Methods of Psychological Research Online 2002, Vol.7, No.1
Internet: http://www.mpr-online.de

Causal Regression Models III: Covariates, Conditional, and Unconditional Average Causal Effects[1]

By Rolf Steyer[2],

Christof Nachtigall, Olivia Wüthrich-Martone, and Katrin Kraus

Abstract

The theory of individual and average causal effects presented in a previous paper is extended introducing conditioning on covariates. From a causal modeling point of view, there are two purposes of including covariates in a regression: (a) to study the conditional average causal effects of X on Y given the values z of the (possibly multi-dimensional) covariate Z, and (b) to adjust for bias in the (unconditional) regression of Y on X and compute the (unconditional) average causal effects of X on Y. One of the examples shows that this adjustment for bias allows analyzing the average causal effects in nonorthogonal analysis of variance designs. This solves a problem that has puzzled methodologists for many decades. The theory presented may be considered the theoretical foundation of the experimental design technique of conditional randomization and of two strategies of data analysis in nonrandomized experiments: (1) trying to include all relevant covariates in the regression that predict the (conditional expectation of the) regressand Y and (2) striving to include all covariates in the regression that determine the indivual assignment probabilities to the treatment conditions x. Conditional randomization and, if successful, both strategies of data analysis in nonrandomized experiments lead to conditional causally unbiased regressions of Y on X given fixed values z of the covariate Z. From these regressions, both the conditional and the unconditional average causal effects of X on Y can be computed. We also study the role of propensity scores in conditional causal regression models. Two examples illustrate the theory.

***Keywords*:** Causality; Covariates; Propensity Scores; Conditional Causal Regression Models; Conditional Randomization; Rubin´s Approach to Causality; Nonorthogonal Analysis of Variance

---

[1] We would like to thank Danny Pfeffermann (Hebrew University, Jerusalem), Siegfried Gabler (ZUMA, Mannheim) and Albrecht Iseler (Free University, Berlin) for extended discussions inspiring parts of this paper.

[2] Address for correspondence: Prof. Dr. Rolf Steyer, Friedrich-Schiller-Universität Jena, Am Steiger 3 – Haus 1, D-07743 Jena, Germany. email: rolf.steyer@uni-jena.de

Introduction

In a previous paper (Steyer, Gabler, von Davier, Nachtigall, & Buhl, 2000a) regression models with discrete units and a discrete (possibly multidimensional) treatment variable were considered. In this framework, the concepts of individual and average causal effects can be studied. These concepts go back to Neyman (1923/1990; 1935). They have been adopted and enriched by Rubin (1973a, b, 1974, 1977, 1978, 1985, 1986, 1990), Holland and Rubin (1983), Holland (1986, 1988a, b), Rosenbaum and Rubin (1983a, b, 1984, 1985a, b), Rosenbaum (1984a, b, c), and Sobel (1994, 1995), for instance.

Regression models with discrete units and a discrete treatment variable consist of three components:

- a probability space $(\Omega, \mathfrak{A}, P)$, which represents the random experiment, i.e., the empirical phenomenon considered,
- a regression $E(Y \mid X)$, the potential causal interpretation of which is focused, where X represents a discrete (but not necessarily univariate) treatment variable and Y a real-valued (and not necessarily continuous) response variable,
- a nonnumeric random variable U, the value of which is the observational unit drawn.

The probability space is chosen such that X, Y, and U are random variables on $(\Omega, \mathfrak{A}, P)$, i.e., X, Y, and U have a joint distribution. In this framework and notation, we presented the basic concepts of the theory of individual and average causal effects: the *individual conditional expected values* $E(Y \mid X = x, U = u)$ pertaining to a treatment condition x, the *individual causal effects* $ICE_u(i, j) := E(Y \mid X = x_i, U = u) - E(Y \mid X = x_j, U = u)$ of a treatment x_i vs. a treatment x_j, the *causally unbiased expected value* [denoted $CUE(Y \mid X = x)$], and the *average causal effect* $ACE(i, j)$, which is the expectation of the individual causal effects over all units in the population. We also showed how these concepts are related to the conditional expected values $E(Y \mid X = x)$ and to the *prima facie effects* $PFE(i, j) = E(Y \mid X = x_i) - E(Y \mid X = x_j)$ that are usually focused in statistical estimation and hypothesis testing. Specifically, it has been shown that both, (a) stochastic independence of U and X (that can be created via random assignment of units to treatment conditions) as well as (b) unit-treatment homogeneity, [i.e., $E(Y \mid X, U) = E(Y \mid X)$] imply causal unbiasedness of $E(Y \mid X = x)$ and of $PFE(i, j)$ [i.e., imply the equations $E(Y \mid X = x) = CUE(Y \mid X = x)$, and $PFE(i, j) = ACE(i, j)$].[3]

[3] In a second paper (Steyer, Gabler, von Davier, & Nachtigall, 2000b) the role of *unconfoundedness* of a

In the present paper, the theory summarized above is extended introducing conditioning on covariates. In ordinary regression models, covariates as additional regressors serve to study *conditional regressive dependencies*, to *reduce the residual variance*, and to learn about the variables predicting the regressand Y. In causal regression models, there are two additional purposes of including covariates in a regression:

(1) to study the *conditional average causal* effects of X on Y given the values z of the (possibly multidimensional) covariate Z,[4] and

(2) to *adjust for bias* of the (unconditional) regression $E(Y \mid X)$ of Y on X and to compute the (unconditional) average causal effects of X on Y.

The first purpose means studying – in the framework of a (conditional) causal regression model – conditional regressive dependencies of Y on X given Z as described by the regression $E(Y \mid X, Z)$. Oftentimes, this is of substantive interest, because conditional dependencies and effects are more informative, since they tell us the average effect for males ($Z =$ males) and the average effect for females ($Z =$ females), for instance, instead of the average effect in the total population. Whereas purpose (1) is well-known from ordinary regression analysis, purpose (2) is more specific for causal analyses. In fact, one of the examples will show that adjustment for bias allows analyzing the average causal effects in nonorthogonal analysis of variance designs. This solves a problem that has puzzled methodologists for many decades (see, e.g., Carlson & Timm, 1974; Gosslee & Lucas, 1965; Keren & Lewis, 1976; Overall & Spiegel, 1969; Overall, Spiegel, & Cohen, 1975; Williams, 1972).

The theory presented may be considered

- the theoretical foundation of the experimental design technique of *conditional randomization*,

and of *two strategies of data analysis*: trying to

- include all relevant covariates in the regression that predict the (conditional expectation) of the response Y, or
- include all covariates in the regression that determine the probabilities of assigning a unit to each of the treatment conditions x.

Conditional random assignment allows that units with different values z_1 and z_2 of Z have different probabilities to be assigned to a treatment condition. Thus it is possible

---

regression $E(Y \mid X)$ in nonrandomized experiments has been studied. However, this concept, although important for causal modeling, is not necessary to understand the present paper.

[4] These conditional average causal effects also play a role in testing propositions on individuals (see, e.g., Iseler, 1997).

to assign a treatment x with a higher probability to units characterized by a specific value z of a covariate Z. In this way, we may respect ethical requirements without giving up the possibility of causal interpretations of the results of such an experiment, even in cases where (unconditional) random assignment is not feasible.

Trying to include all relevant covariates in the regression predicting the (conditional expectation) of the response Y amounts to finding a (possibly multidimensional) covariate Z such that $E(Y \mid X, Z, U) = E(Y \mid X, Z)$, where U denotes the *observational-unit variable*. A value u of U represents the *observational unit* sampled (see Steyer et al. 2000a, p. 43). According to this equation, the individual conditional expected values $E(Y \mid X = x, Z = z, U = u)$ are identical for all units u, i.e., $E(Y \mid X = x, Z = z, U = u) = E(Y \mid X = x, Z = z)$ for each triple (x, z, u). This condition may be called (X, Z)-*conditional regressive independence of* Y *on* U, or alternatively, (Z)-*conditional unit-treatment homogeneity* (see Steyer et al., 2000a, p. 51).

To strive for including in the regression all covariates that determine differential assignment probabilities to the treatment conditions x means finding a (possibly multidimensional) covariate Z such that

$$P(X = x \mid Z, U) = P(X = x \mid Z), \quad \text{for each value } x \text{ of } X.$$

This equation implies that the probabilities of assigning an observational unit u to a treatment condition x are the same for each unit u with a given value $Z(u) = z$. This condition may be called Z-*conditional independence of* X *and* U and be abbreviated $X \perp U \mid Z$. Rosenbaum and Rubin (1983b, 1984) called the values of the conditional assignment probabilities $P(X = x \mid Z)$ the *propensity scores*. These propensity scores may be used as new covariates.

In this paper we will show that conditional random assignment of units to treatment conditions and both strategies of data analysis in nonrandomized experiments described above actually may be used to serve the two purposes (1) and (2) mentioned before, namely to study the *conditional average causal* effects of X on Y given the values z of the (possibly multidimensional) covariate Z, and to *adjust for bias* in the regression $E(Y \mid X)$ and compute the (unconditional) average causal effects of X on Y. One of the examples will show that this adjustment allows analyzing the *average causal effects* of the treatment factor in nonorthogonal analysis of variance designs.

The paper is organized as follows: First, we introduce the basic concepts. Next, we present theorems with sufficient conditions for conditional causal unbiasedness. Third, we present a theorem focusing on how to adjust for bias in the regression $E(Y \mid X)$ if we can assume that the conditional expected values $E(Y \mid X = x, Z = z)$ are $(Z = z)$-conditionally causally unbiased. Fourth, we present a theorem on propensity scores and discuss their role in causal modeling. The fifth section illustrates the concepts and some of the theorems by a detailed example. The sixth section shows how to adjust for bias in the regression $E(Y \mid X)$ in nonorthogonal analysis of variance designs and how to compute the average causal effects in this special case. In the discussion we then summarize what has been gained and hint at some problems yet to be solved.

1. Basic Concepts

Steyer et al. (2000a) defined *causally unbiased conditional expected values* $CUE(Y \mid X = x)$ of Y given $X = x$, c*ausal unbiasedness of the regression* $E(Y \mid X)$ and of its values $E(Y \mid X = x)$, and *average causal effects*. The corresponding concepts will now be defined, conditioning on a value z of a covariate Z. These concepts refer to a random experiment that may be described as follows: Draw a unit u from a population Ω_U of units, register its properties ω_Z which then yield (via a scoring rule) a value z of the (possibly multidimensional) covariate Z (e.g., a variable characterizing the unit, such as gender), register its assignment to a treatment condition $\omega_X \in \Omega_X$, and observe the value $y \in \mathbb{R}$ of the response variable Y. Hence, the set of possible outcomes can be described by the Cartesian product

$$\Omega = \Omega_U \times \Omega_Z \times \Omega_X \times \mathbb{R} \ . \tag{1}$$

The random variables U, X, and Y are defined as described in Steyer et al. (2000a). Note that only Y has to be real-valued with finite expectation in order for the regressions such as $E(Y| X)$ or $E(Y| X, Z)$ to be defined. Furthermore, U is a qualitative random variable, whereas X may or may not transform the qualitative observations into numerical values. The same applies to the regressor Z: $\Omega \rightarrow \Omega'_Z$ with values in a set Ω'_Z. In the simplest case Ω'_Z may only consist of two values such as male and female. In other cases, Z may represent a continuous numerical and possibly fallible pretest, i.e., a pretest with a measurement error component. Again in other cases Z may represent a true-score variable. Finally, it is also possible that Z is a vector $(Z_1, ..., Z_m)$ of several random variables of the types just described. In all these cases, and throughout the paper, we presume that Z refers to an attribute of the sampled unit u *before* it is assigned to one of the treatment conditions, even if Z is only a fallible measure of that attribute. In any case, we preclude that Z could be affected by the treatment variable X.

The following definitions presume that U, Z, X, and Y have a joint distribution and that, for P-almost all values[5] z of Z, there is a conditional probability measure $P_{Z=z}$ and,

[5] The reader unfamiliar with measure theory can ignore this technical detail and just read "for all values z of Z". Bauer (1981) presents the associated probability and measure theory concepts. A proposition holds "for P-almost all values z of Z" if there exists a set Ω_z of potential values z of Z with $P[Z(\omega) \in \Omega_z] = 1$

therefore, a conditional joint distribution of U, X, and Y given $Z = z$. We will assume throughout this paper that $P(U = u, X = x) > 0$, for all pairs (u, x) of values of X and U. The covariate Z will play the role of a „control variable", i.e., the substantive interest is in investigating the dependence of Y on X given the values z of Z.

Note that all concepts introduced below refer to the theorectical or population level and not to a sample. The notations $P_{Z=z}(U = u)$ and $P(U = u \mid Z = z)$ will be used synonymously. We use $P_{Z=z}(U = u)$ instead of $P(U = u \mid Z = z)$, in order to emphasize the analogy to the unconditional case. Similarly, $E_{Z=z}(Y \mid X = x)$ and $E(Y \mid X = x, Z = \mathrm{z})$ are used synonymously. Hence, the conditional expected value $E_{Z=z}(Y \mid X = x)$ denotes the conditional expected value of Y given $X = x$ with respect to the conditional probability measure $P_{Z=z}$.

Also note that the summation in equations (2) and (4) is over *all* units u in the population Ω_U. In those cases in which Z is a (deterministic) function of U, the conditional expected values $E_{Z=z}(Y \mid U = u, X = x)$ simplify to $E(Y \mid U = u, X = x)$ if $Z = f(u) = z$ and they are undefined if $Z = f(u) \neq z$. However, the conditional probabilities $P_{Z=z}(U = u)$ will be zero as well for those values u with $Z = f(u) \neq z$. Hence, the conditional expected values $E_{Z=z}(Y \mid U = u, X = x)$ need not be defined for those cases. The following definitions will also hold if $Z \neq f(U)$, which will be the case if Z represents a fallible pretest, for instance. In this case $Z = f(U) + \varepsilon$, where ε represents a measurement error component.

Definition 1. (i) *The number*

$$CUE_{Z=z}(Y \mid X = x) := \sum_u E_{Z=z}(Y \mid U = u, X = x)\; P_{Z=z}(U = u) \tag{2}$$

is called the $(Z = z)$*-conditional causally unbiased expected value of* Y *given* $X = x$.

(ii) *A conditional expected value* $E_{Z=z}(Y \mid X = x)$ *is called causally unbiased, if*

$$E_{Z=z}(Y \mid X = x) = CUE_{Z=z}(Y \mid X = x). \tag{3}$$

(iii) *The conditional regression* $E_{Z=z}(Y \mid X)$ *is called causally unbiased, if Equation* (3) *holds for each value* x *of* X.

(iv) *The regression* $E(Y \mid X, Z)$ *is called Z-conditionally causally unbiased if the conditional regressions* $E_{Z=z}(Y \mid X)$ *are causally unbiased for P-almost all values* z *of* Z.

(v) *The* $(Z = z)$*-conditional average causal effect of* x_i vs. x_j *on* Y *is defined*:

---

such that the proposition holds for every element z of the set Ω_Z.

$$ACE_{Z=z}(i, j)$$

$$:= \sum_u \left[E_{Z=z}(Y \mid U = u, X = x_i) - E_{Z=z}(Y \mid U = u, X = x_j)\right] P_{Z=z}(U = u). \tag{4}$$

Remark. The conditional average causal effect may also be computed by

$$ACE_{Z=z}(i, j) = CUE_{Z=z}(Y \mid X = x_i) - CUE_{Z=z}(Y \mid X = x_j). \tag{5}$$

2. Sufficient Conditions for Conditional Causal Unbiasedness

The following theorems are extensions of the theorems presented by Steyer et al. (2000a). The proofs are obtained by replacing the (unconditional) probability measure P in Steyer et al. (2000a) by the conditional probability measure $P_{Z=z}$. Similary, the conditional regression $E_{Z=z}(Y \mid X)$ is simply the regression of Y on X with respect to the conditional probability measure $P_{Z=z}$.

Theorem 1. *If X and U are conditionally independent given $Z = z$, i.e., if*

$$P_{Z=z}(U = u \mid X) = P_{Z=z}(U = u) \quad \textit{for each value } u \textit{ of } U, \tag{6}$$

then the conditional regression $E_{Z=z}(Y \mid X)$ is causally unbiased.

Remark. Equation (6) and

$$P_{Z=z}(X = x \mid U) = P_{Z=z}(X = x) \quad \textit{for each value } x \textit{ of } X \tag{7}$$

are equivalent.

According to Theorem 1, causal unbiasedness of a conditional regression $E_{Z=z}(Y \mid X)$ can, in principle, be created by conditional random assignment of units to treatments. By *conditional random assignment* we refer to the following procedure: the experimenter fixes the $(Z = z)$-conditional probabilities $P_{Z=z}(X = x \mid U = u)$ for a unit u with value z of the covariate Z to be assigned to treatment condition x. These probabilities may be different for all pairs (x, z). Conditional random assignment only means that the conditional probabilities $P_{Z=z}(X = x \mid U = u)$ are the same for all units u for a given pair (x, z). Conditional random assignment allows that units with different values z_1 and z_2 of Z have different individual assignment probabilities $P_{Z=z_1}(X = x \mid U = u_1)$ and $P_{Z=z_2}(X = x \mid U = u_2)$. Thus it is possible to assign a specific treatment x to specific units (characterized by a specific value z_1) with a higher probability than to other units (with a specific value z_2). Hence, in the context of therapy research, a person with a high need z_1 for a specific therapy x may be assigned to the treatment condition x with a high probability, and a person with a low need z_2 may be assigned to that treatment condition x with a low probability. In this way, we may respect ethical requirements and still be able to causally interpret the results of such an experiment, even in cases where (unconditional) random assignment is not feasible.

In those cases in which even *conditional* random assignment is not possible, Theorem 1 may be considered the foundation of the following strategy of data analysis: Striving to include all covariates in the regression that determine differential assignment probabilities to the treatment conditions x, i.e., finding a (possibly multidimensional) covariate Z such that X and U are $(Z = z)$-conditionally independent for all values z of Z. A numerical example will be presented in section 5.

Unit-treatment homogeneity, i.e., $E(Y | X, U) = E(Y | X)$, is a second sufficient condition for causal unbiasedness (see Steyer et al. 2000a). Correspondingly, we now have conditional unit-treatment homogeneity as a sufficient condition for conditional causal unbiasedness.

Theorem 2. *If $E(Y | X, U, Z) = E(Y | X, Z)$, almost surely*[6] *(a. s.), then the conditional regressions $E_{Z=z}(Y | X)$ are causally unbiased for P-almost all values z of Z.*

In contrast to the sufficient condition mentioned in Theorem 1, conditional unit-treatment homogeneity cannot be created by the experimenter. However, one may strive to find a (possibly multidimensional) covariate Z such that conditional unit-treatment homogeneity holds.[7]

Theorem 3. *If X and U are $(Z = z)$-conditionally independent for P-almost all values z of Z, or if $E(Y | X, U, Z) = E(Y | X, Z)$, (a. s.), then the conditional regressions $E_{Z=z}(Y | X)$ are causally unbiased for P-almost all values z of Z.*

[6] Conditional expectations are defined under very general conditions. The price, however, is that they not uniquely defined in all cases. For example, if there are continuous regressors, there are many versions of a conditional expectation which are identical but for sets of values with probability zero, hence “almost surely” or “for P-almost all values” (for details, see Bauer, 1981, pp. 309.) The reader unfamiliar with these measure theory concepts should read equations involving conditional expectations involving a continuous regressor as ordinary equations.

[7] In fact one can show that equality of the individual expected values *or* the equality of the assignment probabilities for all persons within a (x, z)-combination is sufficient for causal unbiasedness of the conditional regressions $E_{Z=z}(Y | X)$.

3. Computing Average Causal Effects

The following theorem is crucial for computing the average causal effect $ACE\,(i,\ j)$ in the total population, if one can assume that the conditional regressions $E_{Z=z}(\,Y|\,X\,)$ are causally unbiased. (For a proof see the Appendix.)

Theorem 4. *If*

(a) for P-almost all values z of Z the conditional regression $E_{Z=z}(\,Y|\,X\,)$ is causally unbiased

and

(b) $$E\,(\,Y|\,X,\ U,\ Z\,) = E\,(\,Y|\,X,\ U\,), \quad (a.\ s.), \tag{8}$$

then

$$CUE\,(\,Y|\,X = x) \;= \int E_{Z\,=\,z}(\,Y|\,X = x)\ P^{\,Z}(d\,z) \quad \textit{for each value } x \textit{ of } X. \tag{9}$$

Note that, if Z is discrete, Equation (9) simplifies to: (a. s.)

$$CUE\,(\,Y|\,X = x) \;= \sum\nolimits_z E_{Z\,=\,z}(\,Y|\,X = x)\ P(Z = z) \quad \textit{for each value } x \textit{ of } X. \tag{10}$$

Remarks. (i) Provided that (a) and (b) hold, this theorem shows how to compute the causally unbiased conditional expected values $CUE\,(\,Y|\,X = x)$ from the conditional expected values $E_{Z\,=\,z}(\,Y|\,X = x)$ and the (unconditional or marginal) distribution of Z. The difference $CUE\,(\,Y|\,X = x_i) - CUE\,(\,Y|\,X = x_j)$ of two such unbiased expected values yields the average causal effect $ACE\,(i,\ j)$ in the total population (see also Wüthrich-Martone, 2001).

(ii) Condition (a) is met, e.g., under the conditions mentioned in theorems 1 to 3.

(iii) Condition (b) is met, e.g., if $Z = f\,(\,U\,)$, i.e., if Z is a (deterministic) function of U. Examples are Z_1 = gender, Z_2 = education, and Z_3 = age.

(iv) Condition (b) is met also in other cases. If the within-unit variation of Z is entirely due to measurement error and if we assume that this measurement error does not have itself an effect on Y, then the true-score variable τ_Z is responsible for the effect of Z, i.e.:

$$E\,(\,Y|\,X,\ U,\ Z\,) = E\,(\,Y|\,X,\ U,\ \tau_Z\,). \tag{11}$$

However, this implies :

$$E\,(\,Y|\,X,\ U,\ \tau_Z) = E\,(\,Y|\,X,\ U\,), \tag{12}$$

because the true-score variable τ_Z is by definition a funcion of U (see Steyer & Eid, 2001). Combining these two equations yields condition (b). If Z is a fallible pretest, one can still create causal unbiasedness of the conditional regression $E_{Z=z}(Y \mid X)$, e.g. via conditional random assignment. In this way Theorem 4 will be applicable in practice.

4. Conditioning on the Propensity Score

As mentioned before, Rosenbaum and Rubin (1983b, 1984) called the values of the conditional assignment probabilities $P(X = x_i \mid Z) =: V_i$ the *propensity scores*. Propensity-score variables can be useful, because they may reduce the number of variables with respect to which we have to condition in order to reach causal unbiasedness of the conditional expected values of Y given $X = x_i$. Theorem 6, will show that, if the propensity scores are known, we may use the vector V of propensity-score variables instead of the original vector of covariates Z provided that equations (8) and (13) hold. For a small number of values $x_1, \ldots, x_k$ of X and a large number of covariates in the vector $Z = (Z_1, \ldots, Z_m)$ this may simplify the regression considerably. Equation (8) has been commented before (see Remark (iv) in section 3). Equation (13) will hold if the vector Z contains all covariates that determine the probabilities of the units to be assigned to the specific treatment conditions.

Note that each value x_i of X has its own propensity-score variable V_i. However, if X has k different values $x_1, \ldots, x_k$, then the last propensity-score variable V_k can be computed from the other ones $V_1, \ldots, V_{k-1}$, because probabilities must sum up to one. Hence, for a dichotomous treatment variable with values "treatment" and "control", for instance, we only need a single propensity-score variable V_1.

Theorem 5. *If Z is a covariate such that Equation* (8) *holds and*

$$P(X = x_i \mid Z, U) = P(X = x_i \mid Z), \quad (a.\ s.), \quad \text{for each value } x_i \text{ of } X, \tag{13}$$

then, for $V_i := P(X = x_i \mid Z)$, (a. s.), for each value x_i of X,

$$P(X = x_i \mid Z, U) = P(X = x_i \mid V_i, U), \quad (a.\ s.), \tag{14}$$

$$= P(X = x_i \mid V_i) = V_i, \quad (a.\ s.), \tag{15}$$

and, for P-almost all pairs (x_i, z) of values of X and Z,

$$E(Y \mid X = x_i, Z = z) = E[Y \mid X = x_i, V_i = v_i(z)]. \tag{16}$$

According to Theorem 5 and Theorem 1, the conditional expected values $E[Y \mid X = x_i, V_i = v_i(z)]$ are causally unbiased and identical with the conditional expected values $E(Y \mid X = x_i, Z = z)$ provided that Equation (13) holds.

While Theorem 5 only focusses on a single value x_i of X, the next theorem considers the conditional regression of Y on X. Since this involves all values of $x_1, \ldots, x_k$ of X, we also have to condition on the vector $V := (V_1, \ldots, V_{k-1})$ of propensity-score variables.

Theorem 6. *If equations* (8) *and* (13) *hold, then, for* $V = (V_1, \ldots, V_{k-1})$ *with* $V_i = P(X = x_i \mid Z)$, $i = 1, \ldots, k-1$,

$$P(X = x_i \mid Z, U) = P(X = x_i \mid V, U) = P(X = x_i \mid V), \quad \textit{(a. s.)},$$

$$\textit{for each value } x_i \text{ of } X, \tag{17}$$

and

$$P_{V=v}(U = u \mid X) = P_{V=v}(U = u) \quad \textit{for P-almost all values } (u, v) \textit{ of U and V.} \tag{18}$$

Furthermore, under the same assumptions, for P-almost each value v of V, the conditional regression $E_{V=v}(Y \mid X)$ *is causally unbiased, and*

$$E_{V=v}(Y \mid X) = E_{Z=z}(Y \mid X) \quad \textit{for P-almost each value v of V.} \tag{19}$$

Outside a controlled experiment with conditional random assignment of units to treatment conditions, the propensity scores will usually be unknown. In this case they have to be estimated (see, e.g., Rosenbaum & Rubin, 1984). However, it is not clear if conditioning on *estimated* propensity scores should be preferred over conditioning on the original covariates, which are *not* estimated.

5. Example I: Illustrating the Concepts

We consider the following random experiment. A person u is randomly drawn from a population of six persons, and then assigned to one of two experimental conditions: $X = 1$ for treatment and $X = 0$ for control. Thereafter, the value of the response variable Y is observed. The population considered consists of four males ($Z = 0$) and two females ($Z = 1$). Each person is assumed to have the same probability $P(U = u) = 1/6$ to be sampled.

Table 1 displays the individual conditional expected values $E(Y \mid X = x, U = u)$, the individual causal effects $ICE_u(1, 0)$, and also the individual assignment probabilities

$P(X = 1 \mid U = u)$. Each male has the individual treatment assignment probability 3/4, each female 1/4.

5.1. The Average Causal Effect

The average causal effect is defined to be the average of the individual causal effects. Hence, in this example, it can be computed as follows:

$$ACE(1, 0) = \sum_u ICE_u(1, 0)\ P(U = u)$$

$$= (80 - 68) \cdot \frac{1}{6} + (93 - 81) \cdot \frac{1}{6} + \cdots + (148 - 137) \cdot \frac{1}{6}$$

$$= \frac{1}{6} \cdot (12 + 12 + 14 + 14 + 9 + 11) = 12.$$

In this example, the average causal effect is 12 in the total population of all six units.

Table 1.

Example in which each single person has a positive effect and the treatment assignment probability depends on gender.

Person	$P(U = u)$	$E(Y \mid X = 1, U = u)$	$E(Y \mid X = 0, U = u)$	$ICE_u(1, 0)$	Gender (Z)	$P(X = 1 \mid U = u)$
u_1	1/6	80	68	12	0 (male)	3/4
u_2	1/6	93	81	12	0 (male)	3/4
u_3	1/6	103	89	14	0 (male)	3/4
u_4	1/6	116	102	14	0 (male)	3/4
u_5	1/6	132	123	9	1 (female)	1/4
u_6	1/6	148	137	11	1 (female)	1/4

5.2. The Causally Unbiased Expected Values

Since the weighted sum of a difference is the difference of the two weighted sums,

$$\sum_u ICE_u(1, 0)\ P(U = u) =$$

$$\sum_u E(Y \mid X = 1,\ U = u)\ P(U = u) - \sum_u E(Y \mid X = 0,\ U = u)\ P(U = u),$$

the average causal effect may also be computed as the difference between the two sums

$$CUE(Y \mid X = 1) = \sum_u E(Y \mid X = 1,\ U = u)\ P(U = u)$$

$$= \frac{1}{6} \cdot (80 + 93 + \cdots + 148) = 112$$

and

$$CUE(Y \mid X = 0) = \sum_u E(Y \mid X = 0,\ U = u)\ P(U = u)$$

$$= \frac{1}{6} \cdot (68 + 81 + \cdots + 137) = 100.$$

The terms $CUE(Y \mid X = 1)$ and $CUE(Y \mid X = 0)$ have been defined the *causally unbiased expected values of* Y given $X = 1$ and given $X = 0$, respectively (see Steyer et al., 2000a).

5.3. Expected Values and the Prima Facie Effect

The expected values in the treatment and control conditions can be computed from

$$E(Y \mid X = x) = \sum_u E(Y \mid X = x,\ U = u)\ P(U = u \mid X = x), \tag{20}$$

which is always true for a discrete variable U. In order to use this formula, we need the individual expected values $E(Y \mid X = x,\ U = u)$ which are displayed in Table 1. From the individual treatment assignment probabilities $P(X = 1 \mid U = u)$ given in this table, we may also compute the probabilities $P(U = u \mid X = x)$ according to

$$P(U = u \mid X = x) = P(U = u,\ X = x)\ /\ P(X = x)$$

$$= P(X = x \mid U = u)\ P(U = u)\ /\ P(X = x) \tag{21}$$

and

$$P(X = x) = \sum_u P(X = x \mid U = u)\ P(U = u). \tag{22}$$

Applying Equation (22) to the probabilities given in Table 1 yields

$$P\,(X=1) \;=\; \frac{3}{4}\cdot\frac{1}{6}+\ldots+\frac{3}{4}\cdot\frac{1}{6}+\frac{1}{4}\cdot\frac{1}{6}+\frac{1}{4}\cdot\frac{1}{6}=\frac{7}{12}.$$

Applying Equation (21) for unit u_1 and treatment $X = 1$ we obtain:

$$P\,(U=u_1 |\, X=1) \;=\; \frac{\frac{3}{4}\cdot\frac{1}{6}}{\frac{7}{12}} = \frac{3}{14}.$$

For the other units, these conditional probabilities may be computed analogously.

Using these conditional probabilities, we may now compute the conditional expected values $E\,(Y |\, X = x)$. For treatment $X = 1$ we obtain:

$$\begin{aligned} E\,(Y|\,X=1) &= \frac{3}{14}\cdot(80+93+103+116)+\frac{1}{14}\cdot(132+148) \\ &= \frac{3}{14}\cdot 392+\frac{1}{14}\cdot 280 = 84+20 = 104, \end{aligned}$$

and for control $X = 0$:

$$\begin{aligned} E\,(Y|\,X=0) &= \frac{1}{10}\cdot(68+81+89+102)+\frac{3}{10}\cdot(123+137) \\ &= \frac{1}{10}\cdot 340+\frac{3}{10}\cdot 260 = 34+78 = 112. \end{aligned}$$

The differerence $E(Y|\,X = 1) - E(Y|\,X = 0) = 104 - 112 = -8$ between these two conditional expected values, the *prima facie effect*, is negative, although the average causal effect is positive (namely, 12), and each and every individual causal effect is positive. Hence, the causal interpretation of the prima facie effect in this example would be seriously misleading. Hence, this is another example for the kind of paradox presented in Steyer et al. (2000a).[8]

5.4. Causal Effects Within Males and Females

The causal interpretation of the prima facie effects within gender groups is correct as we will see. We begin looking at the conditional causal effects, i.e., the averages of the individual causal effects, within the gender subpopulations denoted $ACE_{Z=0}(1,\ 0)$ and $ACE_{Z=1}(1,\ 0)$, respectively. The subscript $Z = 0$ indicates the subpopulation (here: „males“) whereas 1 and 0 in the parentheses refer to the two values of *X*, with respect to which the effect is considered. For the *males* this conditional causal effect is:

---

[8] Iseler (1996) has shown that similar examples may also occur if we focus medians. He also showed that, if prima facie effects and individual causal effects would be defined in terms of medians instead of expectations, then a situation with positive individual causal effects for each and every unit, but with a negative prima facie effect can arise even if the random variables X and U are independent.

$$ACE_{Z=0}(1, 0) = \sum\nolimits_u ICE_u(1, 0)\ P_{Z=0}(U = u)$$

$$= \frac{1}{4} \cdot (12 + 12 + 14 + 14) + 0 \cdot (9 + 11) = 13.$$ [9]

This conditional causal effect may also be computed by the difference between the conditional unbiased expectations in the two treatment conditions:

$$CUE_{Z=0}\,(Y|\ X = 1) = \sum\nolimits_u E\,(Y|\ X = 1,\ U = u)\ P_{\,Z=0}\,(U = u)$$

$$= \frac{1}{4} \cdot (80\ + 93 + 103 + 116) + 0 \cdot (132 + 148) = 98$$

and

$$CUE_{Z=0}\,(Y|\ X = 0) = \sum\nolimits_u E\ (Y|\ X = 0,\ U = u)\ \ P_{Z=0}\,(U = u)$$

$$= \frac{1}{4}(68\ + 81 + 89 + 102) + 0 \cdot (123 + 137) = 85.$$

$CUE_{Z=0}(Y|\ X = x)$ denotes the $(Z = 0)$-conditional causally unbiased expected value of Y given $X = x$.

For the *females* the conditional causal effect is:

$$ACE_{Z=1}(1, 0)\ = \sum\nolimits_u ICE_u(1, 0)\ P_{Z=1}\,(U = u)$$

$$= 0 \cdot (12 + 12 + 14 + 14) + \frac{1}{2} \cdot (9 + 11) = 10.$$

This conditional effect may be computed as well by the difference between

$$CUE_{\,Z=1}(Y|\,X = 1) = \sum\nolimits_u E(Y|\ X = 1,\ U = u)\ P_{Z=1}\,(U = u)$$

$$= 0 \cdot (80 + 93 + 103 + 116) + \frac{1}{2} \cdot (132 + 148) = 140$$

and

$$CUE_{\,Z=1}(Y|\,X = 0) = \sum\nolimits_u E(Y|\ X = 0,\ U = u)\ P_{Z=1}(U = u)$$

$$= 0 \cdot (68 + 81 + 89 + 102) + \frac{1}{2} \cdot (123 + 137) = 130.$$

[9] Here, we use the notation $P_{Z=z}\,(U = u)$ instead of $P(U = u \mid Z = z)$, in order to make clear the analogy to the unconditional case. Similarly, we use the notations $E_{Z=z}(Y|\,X = x)$ and $E(Y \mid X = x,\ Z = z)$ synonymously.

5.5. Expected Values Within Gender Populations

In contrast to the $(X = x)$-conditional expected values and their difference *in the total population* the $(X = x)$-conditional expected values and their difference *within the two gender subpopulations* are causally unbiased. As we shall see, this is due to the fact that, within each gender subpopulation, each person has the same treatment assignment probability. Therefore, X and U are $(Z = z)$-*conditionally independent* (see Theorem 1). In other words, in this example

$$P_{Z=z}(U = u \mid X = x) = P_{Z=z}(U = u), \quad \text{for all values } (u, x, z) \text{ of } (U, X, Z). \tag{23}$$

In this example, the four conditional expected values $E_{Z=z}(Y \mid X = x)$ and the corresponding four causally unbiased conditional expected values $CUE_{Z=z}(Y \mid X = x)$ are identical within the two gender subpopulations. This can be seen as follows.

The $(X = x)$-conditional expected values of Y within the gender subpopulations can be computed from

$$E_{Z=z}(Y \mid X = x) = \sum_u E_{Z=z}(Y \mid X = x, U = u)\, P_{Z=z}(U = u \mid X = x), \tag{24}$$

which always holds. Because of $Z = f(U)$, we also have

$$E_{Z=z}(Y \mid X = x, U = u) = E(Y \mid X = x, U = u). \tag{25}$$

Finally, X and U are $(Z = z)$-conditionally independent [see Eq. (23)]. Hence, the conditional probability $P_{Z=z}(U = u \mid X = x)$ in Equation (24) can be replaced by $P_{Z=z}(U = u)$. Inserting equations (23) and (25) into (24) yields $E_{Z=z}(Y \mid X = x) = CUE_{Z=z}(Y \mid X = x)$. Hence, in this example, these conditional expected values can be computed as follows:

$$E_{Z=0}(Y \mid X = 1) = \frac{1}{4} \cdot (80 + 93 + 103 + 116) + 0 \cdot (132 + 148) = 98,$$

$$E_{Z=0}(Y \mid X = 0) = \frac{1}{4} \cdot (68 + 81 + 89 + 102) + 0 \cdot (123 + 137) = 85,$$

$$E_{Z=1}(Y \mid X = 1) = 0 \cdot (80 + 93 + 103 + 116) + \frac{1}{2} \cdot (132 + 148) = 140$$

and

$$E_{Z=1}(Y \mid X = 0) = 0 \cdot (68 + 81 + 89 + 102) + \frac{1}{2} \cdot (123 + 137) = 130.$$

The difference $E_{Z=0}(Y \mid X = 1) - E_{Z=0}(Y \mid X = 0) = 98 - 85 = 13$, i.e., the $(Z = 0)$-conditional prima facie effect, is now exactly the conditional average causal effect for the males, and the difference $E_{Z=1}(Y \mid X = 1) - E_{Z=1}(Y \mid X = 0) = 140 - 130 = 10$, i.e., the

($Z = 1$)-conditional prima facie effect, now yields exactly the conditional average causal effect for the females.

5.6. Average Causal Effect

In section 5.1 we computed the average causal effects based on the individual causal effects that usually cannot be estimated and in section 5.2 we computed the causally unbiased conditional expectations based on the individual conditional expectations that usually cannot be estimated as well. We now use Theorem 4 to compute the unbiased conditional expectations and the average causal effect from the conditional expected values $E_{Z=z}(Y \mid X = x)$ that *can* be estimated. Since these conditional expected values are causally unbiased, we may now use Equation (10), which yields

$$CUE\,(Y| \, X = 1) = \sum\nolimits_z E_{Z\,=\,z}\,(Y| \, X = 1)\; P(Z = z)$$

$$= 98 \cdot 4/6 \; + 140 \cdot 2/6 = 112$$

and

$$CUE\,(Y| \, X = 0) = \sum\nolimits_z E_{Z\,=\,z}\,(Y| \, X = 0)\; P(Z = z)$$

$$= 85 \cdot 4/6 \; + 130 \cdot 2/6 = 100.$$

The difference between these two numbers is the average causal effect of X:

$$ACE\,(1, 0) = CUE\,(Y| \, X = 1) - CUE\,(Y| \, X = 0) = 112 - 100 = 12.$$

5.7. Propensity Scores

In this example, the covariate Z (gender) and the propensity-score variable $V_1 = P\,(X = 1 \mid Z\,)$ are one-to-one functions. Therefore, it is trivial that we may exchange Z by V_1 in this example.

6. Example II: Nonorthogonal Analysis of Variance

Theorem 4 is the theoretical foundation for answering the following question: „How to compute (and test) the main effect of a treatment factor in nonorthogonal analysis of variance?"[10] We will illustrate this point by an example. Let us consider the following

[10] Up to date, the methods of nonorthogonal analysis of variance offered in the program packages SPSS

experiment: A subject is drawn from a population of subjects and his neediness for a therapy is assessed. Suppose we distinguish three levels: *high*, *medium*, and *low neediness*. Assume, these three levels occur with probabilities 1/4, 1/2, and 1/4, respectively.[11] In principle, there are three therapies that might be helpful for the subjects. However, informal knowledge suggests that therapy 1 might be more successful for subjects with a high need than for subjects with medium or low need. Therefore, the experimenter decides to assign therapy 1 with probability 2/3 to the subjects with high need. Subjects with a medium need are assigned to therapy 1 with probability 1/6, and subjects with a low need with probability 1/10. Within the three groups of neediness these assignment probabilities apply for each subject, i.e., they are identical for each subject. Analogously, the assignment probabilities are fixed for the other two treatment conditions. The cell frequencies (see numbers in parentheses in Table 2), correspond to the assignment probabilities fixed by the experimenter. The nine cells of this table also contain the conditional expected values $E(Y \mid X = x,\ Z = z)$, which are assumed to be causally unbiased. Furthermore, as will be shown below, the conditional expected values $E(Y \mid X = x,\ Z = z)$ are choosen such that there are nonzero prima facie effects of the therapy factor, i. e., $E(Y \mid X = x_i) - E(Y \mid X = x_j) \neq 0$, für $i,\ j = 1,\ 2,\ 3$, although the average causal effects $ACE(i,\ j)$ are all equal to 0.

In the last paragraph we already introduced the assumption that the conditional expected values in the nine cells are causally unbiased. This assumption will be fulfilled, for instance, in a design with conditional randomization, i.e., if, for each treatment condition x, each unit with score z of Z has the same chance to be assigned to x. In this example we may also assume that $E(Y \mid X,\ U,\ Z) = E(Y \mid X,\ U)$ holds [see Eq. (8) and Remark iv to Theorem 4]. Hence, the *unbiased expected values* of Y given $X = x$ can be computed according to

$$CUE(Y \mid X = x) = \sum\nolimits_z E(Y \mid X = x,\ Z = z)\ P(Z = z).$$

or SAS, all do *not* test the average causal effect. Wüthrich-Martone, Steyer, Nactigall and Suhl (1999) and Wüthrich-Martone (2001) describe a method and provide a PC-program that *does* test the average causal effect of a treatment factor in nonorthogonal analysis of variance.

[11] In practice the probabilities have to be estimated by the relative frequencies.

Table 2.

Example for a nonorthogonal design, in which there are prima facie effects of the therapy factor, although the corresponding average causal effects are zero.

	neediness						
	high		medium		low		total
therapy	$Z = z_1$		$Z = z_2$		$Z = z_3$		
1 $X = x_1$	120	(40)	110	(20)	60	(6)	(66)
2 $X = x_2$	100	(14)	100	(80)	100	(14)	(108)
3 $X = x_3$	80	(6)	90	(20)	140	(40)	(66)
total		(60)		(120)		(60)	(240)

Note. True cell means and cell frequencies (in parentheses).

For the three values of X we obtain:

$$CUE\,(Y \mid X = x_1) = 120 \cdot 60/240 + 110 \cdot 120/240 + \ 60 \cdot 60/240 = 100,$$

$$CUE\,(Y \mid X = x_2) = 100 \cdot 60/240 + 100 \cdot 120/240 + 100 \cdot 60/240 = 100,$$

$$CUE\,(Y \mid X = x_3) = \ 80 \cdot 60/240 + \ \ 90 \cdot 120/240 + 140 \cdot 60/240 = 100.$$

Hence, all *average causal effects* in the total population are zero in this example, i. e.:

$$ACE\,(1,\,2) = ACE\,(1,\,3) = ACE\,(2,\,3) = 0.$$

The *conditional expected values* $E\,(Y \mid X = x)$ can be computed by:

$$E\,(Y \mid X = x) = \textstyle\sum_z E\,(Y \mid X = x,\, Z = z)\; P\,(Z = z \mid X = x).$$

For the three values of X we obtain:

$$E\,(Y \mid X = x_1) = \ 120 \cdot 40/66 + 110 \cdot 20/66 + \ 60 \cdot 6/66 = 111.52,$$

$$E\,(Y \mid X = x_2) = \ 100 \cdot 14/108 + 100 \cdot 80/108 + 100 \cdot 14/108 = 100,$$

$$E\,(Y \mid X = x_3) = \ 80 \cdot 6/66 + 90 \cdot 20/66 + 140 \cdot 40/66 = 119.39.$$

Hence, the conditional expected values $E\,(Y \mid X = x)$ in this example are causally biased, as well as the prima facie effects

$$PFE(x_1, x_2) = 11.52, \qquad PFE(x_1, x_3) = -7.87, \qquad PFE(x_2, x_3) = -19.39.$$

In contrast to the average causal effects these prima facie effects are *not* zero.

7. Summary and Discussion

In this paper we extended the theory of individual and average causal effects presented in a previous paper (Steyer et al., 2000a) introducing conditioning on covariates. We argued that, from a causal modeling point of view, there are two purposes of including covariates in a regression:

(a) to study the *conditional average causal effects* of X on Y given the values z of the (possibly multidimensional) covariate Z, and

(b) to *adjust for bias* in the regression $E(Y|X)$ and *compute the (unconditional) average causal effects* of X on Y.

We described the kind of random experiments, i.e., the empirical phenomenon to which the theory refers, defined the concepts of *conditional causally unbiased expected values*, *causally unbiased conditional regressions*, and *conditional average causal effects.* Some theorems on sufficient conditions for conditional causal unbiasedness were presented. Another theorem shows how to compute the (unconditional) average causal effects from conditional causally unbiased expected values. Two more theorems dealt with the role of propensity scores and the conditions under which they may replace the covariates. Finally, we presented two examples that illustrate the theoretical results.

The theory presented may be considered the theoretical foundation of the experimental design technique of conditional random assignment of units to treatment conditions and of two strategies of data analysis in nonrandomized experiments: (1) striving to include all relevant covariates in the regression that predict the (conditional expectation) of the regressand Y and (2) striving to include all covariates in the regression that determine differential assignment probabilities to the treatment conditions x. Both are strategies to get unbiased estimates of conditional and − via Theorem 4 − of average causal effects.

Two examples were given. Example I illustrates all concepts introduced. Example II showed that the adjustment for confounding allows analyzing the *average causal effects* of the treatment factor in nonorthogonal analysis of variance designs. This solves a problem that has puzzled methodologists for many decades (see e.g., Carlson & Timm, 1974; Gosslee & Lucas, 1965; Keren & Lewis, 1976; Overall & Spiegel, 1969; Overall, Spiegel, & Cohen, 1975; Williams, 1972). For details see Wüthrich-Martone et al. (1999) and Wüthrich-Martone (2001).

Which are the merits and which are the limitations of the theory presented? A *first merit* is that it provides a clear-cut and relatively simple theoretical foundation for the experimental design technique of conditional randomization and two strategies of data analysis for nonrandomized experiments that were mentioned above.

A *second merit* is that, from a substantive point of view, there is more reason to believe in conditional independence of X and U given a (possibly multidimensional) covariate Z than in unconditional independence of X and U, even in those case in which treatment assignment is not in the hands of the experimenter. Similarly, there is more reason to believe in conditional unit-treatment homogeneity given a covariate Z, than to believe in unconditional unit-treatment homogeneity.

A *third merit* is that the theory is now general enough to cover causal modeling in randomized and nonrandomized experiments, in which the assignment of units to treatment conditions is not random or conditionally random. In our point of view, it reconciles the two traditions of theorizing about causality in statistics identified by Cox (1992) into a single coherent theory.

Which are the limitations? The *first limitation* is the same as in the unconditional case. The claim of unbiasedness of the conditional prima facie effect [i.e., of the difference between two conditional expected values $E_{Z=z}(Y \mid X = x_1) - E_{Z=z}(Y \mid X = x_2)$] though well-defined, is not empirically falsifiable. Postulating that the conditional prima facie effect $PFE_{Z=z}(i, j)$ is equal to the conditional average causal effect $ACE_{Z=z}(i, j)$ or that the conditional expected values $E_{Z=z}(Y \mid X = x)$ are equal to the conditional causally unbiased expected value $CUE_{Z=z}(Y \mid X = x)$ of Y given x does not imply anything one could show to be wrong in an empirical application. The equation $E_{Z=z}(Y \mid X = x) = CUE_{Z=z}(Y \mid X = x)$ cannot be tested, because $CUE_{Z=z}(Y \mid X = x)$ cannot be estimated unless (a) each individual u has the same (Z=z)-conditional probability to be assigned to $X = x$ [i. e., U and X are (Z=z)-conditionally independent such as in the conditionally randomized experiment] or unless (b) $E_{Z=z}(Y \mid X, U) = E_{Z=z}(Y \mid X)$ (i. e., if there is conditional unit-treatment homogeneity). What can be tested, however, are the two sufficient conditions (a) and (b) just mentioned.

Whereas (Z=z)-conditional randomization *guarantees* (Z=z)-conditional independence of U and X (in the population!), things are much more complicated in nonrandomized studies: We may either directly assume $E_{Z=z}(Y \mid X = x) = CUE_{Z=z}(Y \mid X = x)$ and $PFE_{Z=z}(i, j) = ACE_{Z=z}(i, j)$ (conditional unbiasedness) without being able to empirically test this assumption; or, we may assume that the sufficient condition "U and X are (Z=z)-independent and/or (Z=z)-conditional unit-treatment homogeneity" holds and empirically test this assumption. However, since this condition is sufficient but not necessary, rejection of this assumption does not disprove unbiasedness.

The *second limitation* discussed for the unconditional theory in Steyer et al. (2000a) also applies to the conditional case: The concept of causality associated with conditional causally unbiased expected values and conditional average causal effects is rather weak. Even if the conditional average causal effect is positive, there can be observational units and subpopulations for which the individual or conditional average causal effects are negative. As mentioned already in Steyer et al. (2000a), a stronger concept would require invariance of the individual effects (see, e.g., Steyer, 1985, 1992, Ch. 9) or at least invariance of the individual effects in subpopulations (Steyer, 1992, Ch. 14). However, although the search for invariant individual causal effects within subpopulations is certainly a fruitful goal, it should be noted that there is no sufficient condition for such an invariance of individual causal effects *that could deliberately be created by the experimenter*. Nevertheless, from a substantive point of view, the search for interaction or moderator effects is important although it does not replace randomization or conditional randomization. Conditional randomization guarantees that the conditional prima facie effects in the subpopulations are at least average causal effects in the subpopulations, and if the subpopulations were in fact homogeneous, the average effects in the subpopulations were also effects for all the individuals in that subpopulation.

To summarize, the concepts of individual causal effects and average causal effects have been supplemented by the concept of conditional average causal effects. These three are what substantive researchers are looking for in causal modeling. Average causal effects are of interest in the total population, but also in subpopulations which provide more detailed information on the effects of a treatment variable. We have also learned sufficient conditions for the average and conditional average causal effects to be equal to the corresponding prima facie effects: random or conditional random assignment of units to treatment conditions and (conditional) unit-treatment homogeneity. Under these assumptions our well-know statistical procedures for estimating means and their differences also estimate causally unbiased expectations and their differences.

References

[1] Bauer, H. (1981). Probability theory and elements of measure theory. New York: Academic Press.

[2] Carlson, J. E. & Timm, N. H. (1974). Analysis of nonorthogonal fixed-effects designs. *Psychological Bulletin*, *81*, *563-570.*

[3] Cox, D. R. (1992). Causality: Some statistical aspects. *Journal of the Royal Statistical Society, Series A*, *155*, 291-301.

[4] Gosslee, D. G. & Lucas, H. L. (1965). Analysis of variance of disproportionate data when interaction is present. *Biometrics, 21, 115-133*

[5] Holland, P. W. (1988). Causal inference, path analysis, and recursive structural equations models. *Sociological Methodology*, *18*, 449-484.

[6] Holland, P. W. (1988a). Causal inference in retrospective studies. *Evaluation Review*, *13*, 203-231.

[7] Holland, P. W. (1988b). Causal inference, path analysis, and recursive structural equations models. *Sociological Methodology*, *18*, 449-484.

[8] Holland, P. W. & Rubin, D. B. (1983). On Lord's paradox. In H. Wainer & S. Messick (eds.), *Principals of modern psychological measurement* (pp. 3-25). Hillsdale, NJ: Erlbaum.

[9] Holland, P. W. & Rubin, D. B. (1988). Causal inference in retrospective studies. *Evaluation Review*, *23*, 203-231.

[10] Iseler, A. (1996). A paradoxical property of aggregate hypotheses referring to the order of medians. *Methods of Psychological Research-Online*, *1*, 25-40. (http://www.mpr-online.de)

[11] Iseler, A. (1997). Populationsverteilungen von Merkmalen und Geltungsbereiche individuenbezogener Aussagen als Gegenstand der Inferenzstatistik in psychologischen Untersuchungen. In H. Mandl (Hrsg.), *Bericht über den 40. Kongreß der Deutschen Gesellschaft für Psychologie in München 1996* (S. 699-708). Göttingen: Hogrefe.

[12] Keren, G. & Lewis, C. (1976). Nonorthogonal designs: Sample versus population. *Psychological Bulletin, 83, 817-826.*

[13] Nachtigall, C., Steyer, R. & Wüthrich-Martone, O. (2001). Causal Effects in Empirical Research. In M. May & U. Ostermeier (Eds.), *Interdisciplinary perspectives on causality* (pp. 81-100). Norderstedt: Libri Books on Demand.

[14] Neyman, J. (with Iwaszkiewicz, K., and Kolodziejczyk, S.). (1935). Statistical problems in agricultural experimentation (with discussion). *Supplement of Journal of the Royal Statistical Society*, *2*, 107-180.

[15] Neyman, J. (1923/1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science*, *5*, 465-472.

[16] Overall, J. E. & Spiegel, D. K. (1969). Concerning least squares analysis of experimental data. *Psychological Bulletin*, *72*, *311-322*.

[17] Overall, J. E., Spiegel, D. K. & Cohen, J. (1975). Equivalence of orthogonal and nonorthogonal analysis of variance. *Psychological Bulletin*, *82*, *182-186*.

[18] Rosenbaum, P. R. (1984a). Conditional permutation tests and the propensity score in observational studies. *Journal of the American Statistical Association*, *79*, 565-574.

[19] Rosenbaum, P. R. (1984b). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society, Series A*, *147*, 656-666.

[20] Rosenbaum, P. R. (1984c). From association to causation in observational studies: The role of tests of strongly ignorable treatment assignment. *Journal of the American Statistical Association*, *79*, 41-48.

[21] 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*, *45*, 212-218.

[22] Rosenbaum, P. R. & Rubin, D. B. (1983b). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41-55.

[23] 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.

[24] Rosenbaum, P. R. & Rubin, D. B. (1985a). The bias due to incomplete matching. *Biometrics*, *41*, 103-116.

[25] Rosenbaum, P. R. & Rubin, D. B. (1985b). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*, 33-38.

[26] Rubin, D. (1973a). The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics*, *29*, 185-203.

[27] Rubin, D. B. (1973b). Matching to remove bias in observational studies. *Biometrics*, *29*, 159-183.

[28] Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*, 688-701.

[29] Rubin, D. B. (1977). Assignment of treatment group on the basis of a covariate. *Journal of Educational Statistics*, *2*, 1-26.

[30] Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, *6*, 34-58.

[31] Rubin, D. B. (1985). The use of propensity scores in applied Bayesian inference. *Bayesian Statistics*, *2*, 463-472.

[32] Rubin, D. B. (1986). Which ifs have causal answers. *Journal of the American Statistical Association*, *81*, 961-962.

[33] Rubin, D. B. (1990). Comment: Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science*, *5*, 472-480.

[34] Sobel, M. E. (1994). Causal inference in latent variables analysis. In A. von Eye & C. C. Clogg (eds.), *Latent variables analysis* (pp. 3-35). Thousand Oaks, CA: Sage.

[35] Sobel, M. E. (1995). Causal inference in the Social and Behavioral Sciences. In G. Arminger, C. C. Clogg & M. E. Sobel (eds.), *Handbook of Statistical Modeling for the Social and Behavioral Sciences* (pp. 1-38). New York: Plenum.

[36] Steyer, R. (1985). Causal regressive dependencies: an introduction. In J. R. Nesselroade & A. von Eye (ed.), *Individual development and social change: explanatory analysis* (pp. 95-124). Orlando, FL: Academic Press.

[37] Steyer, R. (1992). *Theorie kausaler Regressionsmodelle* [Theory of causal regression models]. Stuttgart: Gustav Fischer.

[38] Steyer, R., Gabler, S., von Davier, A., Nachtigall, C. & Buhl, T. (2000a) Causal regression models I: individual and average causal effects. *Methods of Psychological Research-Online*. (http://www.mpr-online.de)

[39] Steyer, R., Gabler, S., von Davier, A. & Nachtigall, C. (2000b) Causal regression models II: unconfoundedness and causal unbiasedness. *Methods of Psychological Research-Online*. (http://www.mpr-online.de)

[40] Steyer, R. & Eid, M. (2001). *Messen und Testen.* [Measuring and testing. 2nd edition]. Berlin: Springer.

[41] Williams, D. (1972). Two way fixed effects analysis of variance with disproportionate cell frequencies. *Multivariate Behavioral Research, 7, 57-83.*

[42] Williams, D. (1991). *Probability with martingales.* Cambridge: Cambridge University Press.

[43] Wüthrich-Martone, O. (2001). Causal modeling in psychology with qualitative independent variables. Aachen: Shaker.

[44] Wüthrich-Martone, O., Steyer, R., Nachtigall, C. & Suhl, U. (1999). Causality, confounding and unbalanced analysis of variance. In: H. Friedl, A. Berghold and G. Kauermann (eds.), *Statistical Modelling. Proceedings of the 14th International Workshop on Statistical Modelling in Graz, Austria, July 19-23, 1999,* 719-722.

Appendix A: Proofs

Proof of Theorem 4. Following from assumption (b) of Theorem 4

$$E(Y \mid X = x,\ U = u,\ Z = z) = E(Y \mid X = x,\ U = u) \quad \text{for } P\text{-almost all values } z \text{ of } Z.$$

Using this equation and assumption (a), we have:

$$E_{Z=z}(Y \mid X = x) = \textstyle\sum_u E(Y \mid X = x,\ U = u)\ P_{Z=z}(U = u)$$

$$\text{for each value } x \text{ of } X$$

for P-almost all values z of Z. Inserting this equation in the right side of Equation (9), and using the notations $E(Y \mid X = x, Z = z)$ for $E_{Z=z}(Y \mid X = x)$ and $P(U = u \mid Z = z)$ for $P_{Z=z}(U = u)$ we obtain:

$$\int E(Y \mid X = x,\ Z = z)\ P^{Z}(dz)$$

$$= \int \textstyle\sum_u E(Y \mid X = x,\ U = u)\ P(U = u \mid Z = z)\ P^{Z}(dz)$$

$$= \textstyle\sum_u \int E(Y \mid X = x,\ U = u)\ P(U = u \mid Z = z)\ P^{Z}(dz)$$

$$= \textstyle\sum_u E(Y \mid X = x,\ U = u) \int P(U = u \mid Z = z)\ P^{Z}(dz)$$

$$= \textstyle\sum_u E(Y \mid X = x,\ U = u)\ P(U = u),$$

where the theorem of total probability is used in the last row.

Proof of Theorem 5. For the proof of equations (14) and (15) remember that

$$V_i = P(X = x_i \mid Z) = E(I_{x_i} \mid Z), \qquad \text{(a. s)}, \tag{26}$$

where I_{x_i} denotes the indicator (with values 0 and 1) of the event $X = x_i$. Now

$$\begin{aligned}
P(X = x_i \mid V_i) &= E(I_{x_i} \mid V_i), \qquad \text{(a. s)}, \\
&= E[E(I_{x_i} \mid Z) \mid V_i], \qquad \text{(a. s)}, && \text{[Tower property[12]]} \\
&= E(V_i \mid V_i) = V_i, \qquad \text{(a. s)}, \\
&= P(X = x_i \mid Z), \qquad \text{(a. s)}, && \text{[see Eq. (26)]} \\
&= P(X = x_i \mid Z, U), \qquad \text{(a. s)}. && \text{[see Eq. (13)]}
\end{aligned}$$

Since

$$\begin{aligned}
E(I_{x_i} \mid Z) &= E(I_{x_i} \mid Z, U), \qquad \text{(a. s)}, && \text{[see Eq. (13)]} \\
&= E[E(I_{x_i} \mid Z, U) \mid V_i, U] && [E(I_{x_i} \mid Z, U) = V_i \text{ by Eq. (13)}] \\
&= E(I_{x_i} \mid V_i, U) && \text{[Tower property]} \\
&= P(X = x_i \mid V_i, U), \qquad \text{(a. s)}, && (27)
\end{aligned}$$

this proves equations (14) and (15).

For the derivation of Equation (16), note that Equation (13) implies the existence of of $\pi_i(z)$ such that

$$P(U = u \mid X = x_i, Z = z) = \pi_i(z) \qquad \text{for each value } u \text{ of } U, \tag{28}$$

for P-almost all values z of Z, i.e., $\pi_i(z)$ is a constant for given values x_i and z. Let us first consider the left-hand side of Equation (16).

$$\begin{aligned}
E(Y \mid X = x_i, Z = z) &= \textstyle\sum_u E_{Z=z}(Y \mid X = x_i, U = u)\; P(U = u \mid X = x_i, Z = z) \\
&= \textstyle\sum_u E(Y \mid X = x_i, U = u)\; P(U = u \mid X = x_i, Z = z),
\end{aligned} \tag{29}$$

[12] See, e.g., Williams (1991, p. 88). He presents a nice summary of the properties of conditional expectations used in our proofs.

for P-almost all pairs (x_i, z) *of values of* X *and* Z,

because of Equation (8) and Equation (28). We now consider the right-hand side of Equation (16):

$$E\,[Y \mid X = x_i,\, V_i = v_i(z)] = \sum\nolimits_u E_{V_i=v_i(z)}(Y \mid X = x_i,\, U = u)\; P\,[U = u \mid X = x_i,\, V_i = v_i(z)]$$

$$= \sum\nolimits_u E\,(Y \mid X = x_i,\, U = u)\; P\,[U = u \mid X = x_i,\, V_i = v_i(z)] \tag{30}$$

for P-almost all pairs $[x_i, v_i(z)]$ *of values of* X *and* V_i .

Hence, to prove Equation (16), we only have to show that

$$P\,[U = u \mid X = x_i,\, V_i = v_i(z)] = \pi_i(z) \tag{31}$$

for P-almost all pairs $[x_i, v_i(z)]$ *of values of* X *and* V_i .

However, this equation follows immediately from equations (14) and (15).

Proof of Theorem 6. Equation (17) follows immediately from equations (14) and (15), Equation (18) is equivalent to Equation (17), and causal unbiasedness of the conditional regressions $E_{V=v}(Y \mid X)$, for P-almost all values v of V, follows from Theorem 1 and Equation (18). Equation (16) implies $E\,(Y \mid X, Z) = E\,(Y \mid X, V_i)$, (a. s). However

$$E\,(Y \mid X, V) = E\,[E\,(Y \mid X, Z) \mid X, V] =$$

$$= E\,[E\,(Y \mid X, V_i) \mid X, V] = E\,(Y \mid X, V_i), \quad \text{(a. s)},$$

which implies $E\,(Y \mid X, Z) = E\,(Y \mid X, V)$, (a. s), and Equation (19).