According to Corollary 11.16, the conditional expectations $E^{X=x}(Y \mid \pi)$ of the outcome variable Y on the propensity π within treatment conditions are unbiased. Because, in this example, π has only three different values, these conditional expectations can always be can be parameterized as

$$E^{X=x}(Y \mid \pi) = \gamma_0^{(x)} + \gamma_1^{(x)} \cdot I_{\pi=5/8} + \gamma_2^{(x)} \cdot I_{\pi=2/8}, \tag{11.37}$$

where $I_{\pi=5/8}$ and $I_{\pi=2/8}$ denote the indicator variables for $\pi = 5/8$ and $\pi = 2/8$, respectively. In treatment 0, this conditional expectation is

$$\begin{aligned} E^{X=0}(Y \mid \pi) &= \gamma_0^{(0)} + \gamma_1^{(0)} \cdot I_{\pi=5/8} + \gamma_2^{(0)} \cdot I_{\pi=2/8} \\ &= 72 + 25.5 \cdot I_{\pi=5/8} + 39 \cdot I_{\pi=2/8}, \end{aligned} \tag{11.38}$$

and in treatment 1 it is

$$\begin{aligned} E^{X=1}(Y \mid \pi) &= \gamma_0^{(1)} + \gamma_1^{(1)} \cdot I_{\pi=5/8} + \gamma_2^{(1)} \cdot I_{\pi=2/8} \\ &= 83 + 19.5 \cdot I_{\pi=5/8} + 39 \cdot I_{\pi=2/8}. \end{aligned} \tag{11.39}$$

The coefficients of these conditional expectations can be computed from the conditional expectations displayed in Table 11.2 using

$$\begin{aligned} E(Y \mid X{=}x, \pi{=}p) &= \sum_u E(Y \mid X{=}x, \pi{=}p, U{=}u) \cdot P(U{=}u \mid X{=}x, \pi{=}p) \\ &= \sum_u E(Y \mid X{=}x, U{=}u) \cdot P(U{=}u \mid X{=}x, \pi{=}p) \end{aligned}$$

just in the same way as in example 11.11 (see also Exercise 11-4).

We have already observed that the conditional expectations $E^{X=x}(Y \mid \pi)$ are unbiased. Hence, the prima facie effect function

$$\begin{aligned} PFE_{12;\pi} &= E^{X=1}(Y \mid \pi) - E^{X=0}(Y \mid \pi) \\ &= 83 + 19.5 \cdot I_{\pi=5/8} + 39 \cdot I_{\pi=2/8} - (72 + 25.5 \cdot I_{\pi=5/8} + 39 \cdot I_{\pi=2/8}) \\ &= 11 - 6 \cdot I_{\pi=5/8} \end{aligned} \tag{11.40}$$

is also the conditional total effect function $E(\delta_{10} \mid \pi)$. Hence, the conditional total effect $E(\delta_{10} \mid \pi = 6/8)$ of treatment 1 compared to the control on the outcome variable Y given $\pi = 6/8$ is

$$E(\delta_{10} \mid \pi = 6/8) = 11 - 6 \cdot 0 = 11,$$

the conditional causal effect $E(\delta_{10} \mid \pi = 5/8)$ of treatment 1 compared to the control on the outcome variable Y given $\pi = 5/8$ is

$$E(\delta_{10}\,|\,\pi{=}5/8) \;=\; 11-6\cdot 1 \;=\; 5,$$

and

$$E(\delta_{10}\,|\,\pi{=}2/8) \;=\; 11-6\cdot 0 \;=\; 11.$$

According to Theorem 10.2, the unconditional average causal effect is

$$E(\delta_{10}) \;=\; E(PFE_{10;\pi}) \;=\; E(11-6\cdot I_{\pi=5/8}) \;=\; 11-6\cdot E(I_{\pi=5/8}) \;=\; 11-6\cdot 2/6 \;=\; 9,$$

which is the same result as already obtained before.

This example shows how we can use the propensity π_1 as a covariate. The values of the conditional total effect function $E(\delta_{10}\,|\,\pi_1)$ are not only the conditional total effects given $\pi_1 = p$, but we can also apply Theorem 10.5 in order to compute the average total effect. As mentioned before, in contrast to the true propensities, the Z-conditional propensities can be estimated in empirical applications. Hence, we can use this procedure also in quasi-experiments. The drawback is, however, that the conditional expectation $E(Y|X,\pi_1)$ is not necessarily unbiased. It is unbiased, however, if we can assume conditional independence of X and true outcomes given the covariate Z. In a true experiment, this condition can be created by Z-conditional randomized assignment of the unit to a treatment condition, and this is true even if Z is fallible, i. e., even if $Z = f(U) + \varepsilon$, where ε represents the measurement error variable. ◁

11.3 Conditional Propensities for Direct Effects

Now we show that we can use the conditional propensities $P(X{=}x\,|\,Z,M)$ given the covariate Z and an intermediate variable M for adjusting with respect to direct effects. A sufficient assumption is that X and true outcome variables with respect to direct effects τ_{x,t_M} are (Z,M)-conditionally independent.

11.3.1 Theory

We will call the conditional probability $P(X{=}x\,|\,M{=}m,Z{=}z)$ the *propensity score* for $X{=}x$ given $M{=}m$ and $Z{=}z$. Similarly, the function

$$\pi_{x;M} \;:=\; P(X{=}x\,|\,M,Z) \tag{11.41}$$

will be called *the* (Z,M)*-conditional propensity of* x. Again, we can gather the propensities in the J-dimensional vector

$$\pi_M \;:=\; (\pi_{1;M},\ldots,\pi_{J;M}), \tag{11.42}$$

keeping in mind that $\pi_{0;M} = 1-(\pi_{1;M}+\cdots+\pi_{J;M})$. Note, that neither the propensities $\pi_{x;M}$ nor the vector π_M are covariates, i. e., they are not measurable with respect to $\mathscr{C}_X$.

Remark 11.26 (Conditional Independence of X and True Outcome Variables) Again, we assume that X is discrete with values $x = 0, 1, \ldots, J$. We will now consider the true outcome variables *with respect to direct effects*, which are denoted by τ_{x,t_M}. If X is discrete with values $x = 0, 1, \ldots, J$, then $1_{X=x} \perp\!\!\!\perp_P \tau_{x,t_M} | M, Z$ is equivalent to

$$P(X{=}x \mid M, Z, \tau_{x,t_M}) \underset{P}{=} P(X{=}x \mid M, Z). \tag{11.43}$$

If we consider the vector

$$\tau_M := (\tau_{0,t_M}, \tau_{1,t_M}, \ldots, \tau_{J,t_M}) \tag{11.44}$$

of all $J+1$ true outcome variables with respect to direct effects, then $X \perp\!\!\!\perp_P \tau_M | M, Z$ is equivalent to

$$P(X{=}x \mid M, Z, \tau_M) \underset{P}{=} P(X{=}x \mid M, Z) \quad \text{for } x = 0, 1, \ldots, J. \tag{11.45}$$

◁

Theorem 11.27
Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a numerical causality space, let x be a value of X with $P(X{=}x) > 0$, let $\pi_{x;M} := P(X{=}x \mid M, Z)$, and let $\tau_{x,t_M} := E^{X=x}(Y | M, \mathscr{C}_X)$ denote the true outcome variable of x with respect to direct effects.

(i) Then $1_{X=x} \perp\!\!\!\perp_P \tau_{x,t_M} | M, Z$ implies $1_{X=x} \perp\!\!\!\perp_P \tau_{x,t_M} | \pi_{x;M}$ and

$$P(X{=}x | \tau_{x,t_M}, \pi_{x;M}) \underset{P}{=} P(X{=}x | \pi_{x;M}) \underset{P}{=} \pi_{x;M}. \tag{11.46}$$

(ii) If X is discrete with values $x = 0, 1, \ldots, J$, then $X \perp\!\!\!\perp_P \tau_M | M, Z$ implies $X \perp\!\!\!\perp_P \tau_M | \pi_M$ and

$$P(X{=}x | \tau_M, \pi_M) \underset{P}{=} P(X{=}x | \pi_M) \underset{P}{=} \pi_{x;M}. \tag{11.47}$$

(Proof p. 292)

Remark 11.28 (Conditional Independence of X and True Outcome Variables) According to this theorem, $1_{X=x} \perp\!\!\!\perp_P \tau_{x,t_M} | M, Z$ implies $1_{X=x} \perp\!\!\!\perp_P \tau_{x,t_M} | \pi_{x;M}$. Furthermore, $1_{X=x} \perp\!\!\!\perp_P \tau_{x,t_M} | M, Z$ implies

$$\pi_{x;M} := P(X{=}x \mid M, Z) \underset{P}{=} P(X{=}x | \pi_{x;M}). \tag{11.48}$$

Similarly, $X \perp\!\!\!\perp_P \tau_M | M, Z$ implies $X \perp\!\!\!\perp_P \tau_M | \pi_M$, and $X \perp\!\!\!\perp_P \tau_M | M, Z$ implies

$$\pi_M := P(X{=}x \mid M, Z) \underset{P}{=} P(X{=}x | \pi_M). \tag{11.49}$$

Of course, $\tau_{x,t_M} \perp\!\!\!\perp_P 1_{X=x} \mid M, Z$ also implies that the true outcome variable τ_{x,t_M} is *mean independent* from $1_{X=x}$ given $\pi_{x;M}$, i. e., it implies

$$E(\tau_{x,t_M} | 1_{X=x}, \pi_{x;M}) \underset{P}{=} E(\tau_{x,t_M} | \pi_{x;M}). \tag{11.50}$$

This equation is abbreviated by $\tau_{x,t_M} \vdash 1_{X=x} \mid \pi_{x;M}$. Furthermore, if X is discrete with values $x = 0, 1, \ldots, J$, then $\tau_M \perp\!\!\!\perp_P X \mid Z$ implies that each τ_{x,t_M} is *mean independent* from X given π_M, i. e.,

$$E(\tau_{x,t_M} | X, \pi_M) \underset{P}{=} E(\tau_{x,t_M} | \pi_M) \quad \text{for } x = 0, 1, \ldots, J. \tag{11.51}$$

◁

Corollary 11.29 (Conditional Mean Independence)
Let the assumptions of Theorem 11.27 be true.

(i) Then $\tau_{x,t_M} \perp\!\!\!\perp_P 1_{X=x} \mid M, Z$ *implies* $\tau_{x,t_M} \vdash 1_{X=x} \mid \pi_{x;M}$.

(ii) If X *is discrete with values* $x = 0, 1, \ldots, J$, *then* $\tau_M \perp\!\!\!\perp_P X \mid M, Z$ *implies* $\tau_M \vdash X \mid \pi_M$.

(Proof p. 292)

According to Theorem ???, $\tau_{x,t_M} \vdash 1_{X=x} \mid \pi_{x;M}$ implies unbiasedness of the $(X{=}x)$-conditional conditional expectation $E^{X=x}(Y \mid \pi_{x;M})$, and $\tau_M \vdash X \mid \pi_M$ implies unbiasedness of $E^{X=x}(Y \mid \pi_M)$.

Corollary 11.30 (Unbiasedness II)
Let the assumptions of Theorem 11.27 be true. Then $\tau_{x,t_M} \perp\!\!\!\perp_P 1_{X=x} \mid M, Z$ *implies that* $E^{X=x}(Y \mid \pi_{x;M})$ *is unbiased, i. e.,*

$$E^{X=x}(Y \mid \pi_{x;M}) \underset{P}{=} E(\tau_{x,t_M} \mid \pi_{x;M}). \tag{11.52}$$

According to Theorem 10.2, unbiasedness of $E^{X=x}(Y|\pi_{x;M})$ implies that we can use this conditional expectation for identifying the expectation of the true outcome variable τ_{x,t_M} and for identifying the conditional expectation $E(\tau_{x,t_M}|\pi_{x;M})$ of τ_{x,t_M} on any measurable mapping $V_x = f(1_{X=x}, \pi_{x;M})$ of $1_{X=x}$ and $\pi_{x;M}$.

Similarly, unbiasedness of the conditional expectation $E^{X=x}(Y \mid \pi_M)$ implies that we can use it for identifying the expectation of τ_x and for identifying the conditional expectation $E(\tau_{x,t_M}|V)$ of τ_{x,t_M} on any measurable mapping $V = f(X, \pi_M)$ of X and π_M.

Although the conditional expectations $E^{X=x}(Y|\pi_{x;M})$ and $E^{X=x}(Y|\pi_M)$ are both unbiased, they are not necessarily identical, because the σ-algebra generated by $\pi_{x;M}$ is a subset of the σ-algebra generated by π_M. Hence, conditioning on π_M can be more informative.

Corollary 11.31 (Adjusting)
Let the assumptions of Theorem 11.27 be true.

(i) If $\tau_{x,t_M} \perp\!\!\!\perp_P 1_{X=x} \mid M, Z$, *then*

$$E\big[E^{X=x}(Y \mid \pi_{x;M})\big] = E(\tau_{x,t_M}). \tag{11.53}$$

(ii) Let $V_x = f(1_{X=x}, \pi_{x;M})$ *be a measurable mapping of* $(1_{X=x}, \pi_{x;M})$. *Then* $\tau_{x,t_M} \perp\!\!\!\perp_P 1_{X=x} \mid M, Z$ *implies*

$$E\big[E^{X=x}(Y \mid \pi_{x;M}) \,\big|\, V_x\big] \overset{=}{}_P E(\tau_{x,t_M} \mid V_x). \tag{11.54}$$

(iii) Let X *be discrete with values* $x = 0, 1, \ldots, J$. *If* $\tau_M \perp\!\!\!\perp_P X \mid M, Z$, *then*

$$E\big[E^{X=x}(Y \mid \pi_M)\big] = E(\tau_{x,t_M}) \quad \text{for } x = 0, 1, \ldots, J. \tag{11.55}$$

(iv) Let $V = f(X, \pi_M)$ *be a measurable mapping of* (X, π_M). *If* $\tau_M \perp\!\!\!\perp_P X \mid M, Z$, *then*

$$E\big[E^{X=x}(Y \mid \pi_M) \,\big|\, V\big] \overset{=}{}_P E(\tau_{x,t_M} \mid V) \quad \text{for } x = 0, 1, \ldots, J. \tag{11.56}$$

(Proof p. 292)

Remark 11.32 (Identifying Average and Conditional Direct Effects) Because

$$E(\delta_{xx',t_M}) = E(\tau_{x,t_M}) - E(\tau_{x',t_M}),$$

$$E(\delta_{xx',t_M} \mid V_x) \overset{=}{}_P E(\tau_{x,t_M} \mid V_x) - E(\tau_{x',t_M} \mid V_x),$$

and

$$E(\delta_{xx',t_M} \mid V) \overset{=}{}_P E(\tau_{x,t_M} \mid V) - E(\tau_{x',t_M} \mid V),$$

the conditional conditional expectations $E^{X=x}(Y \mid \pi_{x;M})$ and $E^{X=x}(Y \mid \pi_M)$ can be also be used for identifying average and conditional direct effects. We just have to insert the corresponding equations displayed in Corollary 11.31. Note again that if $J > 1$, then considering V-conditional effects may be more informative than V_x-conditional effects; whereas V_x is a mapping of $(1_{X=x}, \pi_{x;M})$, the mapping V is more fine-grained, because it is a mapping of (X, π_M) and therefore may depend on *all* values of X and *all* propensities $\pi_{x;M}$, $x = 1, \ldots, J$. ◁

11.3.2 Methodological Implications

Remark 11.33 (Propensities in Mean Adjustment) If $1_{X=x}$ and the true outcome variable with respect to direct effects, which is denoted τ_{x,t_M}, are conditionally independent given intermediate variable M and covariate Z (i. e., if $\tau_{x,t_M} \perp\!\!\!\perp_P 1_{X=x} \mid M, Z$), we may use the propensity $\pi_{x;M}$ instead of (M, Z) and still have conditional independence of τ_{x,t_M} and $1_{X=x}$ given $\pi_{x;M}$. Similarly, if X and the vector

τ_M of true outcome variables with respect to direct effects are conditionally independent given the covariate (M,Z) (i. e., if $\tau_M \underset{P}{\perp\!\!\!\perp} X \mid M,Z$), we may use the vector π_M of propensities instead of (M,Z) and still have conditional independence of τ_M and X given π_M (see Th. 11.27).

If $\tau_x \underset{P}{\perp\!\!\!\perp} 1_{X=x} | Z$, then there is also mean independence of the true outcome variable τ_x from the indicator $1_{X=x}$ given π_x, and if $\tau \underset{P}{\perp\!\!\!\perp} X \mid Z$, then there is also mean independence of all true outcome variables τ_x from X given π (see Cor. 11.15). Furthermore, $\tau_x \underset{P}{\perp\!\!\!\perp} 1_{X=x} \mid Z$ implies that the conditional expectation $E^{X=x}(Y \mid \pi_x)$ is unbiased and $\tau \underset{P}{\perp\!\!\!\perp} X \mid Z$ implies that all conditional expectations $E^{X=x}(Y \mid \pi)$ are unbiased (see Cor. 11.16). ◁

Remark 11.34 (Reduction of Dimensionality) Using the vector $\pi := (\pi_1,\ldots,\pi_J)$ of propensities instead of the original vector $Z := (Z_1,\ldots,Z_Q)$ of covariates can be advantageous. For a small number $J+1$ of values of X and a large number of covariates $Z_1,\ldots,Z_Q$ this may simplify the conditional expectations $E^{X=x}(Y \mid \pi_x)$ considerably, if compared to the conditional expectations $E^{X=x}(Y \mid Z)$. If, for instance, X represents a treatment variable and there are $J+1=2$ treatment conditions and five covariates $Z_1,\ldots,Z_5$, then conditional independence of τ and X given the covariate $Z := (Z_1,\ldots,Z_5)$ implies that τ and X are also conditionally independent given $\pi_1 := P(X{=}1 \mid Z)$. This fact allows to compute the average total effect $E(\delta_{10})$ not only from the conditional expectations $E^{X=x}(Y \mid Z_1,\ldots,Z_5)$, $x=0,1$, but also from the conditional expectations $E^{X=x}(Y \mid \pi_1)$ and from the π_1-conditional prima facie effect function $PFE_{10;\pi_1}$ (see Theorems 10.2 and 10.5). Furthermore, the π_1-conditional total effect function $E(\delta_{10} \mid \pi_1)$ is identical to the π_1-conditional prima facie effect function $PFE_{10;\pi_1}$.

Note, however, that the conditional total effect function $E(\delta_{10} \mid Z)$ given the covariate Z may be more informative than the conditional total effect function $E(\delta_{10} \mid \pi_1)$ given π_1, because $\pi_1 := P(X{=}1 \mid Z)$ may not be a one-to-one function. Furthermore, the substantive interpretation of π_1 may not be as easy as the interpretation of the covariate Z. If, e. g., Z is a pre-test, then we exactly know the meaning of the conditional treatment effect of a person scoring z on the pre-test. These conditional effects immediately answer questions such as 'Is the treatment effect on the outcome variable Y higher for those having a low score on the pre-test as compared to those having a high pre-test score?' In contrast, the meaning of the conditional treatment effect of a person scoring p on the propensity π_1 is less clear.

Finally, if there are many values of X and few covariates it will be more convenient and more informative to use the conditional expectations $E^{X=x}(Y|Z)$ rather than the conditional expectations $E^{X=x}(Y \mid \pi_x)$ in adjustment procedures with respect to total effects. For example, if there are six treatment conditions and only one single covariate Z for which $\tau \underset{P}{\perp\!\!\!\perp} X \mid Z$ holds, we would need five different propensities $\pi_x := P(X{=}x \mid Z)$ and six conditional expectations $E^{X=x}(Y \mid \pi_x)$ instead of just six conditional expectations $E^{X=x}(Y|Z)$. Because, in empirical applications, the propensities $\pi_x := P(X{=}x|Z)$ often have to estimated, using propen-

sities instead of covariates is much more error-prone in such a case with many values of X and only few covariates. ◁

Remark 11.35 (Experiment) In empirical applications, we distinguish two different cases. In the *true experiment*, the true propensities are fixed by the experimenter, for example, as a function of a measured covariate, such as the rated *neediness* of a person for a therapy. The experimenter may decide that a person with a *high* need receives the treatment with probability 5/6, a person with a *medium* need receives the treatment with probability 3/6, and person with a *low* need receives the treatment with probability 1/6. If a person is categorized to have a *high* need, the experimenter tosses a dice and treats the subject unless he tosses a six, and he proceeds in the same way with persons belonging to the other two categories, except for the different treatment probabilities and associated decision rules: If a person is categorized to have a *medium* need, the experimenter treats the subject unless he tosses a four, five or six. If a person is rated having a low need, the experimenter treats the person if he tosses a one. In the true experiment, the true propensities $\varphi_x := P(X{=}x|\mathscr{C}_X)$ and the Z-conditional propensities $\pi_x := P(X{=}x\,|\,Z)$ are identical if Z denotes *neediness* with values *high*, *medium*, and *low*. ◁

Remark 11.36 (Quasi-Experiment) In a *quasi-experiment*, the propensities have to be estimated. In this case, there are two sources of errors. The *first* one is in the selection of the covariates. The critical question is: 'Did we select those covariates for which treatments and true outcomes are independent given these covariates?' The *second* source of error is in the specification of the function $P(X{=}x|Z)$. While a correct specification is simple if Z has only a few values such as *neediness* in the example above, it is difficult if Z has many values and/or is multivariate, i. e., $Z := (Z_1,\ldots,Z_Q)$. If, e. g., we assume that the propensity is a linear logistic regression, i. e.,

$$P(X{=}x\,|\,Z) \;=\; \frac{exp(\alpha_0+\alpha_1\cdot Z_1+\ldots+\alpha_Q\cdot Z_Q)}{1+exp(\alpha_0+\alpha_1\cdot Z_1+\ldots+\alpha_Q\cdot Z_Q)}, \tag{11.57}$$

the critical question is, whether or not this assumption is correct, i. e.: 'Did we correctly specify the function $P(X{=}x\,|\,Z)$?' Of course, there are many other types of parametric functions for $P(X{=}x\,|\,Z)$, which may include quadratic and higher order terms as well as product terms representing interactions between different covariates. Furthermore, it is also possible that there is no simple parametric function at all that describes $P(X{=}x\,|\,Z)$. In fact, there are many alternatives to logistic conditional expectation for estimating $P(X{=}x\,|\,Z)$, such as neuronal networks (e. g., King & Zeng, 2001), boosted conditional expectations (e. g., McCaffrey et al., 2004), classification trees (e. g., B. K. Lee et al., 2010), and smoothing splines generalized additive models (e. g., Keele, 2008; Woo et al., 2008). ◁

11.4 Weighting the Outcome Variable

11.4.1 Adjusting for Total Effects by Weighting the Outcome Variable

Now we will show how to compute the average total effects by choosing appropriate weights for the outcome variable. The only assumption required is that the conditional expectation $E(Y|X,Z)$ is unbiased with respect to total effects.

11.4.2 Theory

We will formulate the Weighting Theorem for the case in which X has $J+1$ values $x = 0, 1, \ldots, J$. The *weighted outcome variable* referred to in this theorem is:

$$Y_w := Y \cdot \left(\sum_{x=0}^{J} 1_{X=x} \cdot \frac{P(X=x)}{P(X=x|Z)} \right). \tag{11.58}$$

Note that the weighting variable, by which Y is multiplied, is a function of X and Z (see Exercise 11-9). The crucial ingredients of this weighting variable are the Z-conditional propensities $\pi_x := P(X=x|Z)$, the values of which are the conditional probabilities $P(X=x|Z=z)$. The following lemma will be used in proving the Weighting Theorem.

Lemma 11.37
Let Y_w denote the weighted outcome variable defined in Equation (11.58), let $E^{X=x}(Y_w|Z)$ denote the conditional expectation of Y_w on Z w.r.t. the probability measure $P^{X=x}$, and let $P(X=x, Z=z) > 0$, for all pairs of values of X and Z. Then:

$$E^{X=x}(Y_w|Z) \underset{P^{X=x}}{=} E^{X=x}(Y|Z) \cdot \frac{P(X=x)}{P(X=x|Z)} \tag{11.59}$$

(Proof p. 292)

Consider

$$\begin{aligned}
E^{X=x}(Y_w|Z) &\underset{P^{X=x}}{=} E^{X=x}\left(Y \cdot \sum_{x'=0}^{J} I_{X=x'} \cdot \frac{P(X=x')}{P(X=x'|Z)} \,\middle|\, Z \right) && \text{[def. of } Y_w\text{]} \\
&\underset{P^{X=x}}{=} E^{X=x}\left(\sum_{x'=0}^{J} I_{X=x'} \cdot Y \cdot \frac{P(X=x')}{P(X=x'|Z)} \,\middle|\, Z \right) \\
&\underset{P^{X=x}}{=} \sum_{x'=0}^{J} E^{X=x}\left(I_{X=x'} \cdot Y \cdot \frac{P(X=x')}{P(X=x'|Z)} \,\middle|\, Z \right) && \text{[(??) in Box ??]} \\
&\underset{P^{X=x}}{=} E^{X=x}\left(Y \cdot \frac{P(X=x)}{P(X=x|Z)} \,\middle|\, Z \right) \\
&\underset{P^{X=x}}{=} E^{X=x}(Y|Z) \cdot \frac{P(X=x)}{P(X=x|Z)} && \text{[(??) in Box ??]}
\end{aligned}$$

This proves Lemma 11.37.

Theorem 11.38 (Weighting Theorem for Total Effects)
Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a numerical causality space, let Z be a covariate, let τ_x denote the true outcome variable w.r.t. total effects, and let $P(X{=}x \,|\, Z{=}z) > 0$, for all pairs of values of X and Z, and let $E^{X=x}(Y \,|\, Z)$ be unbiased. Then:

$$E(\tau_x) \;=\; E^{X=x}\big[E^{X=x}(Y_w \,|\, Z)\big] \;=\; E(Y_w \,|\, X{=}x). \tag{11.60}$$

(Proof p. 293)

Using Lemma 11.37,

$$\begin{aligned}
E(Y_w \,|\, X{=}x) \;&=\; E^{X=x}(Y_w) \qquad \text{[Lemma \textbf{??}]} \\
&=\; E^{X=x}\big[E^{X=x}(Y_w \,|\, Z)\big] \qquad \text{[(iv) in Box \textbf{??}]} \\
&=\; E^{X=x}\left[E^{X=x}(Y \,|\, Z) \cdot \frac{P(X{=}x)}{P(X{=}x \,|\, Z)}\right] \qquad \text{[Eq. (11.59)]} \\
&=\; \int E(Y \,|\, X{=}x, Z{=}z) \cdot \frac{P(X{=}x)}{P(X{=}x \,|\, Z{=}z)}\, P_{Z \,|\, X=x}\,(dz) \qquad \text{[Lemma \textbf{??} and Th. \textbf{??}]} \\
&=\; \int E(Y \,|\, X{=}x, Z{=}z)\; P_Z\,(dz) \qquad \text{[see Th. \textbf{??}]} \\
&=\; \int E^{X=x}(Y \,|\, Z{=}z)\; P_Z\,(dz) \qquad \text{[see Th. \textbf{??}]} \\
&=\; E\big[E^{X=x}(Y \,|\, Z)\big] \qquad \text{[see Th. \textbf{??}; Rem. \textbf{??}]} \\
&=\; E(\tau_x) \qquad \text{[Th. 10.2]},
\end{aligned}$$

where P_Z and $P_{Z \,|\, X=x}$ denote the distribution of Z (see section **??**) and the conditional distribution of Z given $X{=}x$, respectively. This proves both equations in (11.60).

11.4.3 Substantive Meaning

Remark 11.39 According to Theorem 11.38, the expectation $E(\tau_x)$ of the true outcome variable $\tau_x := E^{X=x}(Y \,|\, \mathscr{C}_X)$ can be identified by the conditional expectations $E(Y_w \,|\, X{=}x)$ of the *weighted outcome variable* Y_w [see the second equation in (11.60)]. Furthermore, according to Equations (11.59) and (11.60), we can identify the expectation $E(\tau_x)$ of the true outcome variable τ_x also by taking the $(X{=}x)$-conditional expectation of the conditional expectation $E^{X=x}(Y \,|\, Z)$ of the outcomes on the covariate Z *weighted* by the function $P(X{=}x)/P(X{=}x \,|\, Z)$. ◁

Remark 11.40 A prerequisite of using the Weighting Theorem for total effects is to know the Z-conditional propensity scores $P(X{=}x|Z{=}z)$. In applications, these functions are known if the experimenter himself determines these conditional treatment probabilities, e. g., as a function of the covariate Z, or as a function of the observational-unit variable U.

In quasi-experiments, in which, by definition of the quasi-experiment, the propensities are *not* under the experimenter's control, the propensities may also be estimated. However, in this case we need (a) a correct choice of the relevant (multivariate) covariate Z such that the conditional expectation $E(Y|X,Z)$ is unbiased, and (b) correct estimates of the propensities $P(X{=}x|Z{=}z)$. ◁

Remark 11.41 Note that using the weighting procedure does neither involve the conditional expectation $E(Y|X,Z)$ nor the conditional expectation $E(Y|X,\pi)$ or the expectations of these regressions. Therefore this procedure may be an option whenever estimating these conditional expectations or their expectations is difficult. What is needed, instead, are the propensities $P(X{=}x|Z)$. It should be emphasized, however, that the Weighting Theorem for total effects does not say anything about efficiency of estimators. Hence, in empirical applications one should also compare the standard errors of the average total effects estimated via the weighting procedure as compared the average total effects estimated via adjusting procedures based on Theorems 10.2 and 10.5. ◁

11.4.4 Numerical Example

Example 11.42 (continued) We will use Table 10.1 to illustrate weighting the outcome variable using the Z-conditional propensities $P(X{=}x|Z)$, where $Z := (Z_1, Z_2)$ with $Z_1 := \textit{sex}$ and $Z_2 := \textit{college}$. Using Equation (11.60), Equation

$$E(Y_w|X{=}x) = \sum_z E(Y_w|X{=}x, Z{=}z) \cdot P(Z{=}z|X{=}x),$$

which is always true if Z is discrete [see Eq. (**??**) in Box 9.2], yields

$$\begin{aligned} E(Y_w|X{=}x) &= \sum_z E(Y_w|X{=}x, Z{=}z) \cdot P(Z{=}z|X{=}x) \\ &= \sum_z E(Y|X{=}x, Z{=}z) \cdot \frac{P(X{=}x)}{P(X{=}x|Z{=}z)} \cdot P(Z{=}z|X{=}x) \qquad \text{[Eq. (11.60)]} \end{aligned}$$

(see also Exercise 11-6). All terms occurring in this equation can be computed from the parameters displayed in Tables 10.1 and 10.2.

For *treatment* 0 we obtain

$$\begin{aligned} E(Y_w|X{=}0) &= \sum_z E(Y|X{=}0, Z{=}z) \cdot \frac{P(X{=}0)}{P(X{=}0|Z{=}z)} \cdot P(Z{=}z|X{=}0) \\ &= 72 \cdot \frac{1/2}{2/8} \cdot 1/6 + 97.50 \cdot \frac{1/2}{3/8} \cdot 1/4 + 111 \cdot \frac{1/2}{6/8} \cdot 1/2 \;=\; 93.50, \end{aligned}$$

and for *treatment* 1:

$$\begin{aligned} E(Y_w|X{=}1) &= \sum_z E(Y|X{=}1, Z{=}z) \cdot \frac{P(X{=}1)}{P(X{=}1|Z{=}z)} \cdot P(Z{=}z|X{=}1) \\ &= 83 \cdot \frac{1/2}{6/8} \cdot 1/2 + 102.50 \cdot \frac{1/2}{5/8} \cdot 5/12 + 122 \cdot \frac{1/2}{2/8} \cdot 1/6 \;=\; 102.50. \end{aligned}$$

Hence, according to Equation (11.60), the average total effect is

$$E(\delta_{10}) \;=\; E(Y_w \,|\, X{=}1) - E(Y_w \,|\, X{=}0) \;=\; 102.50 - 93.50 \;=\; 9.$$

This result is exactly what we obtained before using procedures for adjusting means.

◁

11.4.5 Adjusting for Direct Effects by Weighting the Outcome Variable

Now we will show how to compute the average direct effects by choosing appropriate weights for the outcome variable. If Z denotes a covariate and M an intermediate variable, the crucial assumption is that the conditional expectation $(X{=}x)$-conditional conditional expectation $E^{X=x}(Y \,|\, M, Z)$ is unbiased with respect to direct effects.

11.4.6 Theory

Again we will presume that X has $J+1$ values $x = 0, 1, \ldots, J$. The *weighted outcome variable* is now:

$$Y_w := Y \cdot \left(\sum_{x=0}^{J} 1_{X=x} \cdot \frac{P(X{=}x)}{P(X{=}x \,|\, M, Z)} \right). \tag{11.61}$$

Note that the weighting variable, by which Y is multiplied, is now a function of X, M, and Z. The crucial ingredients of this weighting variable are the (M, Z)-conditional propensities $\pi_{x;M} := P(X{=}x \,|\, M, Z)$, the values of which are the conditional probabilities $P(X{=}x \,|\, M{=}m, Z{=}z)$. Again, we start with a lemma that will be used in proving the Weighting Theorem.

Lemma 11.43
Let Y_w denote the weighted outcome variable defined in Equation (11.61), let $E^{X=x}(Y_w \,|\, M, Z)$ denote the conditional expectation of Y_w on M and Z w.r.t. the probability measure $P^{X=x}$, and let $P(X{=}x, M{=}m, Z{=}z) > 0$, for all pairs of values of X, M, and Z. Then:

$$E^{X=x}(Y_w \,|\, M, Z) \underset{P^{X=x}}{=} E^{X=x}(Y \,|\, M, Z) \cdot \frac{P(X{=}x)}{P(X{=}x \,|\, M, Z)} \tag{11.62}$$

(Proof p. 293)

Consider

$$
\begin{aligned}
E^{X=x}(Y_w \,|\, M,Z) &\underset{P^{X=x}}{=} E^{X=x}\left(Y \cdot \sum_{x'=0}^{J} I_{X=x'} \cdot \frac{P(X=x')}{P(X=x' \,|\, M,Z)} \,\middle|\, M,Z \right) && \text{[def. of } Y_w\text{]} \\
&\underset{P^{X=x}}{=} E^{X=x}\left(\sum_{x'=0}^{J} I_{X=x'} \cdot Y \cdot \frac{P(X=x')}{P(X=x' \,|\, M,Z)} \,\middle|\, M,Z \right) \\
&\underset{P^{X=x}}{=} \sum_{x'=0}^{J} E^{X=x}\left(I_{X=x'} \cdot Y \cdot \frac{P(X=x')}{P(X=x' \,|\, M,Z)} \,\middle|\, M,Z \right) && \text{[(??) in Box ??]} \\
&\underset{P^{X=x}}{=} E^{X=x}\left(Y \cdot \frac{P(X=x)}{P(X=x \,|\, M,Z)} \,\middle|\, M,Z \right) \\
&\underset{P^{X=x}}{=} E^{X=x}(Y \,|\, M,Z) \cdot \frac{P(X=x)}{P(X=x \,|\, M,Z)} && \text{[(??) in Box ??]}
\end{aligned}
$$

This proves Lemma 11.43.

Theorem 11.44 (Weighting Theorem for Direct Effects)
Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a numerical causality space, let Z be a covariate and M be an intermediate variable, let τ_{x,t_M} denote the true outcome variable w.r.t. direct effects, and let $P(X{=}x \,|\, M{=}m, Z{=}z) > 0$, for all pairs of values of X, M, and Z, and let $E^{X=x}(Y \,|\, M,Z)$ be unbiased w.r.t. direct effects. Then:

$$
E(\tau_{x,t_M}) \;=\; E^{X=x}\big[E^{X=x}(Y_w \,|\, M,Z)\big] \;=\; E(Y_w \,|\, X{=}x). \tag{11.63}
$$

(Proof p. 293)

11.4.7 Substantive Meaning

Remark 11.45 According to Theorem 11.44, the expectation $E(\tau_{x,t_M})$ of the true outcome variable $\tau_{x,t_M} := E^{X=x}(Y \,|\, M, \mathscr{C}_X)$ can be identified by the conditional expectations $E(Y_w | X{=}x)$ of the *weighted outcome variable* Y_w [see the second equation in (11.60)]. Furthermore, according to Equations (11.59) and (11.60), we can identify the expectation $E(\tau_{x,t_M})$ of the true outcome variable τ_{x,t_M} also by taking the $(X{=}x)$-conditional expectation of the conditional expectation $E^{X=x}(Y \,|\, M,Z)$ of the outcomes on the intermediate variable M and the covariate Z *weighted* by the function $P(X{=}x)/P(X{=}x \,|\, M,Z)$. ◁

Remark 11.46 A prerequisite of using the Weighting Theorem for direct effects is to know the (M,Z)-conditional propensity scores $P(X{=}x \,|\, M{=}m, Z{=}z)$. Even in true experiments, these functions are not known, because M is an intermediate variable that is unknown at the onset of treatment.

Hence, both in experiments and quasi-experiments, the propensities scores $P(X{=}x \,|\, M{=}m, Z{=}z)$ have to be estimated. However, this needs a correct choice of the relevant (multivariate) covariate Z such that the conditional expectation $E(Y \,|\, X, M, Z)$ is unbiased, and (b) correct estimates of the propensities scores

$P(X{=}x \mid M{=}m, Z{=}z)$. In contrast, the choice of M is uncritical; we only have to make sure that M is an intermediate variable. ◁

Remark 11.47 Note that using the weighting procedure does neither involve the conditional expectations $E^{X=x}(Y|M,Z)$ nor the conditional expectations $E^{X=x}(Y|\pi_{x;M})$ or the expectations of these regressions. Therefore this procedure may be an option whenever estimating these conditional expectations or their expectations is difficult. What is needed, instead, are the propensities $P(X{=}x \mid M,Z)$. It should be emphasized, however, that the Weighting Theorem for direct effects does not say anything about efficiency of estimators. Hence, in empirical applications one should also compare the standard errors of the estimators of average direct effects via the weighting procedure to the estimators of average direct effects using adjustment procedures based on Theorems 10.18 and 10.21. ◁

???

Box und Summary sind noch upzudaten Vielleicht Y_t bzw. Y_w für die gewichteten Variablen?

???

11.5 Summary and Conclusions

In this chapter, we treated several ways to adjust for the bias, some of which are based on the *adjustment theorems*. According to these theorems, Z-conditional unbiasedness of the conditional expectation $E(Y|X,Z)$ does not only allow to identify the conditional total effects, but also yields the average total effects in the total population by averaging the conditional *PFE*s over the distribution of the covariate. According to these theorems, assuming conditional mean independence of the true outcomes from X, also allows identification of the conditional total effects $E(\delta_{xx'} \mid X{=}x^*)$ given a treatment condition x^* (see section 5.2.3 for details and substantive meaning of these concepts).

There are several possibilities for choosing an appropriate covariate. The first possibility is to choose a measured covariate, say Z, assuming that the conditional expectation $E(Y|X,Z)$ is unbiased with respect to total effects. In chapter 13 we will show that traditional analysis of covariance (ANCOVA) and generalized ANCOVA — allowing for interactions between treatment and covariates as well as its extension to latent covariates and outcome variables — are particular cases of this strategy.

A second possibility is to use the true propensities $\pi_x = P(X{=}x|\mathscr{C}_X)$ as covariates. For these functions, the conditional prima facie effects and the conditional expectation $E(Y|X,\pi)$ are always unbiased. The limitation of this procedure is that true propensities have to be known, which means that this procedure can only be applied if the true propensities are fixed by the experimenter.

A third possibility is to choose the Z-conditional propensities, i. e., the conditional propensities $\pi_x = P(X{=}x|Z)$ as 'new' covariates, and assume that the original covariate Z is selected such that treatments and true outcome variables τ_x are

Box 11.1 Adjusting Strategies

Adjusting means

(1) Choose a (possibly multivariate) covariate Z such that $E(Y|X,Z)$ is Z-conditionally unbiased. The conditional prima facie effect functions $PFE_{xx';Z}$ are also the conditional total effect functions $E(\delta_{xx'}|Z)$, and the expectations of the *PFE*-functions $PFE_{xx';Z}$ are also the average total effects $E(\delta_{xx'})$.

(2) Choose a (possibly multivariate) covariate Z such that X and the vector $\pi := (\pi_1,\ldots,\pi_J)$ of the Z-conditional propensities $\pi_x := P(X{=}x|Z)$ are conditionally independent given Z. The conditional prima facie effect functions $PFE_{xx';\pi}$ are also the conditional total effect functions $E(\delta_{xx'}|\pi)$, and the expectations of the functions $PFE_{xx';\pi}$ are also the average total effects $E(\delta_{xx'})$.

Weighting the outcome variable

(3) Choose a (possibly multivariate) covariate Z such that $E(Y|X,Z)$ is Z-conditionally unbiased. Replace the outcome variable Y by

$$Y_w := Y \cdot \sum_{x=0}^{J} 1_{X=x} \cdot \frac{P(X{=}x)}{P(X{=}x|Z)}.$$

The prima facie effect $E(Y_w|X{=}x) - E(Y_w|X{=}x')$ is the average total effect $E(\delta_{xx'})$ of x compared to x' *on the original* outcome variable Y.

conditionally independent given Z. All these procedures can be used to compute the expectations of the true outcome variable τ_x. The differences between these expectations are the average total effects.

While all these procedures are based on the adjustment theorems, the last procedure treated in this chapter is not. Instead, it consists of weighting the outcome variable with weights using the Z-conditional propensity scores $P(X{=}x|Z{=}z)$. Using this procedure either presumes that these probabilities are known — this defines the true experiment — or that they are estimated in quasi-experiments. In the latter case, however, we need the assumption that the conditional expectation $E(Y|X,Z)$ is unbiased with respect to total effects.

11.6 Proofs

Proof of Theorem 11.1

(i) According to PR-Th. 16.26 and PR-Rem. 10.4, $1_{X=x} \perp\!\!\!\perp_P \mathscr{C}_X \,|\, \varphi_x$ is equivalent to Equation (11.3). Hence, we only have to show that Equation (11.3) holds:

$$\begin{aligned}
P(X{=}x \,|\, \varphi_x) &:= E(I_{X=x} \,|\, \varphi_x) && \text{[PR-(10.4)]} \\
&\underset{P}{=} E\big[E(I_{X=x} \,|\, \mathscr{C}_X) \,\big|\, \varphi_x\big] && \text{[PR-Box 10.1 (v)]} \\
&\underset{P}{=} E\big[P(X{=}x \,|\, \mathscr{C}_X) \,\big|\, \varphi_x\big] && \text{[PR-(10.4)]} \\
&\underset{P}{=} P(X{=}x \,|\, \mathscr{C}_X). && \text{[(11.2), Box 10.1 (vii)]}
\end{aligned}$$

(ii) Conditioning on φ instead of conditioning on φ_x and applying the same argument to each value $x = 0, 1, \ldots, J$ proves (ii).

Proof of Corollary 11.2

(i) This proposition immediately follows from Theorem 7.11 (i).
(ii) This proposition immediately follows from Theorem 7.11 (i) and Definition 6.4 (ii).

Proof of Corollary 11.4

Proposition (i) immediately follows from PR-Box 16.2 (vi) and proposition (i) of Theorem 11.1, because τ_x is $\mathscr{C}_X$-measurable. Proposition (ii) follows from PR-Box 16.2 (vi) and proposition (ii) of the same theorem, because τ is also $\mathscr{C}_X$-measurable.

Proof of Theorem 11.13

For proving (11.22) remember that $P(X{=}x \,|\, \tau_x, \pi_x)$ is another notation for the conditional expectation $E(1_{X=x} \,|\, \tau_x, \pi_x)$, where $1_{X=x}$ is the indicator variable of $X{=}x$. Now consider

$$\begin{aligned}
P(X{=}x \,|\, \tau_x, \pi_x) &\underset{P}{=} E\big[P(X{=}x \,|\, Z, \tau_x) \,\big|\, \tau_x, \pi_x\big] && \text{[(??) in Box ??]} \\
&\underset{P}{=} E\big[P(X{=}x \,|\, Z) \,\big|\, \tau_x, \pi_x\big] && \text{[}X \perp\!\!\!\perp_P \tau_x \,|\, Z\text{, Th. 14.64]} \\
&\underset{P}{=} P(X{=}x \,|\, Z) && \text{[def. of } \pi_x\text{, (??) in Box ??]} \\
&:= \pi_x && \text{[def. of } \pi_x\text{]}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
P(X{=}x \,|\, \pi_x) &\underset{P}{=} E\big[P(X{=}x \,|\, Z) \,\big|\, \pi_x\big] && \text{[(??) in Box ??]} \\
&\underset{P}{=} P(X{=}x \,|\, Z) && \text{[def. of } \pi_x\text{, (??) in Box ??]} \\
&:= \pi_x && \text{[def. of } \pi_x\text{]}.
\end{aligned}$$

This proves Equation (11.22). Exactly the same equations hold if we replace π_x by π and τ_x by τ, which also proves Equation (11.23).

Proof of Corollary 11.15

According to Theorem 14.64, mean independence of the true outcome variable τ_x from $1_{X=x}$ given π_x immediately follows from $1_{X=x} \underset{P}{\perp\!\!\!\perp} \tau_x \,|\, \pi_x$. Similarly, mean independence of each true outcome variable τ_x from X given π immediately follows from $X \underset{P}{\perp\!\!\!\perp} \tau_x \,|\, \pi$.

Proof of Corollary 11.17

All four propositions of this Corollary immediately follow from Theorem 10.2.

Proof of Theorem 11.27

For proving (11.46) remember that $P(X{=}x \,|\, \tau_{x,t_M}, \pi_{x;M})$ is another notation for the conditional expectation $E(1_{X=x} \,|\, \tau_{x,t_M}, \pi_{x;M})$, where $1_{X=x}$ is the indicator variable of the event $\{X{=}x\}$. Now consider

$$\begin{aligned}
P(X{=}x \,|\, \tau_{x,t_M}, \pi_{x;M}) &\underset{P}{=} E\big[P(X{=}x \,|\, M, Z, \tau_{x,t_M}) \,\big|\, \tau_{x,t_M}, \pi_{x;M}\big] && [(\textbf{??}) \text{ in Box } \textbf{??}] \\
&\underset{P}{=} E\big[P(X{=}x \,|\, M, Z) \,\big|\, \tau_{x,t_M}, \pi_{x;M}\big] && [1_{X=x} \underset{P}{\perp\!\!\!\perp} \tau_{x,t_M} \,|\, M, Z;\ \text{Th. } 14.64] \\
&\underset{P}{=} P(X{=}x \,|\, M, Z) && [\text{def. of } \pi_{x;M};\ (\textbf{??}) \text{ in Box } \textbf{??}] \\
&:= \pi_{x;M} && [\text{def. of } \pi_{x;M}].
\end{aligned}$$

Similarly,

$$\begin{aligned}
P(X{=}x \,|\, \pi_{x;M}) &\underset{P}{=} E\big[P(X{=}x \,|\, M, Z) \,\big|\, \pi_{x;M}\big] && [(\textbf{??}) \text{ in Box } \textbf{??}] \\
&\underset{P}{=} P(X{=}x \,|\, M, Z) && [\text{def. of } \pi_{x;M};\ (\textbf{??}) \text{ in Box } \textbf{??}] \\
&:= \pi_{x;M} && [\text{def. of } \pi_{x;M}].
\end{aligned}$$

This proves that $1_{X=x} \underset{P}{\perp\!\!\!\perp} \tau_{x,t_M} \,|\, M, Z$ implies Equation (11.46).

Exactly the same equations hold if we replace $\pi_{x;M}$ by π_M and τ_{x,t_M} by τ_M, which also proves Equation (11.47).

Proof of Corollary 11.29

According to Theorem 14.64, mean independence of the true outcome variable τ_{x,t_M} from $1_{X=x}$ given $\pi_{x;M}$ immediately follows from $1_{X=x} \underset{P}{\perp\!\!\!\perp} \tau_{x,t_M} \,|\, \pi_{x;M}$, and mean independence of τ_M from X given π_M immediately follows from $X \underset{P}{\perp\!\!\!\perp} \tau_M \,|\, \pi_M$.

Proof of Corollary 11.31

All four propositions of this Corollary immediately follow from Theorem 10.2.

Proof of Lemma 11.37

Proof of Theorem 11.38

Proof of Lemma 11.43

Proof of Theorem 11.44

Using Lemma 11.43,

$$
\begin{aligned}
E(Y_w \,|\, X{=}x) &= E^{X=x}(Y_w) && \text{[Lemma \textbf{??}]} \\
&= E^{X=x}\big[E^{X=x}(Y_w \,|\, M, Z)\big] && \text{[(iv) in Box \textbf{??}]} \\
&= E^{X=x}\left[E^{X=x}(Y \,|\, M, Z) \cdot \frac{P(X{=}x)}{P(X{=}x \,|\, M, Z)}\right] && \text{[Eq. (11.62)]} \\
&= \int E(Y \,|\, X{=}x, M{=}m, Z{=}z) \cdot \frac{P(X{=}x)}{P(X{=}x \,|\, M{=}m, Z{=}z)}\; P_{M,Z\,|\,X=x}\;[d(m,z)] && \text{[Ths. \textbf{??} and \textbf{??}]} \\
&= \int E(Y \,|\, X{=}x, M{=}m, Z{=}z)\; P_{M,Z}\;[d(m,z)] && \text{[see Th. \textbf{??}]} \\
&= \int E^{X=x}(Y \,|\, M{=}m, Z{=}z)\; P_{M,Z}\;[d(m,z)] && \text{[see Th. \textbf{??}]} \\
&= E\big[E^{X=x}(Y \,|\, M, Z)\big] && \text{[see Th. \textbf{??}; Rem. \textbf{??}]} \\
&= E(\tau_{x,t_M}) && \text{[Th. 10.2]},
\end{aligned}
$$

where $P_{M,Z}$ and $P_{M,Z\,|\,X=x}$ denote the distribution of (M,Z) (see section **??**) and the conditional distribution of (M,Z) given $X{=}x$, respectively. This proves both equations in (11.63).

11.7 Exercises

▷ **Exercise 11-1** Which is the content of Theorem 10.2 (the first adjustment theorem)?

▷ **Exercise 11-2** Why is it not possible to use Equation (10.6) to compute the average causal effect in the total population from the prima facie effects in the two subpopulations for the random experiment presented in Table 6.1?

▷ **Exercise 11-3** Is it possible in applications to know the true treatment probability function $P(X{=}x \,|\, \mathscr{C}_X)$ and why is it useful to know it?

▷ **Exercise 11-4** Compute the coefficients of Equation (11.39) from the conditional expectations presented in Table 11.1.

▷ **Exercise 11-5** Show that the strong ignorability condition $X \underset{P}{\perp\!\!\!\perp} \tau | Z$ holds in the example presented in Table 10.1.

▷ **Exercise 11-6** Show that $E(Y_w \,|\, X{=}x) = E(\tau_x)$ for Y_w defined in Equation (11.58).

▷ **Exercise 11-7** Use Equation (10.6) to compute the average total effect in the total population from the average total effects in the two subpopulations for the random experiment presented in Table 10.1.

▷ **Exercise 11-8** Suppose that conditional randomization is not an option in a specific application with two treatment conditions. Which other research strategies could be followed in order to make it plausible that $U \underset{P}{\perp\!\!\!\perp} X \,|\, Z$ holds?

▷ **Exercise 11-9** Why is the weight variable occurring in Equation (11.58) a function of X and Z?

Solutions

▷ **Solution 11-1** According to the first adjustment theorem, we cannot only compute the conditional, but also the unconditional total effects and the expectations $E(\tau_x)$ of the true outcome variables τ_x from the covariate-treatment conditional expectation $E(Y \,|\, X, Z)$ *provided that it is unbiased with respect to total effects*. Hence, even if the treatment conditional expectation $E(Y \,|\, X)$ is biased, there is the possibility for valid causal inference.

▷ **Solution 11-2** It is not possible to use Equation (10.6) for the random experiment presented in Table 6.1 because, in this example, the covariate-treatment conditional expectation $E(Y \,|\, X, Z)$ and, therefore, the conditional prima facie effects given Z, are biased.

▷ **Solution 11-3** In applications it *is* possible to know the true treatment probability function $P(X{=}x \,|\, \mathscr{C}_X)$, for instance, if the experimenter fixes the individual treatment probabilities $P(X{=}x \,|\, U{=}u)$ himself and makes sure that no other potential confounder determines the treatment probability. These probabilities can be fixed following ethical or economic considerations. They do not have to be identical for all individuals in order to allow for valid causal inference. If the individual treatment probabilities are known, we can use the functions $\pi_x := P(X{=}x \,|\, \mathscr{C}_X) = P(X{=}x \,|\, U)$ as covariates in the adjustment theorems. The conditional expectations $E(Y \,|\, X, \pi)$, with $\pi := (\pi_1, \ldots, \pi_J)$, are always unbiased. Alternatively, we can compute the weighted outcomes [see Eq. (11.58)], using U as the covariate Z. The expectation of these weighted outcomes in treatment x is equal to the expectation $E(\tau_x)$ of the true outcomes under treatment x. Furthermore, the difference between the expectation of the weighted outcomes in treatment $X{=}x$ and the expectation of the weighted outcomes in control $X{=}0$ is the average causal effect of treatment x compared to the control $X{=}0$ (see the Weighting Theorem).

▷ **Solution 11-4** According to Equations (11.37) and (11.39), the conditional expectation $E^{X=1}(Y \,|\, \varphi)$ is

$$E^{X=1}(Y \,|\, \varphi) \;=\; \gamma_0^{(1)} + \gamma_1^{(1)} \cdot I_{\varphi=5/8} + \gamma_2^{(1)} \cdot I_{\varphi=2/8}. \tag{11.64}$$

For $\varphi = 7/8$, this equation and the conditional expectations in Table 11.1 yield

$$E^{X=1}(Y \,|\, \varphi{=}7/8) \;=\; \gamma_0^{(1)} \;=\; 83.00,$$

and for $\varphi = 5/8$ they yield

$$E^{X=1}(Y \,|\, \varphi{=}5/8) \;=\; \gamma_0^{(1)} + \gamma_1^{(1)} \;=\; 96.00.$$

Hence $\gamma_1^{(1)} = 96 - 83 = 13$. Finally, for $\varphi = 2/8$, Equation (11.64) and Table 11.1 yield

$$E^{X=1}(Y|\varphi{=}2/8) \;=\; \gamma_0^{(1)} + \gamma_2^{(1)} \;=\; 122.00.$$

Hence $\gamma_2^{(1)} = 122 - 83 = 39$.

▷ **Solution 11-5** If X is discrete with a finite number of values, each of which has a positive probability, the $X \underset{P}{\perp\!\!\!\perp} \tau \,|\, Z$ is equivalent to

$$P(X{=}1 \,|\, Z, \tau) \;=\; P(X{=}1 \,|\, Z).$$

Looking at Table 10.1 shows that the true treatment probability function $P(X{=}1 \,|\, U)$ is identical with $P(X{=}1|Z)$. Because τ is a function of U this implies $P(X{=}1|Z,\tau) \;=\; P(X{=}1|\, Z)$.

▷ **Solution 11-6**

$$\begin{aligned} E(Y_w \,|\, X{=}x) \;&=\; E^{X=x}(Y_w) \qquad \text{[Lemma \textbf{??}]} \\ &= \sum_z E^{X=x}(Y_w \,|\, Z{=}z) \cdot P^{X=x}(Z{=}z) \\ &= \sum_z E(Y \,|\, X{=}x, Z{=}z) \cdot \frac{P(X{=}x)}{P(X{=}x \,|\, Z{=}z)} \cdot P^{X=x}(Z{=}z) \qquad \text{[Eq. (11.59)]} \\ &= \sum_z E(Y \,|\, X{=}x, Z{=}z) \cdot P(Z{=}z), \end{aligned}$$

using the definitions of the conditional probabilities $P^{X=x}(Z{=}z)$ and $P(X{=}x \,|\, Z{=}z)$. Note that

$$\sum_z E(Y \,|\, X{=}x, Z{=}z) \cdot P(Z{=}z) \;=\; E\big[E^{X=x}(Y \,|\, M, Z)\big].$$

Because the conditional expectation $E(Y|X,Z)$ is unconfounded, $E\big[E^{X=x}(Y \,|\, M, Z)\big]$ is equal to the expectation $E(\tau_x)$ of the true outcomes in treatment x (see Theorem 10.2).

▷ **Solution 11-7** According to Equation (10.6) and because the two-dimensional covariate Z is discrete, we can use

$$E(\delta_{xx'}) \;=\; E(PFE_{xx';Z}) \;=\; \sum_z PFE_{xx';Z=z} \cdot P(Z{=}z).$$

For the example presented in Table 10.1 this yields:

$$\begin{aligned} E(\delta_{10}) \;&=\; PFE_{12;Z=(m,n)} \cdot 1/3 + PFE_{12;Z=(m,y)} \cdot 1/3 + PFE_{12;Z=(f,y)} \cdot 1/3 \\ &=\; 11.00 \cdot 1/3 + 5.00 \cdot 1/3 + 11.00 \cdot 1/3 \;=\; 9. \end{aligned}$$

▷ **Solution 11-8** One could try to observe *all relevant covariates* $Z_1, \ldots, Z_Q$ which determine the treatment probability and gather them in the vector $Z := (Z_1, \ldots, Z_Q)$. One can then analyze $E(Y|X,Z)$, which will be unbiased if Z in fact contains all relevant covariates determining the treatment probability, i. e., if $P(X{=}1 \,|\, c_X) = P(X{=}1 \,|\, Z)$. Alternatively, one can compute the Z-conditional propensity $\pi := P(X{=}1 \,|\, Z)$ and use this function instead of Z as a covariate; the conditional expectation $E(Y \,|\, X, \pi)$ will also be unbiased. Another alternative is to observe all relevant covariates $Z^* := (Z_1^*, \ldots, Z_R^*)$ such that conditional homogeneity, i. e., $E(Y|X,c_X) = E(Y|X,Z^*)$, holds. Note, that the two covariate vectors Z and Z^* need not be identical. Also note that it may not always be easy to estimate $E(Y \,|\, X, Z)$ or $E(Y \,|\, X, Z^*)$ in a small sample, because these conditional expectations may not be linear (or other simple) functions with only a few parameters. Similarly, the Z-conditional

propensity $P(X{=}1 \mid Z)$ may not be a linear logistic function of Z. If the conditional expectation $E(Y \mid X, Z)$ is unbiased with respect to total effects, one may also weight the outcome variable by a function of X and the propensities $P(X{=}x \mid Z)$ [see Eq. (11.58)]. In this case the differences between the expectations $E(Y_w \mid X{=}x)$ and $E(Y_w \mid X{=}x')$ will be equal to the average total effects $E(\delta_{xx'})$ of treatment x compared to treatment x' *on the original outcome variable* Y.

▷ **Solution 11-9** The weight variable in Equation (11.58) is a function of X and Z because it depends on X via the indicator $1_{X=x}$ and on the covariate Z via the Z-conditional propensity $P(X{=}x \mid Z)$.

Chapter 12
Analysis of Change Scores

In chapter 10, we treated the theoretical foundations of a number of data analysis procedures that may be used for estimating and testing conditional and average causal effects, especially in quasi-experimental studies. In this chapter, we will study the assumptions under which the analysis of change scores can be used to estimate conditional and average total effects.

12.1 Theory

Suppose there is a pre-test Z, a treatment variable X with at least two values and an outcome variable (or 'post-test') Y assessed with the same measurement instrument as the pre-test. In such a design it seems natural to analyze the differences $E(Y-Z\,|\,X{=}x) - E(Y-Z\,|\,X{=}x')$ between the conditional expectation values of the difference scores, or, in other words, the conditional expectation $E(Y-Z\,|\,X)$ of the difference between post-test and pre-test given the treatment variable and to test the hypothesis that this conditional expectation is a constant. In fact, this is exactly what we do in a two-factorial ANOVA with the within-factor *pre-post* and the between-factor *treatment*. If X is dichotomous and we choose $Y-Z$ as the dependent variable, the conditional expectation $E(Y-Z\,|\,X)$ is also estimated and tested in the t-test for independent observations, where X is the grouping variable. The intuitive idea is that if the expectation of the change variable $Y-Z$ in the treatment condition is greater than its expectation in the control condition, then this difference in the expected changes must be due to the treatment.

Can we rely on this intuition? Although this idea has some intuitive appeal, it can be shown that causal inference based on such an analysis of change scores rests on two strong assumptions.

Theorem 12.1 (Change Scores Considering two Values of X)
Let $\langle\,(\Omega,\mathscr{A},P),(\mathscr{F}_t,t\in T),X,Y\,\rangle$ be a causality space, let Y be numerical and nonnegative or with finite expectation, and let Z be a covariate of X. If $P(X{=}x),\ P(X{=}x')>0$ and $E(Y\,|\,X)$ is unbiased w.r.t. total effects, then

$$\begin{aligned} &E(Y-Z\,|\,X{=}x) \;-\; E(Y-Z\,|\,X{=}x') \\ &= E(\delta_{xx'}) \;-\; [E(Z\,|\,X{=}x) - E(Z\,|\,X{=}x')]. \end{aligned} \tag{12.1}$$

(Proof p. 304)

Remark 12.2 (Implications for Data Analysis) Hence, if the conditional expectation $E(Y\,|\,X)$ is unbiased w.r.t. total effects, then we can conclude that

$$E(\delta_{xx'}) \;=\; E(Y-Z\,|\,X{=}x) \;-\; E(Y-Z\,|\,X{=}x')$$

only if $E(Z\,|\,X{=}x) = E(Z\,|\,X{=}x')$. In other words, if $E(Y|X)$ is unbiased w.r.t. total effects, then the difference $E(Y-Z\,|\,X{=}x) - E(Y-Z\,|\,X{=}x')$ between the expectations of the difference scores is $\delta_{xx'}$-biased *unless the conditional expectation values* $E(Z\,|\,X{=}x)$ *and* $E(Z\,|\,X{=}x')$ *of the pre-test* Z *are identical*. If the expectations of the pre-test Z *do* depend on X and $E(Y|X)$ is unbiased w.r.t. total effects, then the average total effects can be identified by

$$E(\delta_{xx'}) \;=\; E(Y-Z\,|\,X{=}x) \;-\; E(Y-Z\,|\,X{=}x') \;+\; [E(Z\,|\,X{=}x) - E(Z\,|\,X{=}x')], \tag{12.2}$$

which is identical to $E(\delta_{xx'}) = E(Y|X{=}x) \;-\; E(Y|X{=}x')$. Hence, if the conditional expectation $E(Y|X)$ is unbiased w.r.t. total effects and $E(Z\,|\,X) \underset{P}{=} E(Z)$, then we may analyze change scores or test the *time*×*group* interaction in ANOVA if we aim at testing the hypothesis that the total treatment effect is zero. However, in this case the ordinary t-test for independent groups tests the same hypothesis, and for this test we only have to assume unbiasedness of $E(Y|X)$ and not additionally $E(Z\,|\,X) \underset{P}{=} E(Z)$. ◁

In Theorem 12.1 we considered only two values x and x' of a treatment variable X. Let us generalize this result to the case in which X is discrete with values $0,1,\dots,J$. In this case, we use the random variables $1_{X=x}$ indicating with 1 if the event $\{X{=}x\} = \{\omega \in \Omega\colon X(\omega) = x\}$ occurs. Otherwise $1_{X=x}$ takes on the value 0.

Corollary 12.3 (Change Scores)
Let $\langle\,(\Omega,\mathscr{A},P),(\mathscr{F}_t,t\in T),X,Y\,\rangle$ *be a causality space, let* Y *be numerical and nonnegative or with finite expectation, and let* Z *be a covariate of* X*. Furthermore, let* X *be discrete with a finite number of values* $x = 0,1,\dots,J$*, and let* $P(X{=}x) > 0$ *for all* $x = 0,1,\dots,J$*. Then there are* $\beta_0,\beta_x \in \mathbb{R}$ *such that*

$$E(Y-Z\,|\,X) \;=\; \beta_0 \;+\; \sum_{x=1}^{J} \beta_x\, 1_{X=x}\,. \tag{12.3}$$

and unbiasedness of $E(Y|X)$ *w.r.t. total effects implies*

$$\beta_x \;=\; E(\delta_{x0}) \;-\; [E(Z\,|\,X{=}x) - E(Z\,|\,X{=}0)]\,, \quad x = 1,\dots,J. \tag{12.4}$$

(Proof p. 305)

Remark 12.4 (*Time*×*Group* Interaction in ANOVA) With the *time*×*group* interaction in ANOVA we test the hypothesis

$$H_0 : \beta_x \;=\; 0, \quad x = 1,\dots,J. \tag{12.5}$$

Furthermore, under the assumptions about X specified in Corollary 12.3, the condition

$$E(Z\,|\,X{=}x) \;=\; E(Z\,|\,X{=}0) \quad \text{for all } x = 1,\dots,J,$$

is equivalent to $E(Z\,|\,X) = E(Z)$. Therefore, with the *time*×*group* interaction in ANOVA we also test the hypothesis

$$H_0 :\; E(\delta_{x0}) \;=\; 0, \quad x = 1,\dots,J, \tag{12.6}$$

provided that $E(Y|X)$ is unbiased w.r.t. total effects and $E(Z|X) = E(Z)$. ◁

Corollary 12.3 immediately implies the following proposition.

Corollary 12.5 (A Necessary and Sufficient Condition for Unbiasedness) *Let the assumptions of Corollary 12.3 be true and let $E(Z\,|\,X) = E(Z)$. Then*

$$\beta_x \;=\; E(\delta_{x0}), \quad x = 1,\dots,J, \tag{12.7}$$

if and only if

$$x \;-\; z \;=\; E(\delta_{x0}), \quad x = 1,\dots,J. \tag{12.8}$$

(Proof p. 305)

Remark 12.6 (One-Factorial ANOVA) Note that, under the assumptions about X specified in Corollary 12.3, the conditional expectation $E(Y\,|\,X)$ can be written

$$= \alpha_0 \;+\; \sum_{x=1}^{J} \alpha_x \mathbb{1}_{X=x}\,, \tag{12.9}$$

where

$$\alpha_x \;=\; E(Y\,|\,X{=}x) \;-\; E(Y\,|\,X{=}0), \quad x = 1,\dots,J. \tag{12.10}$$

A one-factorial ANOVA with dependent variable Y and between factor X tests the hypothesis

$$H_0 : \alpha_x \;=\; 0, \quad x = 1,\dots,J. \tag{12.11}$$

Hence, if $E(Z\,|\,X) = E(Z)$ and $E(Y\,|\,X)$ is unbiased, then a one-factorial ANOVA with dependent variable Y and between-factor X also tests the hypothesis of no average total treatment effects. Oftentimes, there is a strong correlation between pre- and post-test. Therefore, statistical power will be higher for the interaction test mentioned in Remark 12.4. ◁

Remark 12.7 (Randomized Experiment) In a randomized experiment, i. e., in an experiment with randomized assignment of the unit to one of the treatment conditions, X and the global covariate C_X are independent, implying that X and Z are independent as well. Then $E(Y\,|\,X)$ is unbiased w.r.t. total effects and $E(Z\,|\,X) = E(Z)$ will hold, i. e., the expectations of the pre-test Z will not differ between treatment conditions. In this case, analyzing change scores yields in fact estimates of average total treatment effects and *tests the hypothesis that the average total treatment effects are zero.* ◁

Remark 12.8 (Mean-Independent Outcome Condition) Suppose that the mean-independent outcome condition $Y \vdash C_X\,|\,X$ holds, i. e., suppose

$$E(Y\,|\,X, C_X) \underset{P}{=} E(Y\,|\,X)\,.$$

Then the conditional expectation $E(Y\,|\,X)$ is unbiased w.r.t. total effects but $E(Y{-}Z\,|\,X)$ will be biased unless $E(Z\,|\,X) = E(Z)$. Hence, unless the expectations of the pre-test do not differ between treatment conditions, the analysis of change scores *does not test the hypothesis that the average total effects of the treatments are zero*, even if $E(Y\,|\,X)$ is unbiased w.r.t. total effects. ◁

Remark 12.9 (Unbiasedness of Z) In Theorem 12.1 and Corollary 12.3 we presuppose that the conditional expectation $E(Y\,|\,X)$ is unbiased w.r.t. total effects. Of course, we can also analyze change scores under the assumption that Z is unbiased w.r.t. total effects. However, in this case we need an additional assumption that is specified in the following theorem. ◁

Theorem 12.10 (Change Scores II)
Let the assumptions of Corollary 12.3 be true. If $E(Y|X,Z)$ is unbiased w.r.t. total effects and

$$E(Y|X,Z) \underset{P}{=} \beta_0 + 1\cdot Z + \sum_{x=1}^{J} \beta_x 1_{X=x}, \tag{12.12}$$

then

$$E(Y-Z|X) = \beta_0 + \sum_{x=1}^{J} \beta_x 1_{X=x}, \tag{12.13}$$

where

$$\beta_x = E(\delta_{x0}|Z) = E(\delta_{x0}), \quad x = 1,\dots,J. \tag{12.14}$$

(Proof p. 305)

Remark 12.11 (A Caveat) Of course, the assumption that Equation (12.12) holds will rarely be true in empirical applications. Neither randomization techniques, nor techniques of covariate selection can be used to secure that this equation holds.

Therefore, it is often a better idea to analyze the conditional expectation $E(Y-Z|X,Z)$, which yields estimates of the same effects of X on Y as the analysis of $E(Y|X,Z)$, because

$$E(Y-Z|X,Z) = E(Y|X,Z) - Z.$$

The analysis of the conditional expectations $E(Y|X,Z)$ and $E(Y-Z|X,Z)$ will be treated in more detail in sections 13.1 and 13.2. ◁

12.2 Numerical Examples

In this section we will present three examples. In the first one, the conditions that are sufficient for unbiased w.r.t. total effectsness of the conditional expectation are $E(Y-Z|X)$ are satisfied. In this case, the analysis of change scores as well as the analysis of the outcome variable yield the average total treatment effect. In the second and third examples, the sufficient conditions are not met and in these cases, the analysis of change scores yields severely biased results.

Example 12.12 (Example 1) In our first example, we consider a treatment with two treatment conditions, $X=0$ and $X=1$, a pre-test Z, and a post-test (outcome variable) Y. In this example, the conditional expectation $E(Y|X)$ is unbiased w.r.t. total effects and the expectations of the covariate Z in the two treatment conditions are identical with each other. This situation occurs in a randomized experiment. Table 12.1 displays the variances, covariances, correlations, and expectations of the variables involved. These parameters are implied by

$$E(Y|X,Z) = 100 + 10\,X + .80\,Z, \tag{12.15}$$

with $E(X) = .50$ and $Var(Y|X) = 100$. Furthermore,

$$E(Z|X) = E(Z) = 100 \tag{12.16}$$

and

$$Var(Z|X) = 100.$$

The filtration $(\mathscr{F}_t, t \in T)$ is specified by

Table 12.1. Covariances, Correlations, and Expectations in Example 1

		Z	X	Y
Covariate	Z	100.00	*.000*	*.582*
Treatment (yes=1, no=0)	X	0.00	0.25	*.364*
Outcome	Y	80.00	2.50	189
Expectations		100.00	0.50	185.00

Note. Correlations (in italics) are rounded. A concrete sample generated from these parameters can be downloaded at *www.causal-effects.de*. The file name is *PC1.Table12.1.dat*.

(a) $\mathscr{F}_1 = \sigma(Z)$,

(b) $\mathscr{F}_2 = \sigma(Z, X)$, and

(c) $\mathscr{F}_3 = \sigma(Z, X, Y)$.

Furthermore, in this example, the covariate Z is identical with the global covariate C_X.

In this example, the conditional expectation

$$\begin{aligned} &= E[E(Y \mid X, Z) \mid X] && \text{[PR-Box 10.2 (v)]} \\ &= E(100 + 10X + .80Z \mid X) && \text{[(12.15)]} \\ &= 100 + 10X + .80E(Z \mid X) && \text{[PR-Box 10.2 (xvi)]} \\ &= 100 + 10X + .80E(Z) && \text{[(12.16)]} \\ &= 180 + 10X && [E(Z) = 100] \end{aligned} \tag{12.17}$$

is unbiased w.r.t. total effects and the average total treatment effect is 10. Therefore, according to Theorem 12.1, the conditional expectation $E(Y-Z \mid X)$ of the difference score on X is unbiased w.r.t. total effects as well. In fact,

$$\begin{aligned} E(Y-Z \mid X) &= E(Y \mid X) - E(Z \mid X) && \text{[PR-Box 10.2 (xvi)]} \\ &= E(Y \mid X) - E(Z) && \text{[Eqs. (12.17), (12.16)]} \\ &= 180 + 10X - 100 \\ &= 80 + 10X. \end{aligned}$$

Hence, the average total treatment effect is 10, and it can be analyzed using $E(Y \mid X)$, but also using the conditional expectation $E(Y-Z \mid X)$ of the change score variable $Y-Z$ on X. Furthermore, in this first example, the t-test for independent groups defined by X with Y or $Y-Z$ as the dependent variable and the ANOVA *time* × *group* interaction test also test the hypothesis of no average total treatment effect.

◁

Example 12.13 (Example 2) In this second example, we again consider a treatment with two conditions, $X=0$ and $X=1$, a pre-treatment variable Z, and an outcome variable Y. In this second example, the conditional expectation $E(Y \mid X)$ is again unbiased w.r.t. total effects, but this time the expectations of the covariate Z in the two treatment conditions

Table 12.2. Covariances, Correlations, and Expectations in Example 2

		Z	X	Y
Covariate	Z	125.00	*.894*	*.400*
Treatment (yes=1, no=0)	X	5.00	0.25	*.447*
Outcome	Y	50.00	2.50	125
Expectations		100.00	0.50	105.00

Note. Correlations (in italics) are rounded. A concrete sample generated from these parameters can be downloaded at *www.causal-effects.de*. The file name is *PC1.Table12.2.dat*.

are not identical with each other, a situation frequently occurring in non-randomized experiments. Table 12.2 displays the variances, covariances, correlations, and expectations of the variables involved. These parameters are implied by

$$= 100 + 10\,X \tag{12.18}$$

with $E(X) = .50$ and $Var(Y\,|\,X) = 100$, as well as

$$E(Z\,|\,X) = 90 + 20\,X \tag{12.19}$$

and $Var(Z\,|\,X) = 25$. The filtration $(\mathscr{F}_t, t \in T)$ and the global covariate C_X are defined as in Example 12.12.

In this second example, the conditional expectation $E(Y|X)$ is unbiased w.r.t. total effects and the average total treatment effect is 10. However, $E(Z\,|\,X) \neq E(Z)$. Therefore, according to Corollary 12.3, the conditional expectation $E(Y-Z|X)$ of the difference score on X is biased. In fact,

$$\begin{aligned} E(Y-Z\,|\,X) &= E(Y|X) - E(Z|X) && \text{[PR-Box 10.2,(xvi)]} \\ &= 100 + 10\,X - (90 + 20\,X) && \text{[Eqs. (12.18), (12.19)]} \\ &= 10 - 10\,X. \end{aligned}$$

Hence, instead of the average total treatment effect of 10, analyzing the difference scores yields the alleged treatment effect of −10. Obviously, this effect is seriously biased.

◁

Example 12.14 (Example 3) In Example 12.13 the outcome variable has been X-conditionally regressively independent of Z (i. e, $Y \vdash Z\,|\,X$), and because we assumed $Z = C_X$, the causality condition $Y \vdash C_X\,|\,X$ has been satisfied. However, if Z represents a pre-test and Y the post-test, this assumption is not realistic. Therefore, we study a third example, which is typical in a quasi-experiment in which pre- and post-tests are assessed: Now Y depends on X *and* Z, and the expectations of Z are different in the two treatment conditions.

Hence, we again consider a treatment with two conditions, $X=0$ and $X=1$, a pre-test Z, and a post-test Y. In this third example, the conditional expectation $E(Y\,|\,X)$ is not unbiased w.r.t. total effects, and again the expectations of the covariate Z in the two treatment

Table 12.3. Covariances, Correlations, and Expectations in Example 3

		Z	X	Y
Covariate	Z	125.00	*.894*	*.795*
Treatment (yes=1, no=0)	X	5.00	0.25	*.770*
Outcome	Y	150.00	6.50	285
Expectations		100.00	0.50	105.00

Note. Correlations (in italics) are rounded. A concrete sample generated from these parameters can be downloaded at *www.causal-effects.de*. The file name is *PC1.Table12.3.dat*.

conditions are not identical with each other. Table 12.3 displays the variances, covariances, correlations, and expectations of the variables involved. Now these parameters are implied by

$$E(Y|X,Z) = 100 + 10X + .80Z, \tag{12.20}$$

with $E(X) = .50$ and $Var(Y|X) = 100$. Furthermore,

$$E(Z|X) = 90 + 20X \tag{12.21}$$

and $Var(Y|X) = 25$. The filtration $(\mathscr{F}_t, t \in T)$ and the global covariate C_X are defined as in Example 12.12.

In this third example, the average total treatment effect is still 10. However, the conditional expectations

$$\begin{aligned} &= E[E(Y|X,Z)|X] && \text{[PR-Box 10.2 (v)]} \\ &= E(100 + 10X + .80Z|X) && \text{[(12.20)]} \\ &= 100 + 10X + .80E(Z|X) && \text{[PR-Box 10.2 (xvi)]} \\ &= 100 + 10X + .80(90 + 20X) && \text{[(12.21)]} \\ &= 172 + 26X \end{aligned} \tag{12.22}$$

and

$$\begin{aligned} E(Y-Z|X) &= E(Y|X) - E(Z|X) && \text{[PR-Box 10.2 (xvi)]} \\ &= 172 + 26X - (90 + 20X) && \text{[(12.22), (12.21)]} \\ &= 82 + 6X \end{aligned}$$

are both seriously biased. Hence, instead of the average total treatment effect of 10, analyzing the conditional expectation of the post-test Y or the conditional expectation of the change score variable $Y-Z$ on treatment variable X yields invalid results about the average total effect of X. ◁

12.3 Summary and Conclusions

In this section, we have shown that the analysis of difference scores yields correct estimates of the average total effects, provided that

(a) the conditional expectation $E(Y|X)$ is unbiased w.r.t. total effects, and

(b) there are no differences in the expectations of the pre-test between treatment conditions.

These prerequisites are met in the perfect randomized experiment, but usually not in other research situations. And of course, if $E(Y|X)$ is unbiased w.r.t. total effects, in order to estimate the total treatment effects, we can also estimate and test the conditional expectation $E(Y|X)$ instead of $E(Y-Z|X)$. However, statistical tests based on the analysis of change score will often have a higher power than the analysis of the outcome variable.

A second set of assumptions under which the analysis of difference scores yields correct estimates of the average total effects is

(c) the conditional expectation $E(Y|X,Z)$ is unbiased w.r.t. total effects, and

(d) $E(Y|X,Z) \underset{P}{=} \beta_0 + 1 \cdot Z + \sum_{x=1}^{J} \beta_x 1_{X=x}$.

However, assumption (b) will usually not hold in empirical applications.

It has also been suggested to use analysis of covariance with difference scores as a dependent variable and the pre-test as the covariate (see, e. g., Jamieson, 2004). This suggestion will be dealt with in section 13.1.

Statistical Programs

The analysis of difference scores can be accomplished with every major statistical program package and with every program for multiple linear regression. Repeated measures ANOVA and, if X is dichotomous, t-tests for independent samples with the dependent variable $Y-Z$ will analyze and test hypothesis about the average total effect of the treatment conditions *if the assumptions (a) and (b) mentioned above are met.*

12.4 Proofs

Proof of Theorem 12.1

If $E(Y|X)$ is unbiased w.r.t. total effects, then by definition,

$$E(Y|X{=}x) = E(\tau_x) \qquad \text{for each value } x \text{ of } X \text{ for which } P(X{=}x) > 0. \tag{12.23}$$

Because

$$\begin{aligned} E(Y-Z|X{=}x) &= E(Y|X{=}x) - E(Z|X{=}x) \\ &= E(\tau_x) - E(Z|X{=}x), && [(12.23)] \\ E(Y-Z|X{=}x') &= E(Y|X{=}x') - E(Z|X{=}x') \\ &= E(\tau_{x'}) - E(Z|X{=}x'), && [(12.23)] \end{aligned}$$

and $E(\delta_{xx'}) = E(\tau_x) - E(\tau_{x'})$, taking the difference $E(Y-Z|X{=}x) - E(Y-Z|X{=}x')$ proves the theorem.

Proof of Corollary 12.3

First we show that Equation (12.3) holds under the assumptions about X. If $X=0$, then Equation (12.3) yields

$$E(Y-Z\,|\,X=0) \;=\; \beta_0, \tag{12.24}$$

and if $X=x$, then it yields

$$E(Y-Z\,|\,X=x) \;=\; \beta_0+\beta_x, \quad x=1,\ldots,J. \tag{12.25}$$

Hence, the conditional expectation $E(Y-Z\,|\,X)$ can always be parameterized by Equation (12.3) and

$$\begin{aligned}\beta_x \;&=\; E(Y-Z\,|\,X=x) \;-\; E(Y-Z\,|\,X=0)\\ &=\; E(Y|\,X=x) \;-\; E(Z\,|\,X=x) \;-\; [E(Y|\,X=0) \;-\; E(Z\,|\,X=0)]\\ &=\; E(\delta_{x0}) \;-\; [E(Z\,|\,X=x) \;-\; E(Z\,|\,X=0)], \quad x=1,\ldots,J,\end{aligned} \tag{12.26}$$

because unbiasedness of $E(Y\,|\,X)$ w.r.t. total effects implies $E(Y|X=x)-E(Y|X=0) \;=\; E(\delta_{x0})$, for all $x=1,\ldots,J$ [see Defs. 6.1 (i) and 6.6 (i)].

Proof of Corollary 12.5

We presume $E(Z\,|\,X)=E(Z)$, which implies $E(Z\,|\,X=x)=E(Z)$ for all $x=0,1,\ldots,J$, and

$$\begin{aligned}\beta_x \;&=\; E(Y-Z\,|\,X=x) \;-\; E(Y-Z\,|\,X=0) && [(12.26)]\\ &=\; E(Y|\,X=x) \;-\; E(Z\,|\,X=x) \;-\; [E(Y|\,X=0) \;-\; E(Z\,|\,X=0)] && [\text{PR-Box 9.2 (xvi)}]\\ &=\; E(Y|\,X=x) \;-\; E(Z) \;-\; [E(Y|\,X=0) \;-\; E(Z)] && [E(Z\,|\,X=x)=E(Z)]\\ &=\; E(Y|\,X=x) \;-\; E(Y|\,X=0).\end{aligned}$$

Hence, $\beta_x=E(\delta_{x0})$ if and only if $E(Y|\,X=x) \;-\; E(Y|\,X=0)=E(\delta_{x0})$.

Proof of Theorem 12.10

First we show that Equation (12.13) follows from Equation (12.12).

$$\begin{aligned}E(Y-Z\,|\,X) \;&=\; E[E(Y-Z\,|\,X,Z)\,|\,X] && [\text{PR-Box 9.2 (v)}]\\ &=\; E[E(Y\,|\,X,Z)-Z\,|\,X] && [\text{PR-Box 9.2 (xvi), (vii)}]\\ &=\; E\big(\beta_0+1\cdot Z \;+\; \sum_{x=1}^{J}\beta_x 1_{X=x} \;-\; Z\,|\,X\big) && [(12.12)]\\ &=\; \beta_0 \;+\; \sum_{x=1}^{J}\beta_x 1_{X=x} && [\text{PR-Box 9.2 (v)}]\end{aligned}$$

Now we show that Equation (12.14) follows from Equation (12.12) and the additional assumption that $E(Y\,|\,X,Z)$ is unbiased w.r.t. total effects. If Equation (12.12) holds, then, according to Theorem PR-15.3 [with $g_i(Z)=\beta_x$],

$$E^{X=0}(Y\,|\,Z) \underset{P^{X=0}}{=} \beta_0+1\cdot Z \tag{12.27}$$

and

$$E^{X=x}(Y|Z) \underset{P^{X=x}}{=} \beta_0 + 1 \cdot Z + \beta_x, \quad x = 1, \ldots, J. \tag{12.28}$$

Therefore,

$$\beta_x \underset{P}{=} E^{X=x}(Y|Z) - E^{X=0}(Y|Z) \underset{P}{=} E(\delta_{x0}|Z), \tag{12.29}$$

the latter equation following from unbiasedness of $E(Y|X,Z)$ w.r.t. total effects (see Def. 6.6). The second equation follows from

$$\begin{aligned} E(\delta_{x0}) &= E[E(\delta_{x0}|Z)] && \text{[PR-Box 9.2 (iv)]} \\ &= E(\beta_x) && \text{[PR-Box 9.2 (i)]} \\ &= \beta_x. \end{aligned}$$

12.5 Exercises

Solutions

Chapter 13
Analysis of Covariance and its Generalizations

In chapter 10, we treated the theoretical foundations of a number of data analysis procedures that may be used for estimating and testing conditional and average causal effects, especially in quasi-experimental studies. In this chapter, we will show how analysis of covariance and its generalization, including latent covariates and latent outcome variables, can be used to estimate causal effects.

13.1 Analysis of Covariance (ANCOVA)

In this section, we show that in traditional analysis of covariance (ANCOVA) (see, e. g., Cochran, 1957; Cohen et al., 2003; Rao, 1973), we estimate the average total effect, provided that (a) the regression $E(Y|X,Z)$ is τ_x-unbiased and (b) the linearity assumptions about the functional form of $E(Y|X,Z)$ hold. Furthermore, under these two assumptions the adjusted means are estimates of the expectations $E(\tau_x)$ of the true-outcome variables. All examples presented in Tables 12.1 to 12.3 can correctly be analyzed with ANCOVA. Another example illustrates how to apply ANCOVA if the covariate is a qualitative nonnumerical variable.

13.1.1 Theory

If, for example, Z is a univariate numerical covariate and X is dichotomous with values 0 and 1, it is assumed in traditional analysis of covariance that

$$E(Y|X,Z) \underset{P}{=} \gamma_{00} + \gamma_{01} Z + \gamma_{10} X. \tag{13.1}$$

(The choices of the indices of the coefficients will be justified in section 13.2.) For $X=0$, this equation yields

$$E^{X=0}(Y|Z) \underset{P^{X=0}}{=} \gamma_{00} + \gamma_{01} Z, \tag{13.2}$$

which shows that the coefficients γ_{00} and γ_{01} are the coefficients of the $(X=0)$-conditional regression of Y on Z. For $X=1$, Equation (13.1) yields:

$$E^{X=1}(Y|Z) \underset{P^{X=1}}{=} (\gamma_{00} + \gamma_{10}) + \gamma_{01} Z. \tag{13.3}$$

If the conditional regressions $E^{X=0}(Y|Z)$ and $E^{X=1}(Y|Z)$ are P-unique, then we may take the difference

$$E^{X=1}(Y|Z) - E^{X=0}(Y|Z) \underset{P}{=} (\gamma_{00} + \gamma_{10}) + \gamma_{01} Z - (\gamma_{00} + \gamma_{01} Z) \underset{P}{=} \gamma_{10}.$$

Hence, according to Theorem 10.5,

$$E(\delta_{x0}\,|\,Z) \;=\; \gamma_{10},$$

i. e., the coefficient γ_{10} is the *conditional total effect of X given each value z of Z* provided that (a) $E(Y\,|\,X,Z)$ is τ_x-unbiased and (b) Equation (13.1) holds. Because this effect is constant for each value z, it is also the *average total effect of X*, i. e.,

$$E(\delta_{x0}) \;=\; \gamma_{10}$$

[see Eq. (10.6)].

In the following theorem we generalize these results for a Q-variate covariate $Z := (Z_1,\ldots,Z_Q)$ and $J+1$ different values $x = 0,1,\ldots,J$ of X, where $1_{X=x}$ again indicates with 1 if the event $\{X{=}x\} = \{\omega\in\Omega\colon X(\omega) = x\}$ occurs. Otherwise $1_{X=x}$ takes on the value 0.

Theorem 13.1 (Traditional ANCOVA)
Let $\langle\,(\Omega,\mathscr{A},P),(\mathscr{F}_t,t\in T),X,Y\,\rangle$ be a numerical causality space and let Z be measurable w.r.t. to C_X. Furthermore, let X be discrete with a finite number of values $x = 0,1,\ldots,J$, and let $P(X{=}x) > 0$ for all $x = 0,1,\ldots,J$. If $E(Y\,|\,X,Z)$ is τ_x-unbiased and

$$E(Y\,|\,X,Z) \;\underset{P}{=}\; \gamma_{00} + \sum_{i=1}^{Q}\gamma_{0i}\,Z_i \;+\; \sum_{x=1}^{J}\gamma_{x0}\,1_{X=x}, \tag{13.4}$$

then

$$E(\tau_0) \;=\; \gamma_{00} + \sum_{i=1}^{Q}\gamma_{0i}\,E(Z_i), \tag{13.5}$$

and, for $x = 1,\ldots,J$,

$$E(\tau_x) \;=\; \gamma_{00} + \gamma_{x0} + \sum_{i=1}^{Q}\gamma_{0i}\,E(Z_i), \tag{13.6}$$

$$E(\delta_{x0}) \;=\; E(\tau_x) - E(\tau_0) \;=\; \gamma_{x0}, \tag{13.7}$$

as well as

$$E(\delta_{x0}\,|\,Z) \;\underset{P}{=}\; \gamma_{x0}. \tag{13.8}$$

(Proof p. 327)

Remark 13.2 (Nonlinear Functions) Equation (13.4) allows for quadratic functions, e. g., because we may define $Z_2 := Z_1^2$ and $Z := (Z_1,Z_2)$. It also allows for interactions between covariates, because we may define $Z_3 := Z_1\cdot Z_2$ and $Z := (Z_1,Z_2,Z_3)$. If we consider a *discrete and non-numerical* covariate with values $z_0,z_1,\ldots,z_Q$, we may still use Equation (13.4) with $Z := (Z_1,\ldots,Z_Q)$ and $Z_1 := 1_{Z=z_1}$, …, $Z_Q := 1_{Z=z_Q}$, where $1_{Z=z_q}$ are the indicators for the values $z_1,\ldots,z_Q$ of the original covariate. ◁

Remark 13.3 (Invariant Slopes of the Covariates) Note that in Equation (13.4), there are no product terms between Z (or functions of Z) and the indicator variables for the values of X. This means that there are *no interactions* (in the sense of analysis of variance) *between X and the covariate*: the $(X{=}x)$-conditional regressions of Y on Z are parallel, i. e., the slope coefficients of the covariates are invariant for different values x of X. For $X{=}0$, Equation (13.4) yields

$$E^{X=0}(Y\,|\,Z) \;\underset{P^{X=0}}{=}\; \gamma_{00} + \sum_{i=1}^{Q}\gamma_{0i}\,Z_i, \tag{13.9}$$

and for $X{=}x, x\geq 1$,

$$E^{X=x}(Y|Z) \underset{P^{X=1}}{=} \gamma_{00} + \gamma_{x0} + \sum_{i=1}^{Q} \gamma_{0i} Z_i. \tag{13.10}$$

These equations show that the slopes of the covariates are invariant and that only the intercepts may change for different values x of X. ◁

Remark 13.4 (Invariant Effects of X) A second way to clarify the assumption of no interaction between X and the covariates in traditional ANCOVA is to condition on the values of the covariates, i. e., to consider

$$E^{Z=z}(Y|X) \underset{P^{Z=z}}{=} \left(\gamma_{00} + \sum_{i=1}^{Q} \gamma_{0i} z_i\right) + \sum_{x=1}^{J} \gamma_{x0} 1_{X=x} \tag{13.11}$$

This equation shows that Z modifies the intercepts (see the term in parentheses) of the $(Z{=}z)$-conditional regressions $E^{Z=z}(Y|X)$ of the outcomes on X, but not the effects of X on the outcomes. In traditional ANCOVA, they are constant, i. e., they do not depend on the values of the covariates. (In section 13.2, we will see (a) how the traditional ANCOVA model can be generalized allowing for interactions between X and the covariates and (b) why the notation for the indices of the γ-coefficients makes sense.) ◁

Remark 13.5 (Adjusted Means) According to Equation (10.2), the expectations of the true outcome variable τ_0 can be computed by

$$\begin{aligned} E(\tau_0) &= E\big(E^{X=0}(Y|Z)\big) && [\tau_0\text{-unbiasedness and Eq. (10.2)}] \\ &= E\big(\gamma_{00} + \sum_{i=1}^{Q} \gamma_{0i} Z_i\big) && [\text{Eq. (13.9)}] \\ &= \gamma_{00} + \sum_{i=1}^{Q} \gamma_{0i} E(Z_i) && [\text{Box PR-}\mathbf{??}\ (\mathbf{??})] \end{aligned} \tag{13.12}$$

and the expectation of the true outcome variable τ_x, $x \geq 1$, by

$$\begin{aligned} E(\tau_x) &= E\big(E^{X=x}(Y|Z)\big) \\ &= E\big(\gamma_{00} + \gamma_{x0} + \sum_{i=1}^{Q} \gamma_{0i} Z_i\big) \\ &= \gamma_{00} + \gamma_{x0} + \sum_{i=1}^{Q} \gamma_{0i} E(Z_i). \end{aligned} \tag{13.13}$$

It is these expectations of the true outcome variables τ_x that are estimated by the *adjusted means* in traditional analysis of covariance, provided that the regression $E(Y|X,Z)$ is τ_x-unbiased and that Equation (13.4) holds. ◁

Remark 13.6 (Conditional and Average Causal Effects) According to Equation (10.8), the *Z-conditional total effect function* of $X{=}x$ compared to $X{=}0$ is:

$$E(\delta_{xx'}|Z) \underset{P}{=} PFE_{x0;Z} = E^{X=x}(Y|Z) - E^{X=0}(Y|Z) = \gamma_{x0} \tag{13.14}$$

[see Eqs. (13.9) and (13.10)]. Hence, this conditional effect function does not depend on Z, which implies that γ_{x0} is also the average total effect $E(\delta_{xx'})$. Therefore, the *hypothesis of no (average and conditional) total effects of X* in ANCOVA is:

$$H_0: \quad \gamma_{10} = \ldots = \gamma_{J0} = 0. \tag{13.15}$$

Of course, one may also be interested in more specific hypotheses such as $\gamma_{x0} = 0$ for a specific value x of X. ◁

13.1.2 Numerical Examples

All examples presented in section 12.2 can be analyzed with traditional ANCOVA yielding the correct Z-conditional total effect function $E(\delta_{xx'}|Z) = 10$ and the average total effect $E(\delta_{xx'}) = 10$. For brevity, we demonstrate this only for Example 12.14. As a second example, we use a new example, in which the covariate is a qualitative variable with three values *low neediness*, *medium neediness*, and *high neediness*.

Example 13.7 (Example 3 – continued) As mentioned before, Example 12.14 is typical in a quasi-experiment in which pre- and post-tests are assessed: Y depends on the treatment variable X and the pre-test Z, and the expectations of Z are different in the two treatment conditions. This kind of examples has been called *non-equivalent group designs* (see, e. g., Reichardt, 1979, 2005).

Note that the example has been constructed such that there is no interaction between treatment variable and the pre-test. Therefore, the regression of Y on X and Z is a linear function of the two regressors, i. e.,

$$E(Y|X,Z) \underset{P}{=} \gamma_{00} + \gamma_{01} Z + \gamma_{10} X \underset{P}{=} 100 + 10\,X + .80\,Z \tag{13.16}$$

(see Exercise ???) and the two $(X{=}x)$-conditional regressions are

$$E^{X=0}(Y|Z) \underset{P^{X=0}}{=} \gamma_{00} + \gamma_{01} Z \underset{P^{X=0}}{=} 100 + .80\,Z.$$

and

$$E^{X=1}(Y|Z) \underset{P^{X=1}}{=} \gamma_{00} + \gamma_{10} + \gamma_{01} Z \underset{P^{X=1}}{=} 110 + .80\,Z.$$

Furthermore, the example has been constructed such that $E(Y|X,Z)$ is unbiased (see Exercise ???). Therefore, the two conditional regressions $E^{X=0}(Y|Z)$ and $E^{X=1}(Y|Z)$ are P-unique and the Z-conditional total effect function is

$$\begin{aligned} E(\delta_{10}|Z) &\underset{P}{=} E^{X=1}(Y|Z) - E^{X=0}(Y|Z) \\ &\underset{P}{=} 110 + .80\,Z - (100 + .80\,Z) \underset{P}{=} 10. \end{aligned}$$

Obviously, this function is a constant. Therefore, taking its expectation

$$E[E(\delta_{10}|Z)] \underset{P}{=} E(\delta_{10}) \underset{P}{=} 10.$$

yields the average total effect 10.

In this example, Z represents a pre-test and Y a post-test. Hence, we may also consider $Y-Z$ as a regressand in ANCOVA, i. e., we may consider the regression $E(Y-Z|X,Z)$. However,

$$\begin{aligned} E(Y-Z|X,Z) &\underset{P}{=} E(Y|X,Z) - E(Z|X,Z) \\ &\underset{P}{=} E(Y|X,Z) - Z && \text{[Box PR-10.2 (vii)]} \\ &\underset{P}{=} \gamma_{00} + (\gamma_{01} - 1)\,Z + \gamma_{10} X && \text{[Eq. 13.16]} \\ &\underset{P}{=} 100 + 10\,X - .20\,Z. \end{aligned}$$

Obviously, the coefficient of X does not change if we replace the regressand Y by the change-score variable $Y-Z$, and this implies that neither the Z-conditional total-effect function nor the average total effect change if we consider $Y-Z$ instead of Y as the regressand. ◁

Table 13.1. Expectations Within Treatment and Neediness Conditions

	neediness			
treatment	low (Z=0)	medium (Z=1)	high (Z=2)	
X=0	120 (20/120)	110 (17/120)	100 (3/120)	(40/120)
X=1	100 (7/120)	90 (26/120)	80 (7/120)	(40/120)
X=2	90 (3/120)	80 (17/120)	70 (20/120)	(40/120)
	(30/120)	(60/120)	(30/120)	

Note. Probabilities $P(X{=}x, Z{=}z)$ (cells), $P(Z{=}z)$ (columns), and $P(X{=}x)$ (rows) in parentheses.

Example 13.8 (ANCOVA With a Qualitative Covariate) Consider Table 13.1, assuming that the covariate-treatment regression $E(Y \mid X, Z)$ is τ_x-unbiased, i. e., assuming that $E(Y \mid X{=}x, Z{=}z) = E(\tau_x \mid Z{=}z)$ holds for each of the nine conditional expectation values of the outcome variable Y given *treatment* x and *neediness* z. In this example, the regression $E(Y \mid X, Z)$ can be parameterized as follows:

$$E(Y \mid X, Z) = \gamma_{00} + \gamma_{01} 1_{Z=1} + \gamma_{02} 1_{Z=2} + \gamma_{10} 1_{X=1} + \gamma_{20} 1_{X=2}. \tag{13.17}$$

In this equation, $1_{Z=1}$ indicates with 1 and 0 whether or not *neediness* is *medium*, and $1_{Z=2}$ with 1 and 0 whether or not *neediness* is *high*. Furthermore, $1_{X=1}$ indicates with 1 and 0 whether or not the unit is in treatment 1 and $1_{X=2}$ with 1 and 0 whether or not the unit is in treatment 2.

Equation (13.17) is a *restrictive parameterization* of the covariate-treatment regression with five γ-parameters describing nine conditional expectation values $E(Y \mid X{=}x, Z{=}z)$. As we shall see, in this specific example, this restriction holds, because there is no interaction between *neediness* and *treatment*.

Looking at the first row in Table 13.1, it is easily seen that the regression of Y on Z in treatment 0 is

$$E^{X=0}(Y \mid Z) = \gamma_{00} + \gamma_{01} 1_{Z=1} + \gamma_{02} 1_{Z=2} = 120 - 10 \cdot 1_{Z=1} - 20 \cdot 1_{Z=2},$$

and the regressions of Y on Z in treatments 1 and 2 are

$$E^{X=1}(Y \mid Z) = (\gamma_{00} + \gamma_{10}) + \gamma_{01} 1_{Z=1} + \gamma_{02} 1_{Z=2} = 100 - 10 \cdot 1_{Z=1} - 20 \cdot 1_{Z=2}$$

and

$$E^{X=2}(Y \mid Z) = (\gamma_{00} + \gamma_{20}) + \gamma_{01} 1_{Z=1} + \gamma_{02} 1_{Z=2} = 90 - 10 \cdot 1_{Z=1} - 20 \cdot 1_{Z=2},$$

respectively. Hence, the covariate-treatment regression is:

$$E(Y \mid X, Z) = 120 - 10 \cdot 1_{Z=1} - 20 \cdot 1_{Z=2} - 20 \cdot 1_{X=1} - 30 \cdot 1_{X=2}.$$

Assuming that the $E(Y \mid X, Z)$ is τ_x-unbiased, the Z-conditional total effect function comparing treatment 1 to treatment 0 is

$$\begin{aligned} E(\delta_{10}|Z) &\underset{P}{=} E^{X=1}(Y|Z) - E^{X=0}(Y|Z) \\ &\underset{P}{=} 100 - 10\cdot 1_{Z=1} - 20\cdot 1_{Z=2} - (120 - 10\cdot 1_{Z=1} - 20\cdot 1_{Z=2}) \\ &\underset{P}{=} -20. \end{aligned}$$

Obviously, this function is a constant. Therefore, taking its expectation yields

$$E[E(\delta_{10}|Z)] \underset{P}{=} E(\delta_{10}) \underset{P}{=} -20,$$

the average total effect of treatment 1 compared to treatment 0.

Similarly, the Z-conditional total effect function comparing treatment 2 to treatment 0 is

$$\begin{aligned} E(\delta_{20}|Z) &\underset{P}{=} E^{X=2}(Y|Z) - E^{X=0}(Y|Z) \\ &\underset{P}{=} 90 - 10\cdot 1_{Z=1} - 20\cdot 1_{Z=2} - (120 - 10\cdot 1_{Z=1} - 20\cdot 1_{Z=2}) \\ &\underset{P}{=} -30. \end{aligned}$$

Again, this function is a constant, and taking its expectation yields

$$E[E(\delta_{10}|Z)] \underset{P}{=} E(\delta_{10}) \underset{P}{=} -30,$$

the average total effect of treatment 2 compared to treatment 0.

◁

13.1.3 Conclusions

In this section, we have shown that traditional analysis of covariance indeed yields the correct estimates of the conditional and the average total effects, provided that we select the covariate Z such that

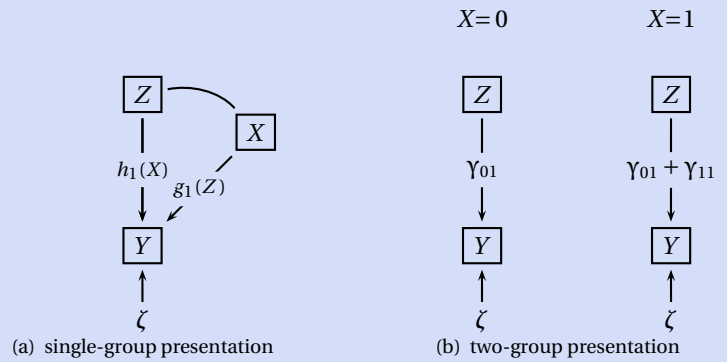
(a) the covariate-treatment regression $E(Y|X,Z)$ is τ_x-unbiased, and

(b) $E(Y|X,Z)$ can be written as a multiple linear regression with no terms for the interaction between X and the covariates [see Eq. (13.4)].

If ANCOVA is applied although these assumptions do not hold, it may neither yield estimates the average total effects, nor of the conditional total effects given the value of the covariate, nor test the hypothesis that these average total effects are zero.

13.1.4 Statistical Programs

ANCOVA can be accomplished with every major statistical program package and with every program for multiple linear regression. In the latter case, if there are more than two values of X, one may use, for instance, an R^2-difference test in order to test the hypothesis (13.15) of no effects of X. If there are just two values of X, then this hypothesis simplifies to $H_0: \gamma_{10} = 0$, which can be tested with a simple t-test. If there is severe heterogeneity of residual variances and, at the same time, severe differences of sample sizes between groups represented by X, we recommend a linear regression multi-group analysis [see Eqs. (13.9) and (13.10)]. These can be accomplished, e. g., with standard software for structural equation modeling such as *EQS* (Bentler, 1995), *LISREL* (Jöreskog & Sörbom, 1996/2001), or *Mplus* (Muthén & Muthén, 2002). Such analyses are greatly facilitated by *EffectLite* (Steyer & Partchev, 2007).



(a) single-group presentation (b) two-group presentation

Figure 13.1. Generalized ANCOVA model with manifest variables

13.2 Generalized ANCOVA

In traditional ANCOVA, we presume that Equation (13.4) holds, which does not allow for interaction between X and covariates. However, the adjustment theorem 10.2 is much more general, because it only requires that the regression $E(Y|X,Z)$ is τ_x-unbiased. Hence, *generalized analysis of covariance* (cf., e. g., Rogosa, 1980a) *does* allow for interaction between X and the covariates. If X is dichotomous with values 0 and 1, the fundamental equation for generalized analysis of covariance is:

$$E(Y|X,Z) \underset{P}{=} g_0(Z) + g_1(Z) \cdot X. \tag{13.18}$$

where the *intercept function* $g_0(Z)$ and the *effect function* $g_1(Z)$, are unknown functions of the (possibly multivariate, numerical or non-numerical) covariate $Z\colon \Omega \to \Omega'$. If Z is univariate and we assume

$$g_0(Z) = \gamma_{00} + \gamma_{01} Z$$

and

$$g_1(Z) = \gamma_{10} + \gamma_{11} Z,$$

then Equation (13.19) simplifies to

$$E(Y|X,Z) \underset{P}{=} (\gamma_{00} + \gamma_{01} Z) + (\gamma_{10} + \gamma_{11} Z) \cdot X. \tag{13.19}$$

13.2.1 Theory

Assuming that X has $J+1$ different values $x = 0, 1, \ldots, J$ and that the variables $1_{X=x}$ indicate with 1 and 0 whether or not X takes on the value x, the fundamental equation for generalized analysis of covariance is:

$$E(Y|X,Z) \underset{P}{=} g_0(Z) + g_1(Z) \cdot 1_{X=1} + \ldots + g_J(Z) \cdot 1_{X=J}, \tag{13.20}$$

where the *intercept function* $g_0(Z)$ and the *effect functions* $g_1(Z), \ldots, g_J(Z)$ are unknown functions of the (possibly multivariate, numerical or non-numerical) covariate $Z\colon \Omega \to \Omega'$.

Note that, if X has $J+1$ different values $x = 0,1,\ldots,J$, then there are always functions $g_0, g_1, \ldots, g_J : \Omega' \to \overline{\mathbb{R}}$ such that this equation true. In other words, Equation (13.20) does not contain any restrictions about the regression $E(Y|X,Z)$ as long as no additional assumptions about the intercept and/or effect functions are introduced.

Theorem 13.9 (Generalized ANCOVA)
Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a numerical causality space and let Z denote a covariate, i.e., let Z be measurable w.r.t. C_X, and let $g_x(Z)$, $x = 0,1,\ldots,J$, denote the functions occurring in Equation (13.20)*. Then τ_x-unbiasedness of the regression $E(Y|X,Z)$ implies:*

$$E(\tau_0) = E[g_0(Z)] \tag{13.21}$$

and, for each $x = 1,\ldots,J$,

$$E(\tau_x) = E[g_0(Z) + g_x(Z)], \tag{13.22}$$

$$E(\delta_{x0}) = E[g_x(Z)], \tag{13.23}$$

and

$$E(\delta_{x0}|Z) \underset{P}{=} g_x(Z). \tag{13.24}$$

(Proof p. 327)

Remark 13.10 ($(Z{=}z)$-Conditional Total Effects) Conditioning on the covariate, Equation (13.20) yields

$$E^{Z=z}(Y|X) \underset{P^{Z=z}}{=} g_0(z) + g_1(z)\mathbb{1}_{X=1} + \ldots + g_J(z)\mathbb{1}_{X=J} \tag{13.25}$$

This equation shows that the effects $g_x(z)$ of X may be different for different values z of the covariate Z.

The functions $g_x(Z)$, $x = 1,\ldots,J$, are called *effect functions*, because a value

$$g_x(z) = E(Y|X{=}x, Z{=}z) - E(Y|X{=}0, Z{=}z) = PFE_{x0;\,Z=z} \tag{13.26}$$

of such a function is the prima facie effect of $X{=}x$ compared to $X{=}0$ given a value z of the covariate Z. Such a conditional prima facie effect is also the conditional total effect $E(\delta_{x0}|Z{=}z)$ of $X{=}x$ compared to $X{=}0$ given the value z of the covariate Z, *provided that $E(Y|X,Z)$ is τ_x-unbiased* [see Eq. (6.3)]. In other words, under τ_x-unbiasedness of $E(Y|X,Z)$,

$$g_x(Z) \underset{P}{=} PFE_{x0;Z} = E(\delta_{x0}|Z). \tag{13.27}$$

Hence, the null hypothesis of *no effects of X* in generalized ANCOVA is:

$$H_0: \quad g_1(Z) = \ldots = g_J(Z) = 0. \tag{13.28}$$

Of course, more specific null hypotheses such as $g_x(Z) = 0$ for a specific value x of X are meaningful as well. Furthermore, we may be interested in the hypothesis that there is *no interaction between X and covariates*, i. e., in the hypothesis

$$H_0: \quad g_1(Z) = \gamma_{10} \quad \ldots \quad g_J(Z) = \gamma_{J0}. \tag{13.29}$$

With this hypothesis we postulate that all Z-conditional total-effect functions are constants. In fact, this hypothesis is equivalent to assuming that the fundamental equation [see Eq.(13.4)] of traditional ANCOVA holds. ◁

Remark 13.11 (Regressions of the Outcomes on Covariates given $X{=}x$) In order to use the formulas of Theorem 10.2 , we again consider the $(X{=}x)$-conditional regressions $E^{X=x}(Y \mid Z)$ of the outcome variable Y on the covariates Z. Equation (13.20) yields:

$$\begin{aligned} E^{X=0}(Y \mid Z) &=_{P^{X=x}} g_0(Z) \\ E^{X=x}(Y \mid Z) &=_{P^{X=x}} g_0(Z) + g_x(Z), \qquad x = 1, \ldots, J. \end{aligned} \tag{13.30}$$

◁

Remark 13.12 (Adjusted Means) In generalized ANCOVA, estimates of the expectations $E\big(E^{X=x}(Y \mid Z)\big)$ of the regressions $E^{X=x}(Y|Z)$ are called *adjusted means*. They are estimates of the expectations of the true outcome variables τ_x, provided that the regression $E(Y \mid X, Z)$ is τ_x-unbiased. ◁

Remark 13.13 (Average Total Effects) According to Equation (10.6), the *average total effect* of $(X{=}x)$ compared to $(X{=}0)$ is obtained by:

$$E(\delta_{x0}) \;=\; E(PFE_{x0;Z}) \;=\; E[g_x(Z)]. \tag{13.31}$$

According to this equation, the expectation of the function $g_x(Z)$ is the average total effect of $(X{=}x)$ compared to $(X{=}0)$ provided that the regression $E(Y \mid X, Z)$ is τ_x-unbiased. Hence, the null hypothesis of *no average total effects of X* in generalized analysis of covariance is:

$$H_0: \quad E[g_1(Z)] \;=\; \ldots \;=\; E[g_J(Z)] \;=\; 0. \tag{13.32}$$

Again, we may also be interested in a more specific null hypotheses such as $E[g_x(Z)] = 0$ for a given value x of X.

Note that the hypotheses *no average total effects of X* and *no effects of X*, i. e.,

$$H_0: \quad g_1(Z) \;=\; \ldots \;=\; g_J(Z) \;=\; 0. \tag{13.33}$$

are *not equivalent* in generalized ANCOVA. While 'no total effects of X' implies 'no *average* total effects of X', the latter hypothesis does not imply the first. Even if the total effects of X are zero on average, they may be different from zero for a given value of the covariate Z. ◁

Remark 13.14 (Parameterization of Intercept and Effect Functions) In empirical applications, the intercept functions $g_0(Z)$ and the effect functions $g_x(Z)$ have to be estimated if we are interested in the average total effects given the values of the covariate. If the covariate Z is discrete with values $z_0, z_1, \ldots, z_K$, the intercept functions can always be parameterized as

$$g_0(Z) \;=\; \gamma_{00} + \gamma_{01}\mathbb{1}_{Z=z_1} + \ldots + \gamma_{0K}\mathbb{1}_{Z=z_K}, \tag{13.34}$$

and the effect functions as

$$g_x(Z) \;=\; \gamma_{x0} + \gamma_{x1}\mathbb{1}_{Z=z_1} + \ldots + \gamma_{xK}\mathbb{1}_{Z=z_K}, \qquad x = 1, \ldots, J, \tag{13.35}$$

where each $\mathbb{1}_{Z=z_k}$ is the indicator for the value z_k of Z. In other cases, a saturated parametrization of the intercepts functions $g_0(Z)$ and the effect functions $g_x(Z)$ may not be estimable with a limited sample size. In these cases, simplifying assumptions about the functional form of these functions have to be introduced in order to be able to estimate them. ◁

Remark 13.15 (Linear Intercept and Effect Functions) If the covariate Z is *numerical and multidimensional*, i. e., if $Z := (Z_1, \ldots, Z_Q)$, and if X has more than two values, we may assume that the intercept and effect functions are linear, i. e.:

$$g_0(Z) = \gamma_{00} + \gamma_{01} Z_1 + \ldots + \gamma_{0Q} Z_Q, \tag{13.36}$$

and

$$g_x(Z) = \gamma_{x0} + \gamma_{x1} Z_1 + \ldots + \gamma_{xQ} Z_Q, \qquad x = 1, \ldots, J. \tag{13.37}$$

Note again that these equations also allow for interactions between covariates, because we may define $Z_3 := Z_1 \cdot Z_2$ and $Z := (Z_1, Z_2, Z_3)$. Similarly, they allow for quadratic functions, e. g., because we may define $Z_2 := Z_1^2$ and $Z := (Z_1, Z_2)$. If 'linear' intercept and effect functions are not appropriate, we may also use nonparametric (see, e. g., Huet, Bouvier, Poursat, & Jolivet, 2004) or semi-parametric regression methods (see, e. g., Fahrmeir & Tutz, 2001; Györfi et al., 2002). ◁

13.2.2 Numerical Examples

All examples treated in this chapter can be analyzed with generalized ANCOVA. In all these examples, the Z-conditional total effect functions are constant and the average total effect is equal the $(Z{=}z)$-conditional total effect of X. Instead of applying generalized ANCOVA to these examples, we revisit the nonorthogonal ANOVA, which has already bee presented in chapter 1.

Example 13.16 (Nonorthogonal ANOVA) Consider Table 1.4 (see p. 10), assuming that the covariate-treatment regression $E(Y|X,Z)$ is τ_x-unbiased, i. e., assuming $E(Y|X{=}x, Z{=}z) = E(\tau_x|Z{=}z)$ for each of the nine conditional expectation values given *treatment* x and *neediness*. In this example, the covariate-treatment regression can be parameterized as follows:

$$\begin{aligned} E(Y|X,Z) &= g_0(Z) + g_1(Z)\,\mathbf{1}_{X=1} + g_2(Z)\,\mathbf{1}_{X=2} \\ &= \gamma_{00} + \gamma_{01}\,\mathbf{1}_{Z=1} + \gamma_{02}\,\mathbf{1}_{Z=2} \\ &\quad + \big(\gamma_{10} + \gamma_{11}\,\mathbf{1}_{Z=1} + \gamma_{12}\,\mathbf{1}_{Z=2}\big)\,\mathbf{1}_{X=1} \\ &\quad + \big(\gamma_{20} + \gamma_{21}\,\mathbf{1}_{Z=1} + \gamma_{22}\,\mathbf{1}_{Z=2}\big)\,\mathbf{1}_{X=2}. \end{aligned} \tag{13.38}$$

In these equations, $\mathbf{1}_{X=1}$ indicates with 1 and 0 whether or not the unit is in treatment 1 and $\mathbf{1}_{X=2}$ with 1 and 0 whether or not the unit is in treatment 2. Furthermore, $\mathbf{1}_{Z=1}$ indicates with 1 and 0 whether or not *neediness* is *medium*, and $\mathbf{1}_{Z=2}$ with 1 and 0 whether or not *neediness* is *high*. Hence, for three treatment conditions, we have one intercept function, $g_0(Z)$, and two effect functions, $g_1(Z)$ and $g_2(Z)$. While the values $g_1(z)$ of the effect function $g_1(Z)$ are the $(Z{=}z)$-conditional mean differences in Y between treatment 1 and control, the values $g_2(z)$ of $g_2(Z)$ are the $(Z{=}z)$-conditional mean differences in Y between treatment 2 and control [see Eq. (13.26)]. Equation (13.38) is a saturated parameterization of the covariate-treatment regression with *nine* γ-parameters describing *nine* conditional expectation values $E(Y|X{=}x, Z{=}z)$.

Using Equation (13.38) and the first row of Table 1.4, the *intercept function* is

$$g_0(Z) = \gamma_{00} + \gamma_{01}\,\mathbf{1}_{Z=1} + \gamma_{02}\,\mathbf{1}_{Z=2} = 120 - 10 \cdot \mathbf{1}_{Z=1} - 60 \cdot \mathbf{1}_{Z=2}.$$

It is easy to see from the other rows of this table that the *effect functions* are

$$g_1(Z) = \gamma_{10} + \gamma_{11}\,\mathbf{1}_{Z=1} + \gamma_{12}\,\mathbf{1}_{Z=2} = -20 + 10 \cdot \mathbf{1}_{Z=1} + 60 \cdot \mathbf{1}_{Z=2} \tag{13.39}$$

and

$$g_2(Z) \;=\; \gamma_{20} + \gamma_{21}\mathbb{1}_{Z=1} + \gamma_{22}\mathbb{1}_{Z=2} \;=\; -40 + 20\cdot\mathbb{1}_{Z=1} + 120\cdot\mathbb{1}_{Z=2}. \tag{13.40}$$

Hence, the covariate-treatment regression is:

$$\begin{aligned} E(Y|X,Z) \;=\; & 120 - 10\cdot\mathbb{1}_{Z=1} - 60\cdot\mathbb{1}_{Z=2} \\ & + (-20 + 10\cdot\mathbb{1}_{Z=1} + 60\cdot\mathbb{1}_{Z=2})\cdot\mathbb{1}_{X=1} \\ & + (-40 + 20\cdot\mathbb{1}_{Z=1} + 120\cdot\mathbb{1}_{Z=2})\cdot\mathbb{1}_{X=2}. \end{aligned} \tag{13.41}$$

In this example, there are *two effect functions* $g_1(Z)$ and $g_2(Z)$, the values of which are the $(Z{=}z)$-conditional causal effects of treatment 1 compared to control and of treatment 2 compared to control, respectively. Hence, we also have *two average total effects*, which can be computed via the expectations $E[g_1(Z)]$ and $E[g_2(Z)]$ of these two effect functions. Because, in this example, the regressions $E^{X=x}(Y|Z)$ are P-unique, we can consider their differences. Hence, $g_1(Z) = E^{X=1}(Y|Z) - E^{X=0}(Y|Z)$ and $g_2(Z) = E^{X=2}(Y|Z) - E^{X=0}(Y|Z)$. Taking the expectations of the functions $g_1(Z)$ and $g_2(Z)$ yields

$$\begin{aligned} E[g_1(Z)] \;&=\; E(-20 + 10\cdot\mathbb{1}_{Z=1} + 60\cdot\mathbb{1}_{Z=2}) && \text{[Eq. (13.39)]} \\ &=\; -20 + 10\cdot E(\mathbb{1}_{Z=1}) + 60\cdot E(\mathbb{1}_{Z=2}) && \text{[Box PR-?? (??)]} \\ &=\; -20 + 10\cdot P(Z{=}1) + 60\cdot P(Z{=}2) && [E(\mathbb{1}_A) = P(A)] \\ &=\; -20 + 10\cdot 1/2 + 60\cdot 1/4 \;=\; 0 && [P(Z{=}1) = 1/2,\ P(Z{=}2) = 1/4] \end{aligned}$$

and

$$\begin{aligned} E[g_2(Z)] \;&=\; E(-40 + 20\cdot\mathbb{1}_{Z=1} + 120\cdot\mathbb{1}_{Z=2}) && \text{[Eq. (13.40)]} \\ &=\; -40 + 20\cdot E(\mathbb{1}_{Z=1}) + 120\cdot E(\mathbb{1}_{Z=2}) && \text{[Box PR-?? (??)]} \\ &=\; -40 + 20\cdot P(Z{=}1) + 120\cdot P(Z{=}2) && [E(\mathbb{1}_A) = P(A)] \\ &=\; -40 + 20\cdot 1/2 + 120\cdot 1/4 \;=\; 0 && [P(Z{=}1) = 1/2,\ P(Z{=}2) = 1/4]. \end{aligned}$$

In this example, both average total effects are zero.

Note that the average total treatment effects are uniquely defined and meaningful. Assuming τ_x-unbiasedness of the covariate-treatment regression, these average total effects are equal to the expectations of the effect functions $g_1(Z)$ and $g_2(Z)$ even though the treatment variable X and the covariate Z are correlated and even if there is interaction between X and Z. Because Z can be qualitative, i. e., a 'factor' in the terminology of analysis of variance, the confusion about how to analyze main effects in nonorthogonal analysis of variance can be dissolved. The only meaningful 'main effects' are the expectations of the effect functions. Under τ_x-unbiasedness of $E(Y|X,Z)$, they are equal to the average total effects. Whether or not these average effects are of interest in a specific application is a matter of substantive judgement in the specific application considered. Remember, an average effect is only an *average* and the conditional effects $g_1(z)$ and $g_2(z)$ usually are more informative. Under τ_x-unbiasedness of $E(Y|X,Z)$, $g_1(z)$ and $g_2(z)$, are the conditional total treatment effects *given the value z of neediness*.

◁

13.2.3 Conclusions

In this section, we have shown that generalized analysis of covariance yields the correct estimates of the average total effects and of the conditional total effects given the values

of a covariate, provided that the covariate Z is such that (a) the regression $E(Y|X,Z)$ is τ_x-unbiased and that (b) $E(Y|X,Z)$ is correctly parameterized [see Eq. (13.20)], including the correct intercept and effect functions [see Eqs. (13.34) to (13.37)]. Generalized ANCOVA will neither necessarily yield correct estimates the average total effects, nor of the conditional total effects given the value of the covariate, nor tests of the hypothesis that these total effects are zero, if these assumptions do not hold.

13.2.4 Statistical Programs

Some goals of Generalized ANCOVA can be accomplished with every program for multiple linear regression. Among these goals are estimating the regression coefficients [see Eqs. (13.34) to (13.37)], as well as testing the hypotheses of *no treatment effects* [see Eq. (13.28)] and of *no interaction between covariates and treatments* [see Eq. (13.29)]. Again, if there are more than two treatment conditions, one may use, for instance, an R^2-difference test for these hypotheses. For just two values of X, these hypotheses can also be tested with simple t-tests. In cases of severe heterogeneity of residual variances and, at the same time, severe differences of sample sizes between treatment groups, we again recommend linear regression multi-group analysis [see Eq. (13.30)]. As mentioned before, these can be accomplished, e. g., with standard software for structural equation modeling such as *LISREL* (Jöreskog & Sörbom, 1996/2001) or *Mplus* (Muthén & Muthén, 2002). Again, such analyses are greatly facilitated by *EffectLite* (Steyer & Partchev, 2007). These programs can also be used to test the hypothesis of *no average total effects* [see Eq. (13.32)]. In standard regression programs, tests of these hypotheses are not available unless (a) the covariate is mean-centered and (b) the covariate is a fixed regressor, implying, for instance, that the sample mean does not change between samples. In most applications this assumption will not hold.

If the covariate is a discrete qualitative variable (a "blocking factor"; see e. g., Maxwell & Delaney, 2004) such as in Example 13.16, one might consider using programs for *nonorthogonal analysis of variance*. However, none of the statistical packages such as SAS, SysStat, or SPSS with their Type I, II, III or IV sums of squares provide correct estimates and tests of the average total effects (or main effects) for such a design. Furthermore, traditional ANCOVA fails as well. For sample data, correct estimates and tests can easily be accomplished with *EffectLite* (Steyer & Partchev, 2007).

13.3 Generalized ANCOVA With Latent Variables

Generalized analysis of covariance described in the last section can also be applied to *latent* covariates and/or *latent* outcome variables. This means that we may also analyze conditional and average total effects given a latent covariate or (a vector of) several latent covariates.

13.3.1 Theory

Applying generalized ANCOVA *with latent covariates* may be important if the outcome variable and the treatment probabilities are not determined by *manifest* covariates, but by one or several *latent* covariates. Note that these treatment probabilities need not necessarily be known. What has to be known (in empirical applications: estimated) in such a

case is the regression

$$E(Y|X,\xi) \underset{P}{=} g_0(\xi) + g_1(\xi)\,1_{X=1} + \ldots + g_J(\xi)\,1_{X=J}, \tag{13.42}$$

where X denotes the cause (e. g., a treatment variable) and ξ represents the latent covariate. The conditional effect functions $g_x(\xi)$ [see Eq. (13.42)] are of interest even if X and ξ are independent, because they are a useful tool to study interaction, i. e., to study how the effects of X are modified by the latent covariate ξ. Once the regression $E(Y|X,\xi)$ is estimated, we can proceed as in section 13.2.

The Random Experiment. For simplicity, let us consider a random experiment in which a unit is sampled, and two pre-tests Z_1 and Z_2 are assessed, which are intended to measure the same latent covariate ξ, say *depression at the onset of treatment*. This latent covariate will be defined to be a function of the observational-unit variable U, representing a unobserved attribute of the unit, the strength of his or her depression at the onset of treatment. Furthermore, let us assume that ξ determines the treatment probabilities $P(X{=}x|U) = P(X{=}x|\xi) \neq P(X{=}x)$, $x = 0,1,\ldots,J$. An appropriate time after treatment, the outcome variable Y is observed.

Filtration and Global Covariate. In this example, the *filtration* $(\mathscr{F}_t, t \in T)$ is constructed as follows: $\mathscr{F}_1 = \sigma(U)$, $\mathscr{F}_2 = \sigma(U, Z_1, Z_2)$, $\mathscr{F}_3 = \sigma(U, Z_1, Z_2, X)$, and $\mathscr{F}_4 = \sigma(U, Z_1, Z_2, X, Y)$. Furthermore, the *global covariate* C_X is defined to be (U, Z_1, Z_2).

Definition of the Latent Covariate As any other random variable, a latent variable has to be defined. In this example we may define

$$\xi := E(Z_1|U) = E(Z_1|U,X), \tag{13.43}$$

i. e., ξ is specified to be the regression of Z_1 on U (see, e. g., Steyer, 2001; Steyer & Eid, 2001) and it is assumed that Z_1 is U-conditionally regressively independent of X. This implies

$$E^{X=x}(Z_1|U) \underset{P^{X=x}}{=} \xi, \quad x = 0,1,\ldots,J, \tag{13.44}$$

i. e., ξ is also the regression of Z_1 on U in the two treatment conditions (see Exercise ???). Hence, if X represents a treatment variable, this allows for mean differences of the latent covariate between different treatment groups:

$$E^{X=x}(\xi) = E^{X=x}[E^{X=x}(Z_1|U)] = E^{X=x}(Z_1), \quad x = 0,1,\ldots,J, \tag{13.45}$$

(see Exercise ???). Obviously, the expectations $E^{X=x}(\xi)$ of ξ in the treatment conditions x are identical with the expectations $E^{X=x}(Z_1)$ of the manifest covariate Z_1 in these treatment conditions.

Measurement Model. Although ξ is well-defined by Equation (13.43) we are not able, for instance, to determine its variance from empirically estimable parameters unless additional assumptions are introduced. The reason is that with Equation (13.43) we decompose Z_1 into two terms:

$$Z_1 = \xi + \delta_1, \tag{13.46}$$

where

$$\delta_1 = Z_1 - E(Z_1|U,X)$$

is the *measurement error variable*. Note that ξ and δ_1 are well-defined but unknown. This implies that we may estimate the variance and the $(X{=}x)$-conditional variances of Z_1, but this does not inform us about the variances and the $(X{=}x)$-conditional variances of ξ and of δ_1. However, the definition of the residual δ_1 does imply

$$E^{X=x}(\delta_1 \,|\, U) \;=\; 0, \quad x = 0, 1, \ldots, J \tag{13.47}$$

(see Exercise ???). Because δ_1 is the residual with respect to the regression $E(Z_1 \,|\, U, X)$, we only know that

$$Var^{X=x}(Z_1) \;=\; Var^{X=x}(\xi) + Var^{X=x}(\delta_1), \quad x = 0, 1, \ldots, J$$

(see Exercise ???). Therefore, we need a measurement model for ξ, which involves at least two manifest variables in order to estimate the $(X{=}x)$-conditional variances and the covariances of ξ with other variables.

For two manifest measures Z_1 and Z_2 of a single latent covariate ξ, a treatment variable X, and an observational-unit variable U, the model of *τ-congeneric variables* is defined by the following additional assumptions:

$$Z_2 \;=\; \lambda_{20} + \lambda_{21}\,\xi + \delta_2, \qquad \lambda_{20}, \lambda_{21} \in \mathbb{R}, \tag{13.48}$$

$$E(\delta_2 \,|\, X, U) \;=\; 0, \tag{13.49}$$

$$Cov(\delta_1, \delta_2 \,|\, X, U) \;=\; 0. \tag{13.50}$$

These assumptions imply

$$E^{X=x}(Z_2 \,|\, U) \underset{P^{X=x}}{=} \lambda_{20} + \lambda_{21}\,\xi, \quad x = 0, 1, \ldots, J, \tag{13.51}$$

and

$$Cov^{X=x}(\delta_1, \delta_2) \;=\; 0, \quad x = 0, 1, \ldots, J \tag{13.52}$$

(see Exercise ???). The intercept λ_{20} and the loading λ_{21} in Equations (13.48) and (13.51) are invariant for different values x of X. This can be seen if we consider the $(X{=}x)$-conditional regressions

$$E^{X=x}(Z_2 \,|\, U) \;=\; \lambda_{20} + \lambda_{21}\,\xi, \quad x = 0, 1, \ldots, J, \tag{13.53}$$

(see Exercise ???).

An additional assumption is

$$Cov(\delta_1, Y \,|\, X) \;=\; Cov(\delta_2, Y \,|\, X) \;=\; 0, \tag{13.54}$$

implying that the measurement error variables δ_1 and δ_2 are uncorrelated with Y in each treatment condition x, i. e.,

$$Cov^{X=x}(\delta_1, Y) \;=\; Cov^{X=x}(\delta_2, Y) \;=\; 0, \quad x = 0, 1, \ldots, J. \tag{13.55}$$

The measurement error variances $Var^{X=x}(\delta_i)$ can be assumed to be invariant or to vary for different values x of X. In the measurement model of *essentially τ-equivalent variables*, the loading λ_{21} is fixed to 1.

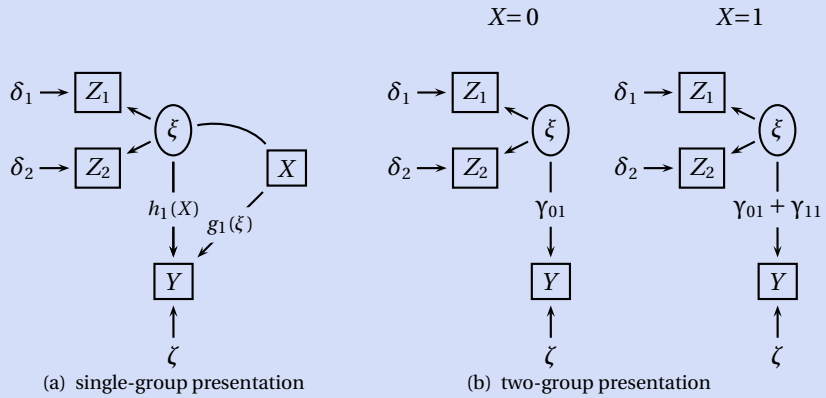


Figure 13.2. Path diagram of a generalized ANCOVA model with latent variables

Path Diagrams. Figure 13.2 represents two path diagrams of the same model with two manifest measures of the latent covariate ξ, a dichotomous variable X, and a single outcome variable Y. Figure 13.2 (a) depicts a single-group presentation including the variable X and the effect function $g_1(\xi)$. In contrast, in Figure 13.2 (b), we have a separate path diagram for each of the two values of X. These two path diagrams show that the slope of the linear regression of the outcome Y on the latent covariate ξ may be different for each value x of X.[1] The two-group presentation also displays two different slope coefficients of the conditional regressions of Y on ξ for X=1 and X= 0. The difference between these two slope coefficients is γ_{11}, the interaction parameter occurring in the equation

$$E(Y|X,\xi) \underset{P}{=} g_0(\xi) + g_1(\xi)\,X \underset{P}{=} (\gamma_{00} + \gamma_{01}\,\xi) + (\gamma_{10} + \gamma_{11}\,\xi)\,X. \tag{13.56}$$

An alternative way to look at this regression is

$$E(Y|X,\xi) \underset{P}{=} h_0(X) + h_1(X)\,\xi \underset{P}{=} (\gamma_{00} + \gamma_{10}X) + (\gamma_{01} + \gamma_{11}X)\,\xi. \tag{13.57}$$

Both kinds of equations have already been considered in the case of a manifest covariate. The only difference is that the covariate is now denoted ξ instead of Z.

Of course, Equations (13.56) and (13.57) can be wrong, because ξ is not dichotomous but assumed to be a latent variable with more than two values. In such a case, the intercept function $g_0(\xi)$ and the effect function $g_1(\xi)$ might be nonlinear functions of ξ (see, e. g., Hoyer & Klein, 2000; Kenny & Judd, 1984; Klein & Moosbrugger, 2000; Klein & Muthén, 2007; Marsh, Wen, & Hau, 2004; Song & Lee, 2006; Wall & Amemiya, 2003 for nonlinear latent variable models.)

More Than two Values of X. How to generalize the parameterization of the regression $E(Y|X,\xi)$ [see Eq. (13.56)] to more than two values of X? In this case, too, we may

[1] Allowing for (a) different slopes for different values x of X distinguishes this model from the traditional analysis of covariance with latent covariates (see, e. g., Klauer, Willmes, & Phye, 2002; Hancock, 2004; Sörbom, 1978), which does not allow for interaction between treatments and covariates.

assume that all effect functions $g_x(\xi)$, $x = 1,\ldots,J$, are linear. Under this assumption, Equation (13.42) can be written:

$$\begin{aligned} E(Y|X,\xi) &\underset{P}{=} g_0(\xi) + g_1(\xi)\,1_{X=1} + \cdots + g_J(\xi)\,1_{X=J} \\ &\underset{P}{=} (\gamma_{00} + \gamma_{01}\xi) + (\gamma_{10} + \gamma_{11}\xi)\,1_{X=1} + \ldots + (\gamma_{J0} + \gamma_{J1}\xi)\,1_{X=J}. \end{aligned} \tag{13.58}$$

Considering the conditional regressions of Y on ξ given $X=0$ yields

$$E^{X=0}(Y|\xi) \underset{P^{X=0}}{=} g_0(\xi) \underset{P^{X=0}}{=} \gamma_{00} + \gamma_{01}\xi.$$

Similarly, the conditional regressions of Y on ξ given $X=x$ are

$$E^{X=x}(Y|\xi) \underset{P^{X=x}}{=} g_0(\xi) + g_x(\xi) \underset{P^{X=0}}{=} (\gamma_{00} + \gamma_{x0}) + (\gamma_{01} + \gamma_{x1})\,\xi.$$

Hence, if the conditional regressions $E^{X=0}(Y|\xi)$ and $E^{X=x}(Y|\xi)$ are P-unique, the regression coefficients occurring in Equation (13.58) can be computed by taking the differences between the group-specific linear regressions of the outcome variable Y on the covariate ξ. This also yields the effect functions:

$$g_x(\xi) = \gamma_{x0} + \gamma_{x1}\xi, \quad x = 1,\ldots,J. \tag{13.59}$$

Now we have specified the intercepts function $g_0(\xi)$ and the effect functions $g_x(\xi)$ and can use the theory of generalized ANCOVA presented in section 13.2.1. Before turning to a numerical example let consider designs in which the outcome variable is latent as well.

Latent Outcome Variables

Latent *outcome variables* can be handled in a similar way as latent covariates. This means that all analyses considered before, i. e., the analysis of conditional and average total effects given manifest covariates or latent covariates, can be conducted also in the case in which the outcome variable is latent. The measurement model for the latent outcome variables is similar to the measurement model for the latent covariates. However, there are also some differences and additional assumptions. Therefore it is worthwhile to go through some aspects of the model in more detail.

The Random Experiment. Again we consider a random experiment in which a unit is sampled, and two pre-tests Z_1 and Z_2 are assessed, which are intended to measure the same latent covariate ξ, say *depression at the onset of treatment*. As before, the latent covariate will be defined to be a function of the observational-unit variable U, representing a unobserved attribute of the unit, the strength of his or her depression at the onset of treatment, and again we assume that ξ determines the treatment probabilities $P(X=x|U) = P(X=x|\xi) \neq P(X=x)$, $x = 0,1,\ldots,J$. An appropriate time after treatment, two manifest outcome variables Y_1 and Y_2 are observed.

Filtration and Global Covariate. In this design, the *filtration* $(\mathscr{F}_t, t \in T)$ is constructed as follows:

(a) $\mathscr{F}_1 = \sigma(U)$,

(b) $\mathscr{F}_2 = \sigma(U, Z_1, Z_2)$,

(c) $\mathscr{F}_3 = \sigma(U, Z_1, Z_2, X)$,

(d) $\mathscr{F}_4 = \sigma(U, Z_1, Z_2, X, \mathscr{M})$,

(e) $\mathscr{F}_5 = \sigma(U, Z_1, Z_2, X, \mathscr{M}, Y_1, Y_2)$.

The σ-algebra $\mathscr{M}$ contains all events that may occur in between treatment and the assessment of the manifest outcome variables Y_1 and Y_2. It will be used in the definition of the latent outcome variable. Furthermore, the *global covariate* C_X is again defined to be $\sigma(U, Z_1, Z_2)$.

The definition of the latent covariate and its measurement model remain unchanged [see Eqs. (13.43) to (13.53)]. However, all equations involving the outcome variable Y have to be changed. Before going into details, let us define the latent outcome variable.

Definition of the Latent Outcome Variable In this design we define

$$\eta := E(Y_1 \,|\, \mathscr{F}_4), \tag{13.60}$$

i. e., η is specified to be the conditional expectation of Y_1 given $\mathscr{F}_4$. The latent outcome variable η is the systematic component of Y_1 that is purged from measurement error. Therefore, it may also be called the *true-score variable* of Y_1. In the definition of η, we condition on the following random variables and events that are pre-ordered to Y_1 with respect to the filtration $(\mathscr{F}_t, t \in T)$:

(a) the latent covariate ξ, because it is measurable with respect to U,

(b) the manifest covariates Z_1 and Z_2,

(c) the treatment variable X, and

(d) the σ-algebra $\mathscr{M}$, which contains all events and, via their generated σ-algebras, all random variables on $(\Omega, \mathscr{A}, P)$ that are *in between* X and Y_1.

The intermediate events and random variables may systematically affect Y_1, and this is why they have to be taken into account in the definition of η.

Measurement Model. Although η is well-defined by Equation (13.60) we are not able to determine its variance and its covariances with other variables from empirically estimable parameters unless additional assumptions are introduced. The reason is that with Equation (13.60) we decompose Y_1 into two terms:

$$Y_1 = \eta + \varepsilon_1, \tag{13.61}$$

where

$$\varepsilon_1 = Y_1 - E(Y_1 \,|\, \mathscr{F}_4)$$

is the *measurement error variable*. The definition of the residual ε_1 does imply, e. g.,

$$E^{X=x}(\varepsilon_1 \,|\, \mathscr{F}_4) = E^{X=x}(\varepsilon_1 \,|\, \eta) = E^{X=x}(\varepsilon_1 \,|\, \xi) = 0, \quad x = 0, 1, \ldots, J, \tag{13.62}$$

$$E^{X=x}(\varepsilon_1 \,|\, Z_1) = E^{X=x}(\varepsilon_1 \,|\, Z_2) = E^{X=x}(\varepsilon_1) = 0, \quad x = 0, 1, \ldots, J, \tag{13.63}$$

[see Box PR-11.1, Rules (vi) and (vii)], as well as

$$Cov^{X=x}(\varepsilon_1, Z_1) = Cov^{X=x}(\varepsilon_1, Z_2) = 0, \quad x = 0, 1, \ldots, J, \tag{13.64}$$

and

$$Cov^{X=x}(\varepsilon_1, \xi) = Cov^{X=x}(\varepsilon_1, \delta_1) = Cov^{X=x}(\varepsilon_1, \delta_2) = 0, \quad x = 0, 1, \ldots, J. \tag{13.65}$$

Because $Cov^{X=x}(\varepsilon_1, \eta) = 0$, we also know that

$$Var^{X=x}(Y_1) = Var^{X=x}(\eta) + Var^{X=x}(\varepsilon_1), \quad x = 0, 1, \ldots, J. \tag{13.66}$$

Although we may estimate the $(X{=}x)$-conditional variance of Y_1, this neither informs us about the $(X{=}x)$-conditional variances of η and of $(X{=}x)$-conditional ε_1 nor about their (unconditional) variances and covariances. Therefore, we need a measurement model for η involving at least two manifest variables in order to estimate the $(X{=}x)$-conditional variances and covariances of η with other variables.

For two manifest measures Y_1 and Y_2 of a single latent outcome variable η, the model of *τ-congeneric variables* is defined by the following additional assumptions:

$$Y_2 = \kappa_{20} + \kappa_{21}\eta + \varepsilon_2, \qquad \kappa_{20}, \kappa_{21} \in \mathbb{R}, \tag{13.67}$$

$$E(\varepsilon_2 \,|\, \mathscr{F}_4) = 0, \tag{13.68}$$

$$Cov(\varepsilon_1, \varepsilon_2 \,|\, \mathscr{F}_4) = 0. \tag{13.69}$$

The definition of η implies

$$E^{X=x}(Y_1 \,|\, \eta) \underset{p^{X=x}}{=} \eta, \quad x = 0, 1, \ldots, J, \tag{13.70}$$

and

$$Cov^{X=x}(\varepsilon_1, \eta) = 0, \tag{13.71}$$

and the assumptions (13.67) to (13.69) imply

$$E^{X=x}(Y_2 \,|\, \eta) \underset{p^{X=x}}{=} \kappa_{20} + \kappa_{21}\eta, \quad x = 0, 1, \ldots, J, \tag{13.72}$$

$$Cov^{X=x}(\varepsilon_2, \eta) = 0, \tag{13.73}$$

and

$$Cov^{X=x}(\varepsilon_1, \varepsilon_2) = 0, \quad x = 0, 1, \ldots, J \tag{13.74}$$

(see Exercise ???). The intercept κ_{20} and the loading κ_{21} in Equations (13.67) and (13.72) are invariant for different values x of X. Other implications are

$$Cov^{X=x}(\varepsilon_i, \delta_j) = 0, \quad i, j = 1, 2. \tag{13.75}$$

The measurement error variances $Var^{X=x}(\varepsilon_i)$ can be assumed to be invariant or to vary for different values x of X. In the measurement model of *essentially τ-equivalent variables*, the loading κ_{21} is fixed to 1, whereas the intercept κ_{20} is not fixed and may differ from 0.

Path Diagrams. Figure 13.3 represents two path diagrams of the same model with two manifest measures of the latent covariate ξ, a dichotomous variable X, and a single latent outcome variable η that is measured by two manifest variables Y_1 and Y_2. Figure 13.3 (a) depicts a single-group presentation including the variable X and the effect functions $g_1(\xi)$ and $h_1(X)$. In contrast, in Figure 13.3 (b), we have a separate path diagram for each of the two values of X. These two path diagrams show that the slope of the linear regression of the latent outcome variable η on the latent covariate ξ may be different for each value

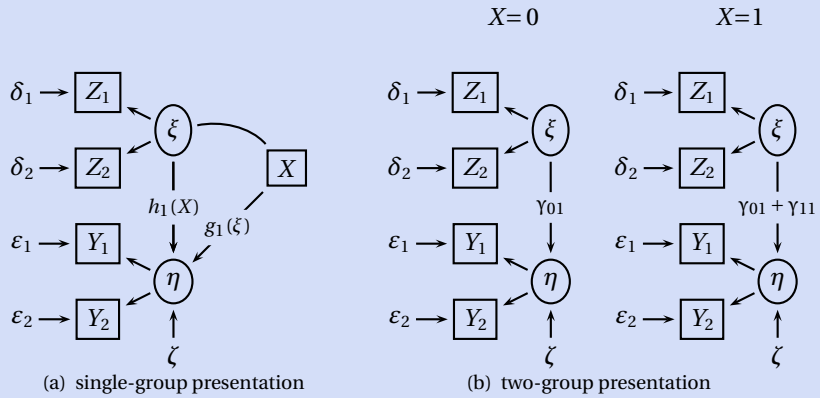


Figure 13.3. Path diagrams of a generalized ANCOVA model with latent variables

x of X.[2] The two-group presentation also displays two different slope coefficients of the conditional regressions of η on ξ for X=1 and X=0. These two slope coefficients are the values of the function $h_1(X)$. The difference between these two slope coefficients is γ_{11}, the interaction parameter occurring in the equation

$$E(\eta\,|\,X,\xi) \underset{P}{=} g_0(\xi) + g_1(\xi)\,X \underset{P}{=} (\gamma_{00} + \gamma_{01}\xi) + (\gamma_{10} + \gamma_{11}\xi)\,X. \tag{13.76}$$

An alternative way to look at Equation

$$E(\eta\,|\,X,\xi) \underset{P}{=} h_0(X) + h_1(X)\,\xi \underset{P}{=} (\gamma_{00} + \gamma_{10}\,X)\xi + (\gamma_{01} + \gamma_{11}X)\,\xi. \tag{13.77}$$

This kind of equation has already been considered in the case of a manifest covariate and a manifest outcome variable. The only difference is that the covariate is now denoted ξ instead of Z and the outcome variable is now the latent variable η. Of course, Equation (**??**) can be wrong, because ξ is not dichotomous but assumed to be a latent variable with more than two values. In such a case, the intercept function $g_0(\xi)$ and the effect function $g_1(\xi)$ might be nonlinear functions of ξ (see, e. g., Hoyer & Klein, 2000; Kenny & Judd, 1984; Klein & Moosbrugger, 2000; Klein & Muthén, 2007; Marsh et al., 2004; Song & Lee, 2006; Wall & Amemiya, 2003 for nonlinear latent variable models.)

13.3.2 Conclusions

Generalized ANCOVA has now been treated for manifest and for latent covariates and outcomes. Which one should be used under which circumstances? Suppose X represents a treatment variable. If assignment to one of the treatment conditions is via self-selection, it is more plausible to assume that one or more *latent* covariates determine the treatment

[2] Allowing for (a) different slopes for different values x of X distinguishes this model from the traditional analysis of covariance with latent covariates (see, e. g., Klauer et al., 2002; Hancock, 2004; Sörbom, 1978), which does not allow for interaction between treatments and covariates.

probability, unless the individual takes a test, say of neuroticism, and — based on the test score obtained — makes up his mind to seek treatment. For self-selection, it is more plausible to assume that the treatment probability is determined by (a) the *subjective need* for a treatment and (b) the *perceived accessibility* of the treatment. Both variables are latent variables that may be assessed via certain items that can be considered to be manifest indicators of the two latent variables. Not choosing a latent covariate in this case may lead to biased estimates of treatment effects.

In contrast, in a conditionally randomized experiment, the treatment probabilities are determined by a manifest covariate. The experimenter gives the subjects a test and determines the treatment probability (the propensity) as a function of the manifest test scores.[3] In this case generalized analysis of covariance with a *manifest* covariate would be more appropriate (see section 13.2).

Choosing a manifest or a latent *outcome variable* only depends on substantive considerations. Measurement error in the observed outcome variable does not lead to bias if these error are not correlated with treatments and the covariate.

13.3.3 Statistical Programs

For statistical analyses of generalized ANCOVA with latent variables, we again recommend the standard software for structural equation modeling such as *LISREL* (Jöreskog & Sörbom, 1996/2001) or *Mplus* (Muthén & Muthén, 2002). Again, such analyses are greatly facilitated by *EffectLite* (Steyer & Partchev, 2007).

13.3.4 Conclusions

Stratification on propensity scores is another procedure for estimating average total effects which may be used instead of generalized ANCOVA based on the original covariate Z. The results are the same: all average total treatment effects are zero in this example. In generalized ANCOVA, however, the emphasis is not only on estimating the average total effects, but also on the conditional total effects *given the values of the covariates*. In fact, one may also use generalized ANCOVA based on propensities. Applying generalized ANCOVA with propensities as conditioning covariates additionally yields the conditional total effects given *the propensity scores*, i. e., the average effects within the propensity strata.

13.3.5 Statistical Programs

In analyzing empirical data, one may first estimate the propensities and add them to the data set. This is possible, because each case in the data set has a vector of covariate values, which uniquely determine the propensity scores for this case. For steps (b) to (d), i. e., for regression estimation, mean adjustment and computing average effects, one may use the programs mentioned in the section on generalized ANCOVA.

---

[3] For simplicity, we do not discuss problems of compliance and the distinction between *treatment probabilities* and *treatment assignment probabilities*.

13.4 Summary and Conclusions

In this chapter, we have treated several data analysis procedures that adjust for bias. The first procedure is traditional *analysis of covariance* (ANCOVA) based on the original (possibly multivariate) covariate Z. Assuming that the regression $E(Y \mid X, Z)$ is τ_x-unbiased and that there is no interaction between X and the covariates, the coefficients of Equation (13.4) are not only the effects of X given any value z of the covariate, but they are also the average total effects.

In generalized ANCOVA we still need the assumption that $E(Y \mid X, Z)$ is τ_x-unbiased. However, we drop the assumption that there is no interaction between X and the covariates, aiming at estimating the regression $E(Y \mid X, Z)$ without restrictive assumptions. This may not always be possible, depending on design and sample size. In these cases, the validity of the total effect estimates will depend on the validity of the assumptions about the functional form of the regression $E(Y \mid X, Z)$.

Generalized ANCOVA can also be applied to cases in which the covariates and/or the outcome variables are latent variables, which are measured each by at least two manifest variables. Such analyzes may be indicated if there is self-selection to treatment conditions.

13.5 Proofs

Proof of Theorem 13.1

Assuming that the true outcomes τ_x exist implies that the extensions $E^{X=x}(Y \mid Z)$ exist as well and that Equations (13.9) and (13.10) also hold for the extensions $E^{X=0}(Y \mid Z)$ and $E^{X=x}(Y \mid Z)$, respectively. Hence,

$$\begin{aligned} E(\tau_0) &= E\big(E^{X=0}(Y \mid Z)\big) && [\tau_0\text{-unbiasedness and Eq. (10.2)}] \\ &= E(\gamma_{00} + \gamma_{01} Z_1 + \ldots + \gamma_{0Q} Z_Q) && [\text{Eq. (13.9)}] \\ &= \gamma_{00} + \gamma_{01} E(Z_1) + \ldots + \gamma_{0Q} E(Z_Q) && [\text{Rule (??) of Box ??}]. \end{aligned}$$

Similarly,

$$\begin{aligned} E(\tau_x) &= E\big(E^{X=x}(Y \mid Z)\big) \\ &= E[(\gamma_{00} + \gamma_{x0}) + \gamma_{01} Z_1 + \ldots + \gamma_{0Q} Z_Q] && [\text{Eq. (13.10)}] \\ &= (\gamma_{00} + \gamma_{x0}) + \gamma_{01} E(Z_1) + \ldots + \gamma_{0Q} E(Z_Q) && [\text{Rule (??) of Box ??}]. \end{aligned}$$

Proof of Theorem 13.9

Assuming that the true outcome variables τ_x exist, implies that the extensions $E^{X=x}(Y \mid Z)$ exist as well. Furthermore, Equation (13.20) implies

$$\begin{aligned} E^{X=0}(Y \mid Z) &= g_0(Z) \quad P^{X=0}\text{-a.s.} \\ E^{X=x}(Y \mid Z) &= g_0(Z) + g_x(Z) \quad P^{X=x}\text{-a.s.} \qquad x = 1, \ldots, J, \end{aligned} \tag{13.78}$$

as well as

$$\begin{aligned} E^{X=0}(Y|Z) &= g_0(Z) \quad a.s. \\ E^{X=x}(Y|Z) &= g_0(Z) + g_x(Z) \quad a.s. \qquad x = 1,\ldots,J. \end{aligned} \tag{13.79}$$

According to Equation (10.2), the expectations of the true outcome variables can then be computed as follows:

$$\begin{aligned} E(\tau_0) &= E\big(E^{X=0}(Y|Z)\big) \qquad [\tau_0\text{-unbiasedness and Eq. (10.2)}] \\ &= E[g_0(Z)] \qquad [\text{Eq. (13.79}] \end{aligned}$$

and

$$\begin{aligned} E(\tau_x) &= E\big(E^{X=x}(Y|Z)\big) \\ &= E[g_0(Z) + g_x(Z)] \qquad x = 1,\ldots,J \qquad [\text{Eq. (13.79}]. \end{aligned}$$

Equation (13.23) immediately follows from Equations (13.21) and (13.22). Finally, Equation (13.24) follows from Equation (10.8), the definition of the prima facie effect function $PFE_{x0;Z} := E^{X=x}(Y|Z) - E^{X=0}(Y|Z)$, and (13.79).

13.6 Exercises

▷ **Exercise 13-1** What are the assumptions under which traditional analysis of covariance yields τ_x-unbiased estimates of the treatment effects given a value z of the covariate Z and of the average treatment effects?

▷ **Exercise 13-2** What are the substantive restrictions in Equation (13.4)?

▷ **Exercise 13-3** Draw a figure representing the regressions $E^{X=0}(Y|Z)$ and $E^{X=1}(Y|Z)$ of the outcomes on the unidimensional and continuous covariate Z within the two treatment conditions [see Eqs. (13.2) and (13.3)].

▷ **Exercise 13-4** Simplify Equation (13.11) for the case of a single univariate covariate Z and a treatment variable X having only two values, 0 and 1.

▷ **Exercise 13-5** Draw a figure representing the regressions $E^{Z=z_1}(Y|X)$ and $E^{Z=z_2}(Y|X)$ of the outcomes on the treatment variable X for two values z_1 and z_2 of the covariate Z [see Eq. (13.80)]. Assume that the treatment variable X has only two values, 0 and 1.

▷ **Exercise 13-6** Use the *causal effects explorer* (*CEE*) provided at *www.causal-effects.de* to generate two data sets according to Table 13.1. For the first one, use sample size 1200 and for the second one, sample size 12000.

▷ **Exercise 13-7 (without solution)** Estimate the nine expectations $E(Y|X{=}x, Z{=}z)$ from the two data sets generated in Exercise 13-6 and then use linear regression to estimate the parameters of Equation (**??**). Compare the results for the two data sets to each other and to theoretical parameters presented in Table 13.1 and Equation (**??**).

▷ **Exercise 13-8** Apply generalized analysis of covariance to the example presented in Table 10.1 and compute the average causal effect of treatment 1 compared to the control.

▷ **Exercise 13-9** Use the *causal effects explorer* (*CEE*) provided at *www.causal-effects.de* to generate a data set according to Table 10.1. Use sample size 1000.

▷ **Exercise 13-10 (without solution)** Estimate the expectations $E(Y|X{=}x, Z{=}z)$ from the data set generated in Exercise 13-9 and then use Generalized ANCOVA to estimate the parameters of Equation (13.81). Compare the results to the theoretical parameters presented in Equation (13.81). See section 13.2.4 for statistical programs.

▷ **Exercise 13-11** Use the *causal effects explorer* (*CEE*) provided at *www.causal-effects.de* to generate a data set according to Table 1.4. Use sample size 10000.

▷ **Exercise 13-12 (without solution)** Estimate the nine expectations $E(Y|X{=}x, Z{=}z)$ from the data set generated in Exercise 13-11 and compare them to the theoretical expectations displayed in Table 1.4. Then use Generalized ANCOVA to estimate the parameters of Equation (13.41). Compare the results to the theoretical parameters presented in Equation (13.41). Also estimate the average total effects and test them to be zero. Compare the results of this test to the results of using (nonorthogonal) ANOVA for all types of sums of squares provided by your statistical program. See section 13.2.4 for statistical programs.

▷ **Exercise 13-13 (without solution)** Use the data generated in Exercise 13-11 to test the hypothesis that the expectations of the weighted covariate $Z_v := Z\,V$ are the same within each treatment. [V is defined by Equation (**??**)]

!!! Achtung: Die Tabellennamen stimmen wohl noch nicht

Solutions

▷ **Solution 13-1** The first assumption is that the regression $E(Y|X,Z)$ is τ_x-unbiased, the second is that the regression $E(Y|X,Z)$ can be parameterized by Equation (13.4).

▷ **Solution 13-2** This equation does not allow for interaction between covariates and treatments. It does allow for interaction between covariates.

▷ **Solution 13-3** The figure should show two parallel lines, one for $E^{X=0}(Y|Z)$ and the other one for the regression $E^{X=1}(Y|Z)$. The horizontal axis represents the values of Z and the vertical axis the values of the regressions $E^{X=0}(Y|Z)$ and $E^{X=1}(Y|Z)$, i. e., the expectations $E(Y|X{=}0, Z{=}z)$ and $E(Y|X{=}1, Z{=}z)$. The vertical distance between the two lines is the treatment effect γ_{01} [see Eq. (13.3)]. Because the two lines are parallel, this distance does not depend on the values of the covariate, and this is what we mean by 'no interaction between covariate and treatments'.

▷ **Solution 13-4** In this case $\mathbb{1}_{X=1} = X$ and Equation (13.11) simplifies to

$$E^{Z=z}(Y|X) \;=\; \gamma_{00} + \gamma_{01}\,z \;+\; \gamma_{10}\,X. \tag{13.80}$$

▷ **Solution 13-5** The figure should show four points, one for each of the four expectations $E^{Z=z_1}(Y|X{=}x)$. On the horizontal axis mark the two values of X, 0 and 1. The vertical axis represents the values of the regressions $E^{Z=z}(Y|X)$. The distance between the two points for $E^{Z=z_1}(Y|X{=}0)$ and $E^{Z=z_2}(Y|X{=}0)$ is $\gamma_{01}(z_1 - z_2)$ and the same is true for the distance between the two points representing $E^{Z=z_1}(Y|X{=}1)$ and $E^{Z=z_2}(Y|X{=}1)$.

▷ **Solution 13-6** In a first step, use the *CEE* to construct a table with three persons, one for each value of the covariate *neediness*. Next, think about how to choose the sampling probabilities $P(U{=}u)$ that are in accordance with the marginal distribution of *neediness*

in Table 13.1. Then choose the nine expectations $E^{X=x}(Y|U=u)$ according to the expectations $E(Y|X=x,Z=z)$ in Table 13.1. Fill in the treatment probabilities $P(X=1|U=u)$ for treatment 1 and the corresponding probabilities $P(X=2|U=u)$ for treatment 2 in such a way that they are in accordance with the probabilities $P(X=x,U=u)$ presented in Table 13.1. (A table created this way is provided at *www.causal-effects.de* under the file name *BookTable7.1.tab*. This table can also be opened by *CEE*.) Now use the option 'generate data' in order to generate the two data sets with sample sizes 1200 and 12000, respectively. (You may use the default option for the residual variance within the nine cells. Within each cell, the data have a normal distribution.) Two such data sets are provided at *www.causal-effects.de* under the file name *BookTable7.1_1200.csv* and *BookTable7.1_12000.csv*.

▷ **Solution 13-8** Using Equations (13.34) and (13.35), the covariate-treatment regression is

$$\begin{aligned} E(Y|X,Z) &= g_0(Z)+g_1(Z)\,1_{X=1} \\ &= \gamma_{00}+\gamma_{01}\,I_{Z=(m,y)}+\gamma_{02}\,I_{Z=(f,y)}+\big(\gamma_{10}+\gamma_{11}\,I_{Z=(m,y)}+\gamma_{12}\,I_{Z=(f,y)}\big)\,1_{X=1} \\ &= 72+25.5\,I_{Z=(m,y)}+39\,I_{Z=(f,y)}+\big(11-6\,I_{Z=(m,y)}+0\,I_{Z=(f,y)}\big)\,1_{X=1}. \end{aligned}$$

In these equations, $1_{X=1}$ indicates with 1 and 0 whether or not the unit is treated, $I_{Z=(m,y)}$ with 1 and 0 whether or not the unit drawn is male and has a college education, and $I_{Z=(f,y)}$ with 1 and 0 whether or not the unit is female and has a college degree.

Applying Equation (13.31) to the last equation, the average total effect of treatment 1 compared to the control can be computed by:

$$\begin{aligned} E[g_1(Z)] &= E\big(11-6\,I_{Z=(m,y)}+0\,I_{Z=(f,y)}\big) \\ &= 11-6\,E\big(I_{Z=(m,y)}\big)+0\,E\big(I_{Z=(f,y)}\big) \\ &= 11-6\,1/3+0\,1/3 \;=\; 9. \end{aligned} \tag{13.81}$$

This is exactly the average total effect already obtained before. Note again that the expectation $E[g_1(Z)]$ of the *PFE*-function $g_1(Z)$ yields the average total effect only because, in this example, the covariate-treatment regression is τ_x-unbiased for the two-dimensional covariate $Z=(Z_1,Z_2)$. It is neither τ_x-unbiased for the covariate Z_1 nor for the covariate Z_2. Unbiasedness of the covariate-treatment regression is the crucial prerequisite in applying the adjustment theorems.

▷ **Solution 13-9** The table for the *CEE* is provided at *www.causal-effects.de* under the file name *BookTable6.1.tab*. Open this table in the *CEE*.) Now use the option 'generate data' in order to generate the data set with sample size 1000. (You may use the default option for the residual variance within the nine cells. Within each cell, the data have a normal distribution.) Such a data set is also provided at *www.causal-effects.de* under the file name *BookTable6.1_1000.csv*.

▷ **Solution 13-11** The table for *CEE* is provided at *www.causal-effects.de* under the file name *BookTable1.4.tab*. Open this table in the *CEE*. Now use the option 'generate data' in order to generate the data set with sample size 10000. (You may use the default option for the residual variance within the nine cells. Within each cell, the data have a normal distribution.) Such a data set is also provided at *www.total-effects.de* under the file name *BookTable1.4.10000.csv*.

Chapter 14
Conclusions of Volume I

In this chapter, we will summarize and discuss the most important issues treated in this book. The specific goals are:

(a) to summarize the theory,

(b) to discuss the most important consequences of the theory of causal effects for the design and analysis of experiments and quasi-experiments, and

(c) to reflect on other issues in causal inference from empirical data, which are not treated in this book.

14.1 A Summary of the Theory

We motivated the theory of causal effects by a version of Simpson's paradox showing that comparing conditional probabilities can lead to contradictory and seriously misleading causal inferences. We argued that this applies also to comparing conditional expectations and illustrated these arguments by another example.

Prima Facie Effects

The differences between the conditional expectations of the outcome variable Y given treatments x and x', respectively, are called *prima facie effects*. They are not necessarily identical with the causal treatment effects. In fact, we presented examples, in which there are *positive* treatment effects *for each and every observational unit* (person) and still a *negative* prima facie effect $E(Y|X=1)-E(Y|X=0)$ (see e. g., Table 4.2, p. 82 or Table 6.1, p. 128).

14.1.1 Basic Ideas

Single-Unit Trials

We continued preparing the stage for the theory of causal effects by describing the kind of empirical phenomena the theory is dealing with. These empirical phenomena are called *single-unit trials*. One of the simplest examples of such a single-unit trial consists of drawing a single observational unit (e. g., a person) from the set of units, exposing or observing exposure to one of at least two treatment conditions and, finally, observing the outcome variable Y, i. e., a variable of interest, the distribution of which might be affected by the treatment variable X.

Causality Spaces

This simple example already shows that there is a time order between observational unit drawn, treatment, and outcome, which is one of the properties distinguishing *causal* dependencies from *ordinary* stochastic dependencies. This time order is most adequately represented by a *filtration* $(\mathscr{F}_t, t \in T)$, which also plays an important role in specifying the *global covariate* σ*-algebra* c_X, a comprehensive covariate that defines the set of the particular covariates such as $Z_1 :=$ *gender*, $Z_2 :=$ *educational status*, etc. These two additional structural components distinguish a *numerical causality space* from a *probability space*. Within the latter, we can deal with ordinary stochastic dependencies. Note, however, that the additional structural components $(\mathscr{F}_t, t \in T)$ and c_X of a numerical causality space do not yet distinguish between a causal and an ordinary stochastic dependency. This distinction depends on the validity of one of several *causality conditions*.

Causality Conditions

The logically weakest of these causality conditions is *unbiasedness*. Unfortunately, this weakest causality condition is not empirically testable. Another, much stronger one, is *independence of the putative cause* X *and all covariates*. This condition can be tested by testing independence of X and *one* of the particular covariates, say W. Even more important is that, if X is a treatment variable, the validity of this causality condition *can be created* by the design technique of random assignment of units to treatment conditions. Still another causality condition is *homogeneity*, which means that there is no covariate W determining the expectations of the outcome variable over and above the putative cause X. This causality condition is fairly realistic only in latent variable models in which X represents the latent variable and the outcome variables are its fallible manifest measures. Another causality condition is *unconfoundedness* which can be defined by postulating that, for each value x of X, there is no covariate W for which both of the following two conditions hold:

(a) $E(Y \mid W) \neq E(Y)$, and

(b) $P(X{=}x \mid W) \neq P(X{=}x)$.

The causality conditions mentioned above refer to the regression $E(Y \mid X)$. However, they have also been formulated for the regression $E(Y \mid X, Z)$, where Z denotes a covariate (see chapters 9 and **??**). These conditional versions of the causality conditions are crucial for the adjustment techniques in quasi-experimental research.

14.1.2 Causal Effects

The theory of causal effects is based on the principle of *atomic stratification*. Since prima facie effects in the total population and in subpopulations can be biased, the basic idea is simply to base the theory on the conditional expectations of Y given $X{=}x$ in the smallest subpopulations. Oftentimes, these smallest subpopulations are the observational units. However, because, in many cases, there is still systematic variability within the observational units, we base the general theory on the conditional expectations of Y given $X{=}x$ in the most fine-grained strata, the so-called *atomic strata*. These atomic strata are obtained by conditioning on all covariates, i. e., on all those variables that are prior or simultaneous to the putative cause X. Technically, this is achieved by conditioning on the global covariate σ-algebra c_X. If there is no systematic variability within observational units, then the

units themselves are already the atomic strata. Those random variables whose values are these conditional expectations of Y given $X{=}x$ in the atomic strata are called the *true-outcome variables*. They are denoted τ_x and their differences $\delta_{xx'} = \tau_x - \tau_{x'}$ for two values x and x' of the putative cause X are called the *true causal effect variables*. Their values are the *true causal effects* in the atomic strata. The expectation $E(\delta_{xx'}) = E(\tau_x) - E(\tau_{x'})$ of these true effect variables over the distribution of the strata defines the *average causal effect* of $X{=}x$ compared to $X{=}x'$. Similarly, various kinds of *conditional causal effects* are defined by the difference between various conditional expectations of the true-outcome variables. Each of these kinds of conditional causal effects provides specific information that might be of interest in empirical causal research and evaluation studies depending on the research questions.

14.1.3 Relationship Between Prima Facie Effects and Causal Effects

Prima-Facie Effects

Next we studied how the expectations $E(\tau_x)$ of the true outcomes and the average causal effects are related to the conditional expectations $E(Y|X{=}x)$ of the outcome variable Y given $X{=}x$ and to the conditional expectations $E(Y|X{=}x, Z{=}z)$ of Y given $X{=}x$ and a value z of a (possibly multivariate) covariate Z. It is these conditional expectations that we traditionally try to estimate in between-group studies. As mentioned before, in general, the differences between these conditional expectations, which are called *prima facie effects*, are neither identical with the average causal treatment effects nor with any other causal effects.

Biases of the Prima Facie Effects

We showed that each prima facie effect is a sum of the average causal effect and two kinds of biases: the *baseline bias* and the *effect bias*. These biases are most easily described if X represents a treatment variable. In this case, the *baseline bias* is induced by systematic selection of the unit to one of the treatment conditions based on the true outcome under control (or whatever is chosen as a baseline or reference condition). In contrast, *the effect bias* is induced by systematic selection based on the true treatment effect (see section 6.3.3). Even though it is not possible to use the true outcomes and effects in treatment assignment, the baseline bias will become effective if exposure to treatment is based on one or several covariates correlating with the true outcomes under control, and similarly, there will be an effect bias if exposure to treatment is based on one or several covariates correlating with the true causal effects.

14.1.4 Identification of Causal Effects

In chapter 10, we showed how to compute average causal effects by specific functions of the conditional expectations $E(Y|X{=}x)$ and/or $E(Y|X{=}x, Z{=}z)$ and explicated the assumptions — which we call *causality conditions* — under which this is possible. If identification of the average causal effect is intended without using covariates, the crucial prerequisite for its identification is *unbiasedness of the regression* $E(Y|X)$, i. e., $E(Y|X{=}x) = E(\tau_x)$, for each value x of X. If, in contrast, identification of the average causal effect may also be based on covariates, the crucial prerequisite is *unbiasedness of the regression* $E(Y|X,Z)$.

Each of these two prerequisites is true under each of several conditions, the *causality conditions*, which are summarized in Box 9.1 (see p. 231). These conditions have been treated in more detail in chapters 9 and **??**.

14.1.5 Experiments and Quasi-Experiments

True Experiments

Some of the conditions implying unbiasedness of the regression $E(Y|X)$ and/or unbiasedness of the regression $E(Y|X,Z)$ can be created in *true experiments*, i. e., in studies in which (a) there are at least two treatment conditions, (b) there is random assignment, or, at least conditional random assignment of the unit to one of the treatment conditions, and (c) assignment also means actual exposure to treatment.

Causal inference can most safely be drawn in true experiments. There, unbiasedness of the treatment regression $E(Y|X)$ and/or unbiasedness of the covariate-treatment regression $E(Y|X,Z)$ can be created by design, namely by some form of random assignment of the unit to one of the treatment conditions. Note, however, that such a design can be jeopardized by systematic attrition of observational units (see, e. g., Shadish et al., 2002; ?, ?). In this case, the study is not a randomized experiment any more, and (conditional) independence of units and treatments (see Box 9.1) is violated. Nevertheless, whenever feasible, the true experiment should be preferred over the quasi-experiment.

Quasi-Experiments

The assumptions under which identification of causal effects is possible may also hold in *quasi-experiments* if appropriate covariates are chosen. These covariates might be manifest or latent. Covariates can also be the Z-conditional propensities $P(X{=}x\,|\,Z)$ assuming conditional independence of units and treatments given Z or at least, assuming unbiasedness of the covariate-treatment regression $E(Y|X,Z)$ (see section 11.4).

Whenever a true experiment is *not* feasible we may conduct a quasi-experiment using a (possibly multivariate) covariate Z. Utilizing one of the adjustment procedures described in chapter 10 *may* still provide correct estimates of the average causal effects. However, the crucial prerequisite is *unbiasedness of the covariate-treatment regression* $E(Y|X,Z)$. For the correct choice of the covariate Z, for which $E(Y|X,Z)$ is unbiased, substantive knowledge is usually indispensable.

14.1.6 Covariate Selection and Causality Tests

Causality Tests

Unfortunately, the weakest causality condition, unbiasedness of the regression $E(Y|X,Z)$, is not empirically testable. What is testable, however, are some of the other causality conditions, all of which are sufficient for unbiasedness. Testing one of these sufficient conditions may be called a *causality test*. Of course, in such a causality test, one should choose the weakest empirically testable condition that still implies unbiasedness. To our knowledge, this is unconfoundedness. Such causality tests can also be used in an empirically guided selection of the covariates. The idea is to start with a covariate Z, for which one assumes that the regression $E(Y|X,Z)$ is unconfounded. This assumption can be tested

choosing a specific covariate W not yet included in the covariate vector $Z := (Z_1, \ldots, Z_Q)$. If, e. g., unconfoundedness of the regression $E(Y \mid X, Z)$ is rejected testing it with the additional covariate W, then $E(Y \mid X, Z^*)$ should be tested for unconfoundedness in the next step of covariate selection, where $Z^* := (Z_1, \ldots, Z_Q, W)$ using another covariate W^* not yet included in Z^* (see section **??** for more details).

14.1.7 Adjustment Techniques

Mean Adjustment vs. Outcome Weighting

We treated the theoretical foundations of two different strategies for estimating average causal effects: (a) *mean adjustment* procedures and (b) procedures based on *weighting the outcome variable*. While mean adjustment procedures in some way or other use Equation (10.2), outcome weighting procedures utilize Equation (11.58) showing how to weight the outcome variable by a function of the conditional treatment probabilities $P(X{=}x \mid Z{=}z)$, the propensity scores. Under independence of X and the true outcomes, the conditional expectations $E(Y_w \mid X{=}x)$ of the weighted outcome variable Y_w are equal to the expectations $E(\tau_x)$ of the true outcome variables [see Eq. (11.60)].

Technically speaking, *mean adjustment* procedures aim at estimating the unconditional expectations of the extensions $E^{X=x}(Y \mid Z)$ of the $(X{=}x)$-conditional regressions $E^{X=x}(Y \mid Z)$ of the outcome variable Y on the covariate Z, i. e., the expectations $E\big[E^{X=x}(Y \mid Z)\big]$ w.r.t. the unconditional or marginal [and not the $(X{=}x)$-conditional] distribution of Z. If the regression $E(Y \mid X, Z)$ is unbiased, then the difference between these expectations is equal to the corresponding average causal effect of $X{=}x$ compared to $X{=}x'$. Depending on the nature of the regressions $E_{X=x}(Y \mid Z)$ and the nature of Z, estimating the expectations $E\big[E^{X=x}(Y \mid Z)\big]$ may be a difficult task.

Traditional analysis of covariance is a well-known and often used mean adjustment procedure that yields valid estimates and tests of the average causal effect if (a) the correct covariates are chosen and (b) the assumptions about the regression $E(Y \mid X, Z)$ are correct (see section 13.1). In chapter 13 we also showed how to generalize traditional analysis of covariance, allowing for interaction between (manifest and/or latent) covariates and treatments (see also Steyer & Partchev, 2007).

In contrast, *outcome weighting*, which is well-known in sampling theory (see, e. g., ?, ?, ?, ?), utilizes Equations (11.58) and (11.60). In this approach, there is no need to estimate the $(X{=}x)$-conditional regressions $E_{X=x}(Y \mid Z)$. The price, however, is to estimate the Z-conditional propensities $P(X{=}x \mid Z)$, which may be difficult as well.

Two Threats to Causal Inference

In quasi-experiments, there are two kinds of threats to valid causal inferences. *First*, the covariate Z may not be chosen such that one of the causality conditions holds (see Box **??**). Observing the covariates determining the probability that a unit is exposed to treatment x and/or the covariates determining the outcome variable are of crucial importance in quasi-experiments. Without the correct covariates fulfilling one of the causality conditions, causal inference will usually fail. Choosing the appropriate covariates is not just statistical routine; it requires substantive knowledge about (a) the process of assigning units to treatment conditions and/or (b) the covariates determining the expectations of the outcome variable Y.

Second, even if the covariate Z is correctly chosen such that the covariate-treatment regression $E(Y|X,Z)$ is unbiased, the regression $E(Y|X,Z)$ may not be modeled correctly.[1] If, e. g., one considers only a *linear* function of X and Z, but the (true) regression $E(Y|X,Z)$ is a *nonlinear* function of X and Z, causal inferences may not be valid. If propensities are used, an additional problem is the correct model for the propensities $P(X{=}x\,|\,Z)$, unless the propensity scores have been fixed by the experimenter.

14.2 Implications for Design

The causality conditions, i. e., the sufficient conditions of unbiasedness treated in detail in chapters 9 and **??**, show which strategies we may follow in the analysis of causal effects in between-group experiments and quasi-experiments. We will discuss randomization procedures, techniques of covariate selection, and strategies of data analysis.

14.2.1 Randomization

Simple Randomization

Random assignment of the unit to one of the treatment conditions creates independence of units and treatments [see condition (1) in Box 9.1], provided, of course, that there is perfect compliance (see, e. g., ?, ?; Greenland, 2000) and no systematic attrition (see, e. g., Abraham & Russell, 2004; Fichman & Cummings, 2003; Graham & Donaldson, 1993; Shadish et al., 2002). According to Theorem **??**, independence of X and covariates implies unbiasedness of the treatment regression $E(Y|X)$. The conditional expectations $E(Y|X{=}x)$ are then equal to the expectations $E(\tau_x)$ of the true outcomes and their differences are the average causal effects $E(\delta_{xx'})$ (see, however, section 14.4.3 on the interpretation problem).

Conditional Randomization

Conditional random assignment of the unit to one of the treatments given the value z of a covariate Z creates Z-conditional independence of X and all covariates [see condition (5) in Box 9.1], which implies unbiasedness of the covariate-treatment regression $E(Y|X,Z)$. Of course, in the conditional case, too, we have to assume that there is perfect compliance and no systematic attrition. Under conditional independence of X and covariates, the expectations of the extensions $E^{X=x}(Y|Z)$ then yield the expectations $E(\tau_x)$ of the true outcomes and the differences $E(\tau_x) - E(\tau_{x'})$ are the average causal effects $E(\delta_{xx'})$.

14.2.2 Covariate Selection

A second class of design techniques is to observe all those covariates in the vector $Z := (Z_1,\ldots,Z_Q)$ such that Z-conditional independence of X and the covariates holds (see Box 9.1). An alternative is to choose a vector Z of covariates such that homogeneity of $E(Y|X,Z)$ holds. Note that the covariate vector Z w.r.t. which there is Z-conditional inde-

[1] Note again that — unlike the linear least-squares regression — the regression $E(Y|X,Z)$ is well-defined without any reference to a specific function of X and Z (see sections **??** and **??**).

pendence of X and the covariates may differ from the covariate vector w.r.t. which there is homogeneity of $E(Y|X,Z)$, and the same is true for all other sufficient conditions mentioned in Box 9.1.

Some, but not all sufficient conditions for unbiasedness of the regression $E(Y|X,Z)$ are useful in covariate selection. Conditional independence of X and true outcomes (i. e., Rubin's strong ignorability) as well as conditional regresssive independence of true outcomes from X, for instance, are useless for covariate selection, because, for a given covariate Z, we cannot check whether or not these conditions hold. They do not imply anything that could prove wrong in an empirical application.

Hence, in covariate selection one uses either *Z-conditional independence of X and the covariates* or *Z-conditional homogeneity of the regression* $E(Y|X,Z)$. Propensity score analysis exclusively draws on the first condition, whereas the second one is often used in modeling the outcome variable. Both conditions *are* empirically testable. Z-conditional independence of X and the covariates is empirically testable, because it implies: $P(X{=}x \mid Z,W) = P(X{=}x \mid Z)$ for each x and each additional covariate W. Similarly, Z-conditional homogeneity of the regression $E(Y|X,Z)$ is empirically testable, because it implies: $E(Y \mid X,Z,W) = E(Y|X,Z)$ for each additional covariate W.

As an example, suppose that Z is a covariate (vector) that does not contain *gender* as one of its components. In this case we may choose W = *gender* and test whether or not $P(X{=}x \mid Z,W) = P(X{=}x \mid Z)$. If this equation does not hold, we can conclude that conditional independence of units and treatments does not hold. Similarly, we may test $E(Y \mid X,Z,W) = E(Y|X,Z)$. If this equation does not hold, we can conclude that Z-conditional homogeneity of the regression $E(Y|X,Z)$ does not hold as well.

Instead of (a) *Z-conditional independence of X and covariates* or (b) *Z-conditional homogeneity of the regression* $E(Y|X,Z)$, we could also use the disjunction [i. e., '(a) or (b)'] of these two conditions, because, if (a) and (b) both imply unbiasedness, then '(a) or (b)' implies unbiasedness as well. Another possibility is to use *Z-conditional unconfoundedness of the regression* $E(Y|X,Z)$ [see condition **??** in Box **??**]. Since this condition is logically weaker than '(a) or (b)', the most economic procedure seems to select those covariates in the vector $Z := (Z_1,\ldots,Z_Q)$ for which unconfoundedness holds. For an empirical test of this condition we may again use an additional covariate W, because Z-conditional unconfoundedness of the regression $E(Y|X,Z)$ implies:

$$P_{Z=z}(X{=}x \mid W) = P_{Z=z}(X{=}x) \qquad \text{or} \qquad E_{X=x,Z=z}(Y \mid W) = E_{X=x,Z=z}(Y)$$

for (almost) all pairs (x,z) of values of X and Z.

14.3 Implications for Data Analysis

As already mentioned before, there are different kinds of data analysis strategies. They may be referred to as *mean adjustment with manifest or latent covariates*, *mean adjustment based on propensities*, and *outcome weighting based on propensities*, respectively. We briefly describe the various strategies and discuss their merits and problems.

14.3.1 Mean Adjustment With Manifest or Latent Covariates

The first data analysis strategy is *mean adjustment with manifest or latent covariates* using those manifest and/or latent variables as covariates, for which one of the sufficient

conditions of Z-conditional unbiasedness of the regression $E(Y|X,Z)$ holds (see Box 9.1). Mean adjustment with *manifest covariates* has its tradition in the analysis of means, analysis of variance, analysis of covariance, and regression analysis. These analysis techniques are traditionally applied without referring to an explicit theory of causal effects. Nevertheless, in favorable situations, such as the perfect randomized experiment, they also estimate and test average causal effects. In other cases, such as nonorthogonal analysis of variance, analysis of covariance with interactions between covariates and treatments, or analysis of covariance with latent variables (see, e. g., Sörbom, 1978; Klauer et al., 2002) and interactions between covariates and treatments (see, e. g., Steyer & Partchev, 2007), mean adjustment procedures need more sophistication than what is provided in traditional statistical program packages.

Mean adjustment with manifest or latent covariates aims at estimating the regressions $E_{X=x}(Y|Z)$ or at least at estimating the expectations of the extensions $E^{X=x}(Y|Z)$ (over the unconditional or marginal distribution of Z) and their differences, where Z is a covariate for which the regression $E(Y|X,Z)$ is unbiased. Under unbiasedness, the expectations $E\left[E^{X=x}(Y|Z)\right]$ are equal to the expectations $E(\tau_x)$ of the true outcomes and the differences between these expectations are equal to the average causal effects.

Merits

A first advantage of using the manifest and/or latent covariates in the vector Z is that, under Z-conditional unbiasedness of the regression $E(Y|X,Z)$, the differences $E^{X=x}(Y|Z) - E^{X=x'}(Y|Z)$ are the conditional causal-effect functions $E(\delta_{xx'}|Z)$, the values of which are the conditional causal effects of $X{=}x$ compared to $X{=}x'$ given $Z{=}z$. These conditional effects are easy to interpret, because the covariates in the vector Z have a clear substantive meaning. If, for instance, $Z :=$ (*gender*, *educational status*) and $z :=$ (*male*, *college*), a value of the function $E(\delta_{xx'}|Z)$ is the conditional causal effect of $X{=}x$ compared to $X{=}x'$ for the male subpopulation with college education. As already mentioned in chapter 5, these conditional causal effects given a value of a covariate are more informative than the corresponding average causal effect in the total population.

A second advantage of using the manifest and/or latent covariates in the vector Z in mean adjustment procedures is that there is an empirically testable criterion that can be used in the selection of the covariates: *Z-conditional unconfoundedness* of the regression $E(Y|X,Z)$. As compared to the other selection criteria, Z-conditional unconfoundedness leads to a covariate vector Z (for which the regression $E(Y|X,Z)$ is unbiased) that has a dimension smaller or equal than those covariate vectors for which the other criteria are met. Using this selection criterion we neither have to select all those covariates determining the true propensities, nor all those covariates determining the conditional expectations of the outcomes. We can be sure that the set of covariates for which Z-conditional unconfoundedness of the regression $E(Y|X,Z)$ holds is always a subset of the set of covariates for which Z-conditional independence of X and covariates or Z-conditional homogeneity of the regression $E(Y|X,Z)$ holds. Hence, selecting the covariates via *unconfoundedness* will often yield a smaller number of covariates as compared to selecting covariates using either Z-conditional independence of X and covariates or Z-conditional homogeneity of the regression $E(Y|X,Z)$ as the selection criterion.

Note that weak causality of the regression $E(Y|X,Z)$ as a selection criterion aims at the causal interpretation of Z-conditional and average prima facie effects. If the goal is to discover all possible interactions and the most fine-grained conditional effects, then one should use Z-conditional homogeneity of the regression $E(Y|X,Z)$ as a criterion for se-

lecting the covariates. If this criterion is fulfilled for a set of covariates, we will have both, the most fine-grained conditional effects and Z-conditional unbiasedness of the regression $E(Y|X,Z)$.

Problems

Using the manifest and/or latent covariates in the vector Z in mean adjustment procedures has also disadvantages. The first problem is that it may be difficult to estimate the regression $E(Y|X,Z)$ correctly if the vector Z consists of many covariates. In this case, a correct parameterization of $E(Y|X,Z)$ might need products such as $Z_1 \cdot Z_2$, $Z_1 \cdot Z_2 \cdot Z_3$, etc. between the covariates. With finite sample sizes such complicated regressions cannot be estimated sufficiently well. In such cases one may also aim at the more modest goal of estimating the expectations $E\left[E^{X=x}(Y|Z)\right]$ and their differences. Although this is certainly easier, it still may be difficult or even impossible with small samples.[2]

A second problem of using the manifest and/or latent covariates in mean adjustment is an ethical issue. Because including or omitting a covariate can change the conditional and average prima facie effects, researchers may be tempted not to select covariates according to the criteria of unconfoundedness or homogeneity of the regression $E(Y|X,Z)$. Instead they may decide to stop including covariates in the regression $E(Y|X,Z)$ as soon as the desired effects of the treatment show up in their analysis. This problem can be avoided in propensity based analyses.

14.3.2 Mean Adjustment Based on Propensities

A second strategy of data analysis consists of selecting a vector $Z := (Z_1,\dots,Z_Q)$ of covariates such that conditional independence of X and true outcomes holds and then using the propensities $\pi_x := P(X{=}x|Z)$, i. e., the Z-conditional probabilities of treatment x, as new covariates. Using this strategy, data analysis consists of two parts:

(a) to estimate the propensities $\pi_x := P(X{=}x|Z)$ in the vector $\pi := (\pi_1,\dots\pi_J)$, and

(b) to estimate the $(X{=}x)$-conditional regressions $E_{X=x}(Y|\pi)$ of the outcome variable Y on π for each value x of X, or at least to estimate the expectations of the extensions $E^{X=x}(Y|\pi)$ or the differences between these expectations for different values of X.

Under Z-conditional independence of X and the covariates (or, at least, of X and true outcomes), the expectations $E\left[E^{X=x}(Y|\pi)\right]$ are equal to the expectations $E(\tau_x)$ of the true outcomes and their differences are equal to the average causal effects.

Merits

As discussed in more detail in chapter 10, for a small number $J+1$ of values of the treatment variable X and a large number of covariates $Z_1,\dots,Z_Q$, using the vector $\pi := (\pi_1,\dots,\pi_J)$ of propensities instead of the original vector $Z := (Z_1,\dots,Z_Q)$ of covariates may simplify estimating the average causal effects considerably. If, for instance, there are $J+1=2$ treatment conditions and five covariates $Z_1,\dots,Z_5$, unbiasedness of the regression $E(Y|X,Z_1,\dots,Z_5)$ implies that the regression $E(Y|X,\pi)$ is also unbiased, where

[2] Note that the discussions of this section refer to the population level. Statistical decision problems are beyond the scope of this book.

$\pi := P(X{=}1 \mid Z_1, \ldots, Z_5)$. Note that, in this case, π is a single univariate random variable, whereas Z consists of five univariate random variables. Therefore, estimating the regressions $E_{X=x}(Y \mid \pi)$, or at least, the expectations $E\left[E^{X=x}(Y \mid \pi)\right]$ and their differences, is a much easier task than estimating the regression $E_{X=x}(Y \mid Z_1, \ldots, Z_5)$, or the expectations $E\left[E^{X=x}(Y \mid Z_1, \ldots, Z_5)\right]$ of their extensions and their differences.

Matching on propensity scores (see, e. g., ?, ?; ?, ?), for instance, aims at estimating the differences between the expectations $E\left[E^{X=x}(Y \mid \pi)\right]$ for different values x of X. If the regression $E(Y \mid X, \pi)$ is unbiased, then these differences are the average causal effects. Stratification on propensity scores (see, e. g., ?, ?) yields estimates (a) of the conditional causal effects given propensity scores *and* (b) of the average causal effects.

A second merit of mean adjustment with propensities is that modeling of the propensity, including the choice of the covariates, can be separated from modeling the outcome variable. If there is, for example, a quasi-experiment on the effects of a drug 1 (as compared to drug 0) on hypertension Y, then the propensity $\pi := P(X{=}1 \mid Z)$ could be estimated without (knowledge or even availability of) the outcome variable Y. In the second step one could analyze the regressions $E_{X=0}(Y \mid \widehat{\pi})$ and $E_{X=1}(Y \mid \widehat{\pi})$ and/or just estimate the expectations $E\left[E^{X=1}(Y \mid \pi)\right]$ and $E\left[E^{X=0}(Y \mid \pi)\right]$, or only their difference, using the estimated propensity $\widehat{\pi}$ instead π itself.

Problems

In *quasi-experiments* the propensities have to be estimated. In this case, there are two sources of errors. The *first* one is in selecting the covariates. The critical question is: 'Did we select those covariates Z for which $E(Y \mid X, \pi)$ is unbiased?' The problem is that there is no way to test unbiasedness of this regression. Therefore one aims at selecting all covariates such that *Z-conditional independence of X and covariates* holds. However, this may lead to an unnecessarily large set of covariates.

Even if the set of covariates is selected such that there is Z-conditional independence of X and covariates, and, therefore, unbiasedness of $E(Y \mid X, \pi)$, there is a *second* source of error: the specification of the functions $\pi_x := P(X{=}x \mid Z)$. While a correct specification is simple if Z only has a few values, it is difficult if Z is (more or less) continuous and/or multivariate. If, for example, we assume that the propensity is a *linear* logistic regression, the critical question is, whether or not this assumption is correct, i. e.: 'Did we correctly specify the functions $P(X{=}x \mid Z)$?' (see section 11.2 on how to answer this question).

If Z is multivariate, we will have similar problems as in the analysis of the outcomes. In this case, a correct parameterization of the propensities $P(X{=}x \mid Z)$ might need products such as $Z_1 \cdot Z_2$, $Z_1 \cdot Z_2 \cdot Z_3$, etc. between the covariates. Again, with finite sample sizes, such complicated regressions are difficult and often impossible to estimate sufficiently well.

14.3.3 Outcome Weighting Based on Propensities

A third data analysis strategy consists of selecting a vector $Z := (Z_1, \ldots, Z_Q)$ of covariates such that the regression $E(Y \mid X, Z)$ is unbiased and then using the propensities $\pi_x := P(X{=}x \mid Z)$ for outcome weighting. Hence, this strategy of data analysis consists of two parts: (a) to estimate the propensities $\pi_x := P(X{=}x \mid Z)$ in the vector $\pi := (\pi_1, \ldots \pi_J)$, and (b) to estimate the conditional expectations $E(Y_w \mid X{=}x)$ of the weighted outcome variable Y_w [see Eq. (11.58)] for each value x of X. Under unbiasedness of the regression $E(Y \mid X, Z)$,

the difference between $E(Y_w \mid X{=}x)$ and $E(Y_w \mid X{=}x')$ is the average causal effect $E(\delta_{xx'})$ (see section 11.4).

Merits

Outcome weighting with propensities shares some of its merits with mean adjustment with propensities, namely the reduction of dimensionality and separating propensity analysis and outcome analysis. A third advantage is that neither the regressions $E_{X=x}(Y \mid \pi)$ or $E_{X=x}(Y \mid Z)$ nor the expectations of their extensions are needed. Instead, we just have to estimate the conditional expectations $E(Y_w \mid X{=}x)$ of the weighted outcome variable.

Problems

The problems of selecting the correct covariate (vector) Z and estimating the correct propensities $P(X{=}x \mid Z)$ already discussed before remain the same.

14.4 Limitations

In this book we focused on the relations between the regression $E(Y \mid X)$, the regression $E(Y \mid X, Z)$, and the regression $E(Y_w \mid X)$ of the weighted outcome variable Y_w on the putative cause X to the various kinds of conditional and average causal effects. However, as already mentioned in the preface, empirical causal research involves several inferences and substantive interpretations. Among these are:

(a) statistical inference, i. e., the inference from sample data to parameters characterizing the distributions of random variables,

(b) causal inference, i. e., the inference from parameters characterizing the distributions of random variables to causal effects,

(c) interpretation of the putative cause,

(d) interpretation of the outcome variable,

(e) interpretation of the random experiment (i. e., the empirical phenomenon) considered.

Most of these issues, which are dealt with in the theory of internal and external validity (see, e. g., Campbell & Stanley, 1963; Cook & Campbell, 1979; Shadish et al., 2002), have not been treated in this book. Instead we only focused on the second point. Briefly considering all these points may help to put the contribution of this book into perspective.

14.4.1 Statistical Inference

Statistical issues and models dealing with sampling processes, estimation and hypothesis testing, e. g., were not treated. We did not present the sampling processes and models allowing to estimate and test the different kinds of causal effects. Under which assumptions is it possible, e. g., to use regression models with fixed regressors, stochastic regressors, or models with errors in the variables for estimating and testing causal effects (see, e. g., ?, ?)? Can we also use random coefficient models (see, e. g., ?, ?, or models with latent variables (see, e. g., ?, ?)? Which of the hypotheses relevant for the analysis of causal effects are linear,

which need methods for testing nonlinear hypotheses? Can we use also nonparametric regression techniques (see, e. g., ?, ?)?

There are also a number of specific statistical issues, e. g., dealing with *non-compliance* and *subject reassignment* (see, e. g., ?, ?, ?, ?; Greenland, 2000; Jo, 2002c; J. M. Robins, 1998; ?, ?), *time-dependent covariates* (see, e. g., ?, ?), and *sensitivity analysis* (see, e. g., ?, ?, ?; Jo, 2002b; ?, ?).

14.4.2 Causal Inference

Causal inference deals with the inference from parameters characterizing the distributions of random variables (such as the expectations of an outcome variable in two treatment conditions) to causal effects. This is what the theory of causal effects in experiments and quasi-experiments presented in this book has been about. In section 14.1 we summarized the most important issues.

14.4.3 Interpretation of the Treatment Variable

The interpretation of the treatment variable is another point at which substantive conclusions can go wrong. Suppose X is a treatment variable. The substantive differences between the different values x of the treatment variable X are often not clear, because there usually are differences *in several aspects*. Suppose, for instance, we study the effects of 'taking drug A' as compared to 'not taking drug A' in a pharmacological experiment. The nominal value 'taking drug A' of the treatment variable consists of several components, such as 'going to a physician', 'taking a pill', and 'ingesting drug A', while the nominal value of X 'not taking drug A' may consist of 'not going to a physician', 'taking no pill', and 'not ingesting the drug'.

Which of these components is the causal agent? Is it the nominal value of X, i. e., is it 'taking drug A' or is it rather 'going to the physician', or 'taking a pill'? Of course, these questions can be answered only if studies are designed such that the interpretation of the values of the treatment variable is unambiguous, e. g., if 'taking drug A' is compared to 'taking a pill' (a placebo) that looks and tastes exactly the same without containing the drug to be investigated and that also involves 'going to a physician' and 'taking a pill', but not 'ingesting drug A'.

It should be emphasized that random assignment does not solve this interpretation problem. As argued above, a treatment usually differs from the control in several components and it is not clear which of these components is the causal agent. Similarly, random assignment neither solves the interpretation problem of the outcome variable, nor the interpretation of the random experiment considered (external validity) (see the following sections).

14.4.4 Interpretation of the Outcome Variable

Another point at which substantive conclusions about causal effects can go wrong is the interpretation of the outcome variable. Here, the problem is that oftentimes there is a difference between what is *intended to measure* and what is *actually measured*. A simple example is *annual income*. What is actually measured is either what subjects report as their annual income or what the tax authorities have in their records. However, the true annual income may differ from both actual measures, self report and tax authorities records. Sim-

ilar problems occur if we measure *intelligence* by a specific *psychological test, academic achievement* by a *citation index, depression* or *life satisfaction* by *specific questionnaires*. In all these cases, we may erroneously believe that what is *actually measured* is what we *intended to measure* (for more details, see Shadish et al., 2002, ch. 3).

14.4.5 Interpretation of the Random Experiment Considered

The individual and average causal effects defined in this book refer to single-unit trials. A simple example is sampling an observational unit (e. g., a person) from a population of observational units, exposing it (or observing its exposure) to one of the treatment conditions, and finally, observing the outcome variable Y for the sampled unit (see chapter 2 for more details). In applications, such a single-unit trial always refers to a specific set of observational units, such as the set of all patients with a given diagnosis, the set of all undergraduate psychology students at your university in a given year, or the set of all consumers of a specific brand at a specific point of time. Furthermore, the single-unit trial also consists of a specific distribution of the observational-unit variable, i. e., the probabilities $P(U=u)$ that the unit u (patient, student, consumer) is sampled. However, what is *not* explicitly represented in the mathematics treated in this book are the side conditions, i. e., the historical context, space, and time under which the single-unit trial is conducted. All these side conditions are implicitly referred to when we say that we are talking about a *specific* single-unit trial, a specific random experiment that is represented by the probability space $(\Omega, \mathcal{A}, P)$.

Consequently, all our propositions about causal effects refer only to the specific random experiment, i. e., the specific single-unit trial considered. What can go wrong in the interpretation of the random experiment considered, for instance, is that we may not be aware of some crucial side conditions without which the effect would not occur. For example, the treatment studied may exert its effects only if the subjects *know that they are a part of a scientific study*. In this case a generalization beyond the specific study would go wrong. Another point is generalizing to a population differing from the population of the units from which the actual sample was drawn. In Psychology, for instance, many experiments are conducted with psychology students. The crucial issue is if and to what extent we may generalize to other populations of subjects. Problems related to the interpretation of the random experiment considered may be called problems of *external validity* (see, e. g., Shadish et al., 2002, ch. 5).

References

Abraham, W. T., & Russell, D. W. (2004). Missing data: A review of current methods and applications in epidemiological research. *Current Opinion in Psychiatry*, *17*, 315–321.

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). Hoboken, NJ: Wiley.

Aiken, L. S., & West, S. G. (1996). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage.

Aitkin, M. (1978). The analysis of unbalanced cross-classifications. *Journal of the Royal Statistical Society, Series A: Statistics in Society*, *141*, 195–223.

Allen, M. P. (1997). *Understanding regression analysis*. New York, NY: Plenum Press.

Appelbaum, M. I., & Cramer, E. M. (1974). Some problems in the nonorthogonal analysis of variance. *Psychological Bulletin*, *81*, 335–343.

Arbuckle, J. L. (2006). AMOS 7.0 user's guide [Computer software manual]. Chicago, IL: SPSS.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Bauer, H. (1996). *Probability theory*. Berlin, Germany and New York, NY: de Gruyter.

Bentler, P. M. (1995). EQS Structural Equations program manual [Computer software manual]. Encino, CA: Multivariate Software.

Bentler, P. M., & Wu, E. J. C. (2002). *EQS 6 for Windows guide*. Encino, CA: Multivariate Software.

Berger, M. P. F., & Wong, W. K. (Eds.). (2005). *Applied optimal designs*. Chichester, England: Wiley.

Biesanz, J. C., Deeb-Sossa, N., Aubrecht, A. M., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods*, *9*, 30–52.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, *53*, 605–634.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation approach*. Hoboken, NJ: Wiley.

Bonney, G. E. (1987). Logistic regression for dependent binary observations. *Biometrics*, *43*, 951–973.

Borooah, V. K. (2001). *Logit and probit: Ordered and multinomial models*. Thousand Oaks, CA: Sage.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203–219. doi: 10.1037/0033-295X.110.2.203

Browne, M. W., & Mels, G. (1998). Path analysis: RAMONA. In *SYSTAT for Windows: Advanced Applications (Version 8) [Computer software manual]*. Evanston, IL: SYSTAT.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago, IL: Rand McNally.

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin.

Carlson, J. E., & Timm, N. H. (1974). Analysis of nonorthogonal fixed-effects designs. *Psychological Bulletin*, *81*, 563–570.

Cartwright, N. (1979). Causal laws and effective strategies. *Noûs*, *13*, 419–437.

Cheng, J., & Small, D. S. (2006). Bounds on causal effects in three-arm trials with non-compliance. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *68*, 815–836.

Cochran, W. G. (1957). Analysis of covariance: Its nature and uses. *Biometrics*, *13*, 261–281.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression / correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago, IL: Rand McNally.

Cox, D. R., & Wermuth, N. (2004). Causality: A statistical view. *International Statistical Review*, *72*, 285–305.

Dawid, A. P. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *41*(1), 1–31.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis*. New York, NY: Wiley.

Dunn, G., Maracy, M., Dowrick, C., Ayuso-Mateos, J. L., Dalgard, O. S., Page, H., ... Wilkinson, G. (2003). Estimating psychological treatment effects from a randomised controlled trial with both non-compliance and loss to follow-up. *British Journal of Psychiatry*, *183*, 323–331.

Fahrmeir, L., & Tutz, G. (2001). *Multivariate statistical modelling based on Generalized Linear Models* (2nd ed.). New York, NY: Springer.

Feller, W. (1968). *An introduction to probability theory and its applications* (2nd ed., Vol. 1). New York, NY: Wiley.

Feller, W. (1971). *An introduction to probability theory and its applications* (2nd ed., Vol. 2). New York, NY: Wiley.

Fichman, M., & Cummings, J. N. (2003). Multiple imputation for missing data: Making the most of what you know. *Organizational Research Methods*, *6*, 282–308.

Fisher, R. A. (1925/1946). *Statistical methods for research workers* (10th ed.). Edinburgh, England: Oliver and Boyd.

Gelman, A., & Hill, J. L. (2007). *Data analysis using regression and multilevel / hierarchical models.* New York, NY: Cambridge University.

Georgii, H.-O. (2008). *Stochastics – Introduction to probability and statistics.* Berlin, Germany: de Gruyter.

Gosslee, D. G., & Lucas, H. L. (1965). Analysis of variance of disproportionate data when interaction is present. *Biometrics, 21*, 115–133.

Graham, J. W., & Donaldson, S. I. (1993). Evaluating interventions with differential attrition: The importance of nonresponse mechanisms and use of follow-up data. *Journal of Applied Psychology, 78*, 119–128.

Green, W. H. (2003). *Econometric analysis* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.

Greenland, S. (2000). Causal analysis in the health sciences. *Journal of the American Statistical Association, 95*, 286–289.

Greenland, S. (2004). An overview of methods for causal inference from observational studies. In A. Gelman & X.-L. Meng (Eds.), *Applied bayesian modeling and causal inference from incomplete-data perspectives* (pp. 3–14). Chichester, England: Wiley.

Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology, 10*, 37–48.

Greenland, S., & Robins, J. M. (1986). Identifiability, exchangeability, and epidemiological confounding. *International Journal of Epidemiology, 15*, 413–419.

Györfi, L., Kohler, M., Krzyzak, A., & Walk, H. (2002). *A distribution-free theory of nonparametric regression.* New York, NY: Springer.

Hancock, G. R. (2004). Experimental, quasi-experimental, and nonexperimental design and analysis with latent variables. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences.* Thousand Oaks, CA: Sage.

Hernán, M. A., Clayton, D., & Keiding, N. (2011). The Simpson's paradox unraveled. *International Journal of Epidemiology*, 1–6. (Advance Access published March 30, 2011) doi: 10.1093/ije/dyr041

Höfler, M. (2005). Causal inference based on counterfactuals. *BMC Medical Research Methodology, 5*, 1–12.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*, 945–960.

Holland, P. W. (1988). Causal inference, path analysis, and recursive structural equation models. *Sociological Methodology, 18*, 449–484.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York, NY: Wiley.

Hoyer, J., & Klein, A. (2000). Self-reflection and well-being: Is there a healthy amount of introspection? *Psychological Reports, 86*, 135–141.

Huet, S., Bouvier, A., Poursat, M.-A., & Jolivet, E. (2004). *Statistical tools for nonlinear regression: A practical guide with S-PLUS and R examples* (2nd ed.).

New York, NY: Springer.

Jamieson, J. (2004). Analysis of covariance ANCOVA with difference scores. *International Journal of Psychophysiology*, *52*, 277–283.

Jennings, E., & Green, J. L. (1984). Resolving nonorthogonal ANOVA disputes using cell means. *Journal of Experimental Education*, *52*, 159–162.

Jo, B. (2002a). Estimation of intervention effects with noncompliance: Alternative model specifications. *Journal of Educational and Behavioral Statistics*, *27*, 385–409.

Jo, B. (2002b). Model misspecification sensitivity analysis in estimating causal effects of interventions with non-compliance. *Statistics in Medicine*, *21*, 3161–3181.

Jo, B. (2002c). Statistical power in randomized intervention studies with noncompliance. *Psychological Methods*, *7*, 178–193.

Jo, B., Asparouhov, T., Muthén, B. O., Ialongo, N. S., & Brown, C. H. (2008). Cluster randomized trials with treatment noncomplicance. *Psychological Methods*, *13*, 1–18.

Jöreskog, K. G., & Sörbom, D. (1996/2001). LISREL 8: User's Reference Guide. [Computer software manual].

Kaplan, D. (2000). *Structural Equation Modeling: Foundations and Extensions*. Thousand Oaks, CA: Sage.

Keele, L. (2008). *Semiparametric regression for the social sciences*. Chichester, England: Wiley.

Kenny, D. A. (1975). Cross-lagged panel correlation: A test for spuriousness. *Psychology Bulletin*, *82*, 887–903.

Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, *96*, 201–210.

Kenny, D. A., & Judd, C. M. (1996). A general procedure for the estimation of interdependence. *Psychological Bulletin*, *119*(1), 138–148.

Keren, G., & Lewis, C. (1976). Nonorthogonal designs: Sample versus population. *Psychological Bulletin*, *83*, 817–826.

King, G., & Zeng, L. (2001). Improving forecasts of state failure. *World Politics*, *53*, 623–658.

Klauer, K. J., Willmes, K., & Phye, G. D. (2002). Inducing inductive reasoning: Does it transfer to fluid intelligence? *Contemporary Educational Psychology*, *27*, 1–25.

Klein, A. G., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*, 457–474.

Klein, A. G., & Muthén, B. O. (2007). Quasi-maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research*, *42*, 647–673.

Klenke, A. (2013). *Probability theory – A comprehensive course* (2nd ed.). London, England: Springer. doi: 10.1007/978-1-4471-5361-0

Kolmogorov, A. N. (1956). *Foundations of the theory of probability* (2nd ed.; N. Morrison, Trans.). New York, NY: Chelsea.

Kramer, C. Y. (1955). On the analysis of variance of a two-way classification with unequal sub-class numbers. *Biometrics*, *11*, 441–452.

Langsrud, Ø. (2003). ANOVA for unbalanced data: Use Type II instead of Type III sums of squares. *Statistics and Computing*, *13*, 163–167.

Lechner, M. (2000, November). A note on the "common support problem" in applied evaluation studies. Retrieved from `http://ssrn.com/abstract=259239`

Lee, B. K., Lessler, J., & Stuart, E. A. (2010). Improving propensity score weighting using machine learning. *Statistics in Medicine*, *29*, 337–346. doi: 10.1002/sim.3782

Lee, S.-Y. (2007). *Structural equation modeling: A bayesian approach*. Chichester, England: Wiley.

Liao, T. F. (1994). *Interpreting probability models: Logit, probit, and other generalized linear models*. London, England: Sage.

Little, T. D., Card, N. A., Bovaird, J. A., Preacher, K. J., & Crandall, C. S. (2007). Structural equation modeling of mediation and moderation with contextual factors. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 207–230). Mahwah, NJ: Lawrence Erlbaum.

Loève, M. (1977). *Probability theory I* (4th ed.). New York, NY: Springer.

Loève, M. (1978). *Probability theory II* (4th ed.). New York, NY: Springer.

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin*, *68*, 304–305.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*, 201–226.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Lawrence Erlbaum.

Marsh, H. W., Wen, Z., & Hau, K. T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods*, *9*, 275–300.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Mayer, A., Dietzfelbinger, L., Rosseel, Y., & Steyer, R. (in press). The effectliter approach for analyzing average and conditional effects. *Multivariate Behavioral Research*.

Mayer, A., Thoemmes, F., Rose, N., Steyer, R., & West, S. G. (2014). Theory and analysis of total, direct, and indirect causal effects. *Multivariate Behavioral Research*, *49*(5), 425–442.

McArdle, J. J. (2001). A latent difference score approach to longitudinal dynamic structural analysis. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 341–380). Lincolnwood, IL: Scientific Software International.

McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational

studies. *Psychological Methods*, *9*(4), 403–425.

McCullagh, P., & Nelder, J. A. (1989). *Monographs on statistics and applied probability: Vol. 37. Generalized linear models* (2nd ed.; D. R. Cox, D. V. Hinkley, N. Reid, D. B. Rubin, & D. V. Silverman, Eds.). Chapman & Hall.

Meredith, M., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107–122.

Mill, J. S. (1843/1865). Of the four methods of experimental inquiry. In *A system of logic, ratiocinative and inductive: Volume 1. Being a connected view of the principles of evidence, and the methods of scientific investigation.* London, England: Longmans, Green, and Co.

Morgan, S. L., & Harding, D. J. (2006). Matching estimators of causal effects: Prospects and pitfalls in theory and practice. *Sociological Methods & Research*, *35*, 3–60. Retrieved from `http://smr.sagepub.com` doi: 10.1177/0049124106289164

Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference. methods and principles for social research.* New York, NY: Cambridge University.

Muthén, L. K., & Muthén, B. O. (1998-2007). Mplus User's Guide (5th ed.) [Computer software manual]. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *9*, 599–620.

Nagengast, B. (2009). *Causal inference in multilevel models* (Unpublished doctoral dissertation). Friedrich-Schiller-Universität Jena, Thüringen, Germany.

Nagengast, B., Kröhne, U., Bauer, M., & Steyer, R. (2007). Causal Effects Explorer: A didactic tool for teaching the theory of individual and average causal effects [Computer software manual]. University of Jena, Thüringen, Germany.

Nelder, J. A., & Lane, P. W. (1995). The computer analysis of factorial experiments: In memoriam – Frank Yates. *The American Statistician*, *49*, 382–385.

OpenMx. (2009). *OpenMx - Advanced Structural Equation Modeling [Computer Software]*. Retrieved from http://openmx.psyc.virginia.edu/.

Overall, J. E., & Spiegel, D. K. (1969). Concerning least squares analysis of experimental data. *Psychological Bulletin*, *72*(5), 311–322.

Overall, J. E., & Spiegel, D. K. (1973a). Comment on "Regression analysis of proportional cell data". *Psychological Bulletin*, *80*, 28–30.

Overall, J. E., & Spiegel, D. K. (1973b). Comments on Rawlings' nonorthogonal analysis of variance. *Psychological Bulletin*, *79*, 164–167.

Overall, J. E., Spiegel, D. K., & Cohen, J. (1975). Equivalence of orthogonal and nonorthogonal analysis of variance. *Psychological Bulletin*, *82*, 182–186.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, *82*(4), 669–688.

Pearl, J. (1998). Graphs, causality, and structural equation models. *Sociological Methods & Research*, *27*, 226–284.

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge, UK.

Pearson, K., Lee, A., & Bramley-Moore, L. (1899). Genetic (reproductive) selection: Inheritance of fertility in man, and of fecundity in thoroughbred racehorses. *Series A, Containing Papers of a Mathematical or Physical Character: Philosophical Transactions of the Royal Society of London*, *192*, 257–330.

Pukelsheim, F. (2006). *Optimal design of experiments*. Philadelphia, PA: SIAM (Society for Industrial and Applied Mathematics).

Rao, C. R. (1973). *Linear statistical inference and its applications* (2nd ed.). New York, NY: Wiley.

Reichardt, C. S. (1979). The statistical analysis of data from nonequivalent group designs. In T. D. Cook & D. C. Campbell (Eds.), *Quasi-experimentation: Design and analysis issues for field settings*. Orlando, FL: Houghton Mifflin.

Reichardt, C. S. (2005). Nonequivalent group design. In B. Everitt & D. Howell (Eds.), *Encyclopedia of behavioral statistics*. New York, NY: Wiley.

Robins, J., & Rotnitzky, A. (2004). Estimation of treatment effects in randomised trials with non-compliance and a dichotomous outcome using structural mean models. *Biometrika*, *91*, 763–783.

Robins, J. M. (1998). Correction for non-compliance in equivalence trials. *Statistics in Medicine*, *17*, 269–302.

Rogosa, D. (1980a). Comparing nonparallel regression lines. *Psychological Bulletin*, *88*, 307–321.

Rogosa, D. (1980b). A critique of cross-lagged correlation. *Psychological Bulletin*, *88*, 245–258.

Rosenbaum, P. R. (2002). Attributing effects to treatment in matched observational studies. *Journal of the American Statistical Association*, *97*(457), 183–192. doi: 10.1198/016214502753479329

Rosenbaum, P. R., & Rubin, D. B. (1983a). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *45*, 212–218.

Rosenbaum, P. R., & Rubin, D. B. (1983b). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–55. doi: 10.1093/biomet/70.1.41

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, *79*, 516–524.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern epidemiology* (3rd ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*, 688–701.

Rubin, D. B. (1984). Assessing the fit of logistic regressions using the implied discriminant-analysis: Comment. *Journal of the American Statistical Association*, *79*, 79–80.

Rubin, D. B. (2004). Direct and indirect causal effects via potential outcomes. *Scandinavian Journal of Statistics*, *31*, 161–170.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, *100*, 322–331. doi: 10.1198/016214504000001880

Rubin, D. B. (2006). *Matched sampling for causal effects*. New York, NY: Cambridge University.

Senn, S. (2006). Change from baseline and analysis of covariance revisited. *Statistics in Medicine*, *25*, 4334–4344.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *13*, 238–241.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University.

Sobel, M. E. (2008). Identification of causal parameters in randomized studies with mediating variables. *Journal of Educational and Behavioral Statistics*, *33*, 230–251. doi: 10.3102/1076998607307239

Song, X. Y., & Lee, S. Y. (2006). Bayesian analysis of structural equation models with nonlinear covariates and latent variables. *Multivariate Behavioral Research*, *41*, 337–365.

Sörbom, D. (1978). An alternative to the methodology for analysis of covariance. *Psychometrika*, *43*, 381–396.

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). Cambridge, MA: MIT.

Splawa-Neyman, J. (1923/1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9 (Reprinted from roczniki nauk rolniczych tom 10, pp. 1–51, 1923). *Statistical Science*, *5*, 465–480.

Spohn, W. (1980). Stochastic independence, causal independence, and shieldability. *Journal of Philosophical Logic*, *9*, 73–99.

Stegmüller, W. (1983). *Erklärung, Begründung, Kausalität: Probleme und Resultate der Wissenschaftstheorie und analytischen Philosophie. [Explanation, justification, causality: Problems and findings of philosophy of science and analytic philosophy]*. Berlin, Germany: Springer.

Steyer, R. (1984). Causal linear stochastic dependencies: An introduction. In J. R. Nesselroade & A. von Eye (Eds.), *Individual development and social change: Explanatory analysis*. New York, NY: Academic Press.

Steyer, R. (1992). *Theorie kausaler Regressionsmodelle [Theory of causal regression models]*. Stuttgart, Germany: Fischer.

Steyer, R. (2001). Classical test theory. In C. Ragin & T. D. Cook (Eds.), *International encyclopedia of the social and behavioural sciences: Logic of inquiry and research design* (pp. 481–520). Oxford, England: Pergamon.

Steyer, R. (2005). Analyzing individual and average causal effects via structural equation models. *Methodology, 1*, 39–54.

Steyer, R., & Eid, M. (2001). *Messen und Testen: Ein Lehrbuch [Measurement and testing]*. Berlin, Germany: Springer.

Steyer, R., Eid, M., & Schwenkmezger, P. (1997). Modeling true intraindividual change: True change as a latent variable. *Methods of Psychological Research Online, 2*, 21–33.

Steyer, R., Gabler, S., von Davier, A. A., & Nachtigall, C. (2000). Causal regression models II: Unconfoundedness and causal unbiasedness. *Methods of Psychological Research Online, 5*, 55–87.

Steyer, R., Mayer, A., Geiser, C., & Cole, D. (2015). A theory of states and traits - revised. *Annual Review of Clinical Psychology, 11*, 71–98.

Steyer, R., & Nagel, W. (in press). *Probability and conditional expectation: Fundamentals for the empirical sciences*. Chichester, England: Wiley.

Steyer, R., & Partchev, I. (2007). EffectLite: User's manual – A program for the uni- and multivariate analysis of unconditional, conditional and average mean differences between groups [Computer software manual]. University of Jena, Thüringen, Germany.

Suppes, P. (1970). *A probabilistic theory of causality*. Amsterdam, Netherlands: North-Holland.

Tisak, J., & Tisak, M. S. (2000). Permanency and ephemerality of psychological measures with application to organizational commitment. *Psychological Methods, 5*, 175–198.

van Breukelen, G. J. P. (2006). ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies. *Journal of Clinical Epidemiology, 59*, 920–925.

von Eye, A., & Schuster, C. (1998). *Regression analysis for social sciences*. San Diego, CA: Academic Press.

Wainer, H. (1991). Adjusting for differential base rates: Lord's paradox again. *Psychological Bulletin, 109*, 147–151.

Wall, M. M., & Amemiya, Y. (2003). A method of moments technique for fitting interaction effects in structural equation models. *British Journal of Mathematical and Statistical Psychology, 56*, 47–63.

Watkins, M. W., Lei, P., & Canivez, G. L. (2007). Psychometric intelligence and achievement: A cross-lagged panel analysis. *Intelligence, 35*, 59–68.

West, S. G., & Aiken, L. S. (2005). Multiple linear regression. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science*. Chichester, England: Wiley.

West, S. G., Biesanz, J. C., & Pitts, S. C. (2000). Causal inference and generalization 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* (pp. 40–84). New York, NY: Cambridge University.

Williams, J. D. (1972). Two way fixed effects analysis of variance with disproportionate cell frequencies. *Multivariate Behavioral Research, 7*, 57–83.

Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, *25*, 659–707.

Wolf, F. M., Chandler, T. A., & Spies, C. J. (1981). A crossed-lagged panel analysis of quality of school life and achievement responsibility. *Journal of Educational Research*, *74*, 363–368.

Woo, M.-J., Reiter, J. P., & Karr, A. F. (2008). Estimation of propensity scores using generalized additive models. *Statistics in Medicine*, *27*, 3805–3816.

Woodward, J. A., & Bonett, G. B. (1991). Simple main effects in factorial designs. *Journal of Applied Statistics*, *18*(2), 255–264. doi: 10.1080/02664769100000019

Wright, S. (1918). On the nature of size factors. *Genetics*, *3*, 367–374.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, *20*(7), 557–585.

Wright, S. (1923). The theory of path coefficients: A reply to Niles's criticism. *Genetics*, *8*, 239–255.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, *5*, 161–215.

Wright, S. (1960a). Path coefficients and path regressions: Alternative or complementary concepts? *Biometrics*, *16*(2), 189–202.

Wright, S. (1960b). The treatment of reciprocal interaction, with or without lag, in path analysis. *Biometrics*, *16*(3), 423–445.

Yule, G. U. (1903). Notes on the theory of association of attributes of statistics. *Biometrika*, *2*, 121–134.