Rolf Steyer
Ulf Kröhne

Probability and Causality

Version: May 26, 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

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

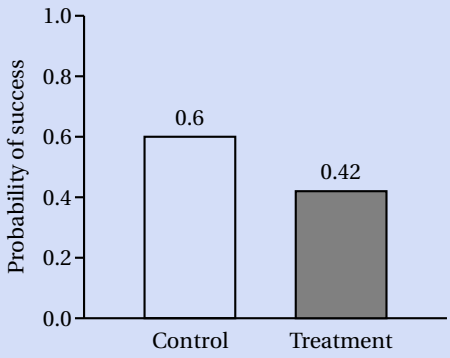

Figure 1.1. Probability of success given treatment

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-*

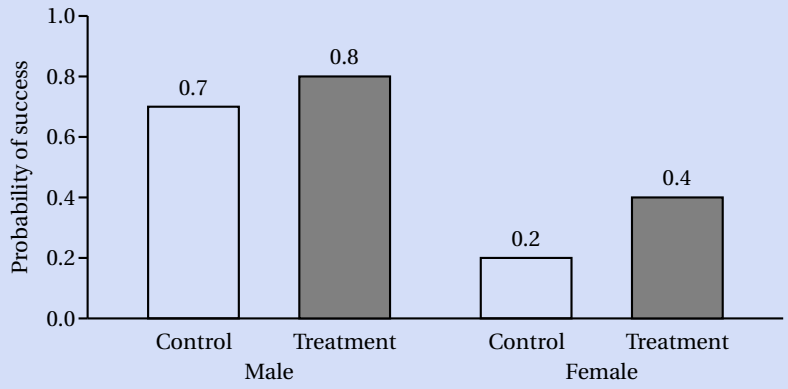

Figure 1.2. Probabilities of success given treatment (and sex)

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:

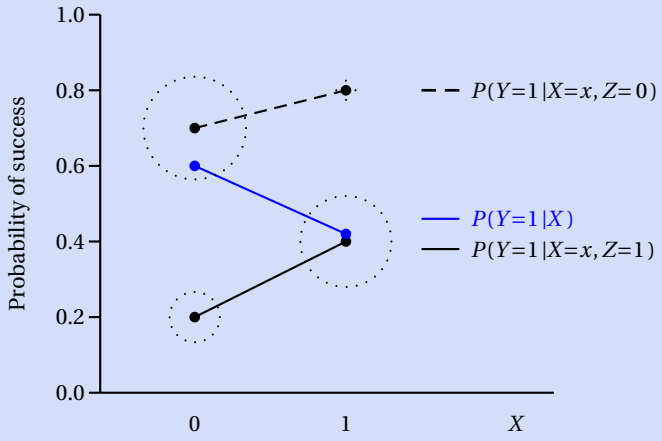

Figure 1.3. Probabilities of success given treatment and sex

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

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

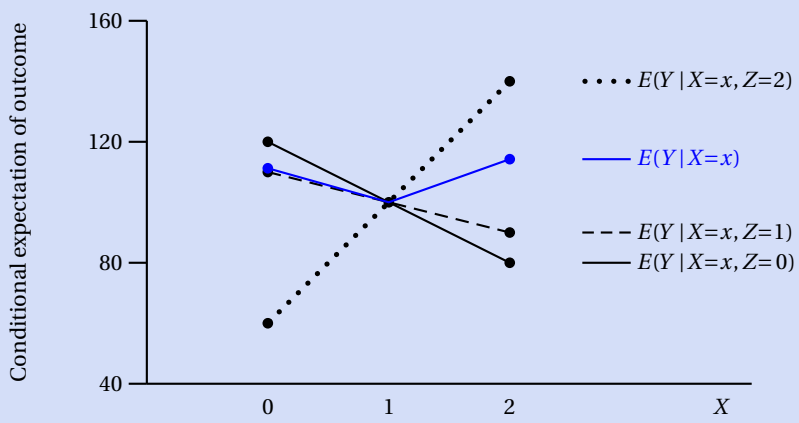

Figure 1.4. Conditional expectation values of Y given treatment and neediness

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

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.

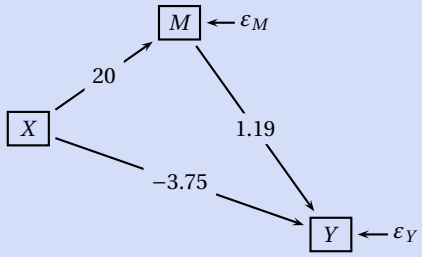

Figure 1.5. Path diagram of $E(M \mid X)$ and $E(Y \mid X, M)$

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?

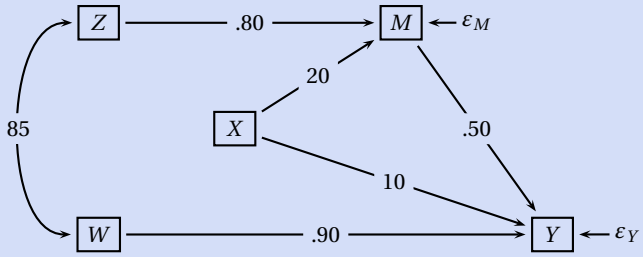

Figure 1.6. Path diagram of $E(M|X, Z, W)$ and $E(Y|X, M, Z, W)$

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

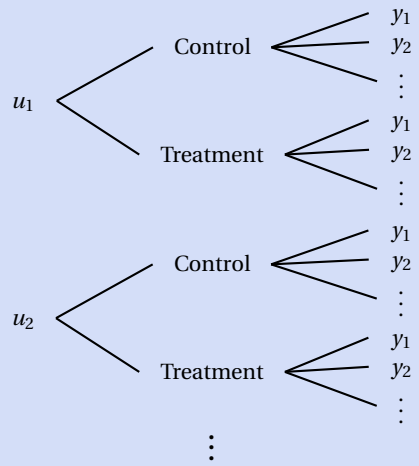

Figure 2.1. A simple experiment or quasi-experiment

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,

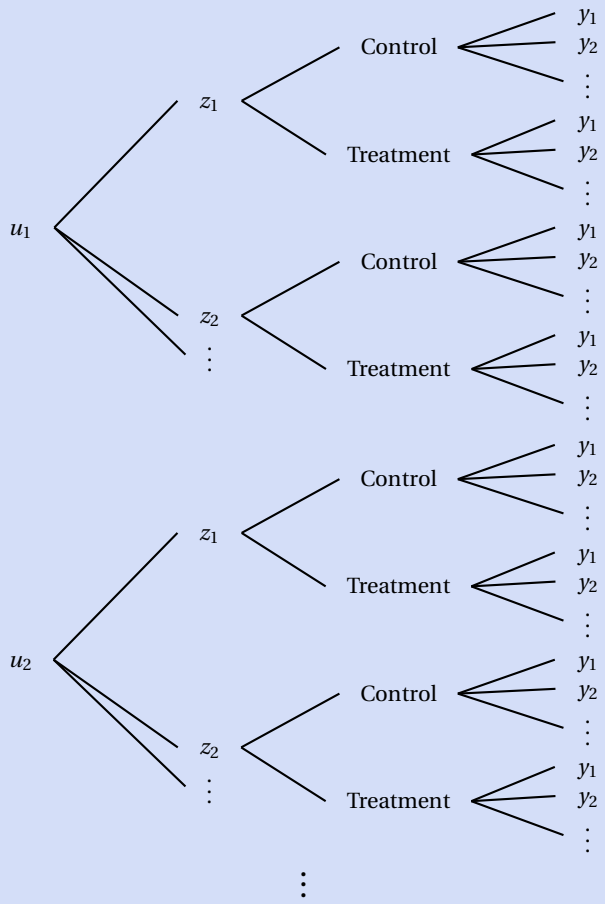

Figure 2.2. Experiment or quasi-experiment with a fallible covariate

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

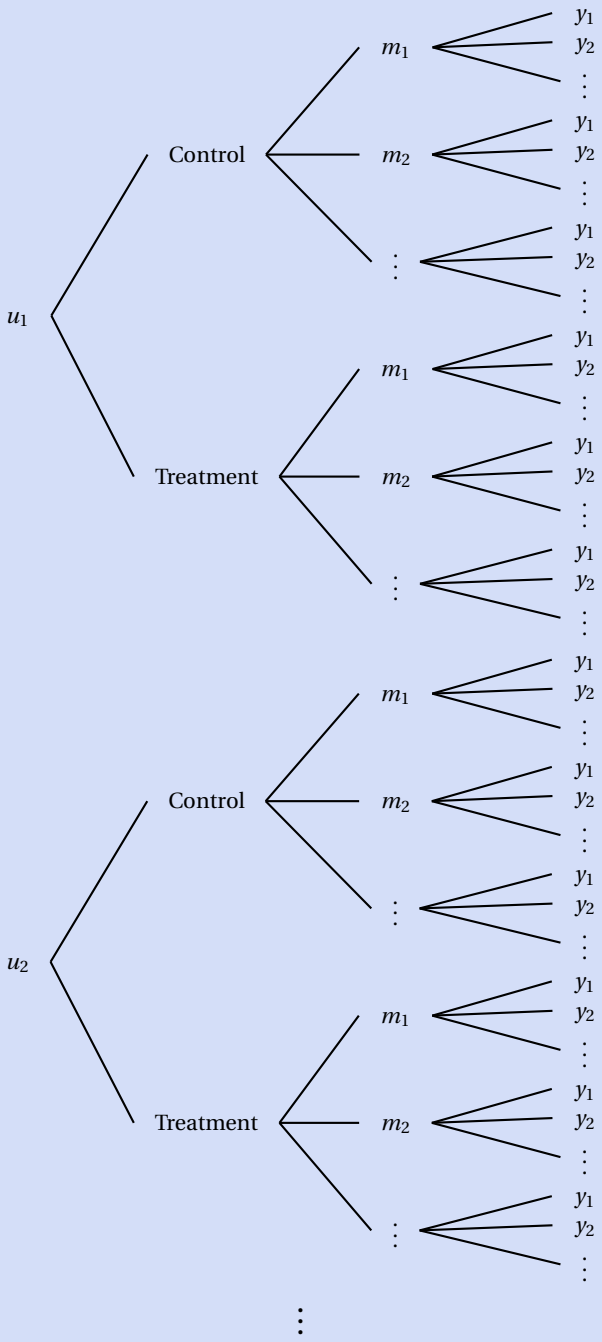

Figure 2.3. Experiment or quasi-experiment with an intermediate variable

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*,

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

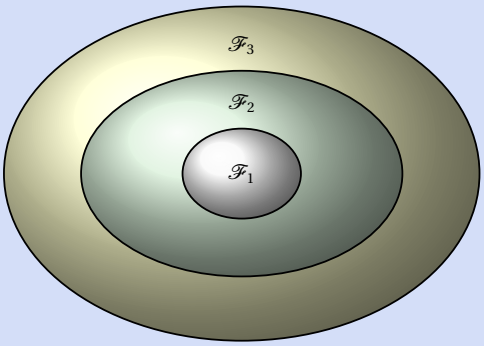

Figure 3.1. Venn-diagram of a filtration with $T = \{1,2,3\}$

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.

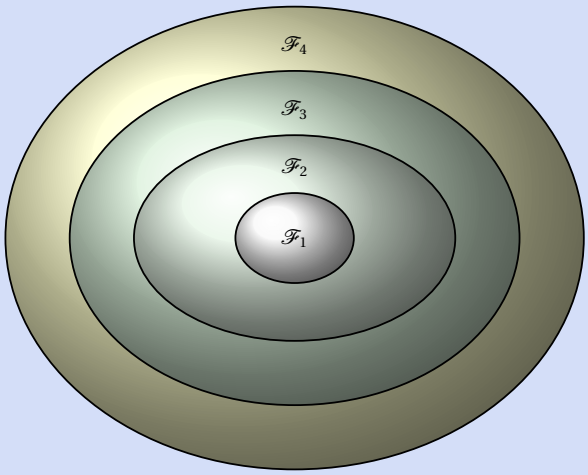

Figure 3.2. Venn-diagram of a filtration with $T = \{1,2,3,4\}$

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. 93)

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. 94)

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 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 4.13).

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. 84), 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*

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

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 4.18.

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

$$E\big(\delta_{xx',t}\big), \tag{5.14}$$

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

$$E\big(\delta_{xx'} - \delta_{xx',t}\big), \tag{5.15}$$

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

$$E\big(\delta_{xx',t} \,\big|\, V=v\big), \tag{5.16}$$

and the $(V=v)$-conditional t-indirect effect of x vs. x' on Y by

$$E\big(\delta_{xx'} - \delta_{xx',t} \,\big|\, V=v\big). \tag{5.17}$$

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 4.19). 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*

$$E\big(\delta_{xx',t}\,|\,Z=z\big) \tag{5.18}$$

of x vs. x' given the value z of a *pre-treatment variable* Z, the *individual t-direct effect*

$$E\big(\delta_{xx',t}\,|\,U=u\big) \tag{5.19}$$

of x vs. x' given the observational unit u, the *$(M=m)$-conditional t-direct effect*

$$E\big(\delta_{xx',t}\,|\,M=m\big) \tag{5.20}$$

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.,

$$E(\delta_{xx'}) \;=\; E(\delta_{xx',t}) + E(\delta_{xx'} - \delta_{xx',t}). \tag{5.21}$$

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	m	1/6	106	114	3/4	8	1/6	1/6
u_6	m	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	m	1/6	106	114	1/4	8	3/10	1/14
u_6	m	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. 4.18).

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. 90) 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).

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.