# Causal Effects Xplorer (CEX)

Rolf Steyer, Benjamin Nagengast,
Ulf Kröhne, and Marcel Bauer

February 23, 2018

## Contents

# 1 Introduction

In this text we present and explain the most important concepts, theorems, and formulas of the theory of causal total effects in a somewhat simplified version. We presume that there is only one single discrete treatment (or intervention) variable $X$ and that there is no *fallible* covariate of $X$. In this case, the observational-unit variable (or person variable) $U$ is a *global potential confounder of* $X$, and this global potential confounder is discrete. The observational-unit variable being discrete implies that all *potential confounders of* $X$ are mappings of $U$ (see Steyer & Nagel, 2017, Corollary 2.53). The general version of the theory of causal effects is presented in Steyer, 2018, where you also find the definitions of the concepts 'potential confounder of $X$', 'covariate of $X$' and 'global potential confounder of $X$'.

The CEX is a learning tool that facilitates getting familiar with the theory. It allows to study examples by specifying input parameters that fix the joint distribution of $X$ and $U$ and the $(X,U)$-conditional expectation of the outcome (or response) variable $Y$. The CEX refers to a random experiment consisting of three subexperiments:

(a) Sampling a unit $u$ from a set of units (persons). (Such a unit is a value of the observational-unit variable $U$ mentioned above.)

(b) Assigning or observing assignment of the unit to one of at least two treatments represented by the values $0, 1, \ldots, J$ of the treatment variable $X$.

(c) Observing a value $y$ of the outcome (response) variable $Y$.

Such a random experiment is the kind of empirical phenomenon that is represented by a probability space $(\Omega, \mathscr{A}, P)$, to which statements about probabilities, expectations, variances, covariances, correlations, conditional expectations, and causal effects refer. Hence, in the examples considered in the CEX, an element of the set $\Omega$ of possible outcomes of such a random experiment is

$$\omega = (u, \textit{treatment}, \textit{response}).$$

The observational unit-variable $U$ then assigns to $\omega$ the value $u$, the observational unit that is sampled. The treatment variable $X$ assigns to $\omega$ one of its values $0, 1, \ldots, J$, depending only on its second component, the *treatment*. Similarly, the outcome variable $Y$ assigns to $\omega$ one of its values, depending only on its third component (the *response*).

Once the input parameters are fixed by the user, the program computes all important parameters, including conditional (see section 4.4) and average causal total effects (see section 4.2) of the treatment variable, and it also checks various causality conditions (see section 6). Furthermore, the CEX can also be used to generate a data sample from the distribution that is specified by the input parameters. For most purposes, the options of the program are self-explanatory. Clicking on the table, opens specific menues, which again are self-explanatory. Just try and play! For the more complete and general theory of causal effects see Steyer, 2018 (in the sequel, abbreviated by RS), and for the probability theory terms see Steyer & Nagel, 2017 (in the sequel, abbreviated by SN).

# 2 Basic concepts and assumptions

## 2.1 Observational-unit (person) variable $U$

The observational-unit (or person) variable is denoted by $U$. This random variable takes on the values $u_1, \dots, u_N$, where $N \geq 1$ is the number of units in the set of units (the "population") from which one unit is drawn (see section 2). This number can be specified by a right click on a unit $u_i$ in the first column of the table. This opens a menue containing the options "Insert new unit before $u_i$" and "Insert new unit after $u_i$".

## 2.2 Sampling probabilities $P(U{=}u_i)$

The first part of the random experiment considered (see section 1) consists of sampling a unit $u_i$ from a set of $N$ units. We assume that all these sampling probabilities $P(U{=}u_i)$ are positive, that is,

$$P(U{=}u_i) \ > \ 0, \quad \text{for all } u_i \in \Omega_U = \{u_1, \dots, u_N\}. \tag{1}$$

These probabilities may be called *sampling probabilities*. They have to be specified by the user. The sum of these sampling probabilities $P(U{=}u_i)$ over all units must be equal to 1. These probabilities need not necessarily be identical for all units.

## 2.3 Treatment variable $X$

We assume that the treatment variable $X$ is discrete with values $x = 0, 1, \dots, J$, where $J \geq 2$. Each of these values must have a positive probability, that is, we assume

$$P(X{=}x) \ > \ 0, \quad \text{for all } x = 0, 1, \dots, J. \tag{2}$$

The probabilities $P(X{=}x)$ are specified indirectly when the user specifies the individual treatment probabilities $P(X{=}x | U{=}u)$ (see section 2.4).

## 2.4 Individual treatment probabilities

The $(U{=}u)$-conditional probabilities

$$P(X{=}x|U{=}u) \ = \ \frac{P(X{=}x, U{=}u)}{P(U{=}u)} \tag{3}$$

are also called the *individual treatment probabilities.* They have to be specified by the user. For simplicity, we assume

$$P(X=x|U=u) \;>\; 0, \quad \text{for all } (x,u) \tag{4}$$

in this introduction.

Note that

$$P(X=0|U=u) \;=\; 1-\sum_{x=1}^{J} P(X=x|U=u). \tag{5}$$

Hence, the user only has to specify the conditional probabilities $P(X=x|U=u)$ for $x = 1,\dots,J$ and all units $u$. This allows the CEX to compute the conditional probabilities $P(X=0|U=u)$ for all units.

If, for example, the conditional probabilities $P(X=x|U=u)$ are specified such that

$$P(X=x|U=u) \;=\; P(X=x), \quad \text{for all } (x,u), \tag{6}$$

then this corresponds to an *experiment with randomized assignment* of the unit to one of the treatment conditions. In other words, in a randomized experiment, the probability of being assigned to treatment $x$ is identical for all units, and this applies to all treatments. However, treatment probabilities may differ between treatments, that is, $P(X=x)$ may differ from $P(X=x')$ for different treatments $x$ and $x'$. If Equation (6) does not hold, then at least one of the treatment probabilities depends on the units and one or more of their attributes, such as their severity of symptoms, and this may induce bias. This bias is also computed by the CEX (see section 5.2).

## 2.5 Covariate $Z$

In the CEX we can also consider a covariate of $X$, which is denoted by $Z$. In the CEX, we presume that $Z$ is a mapping of $U$ so that it can be written as the composition $f(U)$ of $U$ and a mapping $f$. This implies that each observational unit has one and only one value $z$ of $Z$ and that several units can share the same value $z$ of $Z$. A simple example of such a covariate $Z$ is *sex.*

If no covariate is desired, then the corresponding column should be filled with a fixed constant, for example, the constant 1. Because we assume that $Z = f(U)$ is a mapping of $U$, the assumption $P(U=u) > 0$, for all $u$, implies

$$P(Z=z) \;>\; 0, \quad \text{for all values } z \text{ of } Z \tag{7}$$

(see section 2.5). For the same reason, specifying the probabilities $P(U=u)$ for all units, the user indirecty specify the probabilities $P(Z=z)$ for all values of the covariate $Z$, because

$$P(Z=z) \;=\; \sum_{u:\, f(u)=z} P(U=u), \tag{8}$$

that is, because the probability $P(Z{=}z)$ is the sum of the probabilities $P(U{=}u)$ over all values $u$ of $U$ for which $Z = f(U)$ takes on the value $f(u) = z$.

## 2.6 Outcome (response) variable $Y$

The outcome (or response) variable is denoted by $Y$. It has to be real-valued with finite expectation and finite variance. It can be continuous, discrete, or even binary with values 0 and 1. Under the assumption (4), a conditional expectation value $E(Y|X{=}x, U{=}u)$ is uniquely defined by

$$E(Y|X{=}x, U{=}u) := \int Y \, dP^{X=x,U=u} \tag{9}$$

(see SN-Def. 9.2), where $P^{X=x,U=u}$ denotes the $(X{=}x, U{=}u)$-conditional probability measure $P^{X=x,U=u}: \mathscr{A} \to [0,1]$ defined by

$$P^{X=x,U=u}(A) = P(A|X{=}x, U{=}u), \quad \text{for all } A \in \mathscr{A}\,. \tag{10}$$

If $Y$ is discrete and the assumption (4) holds, then Equation (9) simplifies to

$$E(Y|X{=}x, U{=}u) = \sum_y y \cdot P(Y{=}y|X{=}x, U{=}u). \tag{11}$$

Hence, if $Y$ is binary with $y = 0, 1$, then

$$E(Y|X{=}x, U{=}u) = P(Y{=}1|X{=}x, U{=}u). \tag{12}$$

In the CEX, the conditional expectation values $E(Y|X{=}x, U{=}u)$ have to be specified by the user for all pairs of values of $X$ and $U$ (see section 2.7). Neither the values $y$ of $Y$ nor the $(X{=}x, U{=}u)$-conditional distributions of $Y$ need to be known or even be specified. In contrast to Rubin, we do not assume that the values of $Y$ itself are fixed for a given pair $(x, u)$ of values of $X$ and $U$. Instead, only the conditional expectation values $E(Y|X{=}x, U{=}u)$ are fixed (and have to specified in the CEX). In this respect, the theory of causal effects is a generalization of Rubin's 'potential outcome approach' to causality (see, e.g., Rubin, 2005).

## 2.7 True-outcome variables

Let $E(Y|X, U)$ denote the $(X, U)$-conditional expectation of $Y$ in the single-unit trial described in section 1. The definition and the properties of such a conditional expectation are provided in SN-chapters 9, 10, and 11. In the theory of causal effects, the conditional expectation value $E(Y|X{=}x, U{=}u)$ is called the *true outcome of unit $u$ given treatment $x$*.

For $x = 0, 1, \ldots, J$, we define the *partial* $(X{=}x, U)$*-conditional expectation of* $Y$ by

$$E(Y|X{=}x, U)(\omega) \;=\; E(Y|X{=}x, U{=}u), \quad \text{if } \omega \in \{U{=}u\}, \tag{13}$$

where $\{U{=}u\} = \{\omega \in \Omega : U(\omega) = u\}$ denotes the event that unit $u$ is sampled (see again the random experiment described in section 1). Because $X$ is fixed to one of its values $x$, $E(Y|X{=}x, U)$ is a function of the observational-unit variable $U$. Assuming $P(X{=}x) > 0$ implies

$$E(Y|X{=}x, U) \;=\; E^{X=x}(Y|U), \tag{14}$$

where $E^{X=x}(Y|U)$ denotes the $U$-conditional expectation of $Y$ with respect to the conditional-probability measure $P^{X=x} : \mathscr{A} \to [0,1]$ defined by

$$P^{X=x}(A) \;=\; P(A|X{=}x, U{=}u), \quad \text{for all } A \in \mathscr{A}. \tag{15}$$

For the general definition and the properties of such a partial conditional expectation $E(Y|X{=}x, U)$ see SN-chapter 14.)

Under the assumption (made in the CEX) that $U$ is a global potential confounder, the partial conditional expectation $E(Y|X{=}x, U)$ can now be used to define

$$\tau_x \;:=\; E(Y|X{=}x, U), \tag{16}$$

the *true-outcome variable of* $Y$ *given treatment* $x$, $x = 0, 1, \ldots, J$. These true-outcome variables replace the potential outcome variables used in Rubin's approach. According to Equations (13) and (16), the values of the true-outcome variable $\tau_x$ are the conditional expectation values $E(Y|X{=}x, U{=}u)$. More precisely, for all values $x$ of $X$,

$$\tau_x(\omega) \;=\; E(Y|X{=}x, U)(\omega) \;=\; E(Y|X{=}x, U{=}u), \quad \text{if } \omega \in \{U{=}u\}. \tag{17}$$

If $Y$ is a binary outcome variable, then its true-outcome variable $\tau_x$ is also the partial $(X{=}x, U)$-conditional probability of the event $\{Y{=}1\}$ (see SN-Remark 14.30), that is, if $Y$ is binary, then, for $x = 0, 1, \ldots, J$,

$$\tau_x \;=\; P(Y{=}1|X{=}x, U), \tag{18}$$

and Equation (17) can also be written as

$$\tau_x(\omega) \;=\; P(Y{=}1|X{=}x, U)(\omega) \;=\; P(Y{=}1|X{=}x, U{=}u), \quad \text{if } \omega \in \{U{=}u\}. \tag{19}$$

## 2.8 Individual total effects

The difference

$$E(Y|X{=}x, U{=}u) \;-\; E(Y|X{=}0, U{=}u)$$

between the true outcomes under treatment $x$ and under treatment 0 for the same observational unit $u$ is called the *individual total effect of treatment $x$ compared to treatment 0 of unit $u$*. These individual effects are the values of the random variable

$$\delta_{x0} := \tau_x - \tau_0 \tag{20}$$

More precisely, for all $x$, $x \neq 0$,

$$\delta_{x0}(\omega) = E(Y|X=x, U=u) - E(Y|X=0, U=u), \quad \text{if } \omega \in \{U=u\}. \tag{21}$$

The individual total effects are computed by the CEX, once the user has specified the true outcomes for all treatments and all units. They can be inspected on the main screen of the CEX, the same screen on which the user specifies the individual true outcomes.

## 2.9 Conditional probabilities $P(U=u|X=x)$

The conditional probabilities

$$P(U=u|X=x) = \frac{P(X=x|U=u) \cdot P(U=u)}{P(X=x)} \tag{22}$$

are computed by the CEX from the individual treatment probabilities $P(X=x|U=u)$ and the sampling probabilities $P(U=u)$, both of which are specified by the user. These probabilities are also shown on the main screen.

The conditional probabilities $P(U=u|X=x)$ may be called *sampling probabilities given treatment $x$*. If they do not depend on $x$, that is, if

$$P(U=u|X=x) = P(U=u), \quad \text{for all } (x, u), \tag{23}$$

then there is *representative sampling* in each treatment condition $x$. Equation (23) implies that the distributions of all attributes of the units, that is, of all covariates $Z = f(U)$ (i.e., of all mappings of $U$) are identical between treatment conditions. In this sense, the treatments are 'comparable'. In other words, if Equation (23) holds, then we can compare the conditional expectation values $E(Y|X=x)$ and $E(Y|X=x')$ to each other in order make causal inferences, and the same applies to the conditional expectation values $E(Y|X=x, Z=z)$ and $E(Y|X=x', Z=z)$.

## 2.10 Summary of all assumptions made in the CEX

For convenience, now we summarize all assumptions made in the CEX.

(i) The number $N$ of units in the set $\Omega_U = \{u_1, \ldots, u_N\}$ of units, from which one unit is sampled, is greater than or equal to 1, and each unit $u$ in this set of units has a positive probability $P(U=u) > 0$ to be drawn.

(ii) The number $J+1$ of treatment conditions is greater than or equal to 2.

(iii) The covariate $Z$ is a mapping of $U$, that is, there is a mapping $f$ such that $Z = f(U)$. In more precise terms, this means that $Z$ is a composite mapping of $U$ and a mapping $f$ that assigns to each observational unit $u$ a value $f(u)$ in the co-domain of $Z$. This co-domain does not have to be a subset of the real numbers. The covariate *sex* with values *male* and *female* is an example in case.

(iv) The outcome variable $Y$ is assumed to be real-valued with finite expectation and variance. It can be continuous, discrete, or even binary with values 0 and 1.

(v) The joint probabilities $P(X{=}x, U{=}u)$ are positive for all pairs $(x,u)$ of values of $X$ and $U$.

Note that the assumptions in the general theory of causal effects are less restrictive. However, the assumptions (i) to (v) are made in the CEX.

# 3 Unbiasedness of conditional expectations and their values

## 3.1 Unbiasedness of $E(Y|X)$

Let the assumptions specified in section 2.10 hold. Then:

(i) The conditional expectation value $E(Y|X{=}x)$ is called *unbiased* if

$$E(Y|X{=}x) \;=\; E(\tau_x) \;=\; E\big(E(Y|X{=}x,U)\big). \tag{24}$$

(ii) The conditional expectation $E(Y|X)$ is called *unbiased* (with respect to $U$) if Equation (24) holds for all $x = 0, 1, \ldots, J$.

## 3.2 Unbiasedness of $E(Y|X,Z)$

Now we extend the concept of unbiasedness to conditioning on a covariate of $X$, that is, on a random variable that is a mapping of $U$.

Let the assumptions specified in section 2.10 hold and let $Z$ be a random variable such that there is a mapping $f$ satisfying $Z = f(U)$.

(i) The partial conditional expectation $E(Y|X{=}x, Z)$ is called *unbiased* if

$$E(Y|X{=}x,Z) \;=\; E(\tau_x\,|\,Z) \;=\; E\big(E(Y|X{=}x,U)\,\big|\,Z\big). \tag{25}$$

(ii) The partial conditional expectation $E(Y|X, Z{=}z)$ is called *unbiased* if

$$E(Y|X{=}x, Z{=}z) \;=\; E(\tau_x \,|\, Z{=}z) \;=\; E\big(E(Y|X{=}x, U)\,\big|\, Z{=}z\big) \tag{26}$$

holds for all $x = 0, 1, \ldots, J$.

(iii) The conditional expectation $E(Y|X, Z)$ is called *unbiased* if Equation (25) holds for all $x = 0, 1, \ldots, J$.

(iv) The conditional expectation value $E(Y|X{=}x, Z{=}z)$ is called *unbiased* if

$$E(Y|X{=}x, Z{=}z) \;=\; E(\tau_x \,|\, Z{=}z). \tag{27}$$

In these definitions, remember that we are talking about bias *with respect to total effects*, because $U$ is global potential confounder only if we consider *total-effects*, and not direct effects, for example. Second, in the definition of unbiasedness of $E(Y|X, Z)$, it is important to know that we consider $X$ to be the cause and $Z$ the covariate of $X$, and in many cases the variables cannot change their roles. Remember that $X$ — and not $Z = f(U)$ — takes the role of a cause.

The CEX checks whether or not unbiasedness of $E(Y|X, Z)$ holds. If it holds, a green check mark appears in the table of causality conditions on the main screen. The CEX also checks if $E(Y|X, Z{=}z)$ is unbiased. If yes, a green check mark appears in the table of causality conditions on the screen for the $(Z{=}z)$-conditional effects.

# 4 Causal total effects

## 4.1 Expectation of a true-outcome variable

The true-outcome variable $\tau_x$ has been defined in Equation (16). Using Equation (17) for the values of $\tau_x$ yields

$$E(\tau_x) \;=\; E\big(E(Y|X{=}x, U)\big) \;=\; \sum_u E(Y|X{=}x, U{=}u) \cdot P(U{=}u). \tag{28}$$

This expectation is identical to the conditional expectation value of $Y$ given treatment $x$ in a randomized experiment, that is, in an experiment in which Equation (6) holds. The expectation of $\tau_x$ is also identical to the conditional expectation value of $Y$ given treatment $x$ for a representatively sampled observational unit, that is, for a unit sampled in an experiment in which Equation (23) holds.

If $Y$ is binary, then

$$E(\tau_x) \;=\; E\big(P(Y{=}1|X{=}x, U)\big) \;=\; \sum_u P(Y{=}1|X{=}x, U{=}u) \cdot P(U{=}u) \tag{29}$$

is identical to the $(X{=}x)$-conditional probability of $Y{=}1$ that would be estimated in a randomized experiment. In this sense, this number reveals the *causally unbiased expected outcome of $Y$ given treatment $x$*, which is useful if no information about the observational unit is available.

## 4.2 Average causal total effect

Let $\tau_x = E(Y|X{=}x, U)$, $x \in \{0, 1, \ldots, J\}$, denote the true-outcome variables given $x$ and assume $P(X{=}x|U)$, $P(X{=}0|U) > 0$. Then

$$ATE_{x0} := E(\tau_x - \tau_0) = E(\tau_x) - E(\tau_0) \tag{30}$$

is defined to be the *average causal total effect of treatment $x$ compared to treatment* 0 *on the (expectation of the) outcome variable $Y$*.

Note that

$$ATE_{x0} = \sum_u \big(E(Y|X{=}x, U{=}u) - E(Y|X{=}0, U{=}u)\big) \cdot P(U{=}u). \tag{31}$$

This equation can be used for the computation of the $ATE_{x0}$ from the true outcomes for all pairs of units and treatment $x$, which the user has to specify in the CEX.

## 4.3 Identification of the average total treatment effect

If $E(Y|X{=}x, Z)$ and $E(Y|X{=}0, Z)$ are unbiased, then

$$\begin{aligned}
&E\big(E(Y|X{=}x, Z)\big) - E\big(E(Y|X{=}0, Z)\big) \\
&= E\big(E(\tau_x|Z)\big) - E\big(E(\tau_0|Z)\big) && \text{[Eq. (25)]} \\
&= E(\tau_x) - E(\tau_0) && \text{[SN-Box 10.2 (iv)]} \\
&= ATE_{x0}.
\end{aligned} \tag{32}$$

Hence, the average causal total effect of $x$ compared to 0 can be identified (computed) from the expectations $E\big(E(Y|X{=}x, Z)\big)$ and $E\big(E(Y|X{=}0, Z)\big)$, provided that $E(Y|X{=}x, Z)$ and $E(Y|X{=}0, Z)$ are unbiased. Note that these partial conditional expectations and their expectations can be estimated in a data sample of $(X, Y, Z)$.

## 4.4 Conditional expectation value of a true-outcome variable

Let the assumptions of section 2.10 hold and assume that $V = f(U)$ is a covariate of $X$. Then the following equation holds for the $(V{=}v)$-conditional expectation value of a true-outcome variable $\tau_x$:

$$\begin{aligned}
E(\tau_x|V{=}v) &= E\big(E(Y|X{=}x, U) \,\big|\, V{=}v\big) \\
&= \sum_u E(Y|X{=}x, U{=}u) \cdot P(U{=}u|V{=}v).
\end{aligned} \tag{33}$$

Such a conditional expectation value is also called the *causally unbiased* $(V{=}v)$-*conditional expectation value of* $Y$ *in treatment* $x$. The conditional expectation value $E(\tau_x|V{=}v)$ is identical to the expectation of the outcome $Y$ under treatment $x$ for an observational unit with score $v$ on the covariate $V$ in an experiment in which

$$P(X{=}x|U) \ = \ P(X{=}x|V), \quad \text{for all } x, \tag{34}$$

holds. Because we assume $V = f(U)$, we can write $P(X{=}x|U) = P(X{=}x|U,V)$. This implies

$$P(X{=}x|U,V) \ = \ P(X{=}x|V), \quad \text{for all } x, \tag{35}$$

and this equation is equivalent to $V$-conditional independence of $X$ and $U$. Hence, $E(\tau_x|V{=}v)$ is identical to the expectation of $Y$ given treatment $x$, which would be estimated in a *$V$-conditionally randomized experiment*, that is, in an experiment in which Equation (35) holds. In still other words, $E(\tau_x|V{=}v)$ reveals the causally unbiased expected outcome of $Y$ under treatment $x$ given the value $v = f(u)$ of the covariate $V$ for an observational unit $u$.

Again, if $Y$ is binary, then

$$\begin{aligned} E(\tau_x|V{=}v) \ &= \ E\big(P(Y{=}1|X{=}x,U)\,\big|\,V{=}v\big) \\ &= \ \sum_u P(Y{=}1|X{=}x,U{=}u)\cdot P(U{=}u|V{=}v). \end{aligned} \tag{36}$$

In this case, $E(\tau_x|V{=}v)$ is identical to the $(X{=}x, V{=}v)$-conditional probability of $Y{=}1$ that would be estimated in a $V$-conditionally randomized experiment, that is, in a experiment, in which Equation (35) holds. Hence, $E(\tau_x|V{=}v)$ reveals the *causally unbiased probability of* $Y{=}1$ *given* $V{=}v$ *and treatment* $X{=}x$ if no other information than $v$ about the observational unit is available.

## 4.5 Conditional causal total effect

Now we turn to the definition of a conditional causal total effect. In this definition, we refer to a mapping $Z$ of $U$ with respect to which we assume unbiasedness. Additionally, we refer to a mapping $V$ of $Z$ — which therefore is also a mapping of $U$ — with respect to which we define a conditional effect. Because $V$ can also be identical to $Z$ we only need a single definition in order to define a conditional total effect for any mapping of $Z$, including $Z$ itself.

Let the assumption of section 2.10 hold and assume that $V$ is a function of $U$. Then the *conditional causal (total) effect* of the treatment $x$ compared to treatment 0 on the outcome variable $Y$ *given the value* $v$ *of a covariate* $V$ of $X$ is defined by

$$CTE_{V;x0}(v) \ := \ E(\tau_x|V{=}v) - E(\tau_0|V{=}v). \tag{37}$$

This equation defines the $(V{=}v)$-the conditional causal total effect. Equation (38) shows how the conditional causal total effect can be identified from empirically estimable parameters.

Let the assumption of section 2.10 hold and assume:

(a) $Z$ is a mapping of $U$

(b) $V$ is a mapping of $Z$

(c) $E(Y|X{=}x,Z)$ and $E(Y|X{=}0,Z)$ are unbiased.

(d) $P(V{=}v) > 0$.

Then the conditional causal total effect $CTE_{V;x0}(v)$ can identified as follows:

$$CTE_{V;x0}(v) := E\big(E(Y|X{=}x,Z)\,\big|\,V{=}v\big) - E\big(E(Y|X{=}0,Z)\,\big|\,V{=}v\big). \tag{38}$$

Hence, if $E(Y|X{=}x,Z)$ and $E(Y|X{=}0,Z)$ are unbiased and $V$ is a mapping of $Z$, then the $(V{=}v)$-conditional causal total effect of $x$ compared to 0 can be computed from the $(V{=}v)$-conditional expectation values $E\big(E(Y|X{=}x,Z)\,\big|\,V{=}v\big)$ and $E\big(E(Y|X{=}0,Z)\,\big|\,V{=}v\big)$. These values can be estimated in a data sample of $(X,Y,Z)$. (Remember, because $V = f(Z)$, the random variable $V$ can be computed from $Z$.)

If $V = Z$, then Equation (38) simplifies to

$$\begin{aligned} CTE_{Z;x0}(z) &= E\big(E(Y|X{=}x,Z)\,\big|\,Z{=}z\big) - E\big(E(Y|X{=}0,Z)\,\big|\,Z{=}z\big) \\ &= E(Y|X{=}x,Z{=}z) - E(Y|X{=}0,Z{=}z). \end{aligned} \tag{39}$$

If $Z$ is discrete, then

$$\begin{aligned} &CTE_{V;x0}(v) \\ &= E\big(E(Y|X{=}x,Z)\,\big|\,V{=}v\big) - E\big(E(Y|X{=}0,Z)\,\big|\,V{=}v\big) \\ &= \sum_z E(Y|X{=}x,Z{=}z)\cdot P(Z{=}z|V{=}v) - \sum_z E(Y|X{=}0,Z{=}z)\cdot P(Z{=}z|V{=}v) \\ &= \sum_z \big(E(Y|X{=}x,Z{=}z) - E(Y|X{=}0,Z{=}z)\big)\cdot P(Z{=}z|V{=}v). \end{aligned} \tag{40}$$

Under the assumptions of summarized in section 2.10, the partial conditional expectations $E(Y|X{=}x,U)$ *are always unbiased*. Therefore, in the CEX the $CTE_{V;x0}(v)$ can be computed from the true outcomes by

$$CTE_{V;x0}(v) = \sum_u \big(E(Y|X{=}x,U{=}u) - E(Y|X{=}0,U{=}u)\big)\cdot P(U{=}u|V{=}v). \tag{41}$$

## 4.6 Conditional expectation value of the outcome variable

In section 4.1 we showed how to compute the expectation of a true-outcome variable $\tau_x$ and in section 4.4 we showed how to compute its $(V{=}v)$-conditional expectation value. Now we turn to the $(X{=}x)$-conditional expectation value of the outcome variable $Y$ itself, and show how $E(Y|X{=}x)$ can also be computed from the true outcomes. If $P(X{=}x) > 0$, then we define

$$E(Y|X{=}x) := \int Y dP^{X=x}. \tag{42}$$

where $P^{X=x}$ denotes the $(X{=}x)$-conditional probability measure on $\mathscr{A}$ [see Eq. (15)]. If $Y$ is discrete, then

$$E(Y|X{=}x) = \sum_y y \cdot P(Y{=}y|X{=}x). \tag{43}$$

However, $E(Y|X{=}x)$ can also be computed from the true outcomes. The formula is

$$E(Y|X{=}x) = \sum_u E(Y|X{=}x, U{=}u) \cdot P(U{=}u|X{=}x) \tag{44}$$

[see SN-Box 9.2 (ii)]. If $Y$ is binary, this equation can also be written

$$P(Y{=}1|X{=}x) = \sum_u P(Y{=}1|X{=}x, U{=}u) \cdot P(U{=}u|X{=}x). \tag{45}$$

It is worthwhile comparing Equations (28) and (44) to each other. This comparison reveals that $E(Y|X{=}x) = E(\tau_x)$, if

(a) $P(U{=}u|X{=}x) = P(U{=}u)$, for all $u$,

or

(b) $E(Y|X{=}x, U{=}u) = E(Y|X{=}x)$, for all $u$.

This is easily seen in the CEX if you specify the input parameters such that (a) or (b) holds. A mathematical proof is also not difficult. Note, however, that in the CEX condition (a) can be specified only indirectly via specifying

$$P(X{=}x|U{=}u) = P(X{=}x), \quad \text{for all } u. \tag{46}$$

The conditional probabilities

$$P(U{=}u|X{=}x) = P(X{=}x|U{=}u) \cdot \frac{P(U{=}u)}{P(X{=}x)} \tag{47}$$

referred to in (a) are then computed by the CEX.

Each of the conditions (a) and (b) implies *unbiasedness of* $E(Y|X{=}x)$ (see section 3.1). If (a) holds *for all* values $x$ of $X$, then (a) is equivalent to *independence of* $U$

*and X*. In the CEX, condition (b) is called *unit-treatment homogeneity,* referring to the fact that all units respond homogeneously (i.e., in the same way) to treatment $x$ with respect to the (expectation of the) outcome variable $Y$. Each of these conditions is checked by the CEX. If they hold, then a green check mark appears in the corresponding position in the table of causality conditions.

## 4.7 Conditional expectation value of the outcome variable given a treatment and a value of the covariate

If $P(X{=}x, Z{=}z) > 0$, then we define the $(X{=}x, Z{=}z)$-conditional expectation value of a real-valued random variable $Y$ by

$$E(Y|X{=}x, Z{=}z) \;:=\; \int Y \, dP^{X=x,\,Z=z}, \tag{48}$$

where $P^{X=x,\,Z=z}$ denotes the $(X{=}x, Z{=}z)$-conditional probability measure on $\mathscr{A}$ [see Eq. (10)]. If $Y$ is discrete, then its conditional expectation value given treatment $x$ and a value $z$ of the covariate $Z$ can be computed from the values $y$ of $Y$ and their $(X{=}x, Z{=}z)$-conditional probabilities via

$$E(Y|X{=}x, Z{=}z) \;:=\; \sum_y y \cdot P(Y{=}y|X{=}x, Z{=}z). \tag{49}$$

Under the assumptions specified in section 2.10, these conditional expectation values can also be computed from the true outcomes of the observational units using

$$E(Y|X{=}x, Z{=}z) \;=\; \sum_u E(Y|X{=}x, U{=}u) \cdot P(U{=}u|X{=}x, Z{=}z). \tag{50}$$

This equation is a special case of SN-Box 9.2 (ii) for $Z = f(U)$, which implies

$$E(Y|X{=}x, U{=}u, Z{=}z) \;=\; E(Y|X{=}x, U{=}u). \tag{51}$$

Again, it is worthwhile comparing Equations (33) and (50) to each other. This comparison reveals that $E(Y|X{=}x, Z{=}z)$ is *unbiased* if

(a) $P(U{=}u|X{=}x, Z{=}z) \;=\; P(U{=}u|Z{=}z)$, for all $u$,

or

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

If condition (a) holds *for all* values $x$ of $X$, then it is equivalent to $(Z{=}z)$*-conditional independence of U and X*, and this conditional independence is checked by the CEX, when checking the $(Z{=}z)$-conditional causality conditions.

In the CEX, condition (b) is also called $(Z{=}z)$*-conditional unit-treatment homogeneity,* referring to the fact that, given $Z{=}z$, all units respond homogeneously to treatment $x$ with respect to the (expectation of the) outcome variable $Y$. Other causality conditions are treated in section 6.

# 5 More on bias and unbiasedness

## 5.1 Unbiasedness of the $(X{=}x)$-conditional expectation values of the outcome variable $Y$

As mentioned before, under the assumptions summarized in section 2.10, *unbiasedness of the conditional expectation value* $E(Y|X{=}x)$ is defined by

$$E(Y|X{=}x) \;=\; E(\tau_x), \tag{52}$$

where $\tau_x = E(Y|X{=}x, U)$ is the partial $(X{=}x, U)$-conditional expectation of $Y$ (see section 2.7). Once all necessary parameters are specified by the user, this condition is checked by the CEX. If it holds *for all* values $x$ of $X$, then the conditional expectation $E(Y|X)$ is also called *unbiased* and a green check mark appears in the corresponding list of causality conditions.

Because

$$E(Y|X{=}x) \;=\; E(\tau_x|X{=}x) \tag{53}$$

is always true, the difference

$$E(\tau_x) \;-\; E(\tau_x|X{=}x) \tag{54}$$

defines the bias of $E(Y|X{=}x)$.

Unbiasedness in statistics usually refers to *estimators* of a parameter. However, if the expectation

$$E\big(\overline{Y}_x\big) \;=\; E(Y|X{=}x)$$

of the sample mean in treatment $x$ is not identical to the parameter to be estimated — here, this is $E(\tau_x)$ — then $E(Y|X{=}x)$ is also biased if we intend to make causal inferences from $E(Y|X{=}x)$ to $E(\tau_x)$.

## 5.2 Two kinds of bias of the prima facie effect

For the prima facie effect

$$PFE_{x0} \;=\; E(Y|X{=}x) - E(Y|X{=}0)$$

it can be shown that

$$PFE_{x0} \;=\; ATE_{x0} \;+\; \textit{baseline bias}_{x0} \;+\; \textit{effect bias}_{x0}, \quad \text{for all } x, \tag{55}$$

where

$$\textit{baseline bias}_{x0} \;:=\; E(\tau_0|X{=}x) - E(\tau_0|X{=}0) \tag{56}$$

and

$$\textit{effect bias}_{x0} \;:=\; E(\delta_{x0}|X{=}x) \;-\; ATE_{x0}. \tag{57}$$

### 5.2.1 Baseline bias

Reading Equation (56), note that a value of the true-outcome variable $\tau_0$ represents a *(usually unknown) attribute of the observational unit,* namely the individual true outcome $E(Y|X{=}0, U{=}u)$ if unit $u$ *would* receive treatment 0. Remember that we refer to the random experiment described in section 1 that is to be conducted *in the future*. Therefore, a value of the true-outcome variable $\tau_0$ is an attribute of the observational unit just like sex of this unit is one of its attributes. Therefore, just like the random variable sex, the random variable $\tau_0$ is a *pre-treatment variable*, and the same applies to the other true-outcomes variables $\tau_x$. They all are pre-treatment variables whose values are specified for all treatment conditions. In more formal terms, the true-outcomes variables $\tau_x$ are mappings of the observational-unit variable $U$, they are not mappings of the treatment variable $X$. In terms of measure theory, they are *measurable with respect to U* (see SN-chapter 2).

Now, if the conditional expectation value of $\tau_0$ given $X{=}x$ would differ from its conditional expectation value given $X{=}0$ [see Eq. (56)], then this would imply, for example, that a unit $u$ with a high score on $\tau_0$ would tend to be in treatment $x$ rather than in treatment 0, or vice versa. In other words, in this case there would be a selection bias due to the unit's pre-treatment attribute 'true outcome in treatment 0'. If we call the true outcome in treatment 0 — which is the reference treatment, and in some applications it might be the untreated control — the 'baseline', then calling the first kind of bias 'baseline bias' seems justified.

### 5.2.2 Effect bias

Similarly, and for the same reasons that are mentioned for the baseline bias, a value of the individual total effect variable $\delta_{x0} = \tau_x - \tau_0$ represents a pretreatment attribute of the observational unit $u$. If the conditional expectation value of $\delta_{x0}$ given treatment $x$ would deviate from the average total effect $ATE_{x0}$ [see Eq. (5.2.2)], this would mean, for example, that those tending to have a high individual total effect of treatment $x$ compared to treatment 0 would tend to be in treatment $x$ rather than in treatment 0. In other words, in this case, there would be a selection bias due to the total effect *expected* if the unit *would* receive treatment $x$. If units are assigned to treatment $x$ with a high probability, because one correctly expects a high individual total effect for this unit — which is an important goal in medical, educational, and psychological assessment — one would induce such a selection bias.

Note that, in contrast to the CEX examples, in empirical applications, it is *neither* possible to discover bias nor to test unbiasedness directly. However, there are other causality conditions that are *sufficient conditions* for unbiasedness and that *can* be tested empirically (see section 6). Those causality conditions guide us in designing and analyzing experiments and quasi-experiments.

## 5.3 Bias of the conditional expectation value $E(Y|X=x, Z=z)$

Let $x$ be a value of the treatment variable $X$ and assume that $z$ is a value of the covariate $Z$ of $X$ such that $P(X=x, Z=z) > 0$. As mentioned before, *unbiasedness of* $E(Y|X=x, Z=z)$ is then defined by

$$E(Y|X=x, Z=z) \; = \; E(\tau_x|Z=z). \tag{58}$$

Again, once all necessary parameters are specified by the user, this condition is checked by the CEX. If Equation (58) holds *for all* values $x$ of $X$, then the partial conditional expectation $E(Y|Z=z, X)$ is also called *unbiased* and a green check mark appears in the corresponding list of causality conditions on the screen for the specific value $z$ of $Z$ that the user can select via the tabs 'View' and 'Show table conditionally on covariate'.

## 5.4 Two kinds of bias of the conditional prima facie effect

Let $x$ and 0 be two values of $X$ and assume that $z$ is a value of the covariate $Z$ of $X$ such that $P(X=x, Z=z)$, $P(X=0, Z=z) > 0$. Then the $(Z=z)$-conditional prima facie effect

$$PFE_{Z;x0}(z) \; = \; E(Y|X=x, Z=z) - E(Y|X=0, Z=z)$$

is uniquely defined, and it can be shown that

$$PFE_{Z;x0}(z) \; = \; CTE_{Z;x0}(z) \; + \; baseline\ bias_{x0}(z) \; + \; effect\ bias_{x0}(z) \tag{59}$$

where

$$baseline\ bias_{x0}(z) \; := \; E(\tau_0|X=x, Z=z) - E(\tau_0|X=0, Z=z) \tag{60}$$

and

$$effect\ bias_{x0}(z) \; := \; E(\delta_{x0}|X=x, Z=z) \; - \; E(\delta_{x0}|Z=z). \tag{61}$$

If both biases are 0, then the $(Z=z)$-conditional prima facie effect is unbiased. The easiest way to understand the two kinds of conditional biases is to remember that $z$ indicates a subset of observational units (a subpopulation) and to translate what has been said about the unconditional biases to the conditional case.

The CEX computes both kinds of $(Z=z)$-conditional biases for each value $z$ of the covariate $Z$, and it shows a green check mark on the specific screen for $z$ if unbiasedness of $PFE_{Z;x0}(z)$ holds for all values $x$ of $X$.

## 5.5 $(X=x^*)$-conditional expectation value of the true-outcome variable

Let $x, x^*$ be values of $X$ such that $P(U=u, X=x), P(U=u, X=x^*) > 0$. Under the assumptions of section 2.10, the $(X=x^*)$-conditional expectation value of the true-outcome variable $\tau_x$ can be computed by

$$E(\tau_x|X{=}x^*) \;=\; \sum_u E(Y|X{=}x, U{=}u) \cdot P(U{=}u|X{=}x^*). \tag{62}$$

Remember that, for each value $x$ of the treatment variable, $\tau_x$ is a function of the observational-unit variable $U$. Each unit has a value on *all* true-outcome variables $\tau_x$, $x = 0, 1, \ldots, J$.

## 5.6 Treatment-conditional average effects

Let $x, 0, x^*$ be values of $X$ such that $P(U{=}u, X{=}x), P(U{=}u, X{=}0), P(U{=}u, X{=}x^*) > 0$. Then the *conditional total effect of treatment* $x$ *compared to treatment* 0 *given treatment* $x^*$ can be computed by:

$$\begin{aligned} CTE_{X;x0}(x^*) \;:=\; E(\delta_{x0}|X{=}x^*) \;&=\; E(\tau_x|X{=}x^*) - E(\tau_0|X{=}x^*) \\ &= \sum_u \big(E(Y|X{=}x, U{=}u) - E(Y|X{=}0, U{=}u)\big) \cdot P(U{=}u|X{=}x^*). \end{aligned} \tag{63}$$

If there are two treatment conditions $X{=}1$ and $X{=}0$, then, according to Equation (63), we may consider both, $CTE_{X;10}(1)$, the *conditional total effect given treatment,* and $CTE_{X;10}(0)$ the *conditional total effect given control.* These effects are also known as the 'average total effect of the treated' and the 'average total effect of the untreated'.

At first sight, these concepts seem strange. How can we talk about the average total effect (of the treatment) for the untreated? However, remember again that we are not talking about data that resulted from an experiment but about a random experiment that is still to be conducted, that is, we look at the random experiment from the *prefactual perspective.* Hence, we can talk about the individual total effect although the individual is not (yet) treated, and even if it is never treated, just in the same way as we can talk about the probability of flipping 'heads', even if the coin is never flipped. The individual total effect is an attribute of the observational unit even if the random experiment considered is never conducted. Similarly, the conditional or unconditional expectation values of the individual total effects can be considered even if the random experiment is never conducted. Hence, we may interpret the conditional total effect $CTE_{X;10}(0)$ of the untreated as their average effect if they *would be treated.* After all, we prefer the term 'conditional total effect given control' over 'conditional total effect of the untreated', because this term is less prone to mixing up the pre- and postfactual (or counterfactual) perspectives.

The first line of Equation (63) shows that the $(X{=}x^*)$-conditional treatment effects given a treatment condition $x^*$ give more specific information than the (unconditional) average treatment effects, if the $(X{=}x^*)$-conditional expectation values of the individual true-outcome variables $\tau_x$ and $\tau_0$ — they are mappings of the person variable $U$ — *depend on the treatment variable.* If, however, $E(\tau_x|X) = E(\tau_x)$

and $E(\tau_0|X) = E(\tau_0)$, then the conditional average treatment effects given treatment $x^*$ do not differ between the two different treatment conditions and are equal to the (unconditional) average treatment effect, that is, $CTE_{X;10}(0) = CTE_{X;10}(1) = ATE_{10}$. This is the case, for instance, if $U$ and $X$ are stochastically independent, that is, in a randomized experiment.

# 6 Causality conditions

Conditions that imply unbiasedness of conditional expectations such as $E(Y|X)$ or $E(Y|X,Z)$ are called *causality conditions*. If they hold, then this implies that these conditional expectations describe causal dependencies of $Y$ on $X$ and allow us to make causal inferences about the causal total effects of $X$ on $Y$.

## 6.1 Independence and conditional independence of $X$ and $U$

### 6.1.1 Independence of $X$ and $U$

Under the assumptions summarized in section 2.10, the treatment variable $X$ and the observational-unit variable $U$ are *independent* (abbreviated $U \perp\!\!\!\perp X$), if

$$P(X{=}x|U) \ = \ P(X{=}x), \quad \text{for all } x \tag{64}$$

[see SN-Eqs. (5.24) and (16.29)]. And, under these assumptions, it can be shown that Equation (64) implies unbiasedness of $E(Y|X)$ with respect to $U$. If $U \perp\!\!\!\perp X$ holds, then the CEX writes a green check mark in the corresponding place in the table the causality conditions.

**The experimental perspective** If Equation (64) holds, then the probability of receiving treatment $x$ does not depend on the observational units and their attributes such as severity of symptoms, sex, educational status, etc.

**Randomization** Note that independence of $X$ and $U$ can be created by an experimenter via randomized assignment of the unit drawn to one of the treatment conditions. In such a procedure, for example, a coin flip decides whether or not the unit drawn receives treatment. [See step (b) in the random experiment described in section 1.]

**The sampling perspective** Note that $U \perp\!\!\!\perp X$ also holds if

$$P(U{=}u|X) \ = \ P(U{=}u), \quad \text{for all } u, \tag{65}$$

that is, if the probability of sampling unit $u$ is identical between all treatment conditions represented by $X$. If Equation (65) holds, then we may also say that there is *representative random sampling* in each treatment condition, with sampling probabilities that are identical between treatment conditions. In this sense, treatment conditions are then *comparable* with respect to the conditional expectation values $E(Y|X{=}x)$.

**Testability** Note that, if $W = f(U)$ is a mapping of $U$, then Equation (64) implies

$$P(X{=}x|W) \;=\; P(X{=}x), \quad \text{for all } x. \tag{66}$$

If this equation does not hold, then Equation (64) does not hold as well. In other words, if, in an empirical application, we show that (66) does not hold, then (64) is falsified. Hence, in this sense independence of $U$ and $X$ can be tested empirically via testing independence of $W$ and $X$, for example, using binary (if $X$ is binary) or multinomial logistic regression (if $X$ has more than two values).

### 6.1.2 Conditional independence of $X$ and $U$ given $Z$

Under the assumptions summarized in section 2.10, the treatment variable $X$ and the observational-unit variable $U$ are *$Z$-conditionally (stochastically) independent* (abbreviated $U \perp\!\!\!\perp X \,|\, Z$), if

$$P(X{=}x|U,Z) \;=\; P(X{=}x|Z), \quad \text{for all } x. \tag{67}$$

And, under these assumptions, Equation (67) implies unbiasedness of $E(Y|X,Z)$ with respect to $U$.

**The experimental perspective** If Equation (67) holds, then the probability of receiving treatment $x$ only depends on the values of the covariate $Z$. Given a value $z$ of $Z$, it does not depend on other attributes of the observational units. In fact, if $z$ is a value of $Z$ with $P(Z{=}z) > 0$, then Equation (67) implies

$$P(X{=}x|Z{=}z,U) \;=\; P(X{=}x|Z{=}z), \quad \text{for all } x. \tag{68}$$

If Equation (68) holds, then the CEX writes a green check mark in the corresponding place in the table of causality conditions on the specific screen for the value $z$ of $Z$.

**Conditional randomization** Note that $Z$-conditional independence of $X$ and $U$ can be created by an experimenter via $Z$-conditionally randomized assignment of the unit drawn to one of the treatment conditions. For example, if $Z$ represents severity of symptoms, then the experimenter can assign a unit to treatment 1 with probability $P(X{=}1|Z{=}z_1) = 5/6$, if $z_1$ indicates severe symptoms, and with probability $P(X{=}1|Z{=}z_2) = 2/6$, if $z_2$ indicates low severity. In such a procedure, for example, a die toss decides — and not any attribute of the unit other than its value $z = f(u)$ — whether or not the unit drawn receives treatment.

**The sampling perspective** Note that, under the assumptions specified in section 2.10, $U \perp\!\!\!\perp X \,|\, Z$ is equivalent to

$$P(U{=}u|X,Z) \;=\; P(U{=}u|Z), \quad \text{for all } u. \tag{69}$$

If $z$ is a value of $Z$ with $P(Z{=}z) > 0$, then Equation (69) implies

$$P(U{=}u|Z{=}z,X) \;=\; P(U{=}u|Z{=}z), \quad \text{for all } u. \tag{70}$$

Hence, the probability of sampling unit $u$ is identical between all treatment conditions represented by $X$, once we keep constant a value $z$ of $Z$.

If Equation (70) holds, then we also say that there is *representative random sampling* in each combination of treatment and the value $z$ of the covariate, with identical $(Z{=}z)$-conditional sampling probabilities between treatment conditions, and that, given $Z{=}z$, treatment conditions are *comparable* with respect to the conditional expectation values $E(Y|X{=}x, Z{=}z)$.

**Testability** If $W = f(U)$ is a mapping of $U$, then Equation (67) implies

$$P(X{=}x|Z,W) \;=\; P(X{=}x|Z), \quad \text{for all } x. \tag{71}$$

If this equation does not hold, then Equation (67) does not hold as well. In this sense, $Z$-conditional independence of $U$ and $X$ can be falsified in an empirical application, for example, using binary (if $X$ is binary) or multinomial logistic regression (if $X$ has more than two values).

**Covariate selection** In observational studies, we can aim at selecting a (possibly) multivariate covariate $Z = (Z_1, \dots, Z_m)$ such that Equation (67) holds. In such a selection process Equation (71) can play an important role, because if it does not hold, then we can select another covariate — which may or may not include the previous one — for which Equation (67) may hold.

### 6.1.3 Unit-treatment homogeneity

Under the assumptions summarized in section 2.10, the outcome variable $Y$ is $X$-conditionally mean independent from $U$ (abbreviated $Y \vdash U \,|\, X$), if

$$E(Y|X,U) \;=\; E(Y|X). \tag{72}$$

If this equation holds, then the units are *homogeneous* in their response to the treatments represented by $X$. It can be shown that Equation (72) implies unbiasedness of $E(Y|X)$. On the screen for the complete table, the CEX indicates in the table of causality conditions whether (green check mark) or not (red cross) Equation (72) holds.

### 6.1.4 Conditional unit-treatment homogeneity given $Z$

Under the assumptions summarized in section 2.10, the outcome variable $Y$ is $(X, Z)$-conditionally mean independent from $U$ (abbreviated $Y \vdash U|X, Z$), if

$$E(Y|X, Z, U) = E(Y|X, Z). \tag{73}$$

Because we assume that $Z$ is a mapping of $U$, this equation is equivalent to

$$E(Y|X, U) = E(Y|X, Z). \tag{74}$$

It can be shown that, if $P(X{=}x|Z) > 0$ for all values $x$ of $X$, then Equation (73) implies unbiasedness of $E(Y|X, Z)$.

If $P(Z{=}z) > 0$, then Equation (73) implies

$$E(Y|X, U, Z{=}z) = E(Y|X, Z{=}z). \tag{75}$$

If this equation holds, then the units with value $f(u) = z$ are *homogeneous* in their response to the treatments represented by $X$. If it holds, then the CEX writes a green check mark into the table of the causality conditions on the specific screen for the value $z$ of $Z$.

**Testability** If $W = f(U)$ is a mapping of $U$, then Equation (73) implies

$$E(Y|X, Z, W) = E(Y|X, Z). \tag{76}$$

If this equation does not hold, then Equation (73) does not hold as well. In this sense, $Z$-conditional mean independence of $Y$ from $U$ can be falsified in an empirical application, using well-known regression techniques.

**Covariate selection** In observational studies we can aim at selecting a (possibly) multivariate covariate $Z = (Z_1, \dots, Z_m)$ such that Equation (73) holds. In such a selection process Equation (76) can play an important role. As already mentioned, if Equation (76) does not hold, then Equation (73) does not hold and we can select another covariate for which Equation (73) may hold.

### 6.1.5 Independence of $X$ and the true-outcome variables

Let $\tau := (\tau_0, \tau_1, \ldots, \tau_J)$ be the multivariate random variable consisting of the true-outcome variables $\tau_x = E(Y|X{=}x, U)$, $x = 0, 1, \ldots, J$. Under the assumptions specified in section 2.10, the treatment variable $X$ and $\tau$ are *(stochastically) independent* (abbreviated $\tau \perp\!\!\!\perp X$), if

$$P(X{=}x\,|\,\tau) \;=\; P(X{=}x), \quad \text{for all } x. \tag{77}$$

If this equation holds, then the treatment probabilities do not depend on the true-outcome variables, which are specific functions of $U$. Because $\tau$ is a specific function of $U$, Equation (77) follows from $U \perp\!\!\!\perp X$, that is, from Equation (64), but $U \perp\!\!\!\perp X$ does not follow from $\tau \perp\!\!\!\perp X$. It can be shown that Equation (77) implies unbiasedness of $E(Y|X)$. If $\tau \perp\!\!\!\perp X$ holds, then the CEX writes a green check mark into the corresponding place of the table of causality conditions.

### 6.1.6 Conditional independence of $X$ and true-outcome variables

Under the assumptions specified in section 2.10, the treatment variable $X$ and $\tau$ are *$Z$-conditionally (stochastically) independent* (abbreviated $\tau \perp\!\!\!\perp X\,|\,Z$), if

$$P(X{=}x\,|\,Z, \tau) \;=\; P(X{=}x\,|\,Z), \quad \text{for all } x. \tag{78}$$

It can be shown that $\tau \perp\!\!\!\perp X\,|\,Z$ implies unbiasedness of $E(Y|X, Z)$. The conjunction of Equation (78) and

$$P(X{=}x|Z) \;>\; 0, \quad \text{for all } x \tag{79}$$

corresponds to *strong ignorability* in Rubin's terminology (see Rosenbaum & Rubin, 1983). Note that Equation (79) follows from the assumptions specified in section 2.10.

If $P(Z{=}z) > 0$, then Equation (78) implies $(Z{=}z)$-conditional independence of $X$ and $\tau$, which, under the assumptions specified in section 2.10, is equivalent to

$$P(X{=}x\,|\,Z{=}z, \tau) \;=\; P(X{=}x\,|\,Z{=}z), \quad \text{for all } x. \tag{80}$$

This equation is abbreviated by $\tau \perp\!\!\!\perp X\,|\,Z{=}z$. If it holds, then the CEX writes a green check mark into the table of the causality conditions on the specific screen for the value $z$ of $Z$.

## 6.2 Mean independence of $\tau$ from $X$

Let $\tau := (\tau_0, \tau_1, \ldots, \tau_J)$ denote the multidimensional random variable consisting of the true-outcome variables $\tau_x = E(Y|X{=}x, U)$. The true-outcome variables $\tau_x$, $x = 0, 1, \ldots, J$, are called *mean independent from* $X$, if

$$E(\tau_x|X) \;=\; E(\tau_x), \quad \text{for all } x. \tag{81}$$

We abbreviate this equation by $\tau \vdash X$. If $\tau \vdash X$ holds, then $E(Y|X)$ is unbiased with respect to $U$. Note that $\tau \vdash X$ follows from $\tau \perp\!\!\!\perp X$, which itself follows from $U \perp\!\!\!\perp X$.

If $\tau \vdash X$ holds, then the CEX indicates this fact by a green check mark in the table of the causality conditions.

## 6.3 Conditional mean independence of $X$ and $\tau$

Similarly, we define *$Z$-conditional mean independence of the true-outcome variables $\tau_x$, $x = 0, 1, \dots, J$, from the treatment variable $X$* by

$$E(\tau_x|X,Z) \;=\; E(\tau_x|Z), \quad \text{for all } x. \tag{82}$$

We abbreviate this equation by $\tau \vdash X|Z$. If $\tau \vdash X|Z$ holds, then $E(Y|X,Z)$ is unbiased with respect to $U$. Note that $\tau \vdash X|Z$ follows from $\tau \perp\!\!\!\perp X|Z$, which itself follows from $U \perp\!\!\!\perp X|Z$.

If $P(Z{=}z) > 0$, then Equation (82) implies

$$E(\tau_x|X,Z{=}z) \;=\; E(\tau_x|Z{=}z), \quad \text{for all } x. \tag{83}$$

This equation is abbreviated by $\tau \vdash X|Z{=}z$. If $\tau \vdash X|Z{=}z$ holds, then the CEX writes a green check mark into the table of the causality conditions on the specific screen for the value $z$ of $Z$.

## 6.4 Unconfoundedness of $E(Y|X)$

Under the assumptions of section 2.10, the conditional expectation $E(Y|X)$ is called *unconfounded* if, for all values $x$ of $X$, at least one of the following two conditions holds:

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

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

This condition also implies unbiasedness of $E(Y|X)$ with respect to $U$.

This condition is checked by the CEX, and if it holds, then this is indicated by a green check mark in the table of the causality conditions for $E(Y|X)$. Note that, if $X$ is binary and (a) holds for $x{=}1$, then it also holds for $x{=}0$ (see Remark 5.46 of SN). If $X$ is binary, then this implies that $U \perp\!\!\!\perp X$ and unconfoundedness of $E(Y|X)$ are equivalent. This equivalence does not hold anymore if $X$ takes on at least three different values, each one with a positive probability.

## 6.5 Unconfoundedness of $E(Y|X,Z)$

The conditional expectation $E(Y|X,Z)$ is called *unconfounded* if, for all pairs $(x,z)$ of values of $X$ and $Z$, at least one of the following two conditions holds:

(a) $P(X{=}x|Z{=}z,U) = P(X{=}x|Z{=}z)$.

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

This condition implies unbiasedness of $E(Y|X,Z)$ with respect to $U$.

If (a) or (b) holds for a specific value $z$ of $Z$, then this is checked by the CEX and the result is indicated on the specific screen for the value $z$ of $Z$ by a green check mark in the table of the causality conditions.

Again note, that, if $X$ is binary and (a) holds for the pair $(z,1)$ of values of $Z$ and $X$, then it also holds for the pair $(z,0)$ of values of these two random variables (see again SN-Remark 5.46). Just like in the unconditional case, this equivalence does not hold anymore if $X$ takes on at least three different values, each one with a positive probability.

# 7 Conclusion

Under the assumptions summarized in section 2.10, conditional and average total effects of a treatment variable $X$ on a response variable $Y$ can be defined as a specific parameter that exclusively depends on the joint distribution of $X$, $Y$, and the observational-unit variable $U$. There is no need to refer to any manipulations, any philosophical terms, or anything else that is not part of the random experiment considered. There is also no need to assume that the response is fixed given an observational unit $u$ and a treatment condition $x$ as in the definition of Rubin's potential outcome variables. Furthermore, there are several causality conditions, each of which implies unbiasedness of the conditional expectations $E(Y|X)$ or $E(Y|X,Z)$, and some of these causality conditions can be tested in empirical applications and used in covariate selection. The CEX can be used to specify parameter constellations under which one or more of these causality conditions hold or do not hold.

# References


Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70,* 41–55, doi:10.1093/biomet/70.1.41.

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.

Steyer, R. (2018). *Theory of causal effects. Manuscript in preparation.* Jena: FSU Jena.

Steyer, R. & Nagel, W. (2017). *Probability and Conditional Expectation. Fundamentals for the Empirical Sciences.* Chichester, England: Wiley.