Rolf Steyer
Ulf Kröhne

Probability and Causality

Version: June 13, 2016

University of Jena

Contents

List of Figures

List of Tables

Chapter 7

Independent Cause and Mean-Independent Outcome

In chapter 6 we have seen that the conditional expectation values $E(Y|X=x)$ and $E(Y|X=x, Z=z)$ and their differences, the prima facie effects, can be biased, where X denotes a cause and Z a covariate of X. We have defined *unbiasedness* of these conditional expectation values *with respect to total effects* as well as unbiasedness of the conditional expectations $E(Y|X)$ and $E(Y|X,Z)$ with respect to total effects assuming that the cause X is discrete.

Unbiasedness of $E(Y|X)$ and $E(Y|X,Z)$ are a first kind of causality conditions. If $E(Y|X)$ is unbiased with respect to total effects and X is discrete, then the differences $E(Y\,|\,X=x) - E(Y\,|\,X=x')$ are identical to the *average total effects* $E(\delta_{xx'})$. Similarly, unbiasedness with (respect to total effects) of the conditional expectation $E(Y|X,Z)$ implies that the differences $E(Y|X=x, Z=z) - E(Y|X=x', Z=z)$ are equal to the $(Z=z)$-*conditional total effect* $E(\delta_{xx'}\,|Z=z)$, provided that $P(X=x, Z=z),\ P(X=x', Z=z) > 0$. If M denotes an intermediate variable of X and Y, then similar propositions also hold for unbiasedness of the conditional expectations $E(Y|X,M)$ and $E(Y|X,M,Z)$ *with respect to t-direct effects.*

However, the unbiasedness conditions have two limitations. First, they cannot be tested empirically, which implies that they cannot be used for covariate selection. Second, they can be *accidental* in the sense that unbiasedness may hold for $E(Y|X)$ but not for $E(Y|X,Z)$, where Z denotes a covariate. Therefore, in this chapter, we study two other kinds of causality conditions that are less volatile. These conditions will be referred to as the *independent cause conditions* and the *mean-independent outcome conditions* respectively.

Overview

We start with the concepts of independence and conditional independence, and use them to introduce the independent cause conditions. Next we define the mean-independent outcome conditions and study the implications of both kinds of conditions on unbiasedness. Then we illustrate these causality conditions by some examples. Finally, we discuss their methodological implications. In particular, we discuss the role of randomization and conditional randomization as well as covariate selection in the analysis of total and direct effects.

7.1 Independent Cause Conditions

7.1.1 Independence and Conditional Independence

A general definition (including the case in which X is continuous) and many important properties of conditional independence are presented in section PR-16.1.1, which also includes conditional independence of sets of events and of more than two random variables. A number of useful properties of conditional independence are summarized in PR-Box 16.2. (See Dawid, 1979, for an overview of applications of conditional independence in statistics.)

The following corollary is a special case of Theorem PR-16.26 (see Rem. PR-16.27)

Corollary 7.1 (Independence and Conditional Independence: Discrete X) *Let X and V be random variables on the probability space $(\Omega, \mathscr{A}, P)$, let X be discrete with a finite number of values $x = 0, 1, \ldots, J$ and let $P(X{=}x) > 0$ for all $x = 0, 1, \ldots, J$.*

(i) *Then X and V are independent (abbr.: $X \underset{P}{\perp\!\!\!\perp} V$) if and only if*

$$P(X{=}x \mid V) \underset{P}{=} P(X{=}x) \quad \textit{for all } x = 0, 1, \ldots, J. \tag{7.1}$$

(ii) *If Z is a random variable on $(\Omega, \mathscr{A}, P)$ too, then X and V are Z-conditionally independent (abbr.: $X \underset{P}{\perp\!\!\!\perp} V \mid Z$) if and only if*

$$P(X{=}x \mid Z, V) \underset{P}{=} P(X{=}x \mid Z) \quad \textit{for all } x = 0, 1, \ldots, J. \tag{7.2}$$

7.1.2 Independent Cause Conditions for Total Effects

Now let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a causality space, let C_X denote a global covariate of X, and consider

$$P(X{=}x \mid C_X) \underset{P}{=} P(X{=}x), \quad \text{for all } x = 0, 1, \ldots, J, \tag{7.3}$$

which is abbreviated by $X \underset{P}{\perp\!\!\!\perp} C_X$. Next, let Z be a C_X-measurable random variable on $(\Omega, \mathscr{A}, P)$ and consider

$$P(X{=}x \mid C_X) \underset{P}{=} P(X{=}x \mid Z), \quad \text{for all } x = 0, 1, \ldots, J, \tag{7.4}$$

abbreviated by $X \underset{P}{\perp\!\!\!\perp} C_X \mid Z$. These two conditions will be referred to as the *independent cause conditions for total effects*. Note that Equation (7.3) is a special case of (7.4) for Z being a constant mapping, i. e., for $Z(\omega) = const$, for all $\omega \in \Omega$.

If Z is measurable with respect to C_X, i. e., if $\sigma(Z) \subset \sigma(C_X)$, then $\sigma(Z, C_X) = \sigma(C_X)$ and, therefore, $P(X{=}x \mid Z, C_X) \underset{P}{=} P(X{=}x \mid C_X)$ (see sections PR-2.3.2 and PR-10.1).

Remark 7.2 (Propensity) If Z is C_X-measurable, then the conditional probability

$$\pi_x := P(X{=}x \mid Z) \tag{7.5}$$

is also called the *Z-conditional propensity of x*. The propensity plays an important role in adjustment techniques, because Equations (7.4) and (7.5) imply

$$P(X{=}x \mid C_X) \underset{P}{=} P(X{=}x \mid \pi_x) \quad \text{for all } x = 0, 1, \dots, J \tag{7.6}$$

(see Exercise 7-1). Hence, $X \underset{P}{\perp\!\!\!\perp} C_X \mid Z$ implies $1_{X=x} \underset{P}{\perp\!\!\!\perp} C_X \mid \pi_x$, which means that π_x can be used in adjustment techniques instead of the original covariate Z, provided that X is dichotomous (for details, see ch. 11). ◁

7.1.3 Independent Cause Conditions for Direct Effects

Now we turn to direct effects with respect to time t. We consider a $C_{X,t}$-measurable random variable Z on $(\Omega, \mathscr{A}, P)$

$$P(X{=}x \mid C_{X,t}) \underset{P}{=} P(X{=}x \mid Z) \quad \text{for all } x = 0, 1, \dots, J. \tag{7.7}$$

This condition is abbreviated by $X \underset{P}{\perp\!\!\!\perp} C_{X,t} \mid Z$ and called the *independent cause condition for t-direct effects.*

7.1.4 Falsifiability of the Independent Cause Conditions

The independent cause conditions are empirically falsifiable. More specifically, if W is measurable with respect to C_X, then $X \underset{P}{\perp\!\!\!\perp} C_X$ implies

$$P(X{=}x \mid W) \underset{P}{=} P(X{=}x) \quad \text{for all } x = 0, 1, \dots, J. \tag{7.8}$$

Hence, in order to test $X \underset{P}{\perp\!\!\!\perp} C_X$, we simply have to select a C_X-measurable random variable W and see if Equation (7.8) holds.

Similarly, if W is measurable with respect to C_X, then $X \underset{P}{\perp\!\!\!\perp} C_X \mid Z$ implies

$$P(X{=}x \mid Z, W) \underset{P}{=} P(X{=}x \mid Z) \quad \text{for all } x = 0, 1, \dots, J. \tag{7.9}$$

Hence, in order to test $X \underset{P}{\perp\!\!\!\perp} C_X \mid Z$, we can select a C_X-measurable random variable W and see if Equation (7.9) holds.

Finally, if W is measurable with respect to $C_{X,t}$ then $X \underset{P}{\perp\!\!\!\perp} C_{X,t} \mid Z$ implies

$$P(X{=}x \mid Z, W) \underset{P}{=} P(X{=}x \mid Z) \quad \text{for all } x = 0, 1, \dots, J. \tag{7.10}$$

This shows: In order to test $X \underset{P}{\perp\!\!\!\perp} C_{X,t} \mid Z$, we can choose a $C_{X,t}$-measurable random variable W and see if Equation (7.10) holds. Note that all these tests can only *falsify* but not verify these independent cause conditions.

Example 7.3 (First Examples) A first example of independence of the cause X and a global covariate C_X has already been presented in Table 6.2 (p. 129). Furthermore, Table 6.3 (p. 130) displays an example for Z-conditional independence of X and C_X. In both examples U takes the role of a global covariate C_X. In section 7.4 we treat several examples in more detail. ◁

7.2 Mean-Independent Outcome Conditions

Before we study the implications of the independent cause conditions on unbiasedness (see section 7.3), we introduce a second kind of causality conditions, which we will call the *mean-independent outcome conditions*.

7.2.1 Conditional Mean Independence

Instead of conditional independence, we now use the concept of *conditional mean independence*. If X, Y, and Z are random variables on a probability space $(\Omega, \mathscr{A}, P)$ such that Y is nonnegative or with finite expectation $E(Y)$, then

$$E(Y|X,Z) \underset{P}{=} E(Y|X) \tag{7.11}$$

defines *X-conditional mean independence of Y from Z*. (For more details see section PR-10.6.)

7.2.2 Mean-Independent Outcome Conditions for Total Effects

Hence, if $\langle(\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y\rangle$ is a causality space and Y is numerical nonnegative or with finite expectation, then

$$E(Y|X,C_X) \underset{P}{=} E(Y|X) \tag{7.12}$$

defines X-conditional mean independence of Y from C_X, which will be abbreviated by $Y \vdash C_X | X$. Furthermore, if Z is measurable with respect to C_X, then

$$E(Y|X,C_X) \underset{P}{=} E(Y|X,Z) \tag{7.13}$$

defines *(X,Z)-conditional mean independence of Y from C_X*, abbreviated by $Y \vdash C_X | X, Z$.

If $x \in \Omega'_X$ and $P(X{=}x) = 0$, then Equation (7.12) implies

$$E^{X=x}(Y|C_X) \underset{P^{X=x}}{=} E^{X=x}(Y) \tag{7.14}$$

[see PR-(14.82)], and (7.13) implies

$$E^{X=x}(Y|C_X) \underset{P^{X=x}}{=} E^{X=x}(Y|Z) \tag{7.15}$$

[see PR-(14.81)].

The two conditions $Y \vdash C_X \,|\, X$ and $Y \vdash C_X \,|\, X, Z$ will also be referred to as the *mean-independent outcome conditions for total effects.* Assuming that Z is measurable with respect to C_X implies $\sigma(X, Z, C_X) = \sigma(X, C_X)$, which in turn implies

$$E(Y \,|\, X, Z, C_X) \underset{P}{=} E(Y \,|\, X, C_X) \tag{7.16}$$

[see Eq. (10.1)].

Remark 7.4 (Mean-Independent Outcome Conditions Again) Referring to the filtration $(\mathscr{F}_t, t \in T)$, note that Equation (7.12) is equivalent to

$$E(Y \,|\, \mathscr{F}_{t_X}) \underset{P}{=} E(Y \,|\, X) \tag{7.17}$$

(see Rem. 3.23). According to this equation, there are no random variables other than X that are prior or simultaneous to X and determine the conditional expectation of Y, once we condition on X. Intuitively speaking, X is the only cause of Y, at least with respect to all potential causes that are prior or simultaneous to X. Hence, this condition does not exclude other causes of Y that are in between X and Y, that is posterior to X and prior to Y unless T is such that there is no $t \in T$ such that $t_X < t < t_Y$.

Similarly, Equation (7.13) is equivalent to

$$E(Y \,|\, \mathscr{F}_{t_X}) \underset{P}{=} E(Y \,|\, X, Z), \tag{7.18}$$

because $\sigma(X, C_X) = \mathscr{F}_{t_X}$ [see Def. 4.1 (b)], provided of course that Z is C_X-measurable, that is, provided that Z is a covariate of X. According to Equation (7.18), there are no random variables other than X and Z that are prior or simultaneous to X and determine the conditional expectation of Y, once we condition on X and Z. Intuitively speaking, conditioning on Z, the random variable X is the only cause of Y, at least with respect to all potential causes that are prior or simultaneous to X. ◁

7.2.3 *Mean-Independent Outcome Conditions for Direct Effects*

Now we specify the mean-independent outcome conditions *for direct effects.* If Z is measurable with respect to $C_{X,t}$, then

$$E(Y \,|\, X, C_{X,t}) \underset{P}{=} E(Y \,|\, X, Z) \tag{7.19}$$

defines *Z-conditional mean independence of Y from $C_{X,t}$*. It is abbreviated by $Y \vdash C_{X,t} \,|\, X, Z$. This condition will be also be referred to as the *mean-independent outcome condition for t-direct effects.* If X is discrete with values $x = 0, 1, \ldots, J$ and $P(X{=}x) = 0$ for all $x = 0, 1, \ldots, J$, then Equation (7.13) implies

$$E^{X=x}(Y \,|\, C_{X,t}) \underset{P^{X=x}}{=} E^{X=x}(Y \,|\, Z) \tag{7.20}$$

Again, note that measurability of Z with respect to $C_{X,t}$ implies

$$E(Y \,|\, X, Z, C_{X,t}) \underset{P}{=} E(Y \,|\, X, C_{X,t}). \tag{7.21}$$

Example 7.5 (Implications if U is Measurable With Respect to C_X) Assume that X represents a treatment variable and C_X is a global covariate such that the observational-unit variable U is measurable with respect to C_X. Then $Y \vdash C_X \,|\, X$ implies that there are no individual differences between units with respect to their expectation of the outcome variable Y within each treatment condition x, i. e.,

$$E(Y \,|\, X{=}x, U{=}u) \;=\; E(Y \,|\, X{=}x) \quad \text{for all values } x \text{ of } X \text{ and all units } u,$$

provided that $P(X{=}x, U{=}u) > 0$. This equation may more conveniently be written

$$E(Y \,|\, X, U) \underset{P}{=} E(Y \,|\, X).$$

Note that using this equation, we do not have to presume $P(X{=}x, U{=}u) > 0$.

Similarly, $Y \vdash C_X \,|\, X, Z$ implies that no covariate affects the expectations of Y over and above X and the covariate Z. And again, if X represents a treatment variable, this implies that there are no individual differences between units with respect to their conditional expectation values of the outcome variable Y within treatment conditions *given* Z, i. e.,

$$E(Y \,|\, X, Z, U) \underset{P}{=} E(Y \,|\, X, Z).$$

If $P(X{=}x, Z{=}z, U{=}u) > 0$, this equation implies

$$E(Y \,|\, X{=}x, Z{=}z, U{=}u) \;=\; E(Y \,|\, X{=}x, Z{=}z).$$

Table 7.1 (p. 175) displays an example. ◁

7.2.4 *Falsifiability of the Mean-Independent Outcome Conditions*

The mean-independent outcome conditions imply some propositions that lend themselves for falsification. Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ be a causality space and let Y be numerical and nonnegative or with finite expectation. Then the following propositions hold:

(i) If W is measurable with respect to C_X, then $Y \vdash C_X \,|\, X$ implies

$$E(Y \,|\, X, W) \underset{P}{=} E(Y \,|\, X). \tag{7.22}$$

(ii) If Z and W are measurable with respect to C_X, then $Y \vdash C_X \,|\, X, Z$ implies

$$E(Y \,|\, X, Z, W) \underset{P}{=} E(Y \,|\, X, Z). \tag{7.23}$$

(iii) If Z and W are measurable with respect to $C_{X,t}$, then $Y \vdash C_{X,t} \,|\, X, Z$ implies

$$E(Y \,|\, X, Z, W) \underset{P}{=} E(Y \,|\, X, Z). \tag{7.24}$$

Propositions (i) and (ii) immediately follow from measurability of W with respect to C_X, and Proposition (iii) follows from measurability of W with respect to $C_{X,t}$.

Hence, just as the independent cause conditions, the mean-independent outcome conditions can easily be tested empirically, again only in the sense of falsifiability. For example, in order to test $Y \vdash C_X \,|\, X$, we simply have to select a variable W that is measurable with respect to C_X and see if Equation (7.22) holds. If it does not, then the hypothesis $Y \vdash C_X \,|\, X$ is falsified. Similarly, if we want to test $Y \vdash C_X \,|\, X, Z$, we may select a variable W that is measurable with respect to C_X and see if Equation (7.23) holds. If it does not, then we can conclude that $Y \vdash C_X \,|\, X, Z$ is not satisfied. Finally, if we want to test $Y \vdash C_{X,t} \,|\, X, Z$, we may select a variable W that is measurable with respect to $C_{X,t}$ and see if Equation (7.24) holds. If it does not, then we can conclude that $Y \vdash C_{X,t} \,|\, X, W$ is not satisfied.

7.3 Implications on Unbiasedness

Now we present the most important implications of the causality conditions introduced above on unbiasedness of various conditional expectations and prima facie effect functions. The theorems and corollaries presented in this section summarize the most important points with respect to unbiasedness of the conditional expectations $E(Y \,|\, X)$ and $E(Y \,|\, X, Z)$. After treating some examples in section 7.4, we will discuss the methodological implications of our first two causality conditions in section 7.5.

7.3.1 Unbiasedness With Respect to Total Effects: No Covariates

In this section we study the implications of our first two causality conditions on unbiasedness *with respect to total effects*. Let us start without considering covariates.

Conditional Expectations

In our first theorem we consider a single value x of X with $P(X{=}x) > 0$ and the conditional expectation value $E(Y \,|\, X{=}x)$. Reading the following theorem, remember that τ_x denotes the conditional expectation $E^{X=x}(Y \,|\, C_X)$ (see Def. 4.12).

Theorem 7.6 (Unbiasedness of Conditional Expectation Values)
Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ *be a causality space, let* Y *be numerical and nonnegative or with finite expectation, and let* $x \in \Omega'_X$ *be a value of* X *with* $P(X{=}x) > 0$.

(i) *Then* $X \underset{P}{\perp\!\!\!\perp} C_X$ *implies* $P(X{=}x \,|\, C_X) \underset{P}{>} 0$ *and*

$$E(Y \,|\, X{=}x) \;=\; E(\tau_x). \tag{7.25}$$

(ii) $P(X{=}x\,|\,C_X) \underset{P}{>} 0$ *and* $Y \vdash C_X\,|\,X$ *imply Equation* (7.25).

(Proof p. 181)

Remark 7.7 (Randomized Experiment) Remember that Equation (7.25) defines τ_x-unbiasedness of $E(Y\,|\,X{=}x)$. The first sufficient condition for unbiasedness, the independent cause condition $X \underset{P}{\perp\!\!\!\perp} C_X$, can be created by randomized assignment of the unit to one of the treatment conditions. In this case, X and C_X are independent. In contrast, there is no design technique implying that $Y \vdash C_X\,|\,X$ holds. ◁

Remark 7.8 (Less Restrictive Conditions Implying Unbiasedness) If $P(X{=}x) > 0$, then $1_{X=x} \underset{P}{\perp\!\!\!\perp} C_X$ implies $E(Y\,|\,X{=}x) = E(\tau_x)$. Similarly, if $P(X{=}x\,|\,C_X) \underset{P}{>} 0$, then $Y \vdash C_X\,|\,1_{X=x}$ implies $E(Y\,|\,X{=}x) = E(\tau_x)$. This immediately follows from Theorem 7.6 for $X := 1_{X=x}$. ◁

Prima Facie Effects

The following corollary is another immediate implication of Theorem 7.6. In this corollary we specify two conditions under which the prima facie effect

$$PFE_{xx'} := E(Y\,|\,X{=}x) - E(Y\,|\,X{=}x')$$

is $\delta_{xx'}$-unbiased. Reading this corollary, remember that $\delta_{xx'} := \tau_x - \tau_{x'}$ denotes the atomic total-effect variable (see Def. 4.12) and that $E(\delta_{xx'}) = E(\tau_x) - E(\tau_{x'})$.

Corollary 7.9 (Unbiasedness of Prima Facie Effects)
Let $\langle(\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y\rangle$ *be a causality space, let* Y *be numerical and nonnegative or with finite expectation, and let* $x, x' \in \Omega'_X$ *be two values of* X.

(i) *If* $P(X{=}x) > 0$, $P(X{=}x') > 0$, *and* $X \underset{P}{\perp\!\!\!\perp} C_X$, *then*

$$P(X{=}x\,|\,C_X) \underset{P}{>} 0 \quad \textit{and} \quad P(X{=}x'\,|\,C_X) \underset{P}{>} 0 \tag{7.26}$$

and

$$PFE_{xx'} = E(\delta_{xx'}). \tag{7.27}$$

(ii) *If* (7.26) *holds and* $Y \vdash C_X\,|\,X$, *then Equation* (7.27) *holds as well.*

Conditional Expectation

Theorem 7.6 also implies the following corollary concerning the conditional expectation

$$E(Y\,|\,X) := \sum_{x=0}^{J} E(Y\,|\,X{=}x) \cdot 1_{X=x}. \tag{7.28}$$

Corollary 7.10 (Unbiasedness of the Conditional Expectation)
Let $\langle(\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y\rangle$ *be a causality space, let* Y *be numerical and nonnegative or with finite expectation, and let* X *be discrete with values* $x = 0, 1, \ldots, J$.

(i) If $P(X{=}x) > 0$ *for* all $x = 0, 1, \ldots, J$ *and* $X \underset{P}{\perp\!\!\!\perp} C_X$*, then*

$$P(X{=}x \mid C_X) \underset{P}{>} 0 \quad \textit{for all } x = 0, 1, \ldots, J \tag{7.29}$$

and

$$E(Y \mid X{=}x) = E(\tau_x) \quad \textit{for all } x = 0, 1, \ldots, J. \tag{7.30}$$

(ii) If (7.29) *holds and* $Y \vdash C_X \mid X$*, then Equation* (7.30) *holds as well.*

Hence, if X is discrete with values $x = 0, 1, \ldots, J$, then $X \underset{P}{\perp\!\!\!\perp} C_X$ implies τ_x-unbiasedness of the conditional expectation $E(Y \mid X)$ [see Eq. (7.28)], provided that X is discrete with nonzero probabilities for all its values. Similarly, $Y \vdash C_X \mid X$ implies τ_x-unbiasedness of the conditional expectation $E(Y \mid X)$, provided that X is discrete and $P(X{=}x \mid C_X) \underset{P}{>} 0$ holds for all its values x. Remember that $E(Y \mid X)$ denotes the conditional expectation of Y given X, that is, the best fitting function of X, and *not* $Q_{lin}(Y \mid X)$, the best fitting *linear* function of X (see Def. PR-7.2).

7.3.2 Conditioning on a Covariate

Now we extend Theorem 7.6 and its corollaries to the case in which we condition on a covariate Z. Let us start considering the implications of the independent cause condition and mean-independent outcome condition on the conditional expectation $E^{X=x}(Y \mid Z)$ and the conditional expectation values $E(Y \mid X{=}x, Z{=}z)$. Remember, the term $E^{X=x}(Y \mid Z)$ denotes the Z-conditional expectation of Y with respect to the $(X{=}x)$-conditional-probability measure $P^{X=x}$. Hence, if X represents a treatment variable, then $E^{X=x}(Y \mid Z)$ is the Z-conditional expectation of Y in treatment x. Also remember that τ_x*-unbiasedness of* $E^{X=x}(Y \mid Z)$ is defined by

$$E^{X=x}(Y \mid Z) \underset{P}{=} E(\tau_x \mid Z), \tag{7.31}$$

whereas

$$E(Y \mid X{=}x, Z{=}z) = E(\tau_x \mid Z{=}z). \tag{7.32}$$

defines τ_x*-unbiasedness of* $E(Y \mid X{=}x, Z{=}z)$.

Theorem 7.11 (Unbiasedness of the Conditional Expectation $E^{X=x}(Y \mid Z)$)
Let $\langle(\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y\rangle$ *be a causality space, let* Y *be numerical and nonnegative or with finite expectation, let* $x \in \Omega'_X$ *be a value of* X *with* $P(X{=}x \mid C_X) \underset{P}{>} 0$*, and let* Z *be measurable w.r.t.* C_X.

(i) *Then each of* $X \underset{P}{\perp\!\!\!\perp} C_X \,|\, Z$ *and* $Y \vdash C_X \,|\, X, Z$ *implies that* $E^{X=x}(Y \,|\, Z)$ *is* τ_x*-unbiased.*
(ii) *If* x, z *are two values of* X *and* Z *such that* $P(X{=}x, Z{=}z) > 0$, *then each of* $X \underset{P}{\perp\!\!\!\perp} C_X \,|\, Z$ *and* $Y \vdash C_X \,|\, X, Z$ *implies that* $E(Y \,|\, X{=}x, Z{=}z)$ *is* τ_x*-unbiased.*

(Proof p. 183)

Hence, presuming $P(X{=}x \,|\, C_X) \underset{P}{>} 0$, the Z-conditional expectation $E^{X=x}(Y \,|\, Z)$ of Y is unbiased with respect to total effects if $X \underset{P}{\perp\!\!\!\perp} C_X \,|\, Z$ or $Y \vdash C_X \,|\, X, Z$.

Remark 7.12 (Common Support) In Theorem 7.11 we required $P(X{=}x \,|\, C_X) \underset{P}{>} 0$. If Z is C_X-measurable, then this implies $P(X{=}x \,|\, Z) \underset{P}{>} 0$, which in turn implies $P(X{=}x|Z{=}z) > 0$ for all values z of Z that have a positive probability $P(Z{=}z) > 0$. The condition $P(X{=}x \,|\, Z) \underset{P}{>} 0$ may also be called the *common support condition* (cf., e. g., Lechner, 2000). Note that common support refers to conditional *probabilities* and not to relative frequencies. At the level of a concrete sample, we may use the term *data sparseness* if the relative frequencies estimating the conditional probabilities $P(X{=}x \,|\, Z{=}z)$ are zero. ◁

Remark 7.13 (Conditional Randomization) The first condition implying unbiasedness of the $E^{X=x}(Y \,|\, Z)$ with respect to total effects, $X \underset{P}{\perp\!\!\!\perp} C_X \,|\, Z$, can be created in an experiment either by (a) randomized assignment of units to treatment conditions (represented by X) or (b) randomized assignment of units to treatment conditions *conditional on the values* z *of a covariate* Z. However, it may also hold if the covariate Z is properly selected. In contrast, the condition $Y \vdash C_X \,|\, X, Z$ cannot be created by a design technique. Instead we can only try to find a (possibly multivariate) covariate Z such that $Y \vdash C_X \,|\, X, Z$ holds. In other words, $Y \vdash C_X \,|\, X, Z$ can only be used for covariate selection (see section 7.5 for more details). ◁

Prima-Facie-Effect Functions

Remember, the prima facie effect function is defined by $PFE_{xx';Z} := E^{X=x}(Y \,|\, Z) - E^{X=x'}(Y \,|\, Z)$. Hence, unbiasedness of the conditional expectations $E^{X=x}(Y \,|\, Z)$ and $E^{X=x'}(Y \,|\, Z)$ has implications on unbiasedness of the prima-facie-effect function $PFE_{xx';Z}$ and on the $(Z{=}z)$-conditional prima facie effect $PFE_{xx';Z=z}$.

Corollary 7.14 (Unbiasedness of Conditional Prima Facie Effects)
Let $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ *be a causality space, let* Y *be numerical and nonnegative or with finite expectation, let* $x, x' \in \Omega'_X$ *be two values of* X *with* $P(X{=}x \,|\, C_X) \underset{P}{>} 0$ *and* $P(X{=}x' \,|\, C_X) \underset{P}{>} 0$, *let* Z *be measurable w.r.t.* C_X, *and let* $\delta_{xx'} := \tau_x - \tau_{x'}$.

(i) *Each of the two conditions* $X \underset{P}{\perp\!\!\!\perp} C_X \,|\, Z$ *and* $Y \vdash C_X \,|\, X, Z$ *implies*

$$PFE_{xx';Z} := E^{X=x}(Y\,|\,Z) - E^{X=x'}(Y\,|\,Z) \underset{P}{=} E(\delta_{xx'}\,|\,Z).$$

(ii) *If* $P(X=x, Z=z), P(X=x', Z=z) > 0$, *then each of the two conditions* $X \underset{P}{\perp\!\!\!\perp} C_X \,|\, Z$ *and* $Y \vdash C_X \,|\, X, Z$ *implies*

$$PFE_{xx';Z=z} := E(Y\,|\,X=x, Z=z) - E(Y\,|\,X=x', Z=z) = E(\delta_{xx'}\,|\,Z=z).$$

(Proof p. 184)

According to this corollary, the prima-facie-effect function $PFE_{xx';Z}$ is unbiased with respect to total effects if (a) $P(X=x\,|\,C_X) \underset{P}{>} 0$ and $P(X=x'\,|\,Z) \underset{P}{>} 0$ and (b) the independent cause condition $X \underset{P}{\perp\!\!\!\perp} C_X \,|\, Z$ or the mean-independent outcome condition $Y \vdash C_X \,|\, X, Z$, hold. Correspondingly, the prima facie effect $PFE_{xx';Z=z}$ is unbiased with respect to the $(Z=z)$-conditional total effect of x vs. x', provided that $P(X=x, Z=z) > 0$ and $P(X=x', Z=z) > 0$ and at least one of these two causality conditions holds.

Now let us turn to the implications of our first two causality conditions on the conditional expectation $E(Y\,|\,X, Z)$.

Corollary 7.15 (Unbiasedness of the Conditional Expectation $E(Y\,|\,X, Z)$**)**
Let $\langle(\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y\rangle$ *be a causality space, let* Y *be numerical and nonnegative or with finite expectation, let* X *be discrete with a finite number of values* $x = 0, 1, \ldots, J$, *let* Z *be measurable w.r.t.* C_X, *and let* $P(X=x\,|\,C_X) \underset{P}{>} 0$ *for all values* $x = 0, 1, \ldots, J$ *of* X. *Then each of the two conditions* $X \underset{P}{\perp\!\!\!\perp} C_X \,|\, Z$ *and* $Y \vdash C_X \,|\, X, Z$ *implies that* $E(Y\,|\,X, Z)$ *is* τ_x*-unbiased.*

(Proof p. 184)

7.3.3 Unbiasedness With Respect to Direct Effects

Now we study the implications of the independent cause and mean-independent outcome conditions on unbiasedness *with respect to* t*-direct effects*. Remember that $\tau_{x,t}$*-unbiasedness of* $E^{X=x}(Y\,|\,Z)$ is defined by

$$E^{X=x}(Y\,|\,Z) \underset{P}{=} E(\tau_{x,t}\,|\,Z), \tag{7.33}$$

whereas

$$E(Y\,|\,X=x, Z=z) = E(\tau_{x,t}\,|\,Z=z). \tag{7.34}$$

defines $\tau_{x,t}$*-unbiasedness of* $E(Y\,|\,X=x, Z=z)$.

Theorem 7.16 (Unbiasedness of the Conditional Expectation $E^{X=x}(Y|Z)$)
Let $\langle(\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y\rangle$ be a causality space, let Y be numerical and nonnegative or with finite expectation, let $t \in T$ with $t_X \le t < t_Y$, let Z be measurable w.r.t. $C_{X,t}$, and let $x \in \Omega'_X$ be a value of X with $P(X{=}x\,|\,C_{X,t}) \underset{P}{>} 0$.

(i) Then each of $X \underset{P}{\perp\!\!\!\perp} C_{X,t}\,|\,Z$ and $Y \vdash C_{X,t}\,|\,X,Z$ implies

$$E^{X=x}(Y|Z) \underset{P}{=} E(\tau_{x,t}\,|\,Z). \tag{7.35}$$

(ii) If x and z are values of X and Z with $P(X{=}x, Z{=}z) > 0$, then each of $X \underset{P}{\perp\!\!\!\perp} C_{X,t}\,|\,Z$ and $Y \vdash C_{X,t}\,|\,X,Z$ implies

$$E^{X=x}(Y|Z{=}z) \;=\; E(\tau_{x,t}\,|\,Z{=}z). \tag{7.36}$$

(Proof p. 184)

Remark 7.17 Equation (7.35) means that $E^{X=x}(Y|Z)$ is $\tau_{x,t}$-unbiased. Because $P(X{=}x) > 0$,

$$E^{X=x}(Y|Z{=}z) \;=\; E(Y\,|\,X{=}x, Z{=}z) \tag{7.37}$$

(see Remark PR-14.37). Therefore Equation (7.36) is equivalent to

$$E(Y\,|\,X{=}x, Z{=}z) \;=\; E(\tau_{x,t}\,|\,Z{=}z). \tag{7.38}$$

◁

Now we consider the implications of the two causality conditions on the prima facie effect function $PFE_{xx';Z}$ and the prima facie effect $PFE_{xx';Z=z}$.

Corollary 7.18 (Unbiasedness of Conditional Prima Facie Effects)
Let the assumptions of Theorem 7.16 be true, let $x, x' \in \Omega'_X$ be two values of X with $P(X{=}x\,|\,C_{X,t}) \underset{P}{>} 0$ and $P(X{=}x'\,|\,C_{X,t}) \underset{P}{>} 0$, and let $\delta_{xx',t} := \tau_{x,t} - \tau_{x',t}$.

(i) Then each of the conditions $X \underset{P}{\perp\!\!\!\perp} C_{X,t}\,|\,Z$ and $Y \vdash C_{X,t}\,|\,X,Z$ implies

$$PFE_{xx';Z} \;:=\; E^{X=x}(Y|Z) - E^{X=x'}(Y|Z) \underset{P}{=} E(\delta_{xx',t}\,|\,Z).$$

(ii) If $P(X{=}x, Z{=}z), P(X{=}x', Z{=}z) > 0$, then each of the two conditions $X \underset{P}{\perp\!\!\!\perp} C_{X,t}\,|\,Z$ and $Y \vdash C_{X,t}\,|\,X,Z$ implies

$$\begin{aligned} PFE_{xx';Z=z} \;&:=\; E(Y\,|\,X{=}x, Z{=}z) - E(Y\,|\,X{=}x', Z{=}z) \\ &\underset{P}{=}\; E(\delta_{xx',t}\,|\,Z{=}z). \end{aligned}$$

(Proof p. 185)

Hence, if the conditions specified in this corollary are met, then the prima facie effect function $PFE_{xx';Z}$ and the prima facie effect $PFE_{xx';Z=z}$ are $\delta_{xx',t}$-unbiased.

Now we turn to the implications of the two causality conditions on the conditional expectation $E(Y \mid X, Z)$, where Z is assumed to be measurable with respect to $C_{X,t}$. The following corollary immediately follows from Theorem 7.16.

Corollary 7.19 (Unbiasedness of the Conditional Expectation $E(Y \mid X, W)$ **)** *Let* $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ *be a causality space, let* Y *be numerical and nonnegative or with finite expectation, let* $t \in T$ *with* $t_X \le t < t_Y$*, let* Z *be measurable w.r.t.* $C_{X,t}$*, let* X *be discrete with a finite number of values* $x = 0, 1, \ldots, J$*, and assume that* $P(X{=}x \mid C_{X,t}) \underset{P}{>} 0$ *for all* $x = 0, 1, \ldots, J$*. Then each of* $X \underset{P}{\perp\!\!\!\perp} C_{X,t} \mid Z$ *and* $Y \vdash C_{X,t} \mid X, Z$ *implies*

$$E^{X=x}(Y \mid Z) \underset{P}{=} E(\tau_{x,t} \mid Z) \quad \textit{for all } x = 0, 1, \ldots, J. \tag{7.39}$$

Equation (7.39) means that $E(Y \mid X, Z)$ is $\tau_{x,t}$-unbiased [see Def. 6.34 (ii)].

7.4 Examples

Now we illustrate the two causality conditions treated in this chapter by some examples.

7.4.1 Examples for the Independent Cause Conditions

In chapter 4 we introduced a first example with independence of X and a global covariate C_X (see Table 4.2, p. 82). A second example has already been presented in chapter 6 (see Table 6.2, p. 129), and in the same chapter (see Table 6.3, p. 130), there is also an example with Z-conditional independence of X and C_X. In all these examples, the global covariate is identical to the observational-unit variable U, i. e., $C_X = U$, implying

$$E(Y \mid X, C_X) \underset{P}{=} E(Y \mid X, U) \tag{7.40}$$

and

$$P(X{=}1 \mid C_X) \underset{P}{=} P(X{=}1 \mid U) \tag{7.41}$$

for the individual treatment probability function $P(X{=}1 \mid U)$, whose values are the individual treatment probabilities $P(X{=}1 \mid U{=}u)$. If the values u of U represent the observational units *at the onset of treatment*, $C_X = U$ will hold in empirical applications if (a) no fallible covariate is assessed and (b) there is no other variable that is simultaneous with respect to $(\mathscr{F}_t, t \in T)$ to the treatment variable X (such as a second treatment variable).

Example 7.20 (Independent Treatment) Table 6.2 (p. 129) displays an example in which $X \underset{P}{\perp\!\!\!\perp} C_X$ holds, where $C_X = U$. As is easily seen in this table, the individual treatment probabilities are the same for all units, namely

$$P(X{=}1\,|\,U) \;=\; P(X{=}1) \;=\; \frac{3}{4}.$$

$X \perp\!\!\!\perp U$ implies that X and *all* random variables that are measurable with respect to U are independent.

In the same example, $X \underset{P}{\perp\!\!\!\perp} C_X \,|\, Z$ holds as well, where $Z :=$ *sex*, because the $Z{=}z$-conditional treatment probabilities also do not depend on the units, i. e.,

$$P(X{=}1\,|\,U,Z) \;=\; P(X{=}1\,|\,Z) \;=\; P(X{=}1) \;=\; \frac{3}{4}.$$

$X \underset{P}{\perp\!\!\!\perp} U \mid Z$ implies that X and all random variables that are measurable with respect to U are also Z-conditionally independent. This is no coincidence but an implication of the fact that $X \underset{P}{\perp\!\!\!\perp} U$ and measurability of Z with respect to U implies $X \underset{P}{\perp\!\!\!\perp} U \,|\, Z$, which in turn implies Z-conditional independence of X and all random variables that are measurable with respect to U [see PR-Box 16.3 (vi)].

If we assume $C_X = U$, then, according to Corollary 7.10, independence of X and U implies unbiasedness of the treatment-conditional expectation $E(Y\,|\,X)$ with respect to average total effects. Furthermore, because in this example $P(X{=}0) > 0$ and $P(X{=}1) > 0$, independence of X and U also implies τ_x-unbiasedness of the conditional expectation values $E(Y\,|\,X{=}0)$ and $E(Y\,|\,X{=}1)$ and as well as δ_{10}-unbiasedness of the prima facie effect

$$PFE_{10} \;:=\; E(Y\,|\,X{=}1) - E(Y\,|\,X{=}0) \;\approx\; 102.333 - 92.333 \;\approx\; 10$$

(see Th. 7.6 and Cor. 7.9). In other words, $PFE_{10} = E(\delta_{10})$.

As already stated above, $X \underset{P}{\perp\!\!\!\perp} U \mid Z$ holds as well, where $Z :=$ *sex*. Because $C_X = U$ and Z is U-measurable, Theorem 7.11 implies that the conditional expectation $E(Y\,|\,X, Z)$ is unbiased with respect to total effects. Furthermore, because $P(X{=}0, Z{=}z), P(X{=}1, Z{=}z) > 0$ for both values z of Z, this implies τ_x-unbiasedness of all conditional expectation values $E(Y\,|\,X{=}x, Z{=}z)$, and this in turn implies that the prima facie effects

$$PFE_{10;\,Z=m} \;=\; E(Y\,|\,X{=}1, Z{=}m) - E(Y\,|\,X{=}0, Z{=}m) \;\approx\; 92.50 - 83.00 \;\approx\; 9.50$$

and

$$PFE_{10;\,Z=f} \;=\; E(Y\,|\,X{=}1, Z{=}f) - E(Y\,|\,X{=}0, Z{=}f) \;=\; 122 - 111 \;=\; 11$$

(see Theorem 7.11 and Corollary 7.14) are δ_{10}-unbiased.

Also note that

$$PFE_{10} \;=\; E(PFE_{10;\,Z}) \;=\; 9.50 \cdot \frac{4}{6} + 11 \cdot \frac{2}{6} \;=\; 10.$$

This means that the prima facie effect is the expectation of the corresponding conditional prima facie effects. This property, which will be called *expectation stability*, is no coincidence. Instead it follows from $X \underset{P}{\perp\!\!\!\perp} C_X$ (see PR-Th. 15.14 and Rem. 15.17). ◁

Example 7.21 (Conditionally Independent Cause) In Table 6.3 (see p. 130) we already treated an example in which $X \underset{P}{\perp\!\!\!\perp} C_X \mid Z$ holds, whereas $X \underset{P}{\perp\!\!\!\perp} C_X$ does not. Again, $C_X = U$. As is easily seen in this table, the individual treatment probabilities are the same for all units *within each of the two subsets of males and females*, i. e.:

$$P(X{=}1 \mid U{=}u, Z{=}m) \;=\; P(X{=}1 \mid Z{=}m) \;=\; \frac{3}{4} \qquad \text{for each male unit } u$$

and

$$P(X{=}1 \mid U{=}u, Z{=}f) \;=\; P(X{=}1 \mid Z{=}f) \;=\; \frac{1}{4} \qquad \text{for each female unit } u.$$

Hence, the individual treatment probabilities are 3/4 for males and 1/4 for females. These individual treatment probabilities differ for different values of the covariate, but they are invariant *given a value of the covariate Z*. Note that the conditional treatment probability $P(X{=}1|Z{=}m) = 3/4$ is also the individual treatment probability $P(X{=}1 \mid U{=}u)$ for the male units u_1 to u_4, and the conditional treatment probability $P(X{=}1 \mid Z{=}f) = 1/4$ is also the individual treatment probability $P(X{=}1 \mid U{=}u)$ for the female units u_5 and u_6. This follows from

$$\begin{aligned} P(X{=}1 \mid U) &= E[P(X{=}1 \mid U, Z) \mid U] && \text{[PR-Box 10.2 (v)]} \\ &= E[P(X{=}1 \mid Z) \mid U] && [P(X{=}1 \mid U, Z) = P(X{=}1 \mid Z)] \\ &= P(X{=}1 \mid Z). && \text{[PR-Box 10.2 (vii)]} \end{aligned}$$

In Table 6.3 (p. 130) there is only *conditional* independence of units and treatments given Z, i. e., $X \underset{P}{\perp\!\!\!\perp} U \mid Z$. Therefore, in this table, the unconditional prima facie effect

$$PFE_{10} \;=\; E(Y|X{=}1) - E(Y|X{=}0) \;\approx\; 96.715 - 99.800 \;\approx\; -3.085$$

is biased, because the average total effect in this example is $E(\delta_{10}) = 10$ (see Exercise 7-2). However, the $(Z{=}z)$*-conditional* prima facie effects are unbiased. In fact, they are the same as in Table 6.2 (see p. 129 and Example 7.20). Hence, we can use the conditional prima facie effects to compute the average total effect. Remember, if the conditional prima facie effects are unbiased, i. e., if they are equal to the conditional total effects, then the expectation of the conditional prima facie effects is equal to the average total effect. In our example, this expectation is:

$$\begin{aligned} E(\delta_{10}) \;=\; E(PFE_{xx';Z}) &= PFE_{10;Z=m} \cdot \frac{4}{6} + PFE_{10;Z=f} \cdot \frac{2}{6} \\ &= 9.50 \cdot \frac{4}{6} + 11 \cdot \frac{2}{6} \;=\; 10.00. \end{aligned}$$

◁

Example 7.22 (Joe and Ann With Randomized Assignment – continued) In Table 4.5 we presented an example with an intermediate variable M and in Example 4.24 we showed how to compute the t-direct true-outcome variables $\tau_{0,t}$ and

$\tau_{1,t}$ as well as their difference $\delta_{10,t}$, the atomic t-direct-effect variable. Now we illustrate how to compute various total and direct causal effects, including the average t-direct effect and some conditional t-direct effects using the results of section 7.3.

First of all, the average total treatment effect is

$$E(\delta_{10}) = .10 \cdot .50 + .00 \cdot .50 \ = \ .05.$$

Because we assume $C_X = U$ and in this example $X \underset{P}{\perp\!\!\!\perp} U$, according to Theorem 7.6, this effect can also be computed by

$$E(\delta_{10}) \ = \ PFE_{10} \ = \ E(Y \,|\, X{=}1) - E(Y \,|\, X{=}0) \ = \ .50 - .45 \ = \ .05.$$

Next, we compute the U-conditional total-effect function $E(\delta_{10} \,|\, U)$ for the covariate U. The values of the conditional expectation $E(\delta_{10} \,|\, U)$ are the individual total treatment effects $E(\delta_{10} \,|\, U{=}Joe) = .10$ and $E(\delta_{10} \,|\, U{=}Ann) = .00$. Because $X \underset{P}{\perp\!\!\!\perp} C_X \,|\, U$, according to Corollary 7.14, these effects can also be computed by

$$PFE_{10;\,U} \ = \ E^{X=1}(Y \,|\, U) - E^{X=0}(Y \,|\, U),$$

the values of which are $PFE_{10;\,U=Joe} = E^{X=1}(Y \,|\, U{=}Joe) - E^{X=0}(Y \,|\, U{=}Joe) = .10$ for Joe and $PFE_{10;\,U=Ann} = E^{X=1}(Y \,|\, U{=}Ann) - E^{X=0}(Y \,|\, U{=}Ann) = .00$ for Ann.

Next, we consider the atomic t_M-direct-effect function δ_{10,t_M}, whose values are $\delta_{10,t_M}(\omega) = .058$ if $\omega \in \{M{=}0, U{=}Joe\}$ and $\delta_{10,t_M}(\omega) = .091$ if $\omega \in \{M{=}1, U{=}Joe\}$, and $\delta_{10,t_M}(\omega) = .003$ if $\omega \in \{M{=}0, U{=}Ann\}$ and $\delta_{10,t_M}(\omega) = -.023$ if $\omega \in \{M{=}1, U{=}Ann\}$ (see Table 4.5). Because $X \underset{P}{\perp\!\!\!\perp} C_{X,t_M} \,|\, M, U$ and $\delta_{10,t_M} = E(\delta_{10,t_M} \,|\, M, U)$, Corollary 7.18 implies that these effects can also be computed by

$$PFE_{10;\,M,U} \ = \ E^{X=1}(Y \,|\, M, U) - E^{X=0}(Y \,|\, M, U),$$

whose values are $E^{X=1}(Y \,|\, M{=}0, U{=}Joe) - E^{X=0}(Y \,|\, M{=}0, U{=}Joe) = .667 - .609 = .058$ and $E^{X=1}(Y \,|\, M{=}1, U{=}Joe) - E^{X=0}(Y \,|\, M{=}1, U{=}Joe) = .818 - .727 = .091$ for Joe, as well as $E^{X=1}(Y \,|\, M{=}0, U{=}Ann) - E^{X=0}(Y \,|\, M{=}0, U{=}Ann,) = .179 - .176 = .003$ and, finally, $E^{X=1}(Y \,|\, M{=}1, U{=}Ann) - E^{X=0}(Y \,|\, M{=}1, U{=}Ann,) = .227 - .250 = -.023$ for Ann.

Last but not least, the average t_M-direct treatment effect is

$$\begin{aligned} E(\delta_{10,t_M}) &= E(PFE_{10;\,M,U}) \\ &= .058 \cdot (.027 + .042 + .008 + .016) \ + \ .091 \cdot (.063 + .168 + .032 + .144) \ + \\ &\quad\ .003 \cdot (.168 + .036 + .092 + .020) \ - \ .023 \cdot (.072 + .024 + .068 + .020) \\ &= .039147. \end{aligned}$$

Note that this effect can *not* be computed from the conditional expectation $E(Y \,|\, X, M)$ even though this is a randomized experiment such that $X \underset{P}{\perp\!\!\!\perp} C_X$ holds. In this example, the observational-unit variable U is the only covariate for which the prima facie effect function $E^{X=1}(Y \,|\, M, U) - E^{X=0}(Y \,|\, M, U)$ is unbiased, and, in this example, $\delta_{10,t_M} = E^{X=1}(Y \,|\, M, U) - E^{X=0}(Y \,|\, M, U)$. ◁

Table 7.1. Mean-Independent Outcome Condition

Unit u	Sampling probability $P(U=u)$	Covariate sex Z	True outcome variable under control τ_0	True outcome variable under treatment τ_1	True treatment probability $P(X=1\|U)$	True total effect variable $\delta_{10} = \tau_1 - \tau_0$	$P(U=u\|X=0)$	$P(U=u\|X=1)$
u_1	1/6	male	90	95	6/8	5	2/19	6/19
u_2	1/6	male	90	95	7/8	5	1/19	7/19
u_3	1/6	male	90	95	6/8	5	2/19	6/19
u_4	1/6	male	90	95	7/8	5	1/19	7/19
u_5	1/6	female	100	110	1/8	10	7/19	1/19
u_6	1/6	female	100	110	2/8	10	6/19	2/19

$E(\tau_0)$	$E(\tau_1)$	$P(X=1)$	$E(\delta_{10})$
93.333	100.000	29/48	6.667

$E(Y\|X=0)$	$E(Y\|X=1)$	PFE_{10}
96.842	96.552	−0.290

	male	female
Conditional causal effects	5.000	10.000
Conditional *PFE*s	5.000	10.000

7.4.2 Examples for the Mean-Independent Outcome Conditions

Now we illustrate the mean-independent outcome condition by a numerical example.

Example 7.23 Table 7.1 (p. 175) shows an example in which the mean-independent outcome condition $Y \vdash C_X \,|\, X, Z$ with respect to total effects holds. In this example, $C_X = U$, which implies

$$E(Y \,|\, X, C_X) \underset{P}{=} E(Y \,|\, X, U).$$

Hence, in this example, the true outcomes are the individual conditional expectation values $E(Y \,|\, X{=}x, U{=}u)$. As is easily seen in the table, the true outcomes under control and the atomic total effects of treatment 1 vs. 0 are the same for all individuals *within males* and *within females*. For all males, the true outcomes *under control* are

$$E(Y \mid X{=}0, Z{=}m, U{=}u_i) \;=\; E(Y \mid X{=}0, Z{=}m) \;=\; 90, \quad i = 1, \ldots, 4,$$

and for all females they are

$$E(Y \mid X{=}0, Z{=}f, U{=}u_i) \;=\; E(Y \mid X{=}0, Z{=}f) \;=\; 100, \quad i = 5, 6.$$

Under treatment, the true outcomes for all males are

$$E(Y \mid X{=}1, Z{=}m, U{=}u_i) \;=\; E(Y \mid X{=}1, Z{=}m) \;=\; 95, \quad i = 1, \ldots, 4,$$

and for all females they are

$$E(Y \mid X{=}1, Z{=}f, U{=}u_i) \;=\; E(Y \mid X{=}1, Z{=}f) \;=\; 110, \quad i = 5, 6.$$

Obviously, in both treatment groups, the expectations of the outcomes only depend on Z, i. e., $E^{X=x}(Y \mid U, Z) = E^{X=x}(Y \mid U) = E^{X=x}(Y \mid Z)$. Hence, we conclude that in this example

$$E(Y \mid X, C_X) \underset{P}{=} E(Y \mid X, U) \underset{P}{=} E(Y \mid X, Z).$$

Because $C_X = U$, this is the mean-independent outcome condition $Y \vdash C_X \mid X, Z$.

According to Theorem 7.11, $Y \vdash C_X \mid X, Z$ implies that $E(Y \mid X, Z)$ is unbiased with respect to total effects, provided that $P(X{=}1 \mid Z) > 0$. Because, $P(X{=}x, Z{=}z) > 0$ for all values of X and Z, according to Corollary 7.14, this in turn implies that the conditional prima facie effects

$$PFE_{10;Z=m} \;=\; E(Y \mid X{=}1, Z{=}m) - E(Y \mid X{=}0, Z{=}m) \;=\; 95 - 90 \;=\; 5$$

and

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

are unbiased as well, i. e., $PFE_{10;Z=m} = E(\delta_{10} \mid Z{=}m)$ and $PFE_{10;Z=f} = E(\delta_{10} \mid Z{=}f)$. In other words, the $(Z{=}z)$-conditional prima facie effects are equal to the $(Z{=}z)$-conditional total effects.

Another conclusion is that the average of the conditional total effects is equal to the average total effect (see Remark 6.7), i. e.,

$$\begin{aligned} E(\delta_{10}) \;=\; E(PFE_{xx';Z}) \;&=\; PFE_{10;Z=m} \cdot 4/6 + PFE_{10;Z=f} \cdot 2/6 \\ &=\; 5 \cdot 4/6 + 10 \cdot 2/6 \;\approx\; 6.667. \end{aligned}$$

Finally, note that if $Y \vdash C_X \mid X, Z$ holds, the individual treatment probabilities $P(X{=}1 \mid U{=}u)$ that played a crucial role in our examples for independence and Z-conditional independence of X and C_X do not matter any more in the following sense: Even though, in this example, the individual treatment probabilities $P(X{=}1 \mid U{=}u)$ are different within the male and within the female subpopulations, the Z-conditional prima facie effects are unbiased. Just as the independent cause conditions, the mean-independent outcome conditions are sufficient for unbiasedness. ◁

7.5 Methodological Implications

Now we discuss the implications of the independent cause and the mean-independent outcome conditions for the design of experiments and quasi-experiments. Theorem 7.6 — specifically its proposition about the independent cause condition $X \underset{P}{\perp\!\!\!\perp} C_X$ — is the theoretical foundation of the design technique of *randomization* and of the analysis of total treatment effects in experiments by comparing (unadjusted) means between treatment conditions. Theorem 7.11 may also be used in randomized experiments when it comes to analyzing Z-conditional and average total treatment effects, and Theorem 7.16 can be used in randomized experiments as well as in (nonrandomized) quasi-experiments whenever we wish to analyze total, direct, and indirect treatment effects. In the analysis of direct and indirect effects, the independent cause condition $X \underset{P}{\perp\!\!\!\perp} C_{X,t} \,|\, Z$ and the mean-independent outcome condition $Y \vdash C_{X,t} \,|\, X, Z$ may be used for covariate selection. Finally, the causality conditions $X \underset{P}{\perp\!\!\!\perp} C_X \,|\, Z$ and $Y \vdash C_X \,|\, X, Z$ can also be used for *covariate selection* in quasi-experiments aiming only at the analysis of total effects.

Remark 7.24 (Randomization) It is well-known at least since (Fisher, 1925/1946) that *randomization* plays a crucial role in ruling out biases in comparisons of means. In a randomized experiment, there is always an observational-unit variable U whose values u denote the observational units. If the filtration $(\mathscr{F}_t, t \in T)$ is appropriately specified, then U will be prior to X. Therefore, the definition of a global covariate C_X [see Def. **??** (b)] guarantees that U is measurable with respect to C_X. Hence, if X represents the treatment variable and C_X a global covariate, then $X \underset{P}{\perp\!\!\!\perp} C_X$, and this if $P(X{=}x) > 0$, then this implies

$$P(X{=}x \,|\, C_X) \;=\; P(X{=}x \,|\, U) \;=\; P(X{=}x).$$

According to this equation, each unit u has the same probability $P(X{=}x \,|\, U{=}u) = P(X{=}x)$ to be assigned to treatment x.

Remember that we are talking about a *single-unit trial*. A simple example of such a single-unit trial consists of sampling a single unit from a set of units, assessing a number of covariates, assigning the unit (or observing its assignment) to one of the treatment conditions and observing the outcome variable (see ch. 2 for more details and other kinds of single-unit trials).

Note that $P(X{=}x \,|\, U) = P(X{=}x)$ does *not* imply that the probabilities $P(X{=}x)$ are the same for all treatment conditions x. If, e. g., there are two treatment conditions, say 0 and 1, then $P(X{=}1)$ might be 1/4 and $P(X{=}0) = 3/4$. However, $X \underset{P}{\perp\!\!\!\perp} C_X$ implies $P(X{=}1 \,|\, U) = P(X{=}1)$, provided, of course, that we consider a random experiment in which U is C_X-measurable. In other words, $X \underset{P}{\perp\!\!\!\perp} C_X$ implies that the individual treatment probabilities $P(X{=}1 \,|\, U{=}u)$ are identical between units, and they are equal to the unconditional treatment probability $P(X{=}1) = 1/4$. Hence, in a perfect randomized experiment we ensure $X \underset{P}{\perp\!\!\!\perp} C_X$, and this condition implies that the prima facie effect $E(Y \,|\, X{=}x) - E(Y \,|\, X{=}x')$ is identical to the average total effect $E(\delta_{xx'})$ (see Cor. 7.9).

$\triangleleft$

Remark 7.25 (Conditional Randomization) What has just been said about randomization also applies to to *conditional randomization* based on a covariate Z. Within the single-unit trial of an experiment or quasi-experiment, in which a unit u is sampled and a value z of the covariate Z is assessed prior to treatment, *conditional randomization* refers to randomized assignment of the unit u to treatment condition x with probability $P(X{=}x \,|\, Z{=}z)$, where $x = 0, 1, \ldots, J$. In this procedure we have to make sure that treatment assignment only depends on the values of Z, but not on any other attribute of the units or any other covariate. In more formal terms, we create $P(X{=}x \,|\, C_X) = P(X{=}x \,|\, Z)$.

We distinguish two cases. *First*, if Z is measured without measurement error, each value z of Z will represent subset of observational units (see, e. g., Rubin, 1984). Table 6.3 (p. 130) contains an example with independence of units and treatments within each of the two subsets of males and females.

Second, if the covariate Z is measured *with* error, i. e., $Z{=}f(U) + \varepsilon$, then a value z of Z does *not* represent a subset of units, because z is not only determined by u but also by other factors and/or chance. Nevertheless, treatment can be assigned such that the treatment probability depends on the *observed score* z, not on the attribute $f(U)$ of the unit itself. In both cases, $Z{=}f(U)$ and $Z{=}f(U) + \varepsilon$, conditional random assignment of a unit u to one of the treatment conditions given Z secures that a global covariate C_X satisfies $X \underset{P}{\perp\!\!\!\perp} C_X \,|\, Z$, i. e., Z-conditionally independence of X and C_X.

Conditional random assignment allows that units with different values z_1 and z_2 of Z have different probabilities to be assigned to treatment x. Thus it is possible to assign a unit for which we observe a covariate value z_1 (e. g., high need) to treatment x with a higher probability than a unit with covariate value z_2 (e. g., low need). In this way we may respect ethical and/or other requirements without compromising the validity of causal inference in such an experiment. Remember, unconditional random assignment means that each unit has the *same* probability of being assigned to a given treatment, regardless of his or her need (and any other attribute of the unit).

For simplicity, suppose there are just two treatment conditions, $X{=}0$ (control) and $X{=}1$ (treatment). Conditional randomization consists of:

(a) fixing the conditional treatment probabilities $P(X{=}1 | Z)$ as a function of the covariate Z,
(b) sampling a unit u and assessing the value z of the covariate, and
(c) randomized assignment of the unit with probability $P(X{=}1 \,|\, Z{=}z)$ to the treatment.

If, e. g., the covariate has three values, say *high*, *medium*, and *low need*, we may toss a dice and assign a unit with *high need* to treatment if the dice shows less than six dots, and assign it to control otherwise. A unit with *medium need* might be assigned to treatment if the dice shows less than four dots, and to control otherwise. Finally, a unit with *low need* might be assigned to treatment if the dice shows one dot, and to control otherwise.

If the covariate Z is discrete, then the conditional treatment probabilities $P(X{=}1\,|\,Z{=}z)$ may be fixed in a table by assigning a treatment probability to each value z that seems appropriate with respect to ethical and other requirements. In the example above, these were the values 5/6 for *high need*, 3/6 for *medium need* and 1/6 for *low need*. If the covariate is univariate continuous, we may also use a function, such as

$$P(X{=}1\,|\,Z) \;=\; \frac{\exp(\lambda_0 + \lambda_1 \cdot Z)}{1 + \exp(\lambda_0 + \lambda_1 \cdot Z)}$$

with coefficients λ_0 and λ_1 that seem appropriate for the experiment considered.

If $P(Z{=}z) > 0$, then conditional randomization implies that the $(Z{=}z)$-conditional prima facie effects $PFE_{xx';Z=z} := E(Y\,|\,X{=}x, Z{=}z) - E(Y\,|\,X{=}x', Z{=}z)$ are the conditional total effects given a value z of the covariate Z. Hence, studying conditional prima facie effects in a perfect conditionally randomized experiment is tantamount to studying conditional total effects. As has been shown in chapter 5 [see, e. g., Eq. (5.4)], once we know the conditional total effects given the values z of a covariate Z, we can also compute the average total effect.

$\lhd$

Remark 7.26 (Mean-Independent Outcome Condition) Using $Y \vdash C_{X,t}\,|\,X,Z$ instead of $X \underset{P}{\perp\!\!\!\perp} C_{X,t}\,|\,Z$ in covariate selection might be easier. In many cases, a pre-test Z_1 of the outcome variable Y and a pre-test Z_2 of the intermediate variable M may already suffice to satisfy the mean-independent outcome condition $Y \vdash C_{X,t}\,|\,X,Z$ with $Z := (M, Z_1, Z_2)$. As has been shown in the example presented in section 1.3.2, such a pre-test of an intermediate variable is of utmost importance, whenever this pre-test Z_2 and the intermediate variable M are correlated, Z_2 and the pre-test Z_1 of the outcome are correlated, Z_1 and outcome Y correlate, and the treatment variable X affects the intermediate variable M. Note that these conditions are usually satisfied whenever there are systematic and relatively stable inter-individual differences in the outcome variable and in the intermediate variable, and X affects the intermediate variable M. A typical example is $Y :=$ *achievement* and $M :=$ *achievement motivation* in an educational experiment. $\lhd$

Remark 7.27 (Covariate Selection in Randomized Experiments) Selecting the t-covariates in the vector $Z := (Z_1, \ldots, Z_m)$ for which we can hope that $X \underset{P}{\perp\!\!\!\perp} C_{X,t}\,|\,Z$ or $Y \vdash C_{X,t}\,|\,X,Z$ hold, are useful design techniques in the analysis of direct treatment effects. Even if there is independence of X and C_X — e. g., created by randomization — ignoring covariates and just analyzing $E(Y\,|\,X, M)$ will usually not yield the average and conditional direct treatment effects (see Example 7.22 and the example in section 1.3.2). $\lhd$

Remark 7.28 (Covariate Selection in Quasi-Experiments) Covariate selection is also the first step in a number of techniques for the analysis of conditional and average total effects in *quasi-experiments*, in which the experimenter cannot fix the true treatment probabilities himself. The steps to follow distinguish different techniques of analysis (see ch. 10). By definition, there is no randomization in a quasi-experiment. However, even under initial randomization, systematic attrition of subjects may invalidate the independent cause condition $X \underset{P}{\perp\!\!\!\perp} C_X$ (see, e. g., Abraham & Russell, 2004; Fichman & Cummings, 2003; Graham & Donaldson, 1993; Shadish, Cook, & Campbell, 2002). In this case, we will say that randomization failed and the initially randomized experiment turned into a quasi-experiment.

In quasi-experiments, selecting the covariates in the covariate vector $Z := (Z_1, \dots, Z_m)$ for which we can hope that $X \underset{P}{\perp\!\!\!\perp} C_X \,|\, Z$ or $Y \vdash C_X \,|\, X, Z$ holds, is a useful strategy in the analysis of conditional and average total treatment effects. Again, compared to the independent cause condition $X \underset{P}{\perp\!\!\!\perp} C_X \,|\, Z$, it might be easier to find covariates such that the mean-independent outcome condition $Y \vdash C_X \,|\, X, Z$ holds. If, e. g., X is a treatment variable, in many cases a pre-test of the outcome variable Y will already suffice. In contrast, there might be many covariates determining the treatment probabilities. For instance, the *severity of the disorder*, *knowing about the treatment*, and *availability of the treatment* are candidates for such covariates. Often only $Z :=$ *pre-test severity of the disorder* may be important as a covariate to satisfy the mean-independent outcome condition $Y \vdash C_X \,|\, X, Z$ if the outcome variable Y is *post-test severity of the disorder*. However, there is no guarantee that the pre-test is sufficient as a covariate for the mean-independent outcome condition to hold. Whether or not it holds depends on the application considered. ◁

Remark 7.29 (Falsifiability) As already mentioned before, in contrast to some of the other causality conditions treated in the subsequent chapters, both the independent cause and the mean-independent outcome conditions can be tested in empirical applications, at least in the sense that some consequences of these assumptions can be checked. Falsifiability is important, because otherwise we would not have any criterion for covariate selection, i. e., for deciding whether or not a specific covariate should be included in the vector $Z := (Z_1, \dots, Z_m)$ of covariates with respect to which we hope that either the independent cause condition or the corresponding mean-independent outcome condition holds.

Let us summarize how we can falsify the independent cause conditions. If $X \underset{P}{\perp\!\!\!\perp} C_X$, then

$$P(X{=}x \,|\, W) \underset{P}{=} P(X{=}x) \tag{7.42}$$

holds for each W that is measurable with respect to C_X. Examples for such variables W are sex, educational status, but also a fallible pre-test. Similarly, $X \underset{P}{\perp\!\!\!\perp} C_X \,|\, Z$ implies:

$$P(X{=}x \,|\, Z, W) \underset{P}{=} P(X{=}x \,|\, Z), \tag{7.43}$$

for each W that is measurable with respect to C_X. Furthermore, $X \underset{P}{\perp\!\!\!\perp} C_{X,t} \,|\, Z$ implies:

$$P(X{=}x \,|\, Z, W) \underset{P}{=} P(X{=}x \,|\, Z), \tag{7.44}$$

for each W that is measurable with respect to $C_{X,t}$. All these equations are easily tested using standard procedures for conditional probabilities such as logistic (see, e. g., Agresti, 2007; Bonney, 1987; Green, 2003; Hosmer & Lemeshow, 2000) or probit regression (see, e. g., McCullagh & Nelder, 1989; Borooah, 2001; Liao, 1994).

Similarly, if W is measurable with respect to C_X, it is easy to test the mean-independent outcome conditions. For example, in regression analysis, it is well-known how to test

$$E(Y|X,W) \underset{P}{=} E(Y|X), \tag{7.45}$$

which follows from $Y \vdash C_X|X$. The same applies to

$$E(Y|X,Z,W) \underset{P}{=} E(Y|X,Z), \tag{7.46}$$

which follows from $Y \vdash C_X|X,Z$ and to

$$E(Y|X,Z,W) \underset{P}{=} E(Y|X,Z), \tag{7.47}$$

following from $Y \vdash C_{X,t}|X,W$, where Z and W are measurable with respect to $C_{X,t}$. For such tests we can use the well-known techniques of regression analysis (see, e. g., Aiken & West, 1996; Allen, 1997; Cohen, Cohen, West, & Aiken, 2003; Draper & Smith, 1998; Gelman & Hill, 2007; Györfi, Kohler, Krzyzak, & Walk, 2002; von Eye & Schuster, 1998; West & Aiken, 2005). ◁

Remark 7.30 (Covariate Selection for Identifying Total and Direct Effects) Note that $Y \vdash C_X|X,Z$ does not imply $X \underset{P}{\perp\!\!\!\perp} C_X|Z$ and vice versa, $X \underset{P}{\perp\!\!\!\perp} C_X|Z$ does not imply $Y \vdash C_X|X,Z$. Furthermore, $Y \vdash C_{X,t}|X,Z$ does not imply $Y \vdash C_X|X,Z$ and $X \underset{P}{\perp\!\!\!\perp} C_{X,t}|Z$ does not imply $X \underset{P}{\perp\!\!\!\perp} C_X|Z$. Hence, covariate selection for the identification of total effects is a different process than selecting covariates for the identification of direct effects. ◁

Outlook

In the next chapters we will treat further sufficient conditions for unbiasedness that can replace the independent cause or the mean-independent outcome conditions and that will be the foundations of additional design techniques and methods of data analysis for the analysis of causal effects. In chapter 10, we will then show how to identify the various total, direct, and indirect effects from conditional expectations that can be estimated from samples of treatment, covariates, intermediate variables, and outcome variables, provided that one of the causality conditions holds.

7.6 Proofs

Proof of Theorem 7.6

(i) $X \underset{P}{\perp\!\!\!\perp} C_X \Rightarrow E(Y|X{=}x) = E(\tau_x)$:

Box 7.1 Glossary of New Concepts

We assume that $\langle (\Omega, \mathscr{A}, P), (\mathscr{F}_t, t \in T), X, Y \rangle$ is a causality space, that Y is numerical and nonnegative or with finite expectation, and that Z denotes a covariate of X.

Conditions Implying Unbiasedness With Respect to Total Effects

$X \underset{P}{\perp\!\!\!\perp} C_X$ — *Independence of X and C_X.* This condition can be created by randomized assignment of units to treatments x. If $P(X{=}x) > 0$, then it implies $P(X{=}x \,|\, C_X) \underset{P}{>} 0$ and τ_x-unbiasedness of $E(Y \,|\, X{=}x)$.

$X \underset{P}{\perp\!\!\!\perp} C_X \,|\, Z$ — *Conditional independence of X and C_X given Z.* This condition can be created by conditional randomized assignment of units to treatment conditions based on the values z of Z. If $P(X{=}x \,|\, C_X) \underset{P}{>} 0$, then it implies that $E^{X=x}(Y \,|\, Z)$ is τ_x-unbiased. Furthermore, we can try to select the (possibly multivariate) covariate $Z = (Z_1, \ldots, Z_m)$ such that $X \underset{P}{\perp\!\!\!\perp} C_X \,|\, Z$ is satisfied. If $X \underset{P}{\perp\!\!\!\perp} C_X \,|\, Z$, then $1_{X=x} \underset{P}{\perp\!\!\!\perp} C_X \,|\, \pi_x$ holds for the Z-conditional propensity $\pi_x := P(X{=}x \,|\, Z)$, a fact that is used in adjustment techniques based on propensity scores.

$Y \vdash C_X \,|\, X$ — *Mean-independence of Y from C_X given X.* This condition is defined by $E(Y \,|\, X, C_X) \underset{P}{=} E(Y \,|\, X)$. It may or may not hold depending on the empirical phenomenon considered. If $P(X{=}x \,|\, C_X) \underset{P}{>} 0$, then $Y \vdash C_X \,|\, X$ implies that $E(Y \,|\, X{=}x)$ is τ_x-unbiased.

$Y \vdash C_X \,|\, X, Z$ — *Mean-independence of Y from C_X given X and Z.* It is defined by $E(Y \,|\, X, C_X) \underset{P}{=} E(Y \,|\, X, Z)$. We can try to select the (possibly multivariate) variable $Z = (Z_1, \ldots, Z_m)$ such that $Y \vdash C_X \,|\, X, Z$ holds. If $P(X{=}x \,|\, C_X) \underset{P}{>} 0$, then $Y \vdash C_X \,|\, X, Z$ implies that $E^{X=x}(Y \,|\, Z)$ is τ_x-unbiased.

Conditions Implying Unbiasedness With Respect to Direct Effects

Now we assume $t \in T$ with $t_X \le t < t_Y$ and that Z is a t-covariate of X.

$X \underset{P}{\perp\!\!\!\perp} C_{X,t} \,|\, Z$ — *Conditional independence of X and $C_{X,t}$ given Z.* If $t > t_X$, then this condition can *not* be created by any design technique. However, we may try to select the (possibly multivariate) t-covariate $Z = (Z_1, \ldots, Z_m)$ of X such that $X \underset{P}{\perp\!\!\!\perp} C_{X,t} \,|\, Z$ is satisfied. In this case, $1_{X=x} \underset{P}{\perp\!\!\!\perp} C_{X,t} \,|\, \pi_x$ holds for $\pi_x = P(X{=}x \,|\, Z)$, a fact that is used in adjustment techniques based on propensity scores. If $P(X{=}x \,|\, C_{X,t}) \underset{P}{>} 0$, then $X \underset{P}{\perp\!\!\!\perp} C_{X,t} \,|\, Z$ implies that $E^{X=x}(Y \,|\, Z)$ is $\tau_{x,t}$-unbiased.

$Y \vdash C_{X,t} \,|\, X, Z$ — *Mean-independence of Y from $C_{X,t}$ given X and Z.* It is defined by $E(Y \,|\, X, C_{X,t}) \underset{P}{=} E(Y \,|\, X, Z)$. We can try to select a (possibly multivariate) covariate $Z = (Z_1, \ldots, Z_m)$ such that $Y \vdash C_{X,t} \,|\, X, Z$ holds. If $P(X{=}x \,|\, C_{X,t}) \underset{P}{>} 0$, then $Y \vdash C_{X,t} \,|\, X, Z$ implies that $E^{X=x}(Y \,|\, Z)$ is $\tau_{x,t}$-unbiased.

$$\begin{aligned} E(Y|X{=}x) &= E^{X=x}(Y) && \text{[PR-(9.11)]} \\ &= E^{X=x}\big[E^{X=x}(Y|C_X)\big] && \text{[PR-Box 10.2 (iv)]} \\ &= E\big[E^{X=x}(Y|C_X)\big] && \big[X \underset{P}{\perp\!\!\!\perp} C_X,\ \text{PR-(16.40)}\big] \\ &= E(\tau_x). && \text{[(4.8)]} \end{aligned}$$

$Y \vdash C_X | X \Rightarrow E(Y|X{=}x) = E(\tau_x)$: Note that $P(X{=}x \mid C_X) \underset{P}{>} 0$ implies that $E^{X=x}(Y|C_X)$ is P-unique [see Cor. PR-14.48 (a) and (c)]. Furthermore, according to (7.15), $E^{X=x}(Y|Z)$ is a version of $E^{X=x}(Y|C_X)$. Therefore, P-uniquenessof $E^{X=x}(Y|C_X)$ implies that $E^{X=x}(Y|Z)$ is P-unique as well. If Y is nonnegative, then there is a version $E^{X=x}(Y|C_X)$ that is nonnegative [see PR-(10.17)], and if $E(Y) < \infty$, then $E[E^{X=x}(Y|C_X)] < \infty$. Therefore, and because $E^{X=x}(Y|Z)$ is Z-measurable, we can apply PR-Box 10.2 (vii). Hence,

$$\begin{aligned} E(Y|X{=}x) &= E^{X=x}(Y) && \text{[PR-(9.11)]} \\ &= E\big[E^{X=x}(Y)\big] && \text{[PR-Box 6.1 (i)]} \\ &= E\big[E^{X=x}(Y|C_X)\big] && \text{[(7.14), } P\text{-uniqueness of } E^{X=x}(Y|C_X)\text{]} \\ &= E(\tau_x). && \text{[(4.8)]} \end{aligned}$$

(ii) If $P(X{=}x, Z{=}z) > 0$, then $P(X{=}x) > 0$ and $P(Z{=}z) > 0$. Furthermore, according to Remark PR-14.37, $P(X{=}x, Z{=}z) > 0$ implies that $E^{X=x}(Y|Z{=}z)$ is identical to $E(Y|X{=}x, Z{=}z)$. Therefore, Equation (7.32) follows from (7.31). This proves (ii).

Proof of Theorem 7.11

First of all note that $P(X{=}x|C_X) \underset{P}{>} 0$ implies $P(X{=}x) > 0$, because

$$\begin{aligned} E[P(X{=}x|C_X)] &= E[E(1_{X=x}|C_X)] && \text{[PR-(10.4)]} \\ &= E(1_{X=x}) && \text{[PR-Box 10.2 (iv)]} \\ &= P(X{=}x). && \text{[PR-(6.5)]} \end{aligned}$$

Applying Theorem 3.52 (ii) then yields the proposition.

(i) $X \underset{P}{\perp\!\!\!\perp} C_X | Z \Rightarrow E^{X=x}(Y|Z) \underset{P}{=} E(\tau_x|Z)$:

$$\begin{aligned} E^{X=x}(Y|Z) &\underset{P^{X=x}}{=} E^{X=x}\big[E^{X=x}(Y|C_X)\,\big|\,Z\big] && \text{[PR-Box 10.2 (v)]} \\ &\underset{P^{X=x}}{=} E\big[E^{X=x}(Y|C_X)\,\big|\,Z\big] && \big[X \underset{P}{\perp\!\!\!\perp} C_X|Z,\ \text{PR-(16.39)}\big] \\ &\underset{P^{X=x}}{=} E(\tau_x|Z) && \text{[(4.8)]} \\ &\underset{P}{=} E(\tau_x|Z). && [P(X{=}x|C_X) \underset{P}{>} 0,\ \text{Cor. PR-14.48 (a), (c)}] \end{aligned}$$

$Y \vdash C_X | X, Z \Rightarrow E^{X=x}(Y|Z) \underset{P}{=} E(\tau_x|Z)$: Note that $P(X{=}x \mid C_X) \underset{P}{>} 0$ implies that $E^{X=x}(Y|C_X)$ is P-unique [see Cor. PR-14.48 (a) and (c)]. Furthermore, according to (7.15), $E^{X=x}(Y|Z)$ is a version of $E^{X=x}(Y|C_X)$. Therefore, P-uniqueness of $E^{X=x}(Y|C_X)$ implies that $E^{X=x}(Y|Z)$ is P-unique as well. If Y is nonnegative, then there is a version $E^{X=x}(Y|C_X)$ that is nonnegative [see PR-(10.17)], and if $E(Y) < \infty$, then $E[E^{X=x}(Y|C_X)] < \infty$. Therefore, and because $E^{X=x}(Y|Z)$ is Z-measurable, we can apply PR-Box 10.2 (vii). Hence,

$$
\begin{aligned}
E^{X=x}(Y|Z) &\underset{P}{=} E\big[E^{X=x}(Y|Z)\,\big|\,Z\big] && \big[\text{PR-Box 10.2 (vii)}\big] \\
&\underset{P}{=} E\big[E^{X=x}(Y|C_X)\,\big|\,Z\big] && [(7.15),\ P\text{-uniqueness of } E^{X=x}(Y|C_X)] \\
&\underset{P}{=} E(\tau_x|Z). && [(4.8)]
\end{aligned}
$$

(ii) If $P(X=x, Z=z) > 0$, then $P(X=x) > 0$ and $P(Z=z) > 0$. Furthermore, according to Remark PR-14.37, $P(X=x, Z=z) > 0$ implies that $E^{X=x}(Y|Z=z)$ is identical to $E(Y|X=x, Z=z)$. Therefore, Equation (7.32) follows from (7.31). This proves (ii).

Proof of Corollary 7.14

(i) This proposition immediately follows from Theorem 7.11, because

$$
\begin{aligned}
PFE_{xx';Z} &:= E^{X=x}(Y|Z) - E^{X=x'}(Y|Z) \\
&\underset{P}{=} E(\tau_x|Z) - E(\tau_{x'}|Z) && [\text{Th. 7.11 (i)}] \\
&\underset{P}{=} E(\delta_{xx'}|Z).
\end{aligned}
$$

(ii) If $P(X=x, Z=z) > 0$ and $P(X=x', Z=z) > 0$, then proposition (ii) of Theorem 7.11 implies $E^{X=x}(Y|Z=z) = E(\tau_x|Z=z)$ and $E^{X=x'}(Y|Z=z) = E(\tau_{x'}|Z=z)$. Hence,

$$
E^{X=x}(Y|Z=z) - E^{X=x'}(Y|Z=z) \;=\; E(\tau_x|Z=z) - E(\tau_{x'}|Z=z) \;=\; E(\delta_{xx'}|Z=z).
$$

Proof of Corollary 7.15

Corollary 7.15 immediately follows from Theorem 7.11. Also note that Corollary 7.10 is a special case of Corollary 7.15 for Z being a constant.

Proof of Theorem 7.16

First of all note that $P(X=x|C_{X,t}) \underset{P}{>} 0$ implies $P(X=x) > 0$. This can be shown as follows:

$$
\begin{aligned}
E[P(X=x|C_{X,t})] &= E[E(1_{X=x}|C_{X,t})] && [(10.4)] \\
&= E(1_{X=x}) && [\text{PR-Box 10.2 (iv)}] \\
&= P(X=x). && [\text{PR-(6.5)}]
\end{aligned}
$$

Applying Theorem 3.52 (ii) then yields the proposition.

(i) $X \underset{P}{\perp\!\!\!\perp} C_{X,t}|Z \Rightarrow E^{X=x}(Y|Z) \underset{P}{=} E(\tau_{x,t}|Z)$:

$$
\begin{aligned}
E^{X=x}(Y|Z) &\underset{P^{X=x}}{=} E^{X=x}\big[E^{X=x}(Y|C_{X,t})\,\big|\,Z\big] && [\text{PR-Box 10.2 (v)}] \\
&\underset{P^{X=x}}{=} E\big[E^{X=x}(Y|C_{X,t})\,\big|\,Z\big] && \big[X \underset{P}{\perp\!\!\!\perp} C_{X,t}|Z,\ \text{PR-(16.39)}\big] \\
&\underset{P^{X=x}}{=} E(\tau_{x,t}|Z) && [(5.14)] \\
&\underset{P}{=} E(\tau_{x,t}|Z). && [P(X=x|C_{X,t}) \underset{P}{>} 0,\ \text{Cor. PR-14.48 (a), (c)}]
\end{aligned}
$$

$Y \vdash C_{X,t} \,|\, X,Z \;\Rightarrow\; E^{X=x}(Y|Z) \underset{P}{=} E(\tau_{x,t}|Z)$: Note that $P(X{=}x \,|\, C_{X,t}) \underset{P}{>} 0$ implies that $E^{X=x}(Y|C_{X,t})$ is P-unique [see Cor. PR-14.48 (a) and (c)]. Furthermore, according to (7.20), $E^{X=x}(Y|Z)$ is a version of $E^{X=x}(Y|C_{X,t})$. Therefore, P-uniquenessof $E^{X=x}(Y|C_{X,t})$ implies that $E^{X=x}(Y|Z)$ is P-unique as well. If Y is nonnegative, then there is a version $E^{X=x}(Y|C_{X,t})$ that is nonnegative [see PR-(10.17)], and if $E(Y) < \infty$, then $E[E^{X=x}(Y|C_{X,t})] < \infty$. Therefore, and because $E^{X=x}(Y|Z)$ is Z-measurable, we can apply PR-Box 10.2 (vii). Hence,

$$\begin{aligned} E^{X=x}(Y|Z) &\underset{P}{=} E\big[E^{X=x}(Y|Z)\,\big|\,Z\big] && \big[\text{PR-Box 10.2 (vii)}\big] \\ &\underset{P}{=} E\big[E^{X=x}(Y|C_{X,t})\,\big|\,Z\big] && [(7.15),\ P\text{-uniqueness of } E^{X=x}(Y|C_{X,t})] \\ &\underset{P}{=} E(\tau_{x,t}|Z). && [(5.14)] \end{aligned}$$

(ii) If $P(X{=}x, Z{=}z) > 0$, then $P(X{=}x) > 0$ and $P(Z{=}z) > 0$. Furthermore, according to Remark PR-14.37, $P(X{=}x, Z{=}z) > 0$ implies that $E^{X=x}(Y|Z{=}z)$ is identical to $E(Y|X{=}x, Z{=}z)$. Therefore, Equation (7.34) follows from (7.33). This proves (ii).

Proof of Corollary 7.18

The proof is analogous to the proof of Corollary 7.14.

(i) This proposition immediately follows from Theorem 7.16, because

$$\begin{aligned} PFE_{xx';Z} &:= E^{X=x}(Y|Z) - E^{X=x'}(Y|Z) \\ &\underset{P}{=} E(\tau_{x,t}|Z) - E(\tau_{x',t}|Z) && [\text{Th. 7.16 (i)}] \\ &\underset{P}{=} E(\delta_{xx',t}|Z). \end{aligned}$$

(ii) If $P(X{=}x, Z{=}z) > 0$ and $P(X{=}x', Z{=}z) > 0$, then proposition (ii) of Theorem 7.16 implies $E^{X=x}(Y|Z{=}z) = E(\tau_{x,t}|Z{=}z)$ and $E^{X=x'}(Y|Z{=}z) = E(\tau_{x',t}|Z{=}z)$. Hence,

$$E^{X=x}(Y|Z{=}z) - E^{X=x'}(Y|Z{=}z) \;=\; E(\tau_{x,t}|Z{=}z) - E(\tau_{x',t}|Z{=}z) \;=\; E(\delta_{xx',t}|Z{=}z).$$

7.7 Exercises

▷ **Exercise 7-1** Proof that Equation (7.4) and C_X-measurability of Z imply Equation (7.6).

▷ **Exercise 7-2** Compute the conditional expectation values $E(Y|X{=}0)$ and $E(\tau_0)$ in the example presented in Table 6.3 (p. 130) and compare these numbers to each other.

▷ **Exercise 7-3** Assume that X is discrete with a finite number of values $x = 0, 1, \dots, J$, and that $P(X{=}x \,|\, C_{X,t}) \underset{P}{>} 0$ for all $x = 0, 1, \dots, J$. Which unbiasedness condition follows from $X \underset{P}{\perp\!\!\!\perp} C_X$ and which one from $X \underset{P}{\perp\!\!\!\perp} C_X \,|\, Z$?

▷ **Exercise 7-4** Assume that X is discrete with a finite number of values $x = 0, 1, \dots, J$, and that $P(X{=}x \,|\, C_{X,t}) \underset{P}{>} 0$ for all $x = 0, 1, \dots, J$. Which unbiasedness condition follows from $Y \vdash C_X \,|\, X$ and which one from $Y \vdash C_X \,|\, X, Z$?

▷ **Exercise 7-5** Describe randomized assignment of a unit to one of two treatment conditions!

▷ **Exercise 7-6** Describe *conditional* randomized assignment of a unit to one of two treatment conditions given a covariate!

▷ **Exercise 7-7** Show that $E[P(X{=}x \mid U)] = P(X{=}x)$.

▷ **Exercise 7-8** Show that $X \underset{P}{\perp\!\!\!\perp} U$ implies $X \underset{P}{\perp\!\!\!\perp} U \mid Z$, provided that Z is measurable with respect to U. Assume that X is discrete with a finite number of values $x = 0, 1, \ldots, J$ and that $P(X{=}x) > 0$ for all values x of X.

▷ **Exercise 7-9** Show that $X \underset{P}{\perp\!\!\!\perp} C_X$ implies:

$$P(X{=}x \mid W) \underset{P}{=} P(X{=}x) \qquad \text{for each } x = 0, 1, \ldots, J,$$

for each random variable W that is measurable with respect to C_X. Assume that X is discrete with a finite number of values $x = 0, 1, \ldots, J$ and that $P(X{=}x) > 0$ for all values x of X.

▷ **Exercise 7-10** Show that $Y \vdash C_X \mid X, Z$ implies $E(Y \mid X, Z, W) \underset{P}{=} E(Y \mid X, Z)$ if both Z and W are measurable with respect to the global covariate C_X.

Solutions

▷ **Solution 7-1** For $x = 0, 1, \ldots, J$,

$$\begin{aligned}
P(X{=}x \mid \pi_x) &\underset{P}{=} E(1_{X=x} \mid \pi_x) && \text{[PR-(10.4)]}\\
&\underset{P}{=} E[P(X{=}x \mid C_X) \mid \pi_x] && \text{[PR-Box 10.2 (v)]}\\
&\underset{P}{=} E[P(X{=}x \mid Z) \mid \pi_x] && \text{[(7.4)]}\\
&\underset{P}{=} E(\pi_x \mid \pi_x) && \text{[(7.5)]}\\
&\underset{P}{=} \pi_x && \text{[PR-Box 10.2 (vii)]}\\
&\underset{P}{=} P(X{=}x \mid Z) && \text{[(7.5)].}
\end{aligned}$$

Inserting $P(X{=}x \mid \pi_x)$ for $P(X{=}x \mid Z)$ in Equation (7.4) completes the proof.

▷ **Solution 7-2** The conditional expectation value $E(Y \mid X{=}0)$ can be computed via

$$E(Y \mid X{=}x) = \sum_u E(Y \mid X{=}x, U{=}u) \cdot P(U{=}u \mid X{=}x)$$

[see Rule (iv) in PR-Box 9.2]. In this example, this equation yields

$$\begin{aligned}
E(Y \mid X{=}0) &= \sum_u E(Y \mid X{=}0, U{=}u) \cdot P(U{=}u \mid X{=}0)\\
&= (68 + 78 + 88 + 98) \cdot \frac{1}{10} + (106 + 116) \cdot \frac{3}{10} = 99.8.
\end{aligned}$$

In contrast, the expectation $E(\tau_0)$ can be computed via

$$E(\tau_0) = \sum_u E(Y \mid X{=}x, U{=}u) \cdot P(U{=}u)$$
$$= (68+78+88+98+106+116) \cdot \frac{1}{6} = 92.3333.$$

Comparing $E(Y \mid X{=}0) = 99.8$ to $E(\tau_0) = 92.3333$ shows that $E(Y \mid X{=}0)$ is strongly biased.

▷ **Solution 7-3** The independent cause condition $X \underset{P}{\perp\!\!\!\perp} C_X$ implies that $E(Y \mid X)$ is unbiased with respect to total effects, whereas $X \underset{P}{\perp\!\!\!\perp} C_X \mid Z$ implies that $E(Y \mid X, Z)$ is unbiased with respect to total effects.

▷ **Solution 7-4** The mean-independent outcome condition $Y \vdash C_X \mid X$ implies that $E(Y \mid X)$ is unbiased with respect to total effects, whereas $Y \vdash C_X \mid X, Z$ implies that $E(Y \mid X, Z)$ is unbiased with respect to total effects.

▷ **Solution 7-5** In a random experiment, in which a unit u is sampled and assigned to one of two treatment conditions, we may assign the unit by coin toss, for instance. This makes sure that $P(X{=}1 \mid C_X) \underset{P}{=} P(X{=}1)$, i. e., that the treatment probabilities do not depend on any covariate. If the global covariate C_X is constructed such that U is measurable with respect to C_X, then $P(X{=}1 \mid C_X) \underset{P}{=} P(X{=}1)$ implies that each unit has the same probability $P(X{=}1 \mid U{=}u) = P(X{=}1)$ to be in treatment 1, and this implies that each unit has the same probability $P(X{=}0)$ to be assigned to treatment 0 as well.

▷ **Solution 7-6** In a random experiment, in which a unit u is sampled and a value z of the covariate Z is assessed before the unit is assigned to one of the treatment conditions, *conditional random assignment* of a unit to one of the two treatment conditions refers to assigning the unit u to treatment condition 1 with probability $P^{Z=z}(X{=}1 \mid C_X) \underset{P}{=} P^{Z=z}(X{=}1)$. Among other things, this makes sure that all units with the same value z of the covariate Z have the same probability to be assigned to treatment 1. This assignment procedure allows for different treatment probabilities for units for which we observe different values z of the covariate Z.

▷ **Solution 7-7** By definition of the expectation,

$$E[P(X{=}x \mid U)] = \sum_u P(X{=}x \mid U{=}u) \cdot P(U{=}u).$$

Because, according to the theorem of total probability

$$P(X{=}x) = \sum_u P(X{=}x \mid U{=}u) \cdot P(U{=}u)$$

(see Th. PR-4.25), this shows that $E[P(X{=}x \mid U)] = P(X{=}x)$. Note that $E[P(X{=}x \mid U)] = P(X{=}x)$ is also a special case of Rule (iv) in PR-Box 10.2, if we use $P(X{=}x \mid U) := E(1_{X=x} \mid U)$, where $1_{X=x}$ denotes the indicator variable indicating with values 1 and 0 whether or not X takes on the value x.

▷ **Solution 7-8** If $P(X{=}x) > 0$ for all values x of X, then the condition $U \underset{P}{\perp\!\!\!\perp} X$ is equivalent to: $P(X{=}x \mid U) \underset{P}{=} P(X{=}x)$, for each $x = 0, 1, \ldots, J$. If Z is measurable with respect to U, then this equation implies

$$P(X{=}x \mid U, Z) \underset{P}{=} P(X{=}x \mid U) \underset{P}{=} P(X{=}x).$$

Using this equation as well as $P(X{=}x \mid Z) := E(1_{X=x} \mid Z)$ and $E(1_{X=x}) = P(X{=}x)$:

$$
\begin{aligned}
E(\mathit{1}_{X=x}\,|\,Z) &\underset{P}{=} E\big[E(\mathit{1}_{X=x}\,|\,U,Z)\,\big|\,Z\big] && \text{[PR-Box 10.2, (v)]}\\
&\underset{P}{=} E\big[E(\mathit{1}_{X=x}\,|\,U)\,\big|\,Z\big] && [Z \text{ is measurable w.r.t. } U]\\
&\underset{P}{=} E\big[E(\mathit{1}_{X=x})\,\big|\,Z\big] && [U \underset{P}{\perp\!\!\!\perp} X, \text{PR-Box 10.2, (vi)}]\\
&\underset{P}{=} E(\mathit{1}_{X=x}) && \text{[PR-Box 10.2, (i)]}\\
&= P(X{=}x),
\end{aligned}
$$

for each $x = 0, 1, \dots, J$. Hence, $P(X{=}x|U,Z) \underset{P}{=} P(X{=}x|Z)$, for each $x = 0, 1, \dots, J$. This equation is equivalent to $U \underset{P}{\perp\!\!\!\perp} X\,|\,Z$ if X is discrete with a finite number of values $x = 0, 1, \dots, J$.

▷ **Solution 7-9** Note that $P(X{=}x) = E(\mathit{1}_{X=x})$ and $P(X{=}x\,|\,W) := E(\mathit{1}_{X=x}\,|\,W)$. Hence,

$$
\begin{aligned}
E(\mathit{1}_{X=x}\,|\,W) &\underset{P}{=} E\big[E(\mathit{1}_{X=x}\,|\,C_X)\,\big|\,W\big] && \text{[PR-Box 10.2, (v)]}\\
&\underset{P}{=} E\big[P(X{=}x\,|\,C_X)\,\big|\,W\big] && [(10.3)]\\
&\underset{P}{=} E[P(X{=}x)\,|\,W] && [X \underset{P}{\perp\!\!\!\perp} C_X]\\
&\underset{P}{=} P(X{=}x) && \text{[PR-Box 10.2, (i)]}\\
&= E(\mathit{1}_{X=x}),
\end{aligned}
$$

for each $x = 0, 1, \dots, J$. Because $P(X{=}x) = E(\mathit{1}_{X=x})$ and $P(X{=}x\,|\,W) := E(\mathit{1}_{X=x}\,|\,W)$, this proves the proposition.

▷ **Solution 7-10**

$$
\begin{aligned}
E(Y\,|\,X,Z,W) &\underset{P}{=} E\big[E(Y\,|\,X,C_X)\,\big|\,X,Z,W\big] && \text{[PR-Box 10.2, (v)]}\\
&\underset{P}{=} E\big[\,E(Y\,|\,X,Z)\,\big|\,X,Z,W\big] && [Y \vdash C_X\,|\,X,Z]\\
&\underset{P}{=} E(Y\,|\,X,Z) && \text{[Box PR-10.2, (vii)]}.
\end{aligned}
$$

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 non-compliance. *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.