Probabilities of Potential Outcome Types in Experimental Studies: Identification and Estimation Based on Proxy Covariate Information

Ryusei Shingaki, Manabu Kuroki

Graduate School of Engineering Science, Yokohama National University
79-5 Tokiwadai, Hodogaya-ku, Yokohama 240-8501 Japan
shingaki-ryusei-kw@ynu.jp, manabu-kuroki-zm@ynu.ac.jp

Abstract

The concept of potential outcome types is one of the fundamental components of causal inference. However, even in randomized experiments, assumptions on the data generating process, such as monotonicity, are required to evaluate the probabilities of the potential outcome types. To solve the problem without such assumptions in experimental studies, a novel identification condition based on proxy covariate information is proposed in this paper. In addition, the estimation problem of the probabilities of the potential outcome types reduces to that of singular models when they are identifiable through the proposed condition. Thus, they cannot be evaluated by standard statistical estimation methods. To overcome this difficulty, new plug-in estimators of these probabilities are presented, and the asymptotic normality of the proposed estimators is shown.

Introduction

Motivation

Simultaneously deriving the results of the same subjects receiving the experimental treatment and the controlled treatment is an essential part of causal inference. However, even in randomized experiments (Pearl 2009, pp.284–285), this cannot be achieved. Thus, it is difficult to evaluate the likelihood that one event would cause another event.

To understand the importance of the above situation, consider the following statement regarding the phase III randomized clinical trial comparing the COVID-19 vaccine with the placebo, as reported by Goodman, Grabenstein, and Braun (2020)

> "The US Food and Drug Administration (FDA) guidance set as an expectation for licensure that a COVID-19 vaccine would prevent disease or decrease its severity in at least 50% of people who are vaccinated. ... It will also be important to understand whether a vaccine reduces not only mild but also more severe disease, as well as hospitalizations and deaths."

This statement implies that it was uncertain whether the severity would be decreased if unvaccinated healthy subjects would be vaccinated (counterfactually) at that time. To administer the vaccine to the subjects most likely to benefit from it, it is useful to (i) classify the situations of the subjects into four outcome types, which are labeled, "doomed", "causative", "preventive" and "immune", and (ii) evaluate the probabilities of these four outcome types. In the phase III randomized clinical trial described above, the "doomed" situation represents a case where treatment is irrelevant because the disease occurs whether the vaccine or the placebo are received. The "causative" situation represents a case where the disease occurs if and only if subjects receive the placebo. The "preventive" situation represents a case where the disease occurs if and only if subjects receive the vaccine. The "immune" situation occurs when the treatment is irrelevant because the disease does not occur where the vaccine or the placebo are received. Because each subject belongs to one of four outcome types, these outcomes are referred to as "potential outcome types". However, we do not know from the observed data to which type a subject belongs. Here, as an effective vaccination policy, it would be better to target subjects in the "causative" outcome type because the severity in this type of subjects is reduced when they are vaccinated, but not when they are not vaccinated. In contrast, the probability of the "preventive" outcome type would be useful to evaluate the severity of receiving the vaccination because the severity in this type of subjects is decreased when they do not receive the vaccination but not when they receive the vaccination.

Theoretical Background and Contribution

One representative example of "the probabilities of the potential outcome types" is "the probabilities of causation", which probabilistically evaluate the "necessity cause", "sufficiency cause", and "necessity and sufficiency cause" (Cai and Kuroki 2005; Dawid, Musio, and Fienberg 2016; Dawid, Musio, and Murtas 2017; Dawid and Musio 2022; Pearl 2015; VanderWeele 2012). Pearl (2009) and Tian and Pearl (2000) developed formal semantics for the probabilities of causation based on structural causal models. These probabilities are formulated based on the probabilities of the potential outcome types, and thus are not identifiable even in randomized experiments (Pearl 2009, pp.284–285). To solve the problem, Kuroki and Cai (2011) and Tian and Pearl (2000) showed how to bound these quantities from data obtained in experimental and observational studies. Although these bounds provide the range within which the probabili-

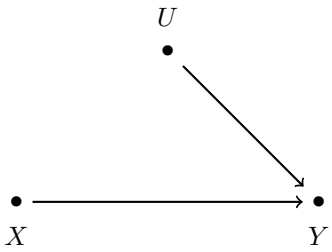


Figure 1: Graphical representation of causal dependencies in a randomized experiment

ties of causation must lie, it has been pointed out that these bounds are too wide to evaluate the probabilities of causation.

To overcome this difficulty, Tian and Pearl (2000) also noted that the probabilities of causation are identifiable if monotonicity can be assumed and causal risks are identifiable. Additionally, Pearl (2009) showed that specific functional relationships between cause and effect lead to the identification of the probabilities of causation. However, when the present assumptions are violated, there has been much less discussion of how to identify the probabilities of the potential outcome types. Referring to the effect restoration by Kuroki and Pearl (2014), Shingaki and Kuroki (2021) provided the identification conditions using the covariate information in observational studies.

In this paper, we provide a novel identification condition for the probabilities of the potential outcome types in a randomized experiment with proxy covariate information. Without relying on the previously used assumptions, the proposed condition enables us to derive a consistent estimator for the probabilities of the potential outcome types under less covariate information than that used in Shingaki and Kuroki (2021). In addition, the estimation problem of the probabilities of the potential outcome types reduces to that of singular models when they are identifiable through the proposed condition. Thus, they cannot be evaluated by standard statistical estimation methods. Although Shingaki and Kuroki (2021) used the augmented Lagrangian method to overcome this difficulty, different from Shingaki and Kuroki (2021), new plug-in estimators of these probabilities are presented in this paper. Given space constraints, the proofs, the details of the statistical estimation method, some numerical experiments and a case study application are provided in the supplementary material.

Problem Description and Notation

To describe our problem, we consider randomized experiments with the purpose of comparing the outcome of an experimental treatment (e.g., the COVID-19 vaccine) with the outcome of a controlled treatment (e.g., the placebo), as shown in Figure 1. For the graph-theoretic terminology and the basic theory of the structural causal models used in this paper, we refer readers to Pearl (2009). In addition, we assume that readers are familiar with the basic theory of causal inference (Imbens and Rubin 2015; Pearl 2009).

Intuitively, in Figure 1, given variables X and Y and the set $\boldsymbol{U}$ of variables, a directed edge from X to Y ($X \to Y$) indicates that X could have an effect on Y. The absence of a directed edge from Y to X ($Y \to X$) indicates that Y cannot be a cause of X. A directed edge from $\boldsymbol{U}$ to Y ($\boldsymbol{U} \to Y$) indicates that some elements of $\boldsymbol{U}$ could have an effect on Y. In addition, the absence of a directed edge from Y to $\boldsymbol{U}$ ($Y \to \boldsymbol{U}$) indicates that Y cannot be a cause of any element of $\boldsymbol{U}$. The absence of edges between X and $\boldsymbol{U}$ ($X \to \boldsymbol{U}$, $X \leftarrow \boldsymbol{U}$, $X \leftrightarrow \boldsymbol{U}$) indicates that there are no cause-effect relationships or associational relationships between X and $\boldsymbol{U}$.

In Figure 1, we assume that X and Y represent the observed dichotomous treatment variable and the observed dichotomous outcome variable, respectively. Here, we let x and y represent the values taken by the variables X and Y, respectively, with the following meanings. $x \in \{x_0, x_1\}$, where x_0 indicates the controlled treatment, and x_1 indicates the experimental treatment; $y \in \{y_0, y_1\}$, where y_0 indicates that there is an occurrence of the disease, and y_1 indicates that there is no occurrence of the disease. In addition, $\boldsymbol{U}$, which is often called a covariate vector, represents the set of all discrete and continuous variables (both observed and unobserved) that cannot be affected by X or Y. Here, some elements in $\boldsymbol{U}$ could have an effect on X and Y in observational studies. However, in randomized experiments, any elements of $\boldsymbol{U}$ are not associated with X because X is randomized. This situation is common in practical science. Here, it is straightforward to extend our results from the case of dichotomous observed variables to the case of multivalued observed variables. In particular, as Balke and Pearl (1997) stated, a multivalued or continuous outcome can be accommodated in the model using the event $Y < y$ as a (dichotomous) outcome variable. For the related discussion, refer to Galhotra, Pradhan, and Salimi (2021) and Kada, Cai, and Kuroki (2013). In addition, when the treatment variable is continuous, according to Balke and Pearl (1997), it is reasonable to assume that there exists a treatment interval around each x, within which a subject's outcome is homogeneous. Under this assumption, it is possible to apply our results.

Let n be the sample size. For $x \in \{x_0, x_1\}$ and $y \in \{y_0, y_1\}$, let $p(X = x, Y = y) = p(x, y)$ be the joint probability of $(X, Y) = (x, y)$, $p(Y = y \,|\, X = x) = p(y \,|\, x)$ be the conditional probability of $Y = y$ given $X = x$, and $p(X = x) = p(x)$ be the marginal probability of $X = x$. A similar notation is used for other probabilities. Then, in principle, for $x \in \{x_0, x_1\}$, the i-th of n subjects has a potential outcome variable $Y_x(i)$ that would have resulted if X had been x for the i-th subject. Here, $Y_x(i) = y$ means that "Y takes the value y when X is experimentally set to x for the i-th subject" or the counterfactual sentence "Y would be y, had X been x for the i-th subject". The potential outcome variable Y_x is observed only if X is x. This property is called the consistency (Robins 1989; Pearl 2009).

In this paper, we assume the stable unit treatment value assumption (Imbens and Rubin 2015), which can be summarized as follows: (i) the treatment status of any subject does not affect the outcomes of the other subjects (i.e., no

interference) and (ii) the treatments of all subjects are comparable (i.e., no variation in treatment). Thus, when the n subjects in the study are considered as random samples from the population of interest, $Y_x(i)$ is referred to as the value of a random variable Y_x.

The causal risk of $X = x$ on $Y = y$ is defined as $p(Y_x = y)$, and the causal risk difference of $X = x_1$ in comparison with $X = x_0$ is defined as $p(Y_{x_1} = y_1) - p(Y_{x_0} = y_1)$. When a randomized experiment is conducted and compliance is perfect, since X is independent of $\{Y_{x_0}, Y_{x_1}\}$, the causal risk $p(Y_x = y)$ is identifiable and is given by

$$p(Y_x = y) = p(y \mid x) \tag{1}$$

from the consistency. Here, "identifiable" means that the causal quantities, such as $p(Y_x = y)$, can be estimated consistently from a joint probability of the observed variables. Although there are other identification conditions of causal risks (e.g., Pearl 2009), they are not covered in this paper due to space constraints. Note that X is considered independent of all covariates in successful randomized experiments.

In contrast, our main interest is to evaluate the probabilities of the potential outcome types in a randomized experiment. Through the pair (Y_{x_0}, Y_{x_1}), "doomed", "causative", "preventive" and "immune", which are described in the section "Motivation", are represented by

$$u_1 = (Y_{x_0} = y_0, Y_{x_1} = y_0), \quad u_2 = (Y_{x_0} = y_0, Y_{x_1} = y_1),$$
$$u_3 = (Y_{x_0} = y_1, Y_{x_1} = y_0), \quad u_4 = (Y_{x_0} = y_1, Y_{x_1} = y_1),$$

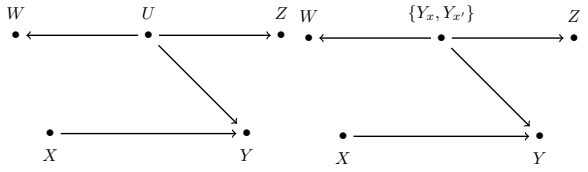
respectively, and the corresponding probabilities we wish to evaluate are given by

$$p(u_1),\ p(u_2),\ p(u_3),\ p(u_4),$$

respectively. These probabilities, which are called probabilities of the potential outcome types, are fundamental components of causal inference in the sense that (i) if they are identifiable, then the causal risk is also identifiable, but not vice versa (Pearl 2009; Tian and Pearl 2000), and (ii) they enable us to evaluate the probabilistic aspects of "necessity cause", "sufficiency cause", and "necessity and sufficiency cause", which are important concepts for providing successful explanations in the field of explainable artificial intelligence (XAI) (Watson et al. 2021). However, these probabilities are not identifiable because Y_{x_1} and Y_{x_0} cannot be simultaneously observed for each subject, even under randomized experiments. Here, under monotonicity, i.e., $p(u_3) = 0$, in randomized experiments, we note that $p(u_1)$, $p(u_2)$ and $p(u_4)$ are identifiable and given by $p(u_1) = p(y_0|x_1)$, $p(u_2) = p(y_1|x_1) - p(y_1|x_0)$, $p(u_3) = 0$ and $p(u_4) = p(y_1|x_0)$ (Pearl 2009; Tian and Pearl 2000).

Identification

Referring to the problem description in the section "Problem Description and Notation", we formulate our problem for evaluating the probabilities of the potential outcome types based on randomized experiments, which are represented by the directed graph shown in Figure 2a. Here, covariates Z and W are measured as proxy variables of the set of covariates $\boldsymbol{U}$. Note that Z and W can be a set of discrete and/or continuous variables. A graphical representation of the data generating process is also shown in this figure.



(a) Unobserved covariates (b) Potential outcomes

Figure 2: Graphical representation of causal dependencies in randomized experiments

$$\begin{aligned} Y &= g_y(X, \boldsymbol{U}, \epsilon_y), \quad X = g_x(\epsilon_x), \\ Z &= g_z(\boldsymbol{U}, \epsilon_z), \quad W = g_w(\boldsymbol{U}, \epsilon_w), \end{aligned} \tag{2}$$

where ϵ_x, ϵ_y, ϵ_z, and ϵ_w are independent random disturbances and are also independent of $\boldsymbol{U}$. When structural equation models, such as Equation (2), are used to represent the data generating process, the corresponding graph, as shown in Figure 2a, is called the causal diagram.

In a situation such as that shown in Figure 2a, $\boldsymbol{U}$ can include the uncertain number of all discrete and continuous covariates that influence the way that a subject responds to treatment. Thus, in many situations, it is reasonable to assume the existence of covariates Z and W that are independent of $\{X, Y\}$ given $\boldsymbol{U}$. Then, irrespective of the complexity of $\boldsymbol{U} \cup \{\epsilon_y, \epsilon_z, \epsilon_w\}$, the impact of $\boldsymbol{U}$ on Y remains restricted to the modification of the functional relationships between X and Y. This yields four functions for two dichotomous variables X and Y; thus, the value taken by $\boldsymbol{U} \cup \{\epsilon_y, \epsilon_z, \epsilon_w\}$ selects one of these four functions (Pearl 2009). Considering these observations, as stated in the section "Motivation", according to Lash et al. (2021, p. 59), the states of $\boldsymbol{U} \cup \{\epsilon_y, \epsilon_z, \epsilon_w\}$ are divided into the following four potential outcome types: "doomed" (u_1), "causative" (u_2), "preventive" (u_3) and "immune" (u_4).

According to the partition of the states of $\boldsymbol{U} \cup \{\epsilon_y, \epsilon_z, \epsilon_w\}$, $\boldsymbol{U}$ is redefined as a variable U taking a value u ($u \in \{u_1, u_2, u_3, u_4\}$). Then, for any x, y, z and w, we assume that Figure 2a can be redescribed, as shown in Figure 2b, and the corresponding recursive factorization of the joint probabilities of Y, Z and W given X, $p(y, z, w \mid x)$ is as follows

$$p(y, z, w \mid x) = \sum_{i=1}^{4} p(y \mid x, u_i) p(z \mid u_i) p(w \mid u_i) p(u_i) \tag{3}$$

according to Figure 2b. When Z has two categories or more and W has three categories or more, i.e., $z \in \{z_1, z_2, \ldots, z_k\}$ ($k \geq 2$) and $w \in \{w_1, w_2, \ldots, w_l\}$ ($l \geq 3$), let

$$Q_{xyw} = \begin{bmatrix} p(y \mid x) & p(y, w \mid x) \\ p(y, z_1 \mid x) & p(y, z_1, w \mid x) \end{bmatrix} \tag{4}$$

for $x \in \{x_0, x_1\}$, $y \in \{y_0, y_1\}$ and $w \in \{w_1, w_2, \ldots, w_l\}$, and

$$R_{x_0 y_0} = \begin{bmatrix} 1 & p(z_1 \mid u_1) \\ 1 & p(z_1 \mid u_2) \end{bmatrix}, \ R_{x_0 y_1} = \begin{bmatrix} 1 & p(z_1 \mid u_3) \\ 1 & p(z_1 \mid u_4) \end{bmatrix},$$

$$R_{x_1y_0} = \begin{bmatrix} 1 & p(z_1 \,|\, u_1) \\ 1 & p(z_1 \,|\, u_3) \end{bmatrix}, R_{x_1y_1} = \begin{bmatrix} 1 & p(z_1 \,|\, u_2) \\ 1 & p(z_1 \,|\, u_4) \end{bmatrix},$$
$$S_{x_0y_0w} = \begin{bmatrix} 1 & p(w \,|\, u_1) \\ 1 & p(w \,|\, u_2) \end{bmatrix}, S_{x_0y_1w} = \begin{bmatrix} 1 & p(w \,|\, u_3) \\ 1 & p(w \,|\, u_4) \end{bmatrix},$$
$$S_{x_1y_0w} = \begin{bmatrix} 1 & p(w \,|\, u_1) \\ 1 & p(w \,|\, u_3) \end{bmatrix}, S_{x_1y_1w} = \begin{bmatrix} 1 & p(w \,|\, u_2) \\ 1 & p(w \,|\, u_4) \end{bmatrix},$$
$$\Delta_{x_0y_0} = \begin{bmatrix} p(u_1) & 0 \\ 0 & p(u_2) \end{bmatrix}, \Delta_{x_0y_1} = \begin{bmatrix} p(u_3) & 0 \\ 0 & p(u_4) \end{bmatrix},$$
$$\Delta_{x_1y_0} = \begin{bmatrix} p(u_1) & 0 \\ 0 & p(u_3) \end{bmatrix}, \Delta_{x_1y_1} = \begin{bmatrix} p(u_2) & 0 \\ 0 & p(u_4) \end{bmatrix}$$

for $w \in \{w_1, w_2, \ldots, w_l\}$. Then, we derive

$$\begin{aligned} Q_{xy_0w_{l_1}} &= R_{xy_0}^\top \Delta_{xy_0} S_{xy_0w_{l_1}}, \\ Q_{xy_0w_{l_2}} &= R_{xy_0}^\top \Delta_{xy_0} S_{xy_0w_{l_2}}, \\ Q_{xy_1w_{l_1}} &= R_{xy_1}^\top \Delta_{xy_1} S_{xy_1w_{l_1}}, \\ Q_{xy_1w_{l_2}} &= R_{xy_1}^\top \Delta_{xy_1} S_{xy_1w_{l_2}}. \end{aligned} \tag{5}$$

for $x \in \{x_0, x_1\}$ and $w_{l_1}, w_{l_2} \in \{w_1, w_2, \ldots, w_l\}$ ($w_{l_1} \neq w_{l_2}$). Here, "$\top$" stands for a transposed vector/matrix. When Q_{xyw} are invertible for $x \in \{x_0, x_1\}$, $y \in \{y_0, y_1\}$ and $w \in \{w_1, w_2, \ldots, w_l\}$, we obtain

$$\begin{aligned} S_{xy_0w_{l_1}} Q_{xy_0w_{l_1}}^{-1} &= S_{xy_0w_{l_2}} Q_{xy_0w_{l_2}}^{-1}, \\ S_{xy_1w_{l_1}} Q_{xy_1w_{l_1}}^{-1} &= S_{xy_1w_{l_2}} Q_{xy_1w_{l_2}}^{-1} \end{aligned} \tag{6}$$

for $x \in \{x_0, x_1\}$ and $w_{l_1}, w_{l_2} \in \{w_1, w_2, \ldots, w_l\}$ ($w_{l_1} \neq w_{l_2}$). Then, we derive the following theorem:

Theorem 1 *Let Z and W be variables that take k (≥ 2) and l (≥ 3) values, e.g., $z \in \{z_1, z_2, \ldots, z_k\}$ for Z and $w \in \{w_1, w_2, \ldots, w_l\}$ for W respectively. Then, the probabilities of the potential outcome types $p(u_1)$, $p(u_2)$, $p(u_3)$ and $p(u_4)$ are identifiable if the following conditions are satisfied:*

(1) Positive probabilities $p(y, z, w|x)$ are available for $x \in \{x_0, x_1\}$, $y \in \{y_0, y_1\}$, $z \in \{z_1, z_2, \ldots, z_k\}$ and $w \in \{w_1, w_2, \ldots, w_l\}$.

(2) Both $W \perp\!\!\!\perp Z \mid U$ and $X \perp\!\!\!\perp \{U, Z, W\}$ hold.

(3) Q_{xyw} are invertible for $x \in \{x_0, x_1\}$, $y \in \{y_0, y_1\}$ and $w \in \{w_1, w_2, \ldots, w_l\}$, and

$$\begin{aligned} &\frac{|Q_{x_0y_0w_{l_1}}|}{|Q_{x_0y_0w_{l_2}}|} \neq \frac{|Q_{x_1y_0w_{l_1}}|}{|Q_{x_1y_0w_{l_2}}|}, \quad \frac{|Q_{x_0y_0w_{l_1}}|}{|Q_{x_0y_0w_{l_2}}|} \neq \frac{|Q_{x_1y_1w_{l_1}}|}{|Q_{x_1y_1w_{l_2}}|}, \\ &\frac{|Q_{x_0y_1w_{l_1}}|}{|Q_{x_0y_1w_{l_2}}|} \neq \frac{|Q_{x_1y_0w_{l_1}}|}{|Q_{x_1y_0w_{l_2}}|}, \quad \frac{|Q_{x_0y_1w_{l_1}}|}{|Q_{x_0y_1w_{l_2}}|} \neq \frac{|Q_{x_1y_1w_{l_1}}|}{|Q_{x_1y_1w_{l_2}}|} \end{aligned} \tag{7}$$

hold for $w_{l_1}, w_{l_2} \in \{w_1, w_2, \ldots, w_l\}$ ($w_{l_1} \neq w_{l_2}$). Here $|\cdot|$ denotes the determinant.

The proof is given in Supplementary Material. Conditions (1) and (3) are testable from observed data. Regarding Condition (2), $X \perp\!\!\!\perp \{U, Z, W\}$ is automatically satisfied in randomized experiments. Here, both Z and W are proxy covariates of U in Figure 2a, but Theorem 1 states that they can be any pair of variables such that both $W \perp\!\!\!\perp Z \mid U$ and $X \perp\!\!\!\perp \{U, Z, W\}$ hold (e.g., $W \to U \to Z$). In addition, although the probabilities that include unobserved variables are not fully identified in Kuroki and Pearl (2014), Theorem 1 shows that the probabilities of the potential outcome types are identifiable. Here, note that the independence assumptions between two observed variables may be affected by partitioning the states of $\boldsymbol{U}$. In addition, although Shingaki and Kuroki (2021) required that both proxy covariates take at least four values, it is sufficient to observe two proxy covariates taking two and three values in Theorem 1.

When Z is a dichotomous variable, Cinelli and Pearl (2021) showed that if $Y_{x_0} \perp\!\!\!\perp Z \mid Y_{x_1}$ holds in a randomized experiment, i.e., $\{Y_0, Y_{x_1}, Z\} \perp\!\!\!\perp X$, then the probabilities of the potential outcome types are identifiable without additional covariate information, e.g., W in Figure 2. Here, note that it is easy to extend the result of Cinelli and Pearl (2021) from an experimental study to an observational study. To understand this, let $x \in \{x_0, x_1\}$, $y \in \{y_0, y_1\}$ and $z \in \{z_1, z_2\}$. When positive probabilities $p(x, y, z)$ are available under the assumption that $p(Y_{x_1} = y_0 \,|\, z_1) \neq p(Y_{x_1} = y_0 \,|\, z_2)$ holds, since $Y_{x_0} \perp\!\!\!\perp Z \mid Y_{x_1}$, we obtain

$$\begin{aligned} &\begin{bmatrix} p(Y_{x_0} = y_1 \,|\, Y_{x_1} = y_0) \\ p(Y_{x_0} = y_1 \,|\, Y_{x_1} = y_1) \end{bmatrix} \\ &= \begin{bmatrix} p(Y_{x_1} = y_0 \,|\, z_1) & p(Y_{x_1} = y_1 \,|\, z_1) \\ p(Y_{x_1} = y_0 \,|\, z_2) & p(Y_{x_1} = y_1 \,|\, z_2) \end{bmatrix}^{-1} \\ &\quad \times \begin{bmatrix} p(Y_{x_0} = y_1 \,|\, z_1) \\ p(Y_{x_0} = y_1 \,|\, z_2) \end{bmatrix}. \end{aligned}$$

Thus, if the conditional causal risk $p(Y_x = y \,|\, z)$ is identifiable by, for example, the back-door criterion (Pearl 2009), in an observational study, then the probabilities of the potential outcome types are also identifiable. Referring to this consideration, it would not be difficult to extend our results from an experimental study to an observational study.

Estimation

When the probabilities of the potential outcome types are identifiable through the proposed condition, as seen from the proof of Theorem 1 (refer to Supplementary Material A), the estimation problem is reduced to that of singular models, and thus, these probabilities cannot be evaluated by standard statistical estimation methods, such as the maximum likelihood estimation method. To solve this problem, we propose new plug-in estimators of the probabilities of the potential outcome types.

Consider the matrices $\widehat{Q}_{xyw}$ that are derived by replacing $p(y \,|\, x)$, $p(w \,|\, x, y)$, $p(z, w \,|\, x, y)$ and $p(z \,|\, x, y)$ of Q_{xyw} with sample probabilities $\widehat{p}(y \,|\, x)$, $\widehat{p}(w \,|\, x, y)$, $\widehat{p}(z, w \,|\, x, y)$ and $\widehat{p}(z \,|\, x, y)$, respectively, for $x \in \{x_1, x_0\}$, $y \in \{y_1, y_0\}$, $z = z_1$ and $w \in \{w_1, w_2\}$. From Equation (6), the consistent estimators of $p(w|u)$ for $w \in \{w_1, w_2\}$ and $u \in \{u_1, u_2, u_3, u_4\}$ are given by

$$\begin{bmatrix} \widehat{p}(w_1 \,|\, u_1) \\ \widehat{p}(w_2 \,|\, u_1) \end{bmatrix} = \begin{bmatrix} 1 & -\dfrac{|\widehat{Q}_{x_0y_0w_1}|}{|\widehat{Q}_{x_0y_0w_2}|} \\ 1 & -\dfrac{|\widehat{Q}_{x_1y_0w_1}|}{|\widehat{Q}_{x_1y_0w_2}|} \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} \widehat{p}(w_1 \,|\, x_0, y_0) - \dfrac{|\widehat{Q}_{x_0y_0w_1}|}{|\widehat{Q}_{x_0y_0w_2}|}\widehat{p}(w_2 \,|\, x_0, y_0) \\ \widehat{p}(w_1 \,|\, x_1, y_0) - \dfrac{|\widehat{Q}_{x_1y_0w_1}|}{|\widehat{Q}_{x_1y_0w_2}|}\widehat{p}(w_2 \,|\, x_1, y_0) \end{bmatrix}, \quad (8)$$

$$\begin{bmatrix} \widehat{p}(w_1 \,|\, u_2) \\ \widehat{p}(w_2 \,|\, u_2) \end{bmatrix} = \begin{bmatrix} 1 & -\dfrac{|\widehat{Q}_{x_0y_0w_1}|}{|\widehat{Q}_{x_0y_0w_2}|} \\ 1 & -\dfrac{|\widehat{Q}_{x_1y_1w_1}|}{|\widehat{Q}_{x_1y_1w_2}|} \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} \widehat{p}(w_1 \,|\, x_0, y_0) - \dfrac{|\widehat{Q}_{x_0y_0w_1}|}{|\widehat{Q}_{x_0y_0w_2}|}\widehat{p}(w_2 \,|\, x_0, y_0) \\ \widehat{p}(w_1 \,|\, x_1, y_1) - \dfrac{|\widehat{Q}_{x_1y_1w_1}|}{|\widehat{Q}_{x_1y_1w_2}|}\widehat{p}(w_2 \,|\, x_1, y_1) \end{bmatrix}, \quad (9)$$

$$\begin{bmatrix} \widehat{p}(w_1 \,|\, u_3) \\ \widehat{p}(w_2 \,|\, u_3) \end{bmatrix} = \begin{bmatrix} 1 & -\dfrac{|\widehat{Q}_{x_0y_1w_1}|}{|\widehat{Q}_{x_0y_1w_2}|} \\ 1 & -\dfrac{|\widehat{Q}_{x_1y_0w_1}|}{|\widehat{Q}_{x_1y_0w_2}|} \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} \widehat{p}(w_1 \,|\, x_0, y_1) - \dfrac{|\widehat{Q}_{x_0y_1w_1}|}{|\widehat{Q}_{x_0y_1w_2}|}\widehat{p}(w_2 \,|\, x_0, y_1) \\ \widehat{p}(w_1 \,|\, x_1, y_0) - \dfrac{|\widehat{Q}_{x_1y_0w_1}|}{|\widehat{Q}_{x_1y_0w_2}|}\widehat{p}(w_2 \,|\, x_1, y_0) \end{bmatrix}, \quad (10)$$

$$\begin{bmatrix} \widehat{p}(w_1 \,|\, u_4) \\ \widehat{p}(w_2 \,|\, u_4) \end{bmatrix} = \begin{bmatrix} 1 & -\dfrac{|\widehat{Q}_{x_0y_1w_1}|}{|\widehat{Q}_{x_0y_1w_2}|} \\ 1 & -\dfrac{|\widehat{Q}_{x_1y_1w_1}|}{|\widehat{Q}_{x_1y_1w_2}|} \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} \widehat{p}(w_1 \,|\, x_0, y_1) - \dfrac{|\widehat{Q}_{x_0y_1w_1}|}{|\widehat{Q}_{x_0y_1w_2}|}\widehat{p}(w_2 \,|\, x_0, y_1) \\ \widehat{p}(w_1 \,|\, x_1, y_1) - \dfrac{|\widehat{Q}_{x_1y_1w_1}|}{|\widehat{Q}_{x_1y_1w_2}|}\widehat{p}(w_2 \,|\, x_1, y_1) \end{bmatrix}. \quad (11)$$

Then, let

$$\begin{aligned} \widehat{S}_{x_0y_0w_1} &= \begin{bmatrix} 1 & \widehat{p}(w_1 \,|\, u_1) \\ 1 & \widehat{p}(w_1 \,|\, u_2) \end{bmatrix}, \\ \widehat{S}_{x_0y_1w_1} &= \begin{bmatrix} 1 & \widehat{p}(w_1 \,|\, u_3) \\ 1 & \widehat{p}(w_1 \,|\, u_4) \end{bmatrix}. \end{aligned} \quad (12)$$

Referring to Equation (6), when we consider

$$\begin{aligned} \widehat{S}_{xy_0w_1}\widehat{Q}^{-1}_{xy_0w_1} &= \widehat{S}_{xy_0w_2}\widehat{Q}^{-1}_{xy_0w_2}, \\ \widehat{S}_{xy_1w_1}\widehat{Q}^{-1}_{xy_1w_1} &= \widehat{S}_{xy_1w_2}\widehat{Q}^{-1}_{xy_1w_2} \end{aligned} \quad (13)$$

for $x \in \{x_0, x_1\}$, the first rows of $\widehat{Q}_{x_0y_0w_1}\widehat{S}^{-1}_{x_0y_0w_1}$ and $\widehat{Q}_{x_0y_1w_1}\widehat{S}^{-1}_{x_0y_1w_1}$ provide the consistent estimators of $(p(u_1), p(u_2))$ and $(p(u_3), p(u_4))$, respectively. In addition, they have the following asymptotic normality:

Theorem 2 *For* $\boldsymbol{\theta}_{ij} = (p(w_1 \,|\, x_i, y_j),\ p(w_2 \,|\, x_i, y_j),\ p(z_1, w_1 \,|\, x_i, y_j),\ p(z_1, w_2 \,|\, x_i, y_j),\ p(z_1 \,|\, x_i, y_j))^\top$ *(denoted as* $(\theta^{ij}_1, \theta^{ij}_2, \theta^{ij}_3, \theta^{ij}_4, \theta^{ij}_5)^\top$*) and* $T^{(n)}_{ij} = (\widehat{p}(w_1 \,|\, x_i, y_j),\ \widehat{p}(w_2 \,|\, x_i, y_j),\ \widehat{p}(z_1, w_1 \,|\, x_i, y_j),\ \widehat{p}(z_1, w_2 \,|\, x_i, y_j),\ \widehat{p}(z_1 \,|\, x_i, y_j))^\top$ $(i, j = 0, 1)$*, suppose that* $\sqrt{n}\left(T^{(n)}_{ij} - \boldsymbol{\theta}_{ij}\right)$ *asymptotically follows the joint normal distribution with a zero mean vector and a covariance matrix* Σ_{ij} *for* $i, j = 0, 1$*. Then,* $\sqrt{n}\Big(\widehat{S}^{-\top}_{x_0y_0w_1}\widehat{Q}^{\top}_{x_0y_0w_1}\boldsymbol{e} - (p(u_1), p(u_2))^\top\Big)$ *asymptotically follows the joint normal distribution with a zero mean vector and a covariance matrix*

$$\begin{bmatrix} \dfrac{\partial S^{-\top}_{x_0y_0w_1}Q^{\top}_{x_0y_0w_1}\boldsymbol{e}}{\partial \boldsymbol{\theta}_{00}} \\ \dfrac{\partial S^{-\top}_{x_0y_0w_1}Q^{\top}_{x_0y_0w_1}\boldsymbol{e}}{\partial \boldsymbol{\theta}_{10}} \\ \dfrac{\partial S^{-\top}_{x_0y_0w_1}Q^{\top}_{x_0y_0w_1}\boldsymbol{e}}{\partial \boldsymbol{\theta}_{11}} \end{bmatrix}^\top \begin{bmatrix} \Sigma_{00} & O & O \\ O & \Sigma_{10} & O \\ O & O & \Sigma_{11} \end{bmatrix} \times \begin{bmatrix} \dfrac{\partial S^{-\top}_{x_0y_0w_1}Q^{\top}_{x_0y_0w_1}\boldsymbol{e}}{\partial \boldsymbol{\theta}_{00}} \\ \dfrac{\partial S^{-\top}_{x_0y_0w_1}Q^{\top}_{x_0y_0w_1}\boldsymbol{e}}{\partial \boldsymbol{\theta}_{10}} \\ \dfrac{\partial S^{-\top}_{x_0y_0w_1}Q^{\top}_{x_0y_0w_1}\boldsymbol{e}}{\partial \boldsymbol{\theta}_{11}} \end{bmatrix},$$

and $\sqrt{n}\left(\widehat{S}^{-\top}_{x_0y_1w_1}\widehat{Q}^{\top}_{x_0y_1w_1}\boldsymbol{e} - (p(u_3), p(u_4))^\top\right)$ *asymptotically follows the joint normal distribution with a zero mean vector and a covariance matrix*

$$\begin{bmatrix} \dfrac{\partial S^{-\top}_{x_0y_1w_1}Q^{\top}_{x_0y_1w_1}\boldsymbol{e}}{\partial \boldsymbol{\theta}_{01}} \\ \dfrac{\partial S^{-\top}_{x_0y_1w_1}Q^{\top}_{x_0y_1w_1}\boldsymbol{e}}{\partial \boldsymbol{\theta}_{10}} \\ \dfrac{\partial S^{-\top}_{x_0y_1w_1}Q^{\top}_{x_0y_1w_1}\boldsymbol{e}}{\partial \boldsymbol{\theta}_{11}} \end{bmatrix}^\top \begin{bmatrix} \Sigma_{01} & O & O \\ O & \Sigma_{10} & O \\ O & O & \Sigma_{11} \end{bmatrix} \times \begin{bmatrix} \dfrac{\partial S^{-\top}_{x_0y_1w_1}Q^{\top}_{x_0y_1w_1}\boldsymbol{e}}{\partial \boldsymbol{\theta}_{01}} \\ \dfrac{\partial S^{-\top}_{x_0y_1w_1}Q^{\top}_{x_0y_1w_1}\boldsymbol{e}}{\partial \boldsymbol{\theta}_{10}} \\ \dfrac{\partial S^{-\top}_{x_0y_1w_1}Q^{\top}_{x_0y_1w_1}\boldsymbol{e}}{\partial \boldsymbol{\theta}_{11}} \end{bmatrix}$$

under the assumptions of Theorem 1. Here the notation "$-\top$*" stands for a transposed inverse matrix and* $\boldsymbol{e} = (1, 0)^\top$.

The proof of Theorem 2 is straightforward by the delta method (e.g., van der Vaart 1998).

Numerical Experiment

In this section, we present a numerical experiment to examine the properties of the proposed estimation method through the "probability of necessity and sufficiency" (PNS) $p(u_2)$, which has been discussed in the context of the probabilities of causation, and the causal risk difference $p(Y_{x_1} = y_1) - p(Y_{x_0} = y_1) = p(u_2) - p(u_3)$. For simplicity, let X, Y, Z, W, and U be discrete variables. Then, we consider the causal diagrams shown in Figure 2, where the joint

	$p(Z \mid U)$		$p(W \mid U)$			$p(Y=0 \mid X,U)$		$p(U)$	$p(X)$	
	$Z=1$	$Z=2$	$W=1$	$W=2$	$W=3$	$X=1$	$X=0$			
$U=1$	9/10	1/10	8/10	3/20	1/20	1	1	1/4	$X=0$	1/2
$U=2$	8/10	2/10	3/20	1/20	8/10	0	1	1/4		
$U=3$	2/10	8/10	7/10	2/10	1/10	1	0	1/4	$X=1$	1/2
$U=4$	1/10	9/10	2/10	1/10	7/10	0	0	1/4		

Table 1: Conditional probabilities

probabilities of (X, Y, Z, W, U) are given according to Table 1. Under the situation where (X, Y, Z, W) can be observed but U cannot, the properties of the proposed estimators $\widehat{p}(u_2)$ and $\widehat{p}(u_2) - \widehat{p}(u_3)$ of $p(u_2)$ and $p(u_2) - p(u_3)$, respectively, are verified in a numerical experiment using the setting with sample sizes $n = 500$, 1000, 5000 and 10000. In this situation, since $p(u_2)$ and $p(u_2) - p(u_3)$ are 0.25 and 0.00, respectively, the sample means of $\widehat{p}(u_2)$ and $\widehat{p}(u_2) - \widehat{p}(u_3)$ are expected to be close to 0.25 and 0.00, respectively. Table 2 and Figure 3 show the basic statistics and the box plots of $\widehat{p}(u_2)$ and $\widehat{p}(u_2) - \widehat{p}(u_3)$ for 1000 replications with the given sample size n, respectively. Note that if the determinants of $\widehat{Q}_{xyw}$ are close to zero, then Condition (3) in Theorem 1 may not be satisfied. Thus, Table 2 and Figure 3 are considered with the exception of a maximum of 85 cases. The horizontal lines in Figure 3 show the true values of $p(u_2)$ and $p(u_2) - p(u_3)$. Here, since we focus on the randomized experiments, $p(u_2) - p(u_3)$ can be estimated by $\widehat{p}(y_1|x_1) - \widehat{p}(y_1|x_0)$ without the proposed estimation method. Although we know that $\widehat{p}(y_1|x_1) - \widehat{p}(y_1|x_0)$ provides better estimation accuracy than the proposed estimation method due to the smaller value of $|\widehat{Q}_{xyw}|$ in Equations (8)-(11), we estimate $p(u_2) - p(u_3)$ by the proposed estimation method for our purpose.

From Table 2, the sample means of $\widehat{p}(u_2)$ and $\widehat{p}(u_2) - \widehat{p}(u_3)$ are close to the true values and the sample standard errors (s.e.) are smaller as the sample size is larger. Thus, it seems that the proposed estimation method provides consistent estimators for $p(u_2)$ and $p(u_2) - p(u_3)$. In addition, from Figure 3, the interquantile ranges for $\widehat{p}(u_2)$ and $\widehat{p}(u_2) - \widehat{p}(u_3)$ are narrower and still include the true values even if the sample size is large. In contrast, it seems that $\widehat{p}(u_2) - \widehat{p}(u_3)$ is symmetrically distributed, and thus, the asymptotic normality holds, but $\widehat{p}(u_2)$ may not hold for the smaller sample size. This is because the true value of $p(u_2)$ is relatively close to zero for the finite sample size. Actually, the boxplot of $\widehat{p}(u_2)$ for $n = 10000$ is more symmetrical than those for $n \leq 5000$. Here, note that there are many outliers in each sample size. Outliers occur when it is difficult to judge that the conditions of Theorem 1 hold from the observed data. For further discussion of the simulation experiments, see the supplementary material.

Discussion

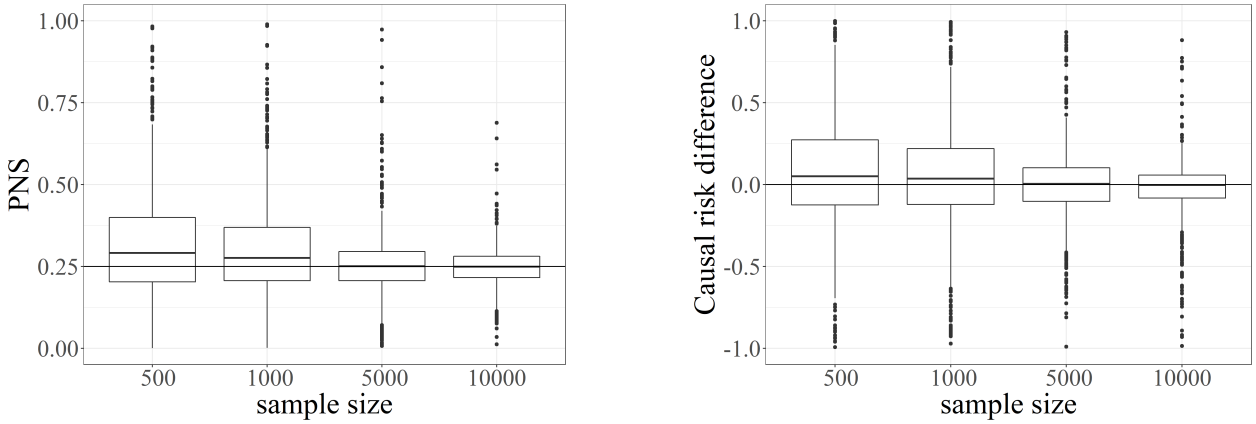
The probabilities of the potential outcome types often appear in unit selection problems (Li and Pearl 2019), the impact evaluation problem of social programs (Heckman, Smith, and Clements 1997), the non-compliance problem of treatment effects (Angrist, Imbens, and Rubin 1996; Balke and Pearl 1997), the principal stratification (Frangakis and Rubin 2002), the identification problems of natural direct and indirect effects (Pearl 2001) and prevented and preventable proportions (Yamada and Kuroki 2017), traffic conflict (Yamada and Kuroki 2019), the explainability problem of artificial intelligence (Watson et al. 2021), and causal classification (Fernández-Loría and Provost 2022). Therefore, the identification and estimation problems of these probabilities have been an important topic in causal inference. To solve these problems in experimental studies, we have proposed a novel identification condition of the probabilities of the potential outcome types. While Miao, Geng, and Tchetgen Tchetgen (2018) and Lee and Bareinboim (2021) discussed the identification problem of causal effects with proxy variables, we have rather discussed the identification problem of the probabilities of the potential outcome types. In addition, the proposed condition enables us to derive a consistent estimator for the probabilities of the potential outcome types under less covariate information than that used in Shingaki and Kuroki (2021). When the probabilities of the potential outcome types are identifiable through the proposed condition, the estimation problem is reduced to that of singular models. To overcome this difficulty, we proposed new plug-in estimators of these probabilities. Here, the estimation problem of singular models often appears in the context of causal inference, but it seems that no solution has been given in the literature. Thus, the results of this paper extend the range of solvable evaluation problems in causal inference under nonparametric causal models.

To estimate the joint probabilities of potential outcomes, the proposed method needs to input two proxy variables that satisfy the conditions in Theorem 1. As stated in the section "Identification", $\boldsymbol{U}$ can include the uncertain number of all discrete and continuous covariates that influence the way that a subject responds to treatment. Thus, in many situations, it is reasonable to assume the existence of covariates Z and W that are independent given $\boldsymbol{U}$.

As seen from Figure 3, the proposed estimation methods could be unstable in practice due to matrix inversion. Although the paper focuses on the identification and estimation problems, the problem of the variance estimation is also an important topics in causal inference. Empirically, we note that the unstable estimations would be caused by (1) the lower association between $\boldsymbol{U}$ and the proxy covariates and (2) the small sample size. We leave the problem of such

	(a) $\widehat{p}(u_2)$				(b) $\widehat{p}(u_2) - \widehat{p}(u_3)$			
	$n = 500$	$n = 1000$	$n = 5000$	$n = 10000$	$n = 500$	$n = 1000$	$n = 5000$	$n = 10000$
Minimum	-9.609	-15.024	-2.530	-0.480	-19.961	-14.944	-4.966	-12.621
1st Quantile	0.164	0.177	0.194	0.213	-0.157	-0.135	-0.110	-0.084
Median	0.275	0.267	0.247	0.247	0.051	0.044	0.003	-0.003
Mean	0.314	0.291	0.284	0.243	0.067	0.084	0.050	-0.026
3rd	0.406	0.369	0.295	0.280	0.325	0.255	0.110	0.059
Maximum	17.233	8.404	15.692	2.053	17.236	8.196	15.494	3.194
s.e.	1.047	0.921	0.764	0.116	1.603	1.285	0.932	0.494

Table 2: Basic statistics of estimates based on the proposed method



(a) Boxplots for PNS

(b) Boxplots for causal risk difference

Figure 3: Boxplots of estimates based on the proposed method

variance estimation as future work.

Finally, we assume that both a treatment variable and an outcome variable are dichotomous. As we stated in the section "Problem Description and Notation", it is straightforward to extend our results from the case of dichotomous observed variables to the case of multivalued observed variables under certain assumptions. Here, note that in such cases, it may be difficult to obtain reliable statistics of the recovered probabilities due to data sparseness. This problem has been left for future work.

Acknowledgments

We would like to acknowledge the helpful comments of the four anonymous reviewers. This research was partially funded by Mitsubishi Electric Corporation and Japan Society for the Promotion of Science (JSPS), Grant Number 19K11856 and 21H03504.

References

Angrist, J. D.; Imbens, G. W.; and Rubin, D. B. 1996. Identification of Causal Effects Using Instrumental Variables. *J. Amer. Statist. Assoc.*, 91(434): 444–455.

Balke, A.; and Pearl, J. 1997. Bounds on Treatment Effects From Studies With Imperfect Compliance. *J. Amer. Statist. Assoc.*, 92(439): 1171–1176.

Cai, Z.; and Kuroki, M. 2005. Variance estimators for three "probabilities of causation". *Risk Analysis: An International Journal*, 25(6): 1611–1620.

Cinelli, C.; and Pearl, J. 2021. Generalizing experimental results by leveraging knowledge of mechanisms. *European Journal of Epidemiology volume*, 36: 149–164.

Dawid, A. P.; and Musio, M. 2022. Effects of Causes and Causes of Effects. *Annual Review of Statistics and Its Application*, 9(1): 261–287.

Dawid, A. P.; Musio, M.; and Fienberg, S. E. 2016. From statistical evidence to evidence of causality. *Bayesian Analysis*, 11(3): 725–752.

Dawid, A. P.; Musio, M.; and Murtas, R. 2017. The probability of causation. *Law, Probability and Risk*, 16(4): 163–179.

Fernández-Loría, C.; and Provost, F. 2022. Causal Classification: Treatment Effect Estimation vs. Outcome Prediction. *Journal of Machine Learning Research*, 23(59): 1–35.

Frangakis, C. E.; and Rubin, D. B. 2002. Principal Stratification in Causal Inference. *Biometrics*, 58(1): 21–29.

Galhotra, S.; Pradhan, R.; and Salimi, B. 2021. Explaining black-box algorithms using probabilistic contrastive counterfactuals. In *Proceedings of the 2021 International Conference on Management of Data*, 577–590.

Goodman, J. L.; Grabenstein, J. D.; and Braun, M. M. 2020. Answering key questions about COVID-19 vaccines. *Journal of the American Medical Association*, 324(20): 2027–2028.

Heckman, J. J.; Smith, J.; and Clements, N. 1997. Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts. *The Review of Economic Studies*, 64(4): 487–535.

Imbens, G. W.; and Rubin, D. B. 2015. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Kada, A.; Cai, Z.; and Kuroki, M. 2013. Medical diagnostic test based on the potential test result approach: bounds and identification. *Journal of Applied Statistics*, 40(8): 1659–1672.

Kuroki, M.; and Cai, Z. 2011. Statistical Analysis of 'Probabilities of Causation' Using Co-variate Information. *Scand. J. Statist.*, 38(3): 564–577.

Kuroki, M.; and Pearl, J. 2014. Measurement bias and effect restoration in causal inference. *Biometrika*, 101(2): 423–437.

Lash, T.; VanderWeele, T.; Haneuse, S.; and Rothman, K. 2021. *Modern Epidemiology*. Lippincott Williams & Wilkins, 4th edition.

Lee, S.; and Bareinboim, E. 2021. Causal Identification with Matrix Equations. In Ranzato, M.; Beygelzimer, A.; Dauphin, Y.; Liang, P.; and Vaughan, J. W., eds., *Advances in Neural Information Processing Systems*, volume 34, 9468–9479. Curran Associates, Inc.

Li, A.; and Pearl, J. 2019. Unit Selection Based on Counterfactual Logic. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, 1793–1799. International Joint Conferences on Artificial Intelligence Organization.

Miao, W.; Geng, Z.; and Tchetgen Tchetgen, E. J. 2018. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4): 987–993.

Pearl, J. 2001. Direct and Indirect Effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, 411–420. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Pearl, J. 2009. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2nd edition.

Pearl, J. 2015. Causes of effects and effects of causes. *Sociological Methods & Research*, 44(1): 149–164.

Robins, J. 1989. The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In Sechrest, L.; Freeman, H.; and Mulley, A., eds., *Health Service Research Methodology: A Focus on AIDS*, 113–159. U.S. Public Health Service, Washington, DC.

Shingaki, R.; and Kuroki, M. 2021. Identification and Estimation of Joint Probabilities of Potential Outcomes in Observational Studies with Covariate Information. In Ranzato, M.; Beygelzimer, A.; Dauphin, Y.; Liang, P.; and Vaughan, J. W., eds., *Advances in Neural Information Processing Systems*, volume 34, 26475–26486. Curran Associates, Inc.

Tian, J.; and Pearl, J. 2000. Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence*, 28(1): 287–313.

van der Vaart, A. W. 1998. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

VanderWeele, T. J. 2012. The sufficient cause framework in statistics, philosophy and the biomedical and social sciences. In Berzuini, C.; Dawid, P.; and Bernardinell, L., eds., *Causality: Statistical Perspectives and Applications*, chapter 13, 180–191. Wiley.

Watson, D. S.; Gultchin, L.; Taly, A.; and Floridi, L. 2021. Local explanations via necessity and sufficiency: unifying theory and practice. In de Campos, C.; and Maathuis, M. H., eds., *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, volume 161 of *Proceedings of Machine Learning Research*, 1382–1392. PMLR.

Yamada, K.; and Kuroki, M. 2017. Counterfactual-Based Prevented and Preventable Proportions. *Journal of Causal Inference*, 5(2): 20160020.

Yamada, K.; and Kuroki, M. 2019. New Traffic Conflict Measure Based on a Potential Outcome Model. *Journal of Causal Inference*, 7(1): 20180001.