Statistical Science
2021, Vol. 36, No. 4, 493–508
https://doi.org/10.1214/20-STS810

Identification of Causal Effects Within Principal Strata Using Auxiliary Variables

Zhichao Jiang and Peng Ding

Abstract. In causal inference, principal stratification is a framework for dealing with a posttreatment intermediate variable between a treatment and an outcome. In this framework, the principal strata are defined by the joint potential values of the intermediate variable. Because the principal strata are not fully observable, the causal effects within them, also known as the principal causal effects, are not identifiable without additional assumptions. Several previous empirical studies leveraged auxiliary variables to improve the inference of principal causal effects. We establish a general theory for the identification and estimation of principal causal effects with auxiliary variables, which provides a solid foundation for statistical inference and more insights for model building in empirical research. In particular, we consider two commonly used assumptions for principal stratification problems: principal ignorability and the conditional independence between the auxiliary variable and the outcome given principal strata and covariates. Under each assumption, we give nonparametric and semiparametric identification results without modeling the outcome. When neither assumption is plausible, we propose a large class of flexible parametric and semiparametric models for identifying principal causal effects. Our theory not only establishes formal identification results of several models that have been used in previous empirical studies but also generalizes them to allow for different types of outcomes and intermediate variables.

Key words and phrases: Augmented design, auxiliary independence, identification, principal ignorability, principal stratification.

1. INTRODUCTION

Complications arise in causal inference with an intermediate variable between the treatment and the outcome. Cochran (1957), Rosenbaum (1984) and Frangakis and Rubin (2002) pointed out that naively conditioning on the observed intermediate variable does not yield valid causal interpretations in general. Frangakis and Rubin (2002) proposed to use principal stratification, the joint potential values of the intermediate variable under both the treatment and control, to define subgroup causal effects, because it acts as a pretreatment covariate vector unaffected by the treatment. Principal stratification has a wide range of applications with meanings varying in different scientific contexts. In noncompliance problems where the treatment received might differ from the treatment assigned, principal stratification represents individual potential compliance behavior (Angrist, Imbens and Rubin, 1996). In truncation-by-death problems where some units die before the measurement time point of their outcomes, principal stratification represents individual potential survival status (Rubin, 2006). In surrogate evaluation problems, principal stratification helps to clarify criteria for good surrogate endpoints (Frangakis and Rubin, 2002, Gilbert and Hudgens, 2008). In mediation analysis, principal stratification with respect to the mediator represents different causal mechanisms from the treatment to the outcome (Rubin, 2004, Gallop et al., 2009, Elliott, Raghunathan and Li, 2010, Mattei and Mealli, 2011). VanderWeele (2008) and Forastiere, Mattei and Ding (2018) linked the principal stratification approach with the direct and indirect effect approach and Jo (2008) linked the principal stratification approach with structural equation model for mediation analysis. These problems with intermediate variables concern the average causal ef-

Zhichao Jiang is Assistant Professor, Department of Biostatistics and Epidemiology, University of Massachusetts, Amherst, Massachusetts 01003, USA (e-mail: zhichaojiang@umass.edu). Peng Ding is Associate Professor, Department of Statistics, University of California, Berkeley 94720, USA (e-mail: pengdingpku@berkeley.edu).

fects within principal strata, which are also known as the principal causal effects (PCEs).

Because we cannot simultaneously observe the potential values of the intermediate variable under the treatment and control, we do not know the principal stratum of every individual, and thus cannot identify the PCEs without additional assumptions. For a binary intermediate variable, Zhang and Rubin (2003), Cheng and Small (2006) and Imai (2008) derived large sample bounds, which can be too wide to be informative. Angrist, Imbens and Rubin (1996), Little and Yau (1998), Zhang, Rubin and Mealli (2009) and Frumento et al. (2012) imposed additional structural or modeling assumptions to achieve identification. When the intermediate variable is continuous, identification becomes more difficult because of the infinitely many principal strata. To estimate the PCEs, Gilbert and Hudgens (2008) assumed parametric models and used a likelihood approach. Jin and Rubin (2008), Schwartz, Li and Mealli (2011), and Zigler and Belin (2012) proposed different forms of parametric and semiparametric Bayesian approaches. However, the identifiability of their models is not formally established. Without identifiability, the likelihood function may be flat over a region of some parameters, and the Bayesian inference can be sensitive to prior specifications. See Gustafson (2009) and Ding and Li (2018) for more discussion on identifiability.

Identification is sometimes achievable with a pretreatment auxiliary variable satisfying some conditional independence assumptions. We focus on two categories. The first category assumes that the outcome is independent of the principal strata given the auxiliary variable. This assumption is known as *principal ignorability* (Jo et al., 2011, Ding and Lu, 2017). Under principal ignorability, Jo and Stuart (2009) and Stuart and Jo (2015) used principal scores to analyze data with one-sided noncompliance, and Joffe, Small and Hsu (2007) suggested using principal scores to estimate general PCEs. Ding and Lu (2017) established formal identification results for PCEs with a binary intermediate variable in randomized experiments. The other category assumes the conditional independence between the outcome and the auxiliary variable within principal strata. We will refer to this conditional independence as *auxiliary independence*. This assumption motivates several identification and estimation strategies in different contexts. For a binary intermediate variable indicating the survival status, Ding et al. (2011) used the baseline quality of life as an auxiliary variable to help to identify the effect of a treatment on the quality of life which is truncated by death. Under monotonicity, Mealli and Pacini (2013) relaxed Ding et al.'s (2011) assumptions and discussed bounds and identification of the PCEs with a binary secondary outcome. Wang, Zhou and Richardson (2017) extended the strategy to observational studies and relaxed monotonicity in a sensitivity analysis. In a study with multiple independent trials, Jiang, Ding and Geng (2016) used the trial number as an auxiliary variable and proposed strategies to identify the PCEs. Yuan, Feller and Miratrix (2019) weakened the identification assumptions and applied the methodology to a multisite trial in education. Similar ideas have also been used to deal with continuous intermediate variables. In assessing the effect of an HIV vaccine on infection rate through immune response, Follmann (2006) used the baseline immune response to the rabies vaccine as an auxiliary variable. Qin et al. (2008) extended this idea to deal with time-to-event endpoints under a case-cohort sampling. Gilbert and Hudgens (2008) and Huang and Gilbert (2011) proposed approaches to evaluating biomarkers based on principal stratification by incorporating baseline covariates as auxiliary variables to predict the biomarkers. These strategies also provided insights for better experimental designs. In particular, Gabriel and Follmann (2016) proposed the augmented treatment run-in design and used a baseline measure as a predictor of the potential values of the intermediate variable. However, under auxiliary independence, formal identification results are established only for binary intermediate variables (Ding et al., 2011, Mealli and Pacini, 2013, Jiang, Ding and Geng, 2016).

This paper discusses the identification of PCEs defined by a general intermediate variable with auxiliary variables. We first generalize the identification results under principal ignorability in Ding and Lu (2017) to general intermediate variables in both randomized experiments and observational studies, and then study the identification under auxiliary independence in various scenarios. With auxiliary independence, we establish nonparametric identification results for discrete intermediate variables and semiparametric identification results for continuous intermediate variables. These results do not require modeling the outcome. Without principal ignorability or auxiliary independence, we propose a large class of parametric models to identify the PCEs, which have not been formally established before. Compared with models used in previous empirical studies, our models require weaker assumptions and can deal with different types of data.

Identifiability is a cornerstone for both frequentists' (Bickel and Doksum, 2015) and Bayesian (Gustafson, 2015) inferences. Our results provide theoretical bases to check the identifiability of PCEs. Practitioners can use our results to guide model building for principal stratification problems. Our results imply that some existing models are identifiable but some are not (e.g., Follmann, 2006, Gilbert and Hudgens, 2008, Zigler and Belin, 2012). Moreover, our results reveal that some existing models invoked unnecessary assumptions for identification, for example, restricting the parameter space or imposing informative priors, although these assumptions can improve finite-sample inference.

The paper uses the following notation. Let i.i.d. denote "independently and identically distributed," $A \perp\!\!\!\perp B \mid C$ denote the conditional independence of A and B given C, and $A \overset{\mathrm{d}}{=} B$ denote that A has the same distribution as B. Let $\mathbf{1}(\cdot)$ be the indicator function, $\mathbb{P}(\cdot)$ be the probability mass or density function, and $\Phi(\cdot)$ be the cumulative distribution function of the standard Normal distribution. We say that functions $\{f_1(x), \ldots, f_J(x)\}$ are *linearly independent* if $c_1 f_1(x) + \cdots + c_J f_J(x) = 0$ for all x implies $c_1 = \cdots = c_J = 0$. We say that a family $\mathcal{Q}$ of probability distributions is *complete* if $\int f(v)Q(\mathrm{d}v) = 0$ for all $Q \in \mathcal{Q}$ implies $f(v) = 0$ a.e. $\mathcal{Q}$ (Lehmann and Romano, 2005).

2. NOTATION AND ASSUMPTIONS

Let Z be a binary treatment indicator with $Z = 1$ for the treatment and 0 for the control, Y be an outcome of interest, and S be an intermediate variable between the treatment and outcome. Let S_{iz} and Y_{iz} be the potential values of the intermediate variable and the outcome if unit i were to receive treatment z ($z = 0, 1$). The observed values of the intermediate variable and the outcome are

$$S_i = Z_i S_{i1} + (1 - Z_i)S_{i0}, \quad Y_i = Z_i Y_{i1} + (1 - Z_i)Y_{i0}.$$

Assume that $\{Z_i, S_{i1}, S_{i0}, Y_{i1}, Y_{i0} : i = 1, \ldots, n\}$ are i.i.d. samples drawn from an infinite superpopulation, and thus the observed $\{Z_i, S_i, Y_i : i = 1, \ldots, n\}$ are also i.i.d. As a result, we will drop the subscript i for notational simplicity when no confusion would arise.

Frangakis and Rubin (2002) defined principal stratification as $U_i = (S_{i1}, S_{i0})$, the joint potential values of the intermediate variable, and the PCEs as

$$\tau_{s_1 s_0} = \mathbb{E}\{Y_1 - Y_0 \mid U = (s_1, s_0)\}$$

for all s_1, s_0. The PCEs are not identifiable because U is latent in general. It is common to exploit a pretreatment auxiliary variable for identifying the PCEs. Let W_i denote this variable with meanings varying in different settings. We start with the following basic assumption.

ASSUMPTION 1. $Z \perp\!\!\!\perp (Y_1, Y_0, S_1, S_0) \mid W$.

Assumption 1 is sometimes guaranteed by design. In completely randomized experiments, Assumption 1 holds because $Z \perp\!\!\!\perp (Y_1, Y_0, S_1, S_0, W)$. In a multicenter experiment with W being the center number, Assumption 1 holds because Z is randomized in each center.

We consider two different assumptions for identification. The first assumption is the conditional independence between the potential outcome Y_z and the principal stratum U given the auxiliary variable W.

ASSUMPTION 2 (Principal ignorability). $Y_z \perp\!\!\!\perp U \mid W$ for $z = 0, 1$.

Assumption 2 means that given auxiliary variable W, the principal stratification variable is randomly assigned with respect to the potential outcomes. It requires that no difference exists between the distributions of the potential outcomes across principal strata given the auxiliary variable. Many applied researchers have invoked it to estimate the PCEs (Follmann, 2000, Jo and Stuart, 2009, Jo et al., 2011, Stuart and Jo, 2015). Assumption 2 can be weakened (Ding and Lu, 2017, Forastiere, Mattei and Ding, 2018), but we present it for simplicity. To make Assumption 2 more plausible, researchers often include all pretreatment covariates in W. We provide two examples below.

EXAMPLE 1. Follmann (2000) studied the effect of a multifactor intervention on mortality due to coronary heart disease, where Z is the indicator of the intervention and Y is the survival time of the patients. One-sided noncompliance occurred in the experiment, where patients assigned to the treatment group might not actually take the treatment. Let S denote the actual treatment, which can be different from Z. Then, the principal stratification variable characterizes the compliance behavior of the patients. Follmann (2000) argued that the potential survival time of the patients with different compliance behavior would be similar conditional on pretreatment covariates W.

EXAMPLE 2. Ding and Lu (2017) gave an example of a randomized experiment with truncation-by-death, where Z is the treatment indicator, S is the binary survival status, and Y is the health-related quality of life. Because the outcome is only well-defined for the survived patients, the parameter of interest is the PCE within the stratum of the patients who would survive regardless of the treatment. They used all the covariates as the auxiliary variables and invoked principal ignorability in their analysis, which requires that the health-related quality of life for always survived patients would be identical to that for other patients given the covariates.

The second identification assumption is the conditional independence between the potential outcome Y_z and the auxiliary variable W given the principal stratum U.

ASSUMPTION 3 (Auxiliary independence). $Y_z \perp\!\!\!\perp W \mid U$ for $z = 0, 1$.

Assumption 3 requires the units with different values of the auxiliary variable to have the same distribution of potential outcomes if they are in the same principal stratum. Under Assumption 1, we can show that Assumption 3 is equivalent to $Y \perp\!\!\!\perp W \mid (Z, U)$, that is, the auxiliary variable is independent of the outcome conditional on the treatment and principal strata. Including additional pretreatment covariates can make this assumption more

plausible. However, for notational simplicity, we condition on such covariates implicitly and omit them below. In some situations, Assumption 3 is justifiable by design. We illustrate it using two examples.

EXAMPLE 3. Follmann (2006) introduced an augmented design to assess immune response in vaccine trials, where Z is the indicator of an HIV vaccine injection, S is the immune response to this vaccine, and Y is the infection indicator. Before the randomization of Z, all patients receive the rabies vaccine. Let W denote the immune response to the rabies vaccine, which is correlated with S. Because the rabies vaccine is irrelevant to the HIV infection, the potential HIV infection status should depend only on the immune response to the HIV vaccine but not the rabies vaccine. This justifies auxiliary independence.

EXAMPLE 4. Jiang, Ding and Geng (2016) proposed approaches to identifying the PCEs by multiple independent trials, where Z is the treatment indicator, S is the indicator of three-year cancer reoccurrence, and Y is the five-year survival status. The data are from multiple trials with the trial number denoted by W. Jiang, Ding and Geng (2016) argued that the principal stratification variable is a measure of physical status, and assumed that the potential survival status does not depend on the trial number W given the patient's physical status. So auxiliary independence is plausible in their study.

When S is binary as in Example 4, Jiang, Ding and Geng (2016) showed the identifiability of PCEs. With a general S as in Example 3, formal identification results have not been established although several parametric or semiparametric models have been used in empirical studies.

In the following two sections, we will give a unified theory for the identification of the PCEs with an auxiliary variable under various scenarios. We divide the discussion into two sections depending on whether or not S_0 is constant. Within each section, the theoretical results depend on two factors: (1) whether or not the intermediate variable S is discrete or continuous, and (2) whether or not Assumption 2 or 3 holds. Table 1 presents an overview of the key results in our paper.

3. CONSTANT CONTROL INTERMEDIATE VARIABLE

We start with the case with a constant intermediate variable under control. Under this assumption, the distribution of principal strata is identifiable, which greatly simplifies the identification strategies. We will study the case without this assumption in the next section.

ASSUMPTION 4. $S_{i0} = c$ for all i, where c is a constant.

In some vaccine trials (e.g., Follmann, 2006, Hudgens and Gilbert, 2009), Assumption 4 is plausible because vaccine antigens must be present to induce a specific immune response, which is absent in the control group. For a binary S, Assumption 4 with $c = 0$ is called *strong monotonicity*, which holds in the one-sided noncompliance setting because individuals assigned to the control group do not have access to the treatment (Sommer and Zeger, 1991, Imbens and Rubin, 2015). Under Assumption 4, S_0 is constant, and therefore it is not necessary to include it in U, simplifying the PCEs to

$$\begin{aligned}\tau_{s_1} &= \mathbb{E}(Y_1 - Y_0 \mid S_1 = s_1) \\ &= \mathbb{E}(Y_1 \mid S_1 = s_1) - \mathbb{E}(Y_0 \mid S_1 = s_1).\end{aligned}$$

Because S_1 is observed in the treatment group, we can identify $\mathbb{E}(Y_1 \mid S_1 = s_1) = \mathbb{E}\{Y_1 \mathbf{1}(S_1 = s_1)\}/\mathbb{P}(S_1 = s_1)$ by the standard formula under Assumption 1, for example,

$$\begin{aligned}&\frac{\mathbb{E}\left[\mathbb{E}\{Y\mathbf{1}(S = s_1) \mid Z = 1, W\}\right]}{\mathbb{E}\{\mathbb{P}(S = s_1 \mid Z = 1, W)\}} \\ &= \frac{\mathbb{E}\left[\mathbb{P}(S = s_1 \mid Z = 1, W)\mathbb{E}\{Y \mid Z = 1, S_1 = s_1, W\}\right]}{\mathbb{E}\{\mathbb{P}(S = s_1 \mid Z = 1, W)\}}.\end{aligned}$$

TABLE 1

Overview of the sufficient conditions for identifying PCEs. Note that the results with a nonconstant S_0 require the identification of $\mathbb{P}(S_1, S_0 \mid W)$

	Assumptions	Type of S	Requirement for W	Outcome model
		Constant S_0		
Section 3.1	1, 2 and 4	General	No	No
Section 3.2	1, 3 and 4	Discrete	More categories than S	No
Section 3.3	1, 3 and 4	General	Completeness	No
Section 3.4	1 and 4	General	Depends on the model of S	Yes
		Nonconstant S_0		
Section 4.2	1 and 2	General	No	No
Section 4.3	1 and 3	Discrete	More categories than S	No
Section 4.4	1 and 3	General	Completeness	No
Section 4.5	1	General	Depends on the model of S	Yes

Thus, we need only to identify $\mathbb{E}(Y_0 \mid S_1 = s_1)$. Because S_1 is missing in the control group, the PCEs are not identifiable without additional assumptions. Below we will discuss the identification of PCEs under Assumption 2 or 3.

3.1 Principal Ignorability

Ding and Lu (2017) identify the PCEs for a binary S under principal ignorability using the principal score, which is the probability of principal strata conditional on the auxiliary variable. We extend it to a general S:

$$e_{s_1,s_0}(W) = \mathbb{P}(S_1 = s_1, S_0 = s_0 \mid W).$$

Under Assumption 4, the principal score simplifies to $e_{s_1}(W) = \mathbb{P}(S_1 = s_1 \mid W)$, which is identified by $e_{s_1}(W) = \mathbb{P}(S = s_1 \mid Z = 1, W)$ under Assumption 1. The proportions of principal strata are then identified by $e_{s_1} = \mathbb{P}(S_1 = s_1) = \mathbb{E}\{e_{s_1}(W)\}$. The following theorem gives the identification results for the PCEs.

THEOREM 1. *Under Assumptions 1, 2, and 4, the PCEs are identified by*

$$\begin{aligned}\tau_{s_1} = {} & \mathbb{E}\left\{\frac{e_{s_1}(W)}{e_{s_1}} \cdot \frac{ZY}{\pi(W)}\right\} \\ & - \mathbb{E}\left\{\frac{e_{s_1}(W)}{e_{s_1}} \cdot \frac{(1-Z)Y}{1-\pi(W)}\right\},\end{aligned} \tag{1}$$

where $\pi(W) = \mathbb{P}(Z = 1 \mid W)$ is the propensity score.

Theorem 1 shows that $\mathbb{E}(Y_z \mid S_1 = s_1)$ can be identified by the average of the outcomes in a weighted sample, with the weights depending on both the principal score and the propensity score. The principal score accounts for the relationship between the principal stratum membership and the covariates, whereas the propensity score accounts for the relationship between the treatment and the covariates. The result of Ding and Lu (2017) holds with a binary S in randomized experiments, while Theorem 1 allows for a general S in observational studies. Theorem 1 motivates simple moment estimators for the PCEs with the expectations replaced by the sample averages and $\{e_{s_1}(W), \pi(W)\}$ replaced by their fitted values.

3.2 Auxiliary Independence with a Discrete Intermediate Variable

Suppose $S \in \{s_1, \ldots, s_K\}$ and $W \in \{w_1, \ldots, w_L\}$. Let M denote the $K \times L$ matrix with the (k, l)th element $\mathbb{P}(S = s_k \mid Z = 1, W = w_l)$.

THEOREM 2. *Under Assumptions 1, 3, and 4, if* $\operatorname{rank}(M^\top M) = K$, *then the PCEs are identifiable.*

From Theorem 2, a necessary condition for identification is $L \geq K$, that is, W must have more categories than S. Because M depends only on the distribution of the observed data, the condition $\operatorname{rank}(M^\top M) = K$ is testable. The following example from Jiang, Ding and Geng (2016) illustrates the identifiability for the case with a binary intermediate and auxiliary variable.

EXAMPLE 5. Consider binary S and W. First, from the observed distribution and Assumption 1, we can identify $\theta_{sw} = \mathbb{P}(S_1 = s \mid W = w) = \mathbb{P}(S = s \mid Z = 1, W = w)$ and $\delta_w = \mathbb{E}(Y_0 \mid W = w) = \mathbb{E}(Y \mid Z = 0, W = w)$ for $s, w = 0, 1$. Second, under Assumption 3,

$$\begin{aligned}\delta_1 &= \mathbb{E}(Y_0 \mid S_1 = 1)\theta_{11} + \mathbb{E}(Y_0 \mid S_1 = 0)\theta_{01}, \\ \delta_0 &= \mathbb{E}(Y_0 \mid S_1 = 1)\theta_{10} + \mathbb{E}(Y_0 \mid S_1 = 0)\theta_{00},\end{aligned}$$

which are two linear equations of $\mathbb{E}(Y_0 \mid S_1 = 1)$ and $\mathbb{E}(Y_0 \mid S_1 = 0)$. If $\operatorname{rank}(M^\top M) = 2$, the above linear equations have unique solutions

$$\begin{aligned}\mathbb{E}(Y_0 \mid S_1 = 1) &= \frac{\delta_1\theta_{00} - \delta_0\theta_{01}}{\theta_{11}\theta_{00} - \theta_{10}\theta_{01}}, \\ \mathbb{E}(Y_0 \mid S_1 = 0) &= \frac{\delta_1\theta_{10} - \delta_0\theta_{11}}{\theta_{11}\theta_{00} - \theta_{10}\theta_{01}}.\end{aligned}$$

Therefore, the PCEs are identifiable. In this example, the condition $\operatorname{rank}(M^\top M) = 2$ is equivalent to $S \not\perp\!\!\!\perp W \mid Z = 1$ or $\theta_{11}\theta_{00} - \theta_{10}\theta_{01} \neq 0$.

3.3 Auxiliary Independence with a General Intermediate Variable

Identification is more difficult with a continuous intermediate variable, which generates infinitely many principal strata. Let $\mathcal{W}$ be the support of W, and

$$\mathcal{P}_{\mathcal{W}} = \{\mathbb{P}(S_1 \mid W = w) : w \in \mathcal{W}\}$$

be the family of probability distributions indexed by w. Based on the definition of completeness, we give a sufficient condition for identification.

THEOREM 3. *Under Assumptions 1, 3, and 4, if $\mathcal{P}_{\mathcal{W}}$ is complete, then the PCEs are identifiable.*

As discussed before, the key to identify the PCEs is to identify $\mathbb{E}(Y_0 \mid S_1)$. Under Assumptions 1 and 3, we have

$$\begin{aligned}\mathbb{E}(Y \mid Z = 0, W = w) &= \mathbb{E}(Y_0 \mid W = w) \\ &= \mathbb{E}\{\mathbb{E}(Y_0 \mid S_1) \mid W = w\} \\ &= \int \mathbb{E}(Y_0 \mid S_1 = s)Q(\mathrm{d}s)\end{aligned} \tag{2}$$

for any probability measure $Q(s) = \mathbb{P}(S_1 \leq s \mid W = w)$ in $\mathcal{P}_{\mathcal{W}}$. The left-hand side of (2) is directly estimable from the observed data, and the distributions in $\mathcal{P}_{\mathcal{W}}$ are identified by $\mathbb{P}(S_1 \mid W) = \mathbb{P}(S \mid Z = 1, W)$. Therefore, (2) is an integral equation for $\mathbb{E}(Y_0 \mid S_1 = s)$. As a result, $\mathbb{E}(Y_0 \mid S_1 = s)$ is identifiable if it can be uniquely determined by (2), which is guaranteed by the completeness of $\mathcal{P}_{\mathcal{W}}$. When S is discrete, the integral in (2) becomes summation, and the completeness is the same as the rank condition in Theorem 2.

Theorem 3 is general but abstract. From the well-known completeness property of an exponential family (Lehmann and Romano, 2005), we have a more interpretable sufficient condition for identifying PCEs.

THEOREM 4. *Under Assumptions 1, 3, and 4, we further assume*

$$\mathbb{P}(S_1 = s_1 \mid W = w) = h(s)g(w)\exp\{\boldsymbol{\eta}^\top(w)\boldsymbol{t}(s_1)\},$$

where $s_1 \to \boldsymbol{t}(s_1)$ is a one-to-one mapping and $\{\boldsymbol{\eta}(w) : w \in \mathcal{W}\}$ contains an open set in $\mathbb{R}^d$ where d is the dimension of the vector function $\boldsymbol{\eta}(w)$. The PCEs are identifiable.

Theorem 4 requires that the distribution of S_1 conditional on W belongs to the exponential family, but it does not require any models for the potential outcome Y_z. Therefore, Theorem 4 guarantees semiparametric identifiability and allows for different types of outcomes. Below we give an example with Normal (S_1, W).

COROLLARY 1. *Under Assumptions 1, 3, and 4, if (S_1, W) follows a bivariate Normal distribution, then the PCEs are identifiable.*

REMARK 1. For a binary outcome, Follmann (2006) assumes that the outcome follows a Probit model and (S_1, W) follows a bivariate Normal distribution, which is a special case of Corollary 1. Thus, Follmann's (2006) model is semiparametrically identified even without the outcome model, and his parametric outcome model is invoked only for convenience in the finite-sample inference.

To further improve the applicability of Theorem 3, we review the following lemma (Hu and Shiu, 2018, Lemma 4) on the completeness of a class of location-scale distribution families, which works for nonexponential distributions.

LEMMA 1. *Suppose the support of W has an interior point, and $S_1 \overset{\mathrm{d}}{=} h(W) + \sigma(W)\epsilon$ with continuously differentiable $h(w)$ and $\sigma(w)$ and $\epsilon \perp\!\!\!\perp W$. Then, $\mathcal{P}_{\mathcal{W}}$ is complete if the characteristic function and density function of ϵ, $\phi(t)$ and $f(\epsilon)$, satisfy the following conditions*:

(a) $0 < |\phi(t)| < C\exp(-\delta|t|)$ *for all $t \in \mathbb{R}$ and some constants $C, \delta > 0$;*

(b) *$f(\epsilon)$ is continuously differentiable, and*

$$\int_{-\infty}^{+\infty} |xf'(x)|\,\mathrm{d}x < +\infty, \quad \int_{-\infty}^{+\infty} f^2(x)\,\mathrm{d}x < +\infty;$$

(c) *for any positive integer J, the following functions are linearly independent,*

$$\left\{f\left(\frac{x-h_1}{\sigma_1}\right), \ldots, f\left(\frac{x-h_J}{\sigma_J}\right)\right\},$$

where the (h_j, σ_j)'s are distinct.

The existence of the interior point required by Lemma 1 holds automatically for continuous W but fails for discrete W. Conditions (a) and (b) in Lemma 1 are technical requirements on the distribution of the error term ϵ. Condition (c) means that the finite location-scale mixture of the distribution of ϵ is identifiable, which holds for many distributions (Everitt and Hand, 1981). For example, Appendix B.1 shows that Conditions (a)–(c) hold when ϵ follows a Normal, t or Logistic distribution. Combining Theorem 3 and Lemma 1 yields the following theorem for the location-scale distribution families.

THEOREM 5. *Suppose that W is continuous, Assumptions 1, 3, and 4 hold, $S_1 \overset{\mathrm{d}}{=} h(W) + \sigma(W)\epsilon$ with continuously differentiable $h(w)$ and $\sigma(w)$, and $\epsilon \perp\!\!\!\perp W$. If ϵ satisfies Conditions (a)–(c) in Lemma 1, then the PCEs are identifiable.*

Theorem 5 guarantees the identifiability of PCEs in many models involving distributions that do not belong to an exponential family. It allows for heteroscedastic errors and enables flexible model choices. For example, if we replace the bivariate Normal distribution assumption of (S_1, W) with $S_1 \mid W = w \sim N(\mu(w), \sigma^2(w))$, then Theorem 4 and Corollary 1 cannot be applied because $\{\boldsymbol{\eta}(w) = (1/\sigma^2(w), \mu(w)/\sigma^2(w)) : w \in \mathcal{W}\}$ is a line in $\mathbb{R}^2$. In contrast, Theorem 5 is still applicable in this example which ensures that the PCEs are identifiable.

3.4 Without Conditional Independence

The conditional independence in Assumption 2 or 3 may be violated. In Example 2, covariates may not be sufficient to account for the difference in the health-related quality of life across principal strata, which makes Assumption 2 implausible; in Example 4, different centers may have different qualities of services, which makes Assumption 3 implausible. Without conditional independence, W does not help to achieve nonparametric or semiparametric identification. One solution is to conduct sensitivity analysis, which, however, requires to use sensitivity parameters to characterize the violation of the assumptions and further requires to specify their ranges. Sensitivity analysis gives a range of estimates rather than a point estimate, and it often depends on additional model assumptions. We will not pursue this direction. Instead, we propose to identify the PCEs by exploiting the role of W with some parametric models for the outcome. We can also include other covariates $\boldsymbol{X}$ in our models, but do not require any modeling assumptions for $\boldsymbol{X}$. So, again, we condition on $\boldsymbol{X}$ implicitly. The results in this subsection ensure the identifiability of the PCEs under many models that have been used in previous empirical studies and generalize some models to account for different types of outcomes and intermediate variables.

PROPOSITION 1. *Under Assumptions 1 and 4, assume that (S_1, Y_0) follow additive models*:

$$S_1 = g(W) + \sigma_1(W)\epsilon_{S_1}, \tag{3}$$

$$Y_0 = \beta_0 + \alpha S_1 + \sum_{j=1}^{J} \beta_j f_j(W) + \sigma_2(W)\epsilon_{Y_0}, \tag{4}$$

where $\mathbb{E}(\epsilon_{S_1} \mid W) = \mathbb{E}(\epsilon_{Y_0} \mid W) = 0$, *and* $g(w)$ *and* $\sigma_1(w)$ *can be unknown functions. If* $\{1, g(w), f_1(w), \ldots, f_J(w)\}$ *are linearly independent, then the PCEs are identifiable.*

We do not need to specify $g(w)$ and $\sigma_1(w)$ because they are identifiable from the observed distribution $\mathbb{P}(S, W \mid Z = 1)$ under Assumption 1. In contrast, we need to specify the $f_j(w)$'s and $\sigma_2(w)$ in the model of Y_0.

Intuitively, replacing S_1 in (4) by (3), we obtain an additive model of Y_0 on W, and the linear independence condition in Proposition 1 allows us to disentangle the coefficients of different terms involving W. For example, if $g(w)$ is quadratic in w in (3) and $\{J = 1, f_1(w) = w\}$ in (4), then the linear independence assumption holds in Proposition 1. However, if $g(w)$ is linear in w, then the linear independence assumption fails.

If $f_j(w) = 0$ for all $j = 1, \ldots, J$, then Proposition 1 becomes a special case of Theorem 5. Proposition 1 guarantees the identifiability of PCEs in additive models without specifying the distributions of the error terms.

In the model of Y_0, we require S_1 to have a linear form. Identification may also be possible for other forms of S_1, but will require the knowledge of the distributions of the error terms.

For binary outcomes, we show an identification result below for the Probit model.

PROPOSITION 2. *Under Assumptions 1 and 4, assume that* S_1 *follows an additive model with a Normal error term and* Y_0 *follows a Probit model*:

$$S_1 = g(W) + \epsilon_{S_1}, \quad \epsilon_{S_1} \perp\!\!\!\perp W, \quad \epsilon_{S_1} \sim N(0, \sigma^2),$$

$$\mathbb{P}(Y_0 = 1 \mid S_1 = s, W = w) = \Phi\left\{\beta_0 + \alpha s + \sum_{j=1}^{J} \beta_j f_j(w)\right\},$$

where $g(w)$ *can be unknown. If* $\{1, g(w), f_1(w), \ldots, f_J(w)\}$ *are linearly independent, then the PCEs are identifiable.*

The model of S_1 in Proposition 2 requires the variance of the error term ϵ_{S_1} not depend on W, which is different from Proposition 1. Identification may also be possible with the variance depending on W, but will rely on the functional form of $\operatorname{var}(S_1 \mid W)$.

REMARK 2. Our result does not contradict Follmann (2006). Without Assumption 3, Follmann (2006) assumed a bivariate Normal distribution for (S_1, W) and used the following Probit model for Y:

$$\mathbb{P}(Y = 1 \mid Z, S_1, W) = \Phi(\beta_0 + \beta_1 Z + \beta_2 S_1 + \beta_3 W + \beta_4 Z S_1). \tag{5}$$

Under Assumption 1, (5) is equivalent to

$$\mathbb{P}(Y_1 = 1 \mid S_1, W) = \Phi\{\beta_0 + \beta_1 + (\beta_2 + \beta_4) S_1 + \beta_3 W\},$$

$$\mathbb{P}(Y_0 = 1 \mid S_1, W) = \Phi(\beta_0 + \beta_2 S_1 + \beta_3 W).$$

From Proposition 2, the PCEs are not identifiable without the model of Y_1 because the linear independence condition is violated. The identifiability comes from the parallel model assumption that restricts the coefficients of W be the same in the models of Y_1 and Y_0.

REMARK 3. Without the linear independence condition, researchers often use additional information on the parameters to improve identification. Using a Bayesian approach, Zigler and Belin (2012) imposed informative priors on α. In a similar setting with a time-to-event outcome, Qin et al. (2008) imposed the principal ignorability $Y_0 \perp\!\!\!\perp S_1 \mid W$, or, equivalently, $\alpha = 0$.

4. NONCONSTANT CONTROL INTERMEDIATE VARIABLE

When Assumption 4 does not hold, we can never simultaneously observe S_1 and S_0, making it challenging to identify the joint distribution of (S_1, S_0) in the first place, let alone the PCEs. Below we first use a copula model for the joint distribution of (S_1, S_0), and then discuss identification of the PCEs.

4.1 A Copula Model for $\mathbb{P}(S_1, S_0 \mid W)$

Under Assumption 1, $\mathbb{P}(S_z \mid W) = \mathbb{P}(S \mid Z = z, W)$, and thus the marginal distributions of S_z given W are identifiable from the observed data. To recover the joint distribution of (S_1, S_0) given W from the marginal distributions, we need some prior knowledge about the association between S_1 and S_0 conditional on W. For a binary S, a commonly used assumption to recover the joint distribution of (S_1, S_0) is the monotonicity assumption that $S_1 \geq S_0$. Under this assumption, the joint distribution is identifiable:

$$\mathbb{P}(S_1 = 1, S_0 = 1 \mid W) = \mathbb{P}(S = 1 \mid Z = 0, W),$$

$$\mathbb{P}(S_1 = 0, S_0 = 0 \mid W) = \mathbb{P}(S = 0 \mid Z = 1, W),$$

$$\mathbb{P}(S_1 = 1, S_0 = 0 \mid W) = \mathbb{P}(S = 1 \mid Z = 1, W) - \mathbb{P}(S = 1 \mid Z = 0, W).$$

For a continuous S, Efron and Feldman (1991) and Jin and Rubin (2008) discussed the equipercentile equating assumption, that is, $F_1(S_1 \mid W) = F_0(S_0 \mid W)$, where $F_z(\cdot \mid W)$ is the cumulative distribution function of S_z given W for $z = 0, 1$. Under this assumption, S_z determines S_{1-z} via $F_1(\cdot \mid W)$ and $F_0(\cdot \mid W)$ for $z = 0, 1$.

The monotonicity and equipercentile equating assumptions are special cases of the copula approach (Nelsen, 2006), which is a general strategy to obtain the joint distribution from marginal distributions. Various copula models have been proposed to model principal strata (Roy, Hogan and Marcus, 2008, Bartolucci and Grilli, 2011, Schwartz, Li and Mealli, 2011, Daniels et al., 2012,

Conlon, Taylor and Elliott, 2017, Yang and Ding, 2018, Kim et al., 2020). Assume

$$\begin{aligned}(6)\quad &\mathbb{P}(S_1, S_0 \mid W = w)\\ &= C_\rho\{\mathbb{P}(S_1 \mid W = w), \mathbb{P}(S_0 \mid W = w)\},\end{aligned}$$

where $C_\rho(\cdot,\cdot)$ is a copula and ρ is a measure of the association between S_1 and S_0. If we know ρ, then we can identify $\mathbb{P}(S_1, S_0 \mid W = w)$ from the marginal distributions by (6). Otherwise, we can view ρ as a sensitivity parameter.

4.2 Principal Ignorability

Assume that the principal score $e_{s_1,s_0}(W) = \mathbb{P}(S_1, S_0 \mid W)$ is identifiable. So the density of the principal strata equals $e_{s_1,s_0} = \mathbb{E}\{\pi_{s_1,s_0}(W)\}$. Similar to Section 3.1, PCEs are identifiable as shown below.

THEOREM 6. *Under Assumptions 1 and 2, if $e_{s_1,s_0}(W)$ is identifiable, then the PCEs are identified by*

$$\begin{aligned}\tau_{s_1s_0} = &\mathbb{E}\left\{\frac{e_{s_1,s_0}(W)}{e_{s_1,s_0}} \cdot \frac{ZY}{\pi(W)}\right\}\\ &- \mathbb{E}\left\{\frac{e_{s_1,s_0}(W)}{e_{s_1,s_0}} \cdot \frac{(1-Z)Y}{1-\pi(W)}\right\}.\end{aligned}$$

Theorem 6 generalizes Theorem 1 to the case with nonconstant control intermediate variables. It shows that $\mathbb{E}(Y_z \mid S_1 = s_1, S_0 = s_0)$ can be identified by the average of the outcomes in a weighted sample, with the weight depending on both the principal score and the propensity score.

4.3 Auxiliary Independence with a Discrete Intermediate Variable

We give the identification results for discrete intermediate variables. Suppose $S \in \{s_1, \ldots, s_K\}$ and $W \in \{w_1, \ldots, w_L\}$. Let M_{s_0} denote the $K \times L$ matrix with (k,l)th element $\mathbb{P}(S_1 = s_k \mid S_0 = s_0, W = w_l)$, and M_{s_1} denote the $K \times L$ matrix with (k,l)th element $\mathbb{P}(S_0 = s_k \mid S_1 = s_1, W = w_l)$.

THEOREM 7. *Suppose that Assumptions 1 and 3 hold, and $\mathbb{P}(S_1, S_0 \mid W)$ is identifiable.*

(a) *For a fixed s_0, if $\operatorname{rank}(M_{s_0}^\top M_{s_0}) = K$, then $\mathbb{P}(Y_0 \mid S_1, S_0 = s_0)$ is identifiable.*
(b) *For a fixed s_1, if $\operatorname{rank}(M_{s_1}^\top M_{s_1}) = K$, then $\mathbb{P}(Y_1 \mid S_1 = s_1, S_0)$ is identifiable.*
(c) *If $\operatorname{rank}(M_{s_0}^\top M_{s_0}) = \operatorname{rank}(M_{s_1}^\top M_{s_1}) = K$ for all s_1 and s_0, then the PCEs are identifiable.*

Theorem 7 extends Theorem 2. As a special case of Theorem 7, for a binary intermediate variable under monotonicity, Ding et al. (2011) and Jiang, Ding and Geng (2016) gave the identification results, and the rank conditions in Theorem 7 simplify to testable conditions $S_1 \not\perp\!\!\!\perp W \mid S_0$ and $S_0 \not\perp\!\!\!\perp W \mid S_1$.

4.4 Auxiliary Independence with a General Intermediate Variable

Recalling that $\mathcal{W}$ is the support of W. For fixed s_0 and s_1, let

$$\begin{aligned}\mathcal{P}_{\mathcal{W},s_0} &= \{\mathbb{P}(S_1 \mid S_0 = s_0, W = w) : w \in \mathcal{W}\},\\ \mathcal{P}_{\mathcal{W},s_1} &= \{\mathbb{P}(S_0 \mid S_1 = s_1, W = w) : w \in \mathcal{W}\}\end{aligned}$$

be the families of the distributions indexed by w given $S_0 = s_0$ and $S_1 = s_1$, respectively. Similar to Section 3.2, the identifiability of PCEs reduces to the completeness of $\mathcal{P}_{\mathcal{W},s_0}$ and $\mathcal{P}_{\mathcal{W},s_1}$.

THEOREM 8. *Suppose that Assumptions 1 and 3 hold, and $\mathbb{P}(S_1, S_0 \mid W)$ is identifiable.*

(a) *If $\mathcal{P}_{\mathcal{W},s_0}$ is complete for all s_0, then $\mathbb{P}(Y, S_1, S_0, W \mid Z = 0)$ is identifiable.*
(b) *If $\mathcal{P}_{\mathcal{W},s_1}$ is complete for all s_1, then $\mathbb{P}(Y, S_1, S_0, W \mid Z = 1)$ is identifiable.*
(c) *If (a) and (b) above hold, then the PCEs are identifiable.*

Similar to Theorem 3, Theorem 8 does not require any models for the distribution of Y_z $(z = 0, 1)$, which guarantees the nonparametric or semiparametric identification of PCEs. Based on the completeness of the location-scale distribution families in Lemma 1, we can obtain identification results for some widely used models with an example below.

COROLLARY 2. *For a continuous W, suppose that Assumptions 1 and 3 hold. If*

$$(7)\quad (S_1, S_0) \mid W = w \sim \boldsymbol{N}_2\left\{\begin{pmatrix}\mu_1(w)\\ \mu_0(w)\end{pmatrix}, \boldsymbol{\Sigma}(w)\right\},$$

where

$$\boldsymbol{\Sigma}(w) = \begin{pmatrix}\sigma_1^2(w) & \rho(w)\sigma_1(w)\sigma_0(w)\\ \rho(w)\sigma_1(w)\sigma_0(w) & \sigma_0^2(w)\end{pmatrix}$$

with a known $\rho(w)$, then the PCEs are identifiable.

Corollary 2 does not need any models for the outcome, but requires the auxiliary variable to be continuous. In Corollary 2, with a known $\rho(w)$, we can identify the joint distribution of (S_1, S_0) given W from the marginal distributions of S_z given W. Therefore, the PCEs are identifiable from Theorem 8. To apply Corollary 2, we need to specify the correlation coefficient $\rho(w)$, which is a sensitive parameter in practice.

4.5 Without Conditional Independence

Similar to the case with a constant control intermediate variable, we propose some useful parametric models for identifying the PCEs using the auxiliary variable W when Assumptions 2 or 3 fails.

PROPOSITION 3. *For a binary S with monotonicity $S_1 \geq S_0$, suppose that Assumption 1 holds, and Y_1 and Y_0 follow linear models*

$$\mathbb{E}(Y_z \mid S_1, S_0, W) = \beta_{z0} + \beta_{z1} S_1 + \beta_{z2} S_0 + \beta_{z3} W, \quad (z = 0, 1). \tag{8}$$

If neither

$$\frac{\mathbb{P}(S=1 \mid Z=1, W=w)}{\mathbb{P}(S=1 \mid Z=0, W=w)} \quad \text{nor} \quad \frac{\mathbb{P}(S=0 \mid Z=1, W=w)}{\mathbb{P}(S=0 \mid Z=0, W=w)} \tag{9}$$

is constant in w, then the PCEs are identifiable.

We can use observed data to check whether the two terms in (9) are constant in w. For a binary W, the only restriction of (8) is no interaction term among (S_1, S_0, W) in the model of Y, which is similar to some existing no-interaction or homogeneity assumption (Ding et al., 2011, Wang, Zhou and Richardson, 2017).

For a continuous intermediate variable, we give the following proposition.

PROPOSITION 4. *Suppose that Assumption 1 holds, (S_1, S_0) given W follows (7) with a known $\rho(w)$, and Y_1 and Y_0 follow additive models*:

$$Y_1 = \beta_0 + \alpha_1 S_1 + \alpha_0 S_0 + \sum_{j=1}^{J_1} \beta_j f_j(W) + \sigma_1^2(W)\epsilon_{Y_1},$$

$$Y_0 = \beta_0' + \alpha_1' S_1 + \alpha_0' S_0 + \sum_{j=1}^{J_0} \beta_j' h_j(W) + \sigma_0^2(W)\epsilon_{Y_0},$$

$$(\epsilon_{Y_1}, \epsilon_{Y_2}) \perp\!\!\!\perp (S_1, S_0, W).$$

The PCEs are identifiable if the following two conditions hold:

(a) $\{1, s_1, \mathbb{E}(S_0 \mid S_1 = s_1, W = w), f_1(w), \ldots, f_{J_1}(w)\}$ *are linearly independent as functions of* (s_1, w);
(b) $\{1, s_0, \mathbb{E}(S_1 \mid S_0 = s_0, W = w), h_1(w), \ldots, h_{J_0}(w)\}$ *are linearly independent as functions of* (s_0, w).

Proposition 4, as an extension of Proposition 1, is mostly useful for continuous outcomes. The Normality in (7) implies a linear relation of S_0 on S_1 given W, that is, $S_0 = a_0(W)S_1 + b_0(W)\epsilon_{S_0}$ with $a_0(w)$ and $b_0(w)$ determined by the distribution of (S_1, S_0) given W. Then, in Proposition 4, we can obtain an additive model of Y_1 on S_1 and W by replacing S_0 in the model of Y_1. The linear independence condition (a) allows us to disentangle the coefficients of different terms involving S_1 and W. Similar discussion applies to condition (b).

The Normality in (7) is also helpful for binary outcomes. The following proposition gives the identification result under the Probit model for Y_z.

PROPOSITION 5. *Suppose that Assumption 1 holds, and (S_1, S_0) given W follows (7) with a known $\rho(w)$. Suppose Y_1 and Y_0 follow Probit models*:

$$\mathbb{P}(Y_1 = 1 \mid S_1 = s_1, S_0 = s_0, W = w) = \Phi\left\{\beta_0 + \alpha_1 s_1 + \alpha_0 s_0 + \sum_{j=1}^{J_1} \beta_j f_j(w)\right\}, \tag{10}$$

$$\mathbb{P}(Y_0 = 1 \mid S_1 = s_1, S_0 = s_0, W = w) = \Phi\left\{\beta_0' + \alpha_1' s_1 + \alpha_0' s_0 + \sum_{j=1}^{J_0} \beta_j' h_j(w)\right\}. \tag{11}$$

If Conditions (a) and (b) in Proposition 4 hold, then the PCEs are identifiable.

REMARK 4. Using a Bayesian approach, Zigler and Belin (2012) assumed a trivariate Normal distribution for (S_1, S_0, W) with a sensitivity parameter to characterize the correlation between S_1 and S_0, and Probit models for Y_z with $f_j(w)$ and $h_j(w)$ linear in w. Under their models, the conditional expectation $\mathbb{E}(S_0 \mid S_1 = s_1, W = w)$ is linear in both s_1 and w, and $\mathbb{E}(S_1 \mid S_0 = s_0, W = w)$ is linear in both s_0 and w. Thus, the linear independence condition is violated, and the parameters are not identifiable. To mitigate the inferential difficulties, Zigler and Belin (2012) imposed informative priors on $\alpha_1 - \alpha_1'$ and $\alpha_0 - \alpha_0'$.

5. NUMERICAL EXAMPLES

In the frequentists' inference, nonidentifiability renders the likelihood function flat over a region for some parameters, and the classical repeated sampling theory of the maximum likelihood estimates do not apply (Bickel and Doksum, 2015). Computationally, the Bayesian machinery is still applicable as long as the priors are proper. The simulation below, however, highlights the importance of identifiability in the Bayesian inference. In both cases with a constant and nonconstant control intermediate variable, we use two models to estimate the PCEs under several data generating processes (DGPs). The two models seem similar in form but have different identifiability. We use the Gibbs Sampler to simulate the posterior distributions of the PCEs with 20,000 iterations and the first 4000 iterations as the burn-in period. The Markov chains mix very well with the Gelman–Rubin diagnostic statistics close to one based on multiple chains.

5.1 Constant Control Intermediate Variable

We generate data from DGP 1:

$$Z \sim \text{Bernoulli}(0.5), \quad W \sim N(0,1), \quad Z \perp\!\!\!\perp W,$$

$$S_1 \mid W \sim N(\gamma_0 + \gamma_1 W, \sigma^2),$$

$$\mathbb{P}(Y_z = 1 \mid S_1, W) = \Phi(\beta_{z0} + \beta_{z1} S_1 + \beta_{z2} W),$$

with parameters $(\beta_{00}, \beta_{01}, \beta_{02}) = (1, -0.5, 0.5)$, $(\beta_{10}, \beta_{11}, \beta_{12}) = (0.5, 1, 1.5)$ and $(\gamma_0, \gamma_1, \sigma) = (1, 0.5, 1)$. We name the model corresponding to DGP 1 as model 1. Under model 1, Assumption 1 holds but the conditions in Proposition 2 do not. Therefore, model 1 is not identifiable.

In DGP 2, Z, W and Y_z are the same as DGP 1, but

$$S_1 \mid W \sim N(\gamma_0 + \gamma_1 W + \gamma_2 W^2, \sigma^2),$$

where $(\gamma_0, \gamma_1, \gamma_2, \sigma) = (1, 0.5, 1, 1)$. We name the model corresponding to DGP 2 as model 2. Because $(1, \gamma_0 + \gamma_1 W + \gamma_2 W^2, W)$ are linearly independent, the PCEs are identifiable based on Proposition 2.

For both DGPs 1 and 2, we use the true models to analyze the generated data with sample size 1000. We choose the following two sets of priors to assess the sensitivity of the inference based on posteriors:

(A) $(\beta_{z0}, \beta_{z1}, \beta_{z2}) \sim \boldsymbol{N}_3(\mathbf{0}_3, \text{diag}(1, 1, 1)/10^{-2})$ for $z = 0, 1$, $p(\sigma^2) \propto 1/\sigma^2$, and $(\gamma_0, \gamma_1) \sim \boldsymbol{N}_2(\mathbf{0}_2, \text{diag}(1, 1)/10^{-2})$ for model 1 (correspondingly, $(\gamma_0, \gamma_1, \gamma_2) \sim \boldsymbol{N}_3(\mathbf{0}_2, \text{diag}(1, 1, 1)/10^{-2})$ for model 2).

(B) $(\beta_{z0}, \beta_{z1}, \beta_{z2}) \sim \boldsymbol{N}_3(\mathbf{0}_3, \text{diag}(1, 1, 1))$ for $z = 0, 1$, $p(\sigma^2) \propto 1/\sigma^2$, and $(\gamma_0, \gamma_1) \sim \boldsymbol{N}_2(\mathbf{0}_2, \text{diag}(1, 1)/10^{-2})$ for model 1 (correspondingly, $(\gamma_0, \gamma_1, \gamma_2) \sim \boldsymbol{N}_3(\mathbf{0}_2, \text{diag}(1, 1, 1)/10^{-2})$ for model 2).

The prior for $(\beta_{z0}, \beta_{z1}, \beta_{z2})$ is much less diffused in prior (B) than in prior (A).

Figure 1 shows the posterior distributions of $(\beta_{01}, \beta_{02}, \beta_{11}, \beta_{12})$. For model 2, the posterior 95% credible intervals cover the true parameters under both priors. For model 1, the posterior distributions of β_{01} and β_{02} differ greatly under the two priors. Their posterior distributions deviate greatly from the true values under prior (A), which shows strong evidence of nonidentifiability or weakly identifiability of model 1.

5.2 Nonconstant Control Intermediate Variable

Similar to Section 5.1, we describe two DGPs with different identifiability and evaluate the finite-sample performance of Bayesian inference under each DGP. We choose two models corresponding to two nested DGPs so that we can go beyond Section 5.1 to assess the performance of the Bayesian inference with a misspecified model.

We first specify the two DGPs. For DGP 3, $W \sim$ Bernoulli(0.5) and $Z \mid W = w \sim$ Bernoulli(α_w), where $(\alpha_1, \alpha_2) = (0.5, 0.5)$. We then generate $U = (S_1, S_0)$ from categorical distributions conditional on W, and Y from Bernoulli distributions conditional on Z and U with true values of the parameters in Table 2(a). We name the model corresponding to DGP 3 as model 3. For model 3, Assumptions 1 and 3 hold. Because the stratum $(S_1, S_0) = (0, 1)$ does not exist, monotonicity holds and thus the distribution of (S_1, S_0) given W is identifiable. From Theorem 7, the PCEs are identifiable.

For DGP 4, we generate W and Z in the same way as DGP 4. We then generate $U = (S_1, S_0)$ from categori-

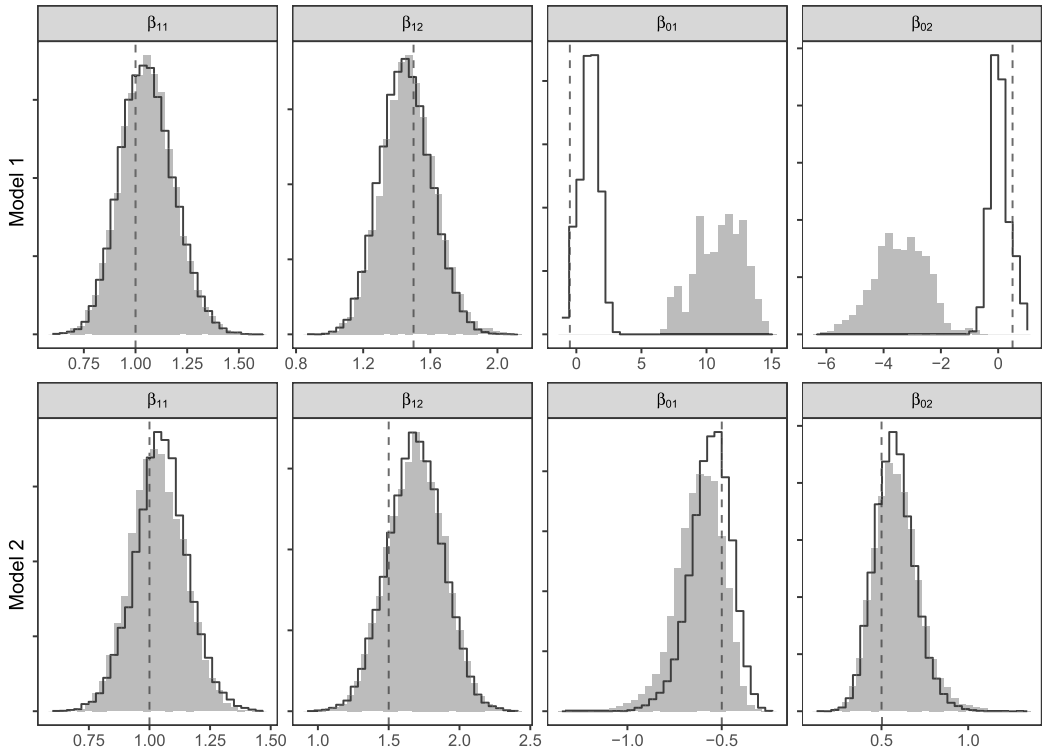

FIG. 1. *Posterior distributions of the parameters in Section 5.1. The grey histograms are the results with prior (A), and the white histograms are the results with prior (B). The vertical dashed lines represent the true values of the parameters.*

TABLE 2
True values of the parameters under DGP 3 *and DGP* 4

(a) DGP 3 with $\tau_{11} = 0.3$, $\tau_{10} = 0.4$ and $\tau_{00} = 0.5$

	$u = (1, 1)$	$u = (1, 0)$	$u = (0, 0)$
$\mathbb{P}(U = u \mid W = w)$			
$w = 1$	0.5	0.3	0.2
$w = 2$	0.2	0.3	0.5
$\mathbb{P}(Y = 1 \mid Z = z, U = u)$			
$z = 1$	0.8	0.7	0.6
$z = 0$	0.5	0.3	0.1

(b) DGP 4 with $\tau_{11} = 0.3$, $\tau_{10} = 0.4$, $\tau_{00} = 0.5$ and $\tau_{01} = -0.3$

	$u = (1, 1)$	$u = (1, 0)$	$u = (0, 0)$	$u = (0, 1)$
$\mathbb{P}(U = u \mid W = w)$				
$w = 1$	0.5	0.3	0.1	0.1
$w = 2$	0.1	0.3	0.5	0.1
$\mathbb{P}(Y = 1 \mid Z = z, U = u)$				
$z = 1$	0.8	0.7	0.6	0.2
$z = 0$	0.5	0.3	0.1	0.5

cal distributions conditional on W, and Y from Bernoulli distributions conditional on Z and U with true values of the parameters in Table 2(b). We name the model corresponding to DGP 4 as model 4. For model 4, stratum $(S_1, S_0) = (0, 1)$ exists, and monotonicity does not hold. Without monotonicity, the distribution of $(S_1, S_0) \mid W$ is not identifiable, and thus the PCEs are not identifiable.

We first use models 3 and 4 to analyze the data simulated from DGP 3. Because model 4 is a generalization of model 3, they are both correctly specified under DGP 3. However, the true value of τ_{01} in model 4 is not well-defined.

We choose two sample sizes 1000 and 50,000. For model 3, we choose the following priors: $\mathbb{P}(W = 1) \sim$ Beta(1, 1), $\alpha_w \sim$ Beta(1, 1), and $(\pi_{11,w}, \pi_{10,w}, \pi_{00,w}) \sim$ Dirichlet(1, 1, 1) for $w = 1, 2$. We choose two different priors for the parameters δ_{u,s_1s_0}. One is the uniform prior Beta(1, 1) and the other is Beta(0.5, 0.5). For model 4, all the priors are the same except that the prior for $(\pi_{11,w}, \pi_{10,w}, \pi_{00,w}, \pi_{01,w})$ is Dirichlet(1, 1, 1, 1).

Figure 2(a) shows the posterior distributions of τ_{11}^{m}, τ_{11} and τ_{01}, where τ_{11}^{m} is the PCE within the stratum

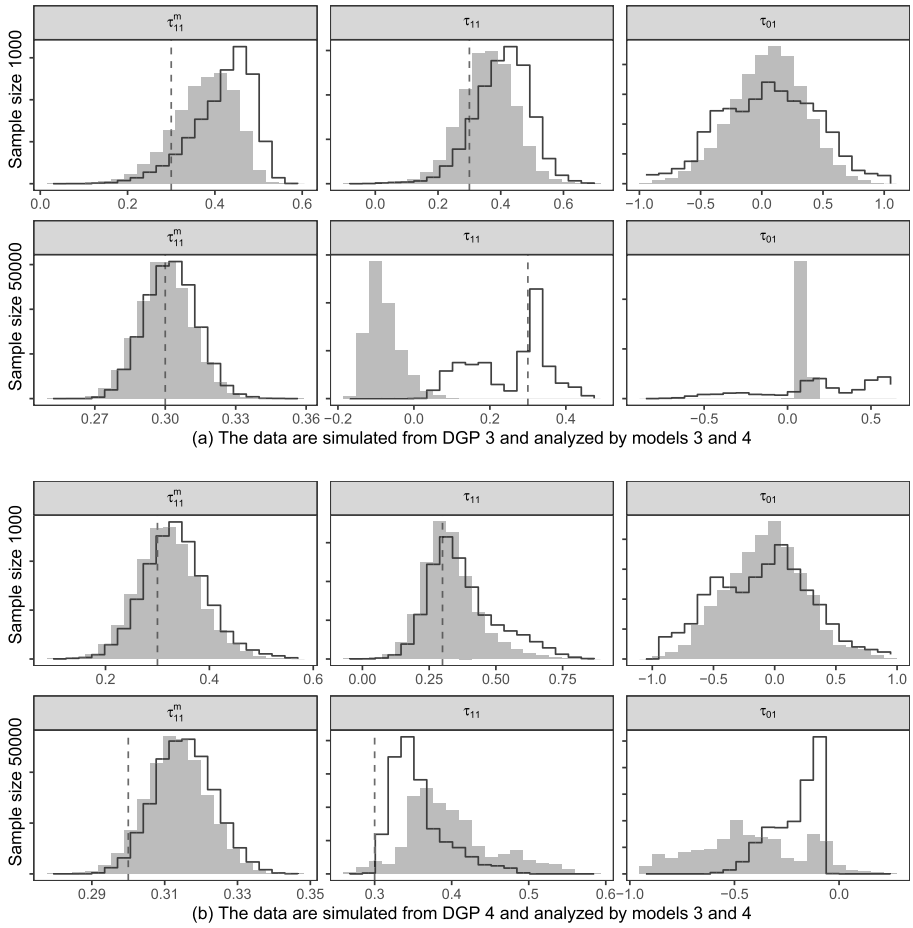

FIG. 2. *Posterior distributions of the PCEs in Section 5.2.* τ_{11}^{m} *is the PCE within the stratum* $(S_1, S_0) = (1, 1)$ *under model* 3; τ_{11} *and* τ_{01} *are the PCEs within the strata* $(S_1, S_0) = (1, 1)$ *and* $(0, 1)$ *under model* 4. *The grey histograms are the results with prior* Beta(1, 1) *for* δ_{u,s_1s_0}, *and the white histograms are the results with prior* Beta(0.5, 0.5) *for* δ_{u,s_1s_0}. *The vertical dashed lines represent the true values of the parameters.*

$(S_1, S_0) = (1, 1)$ under model 3, and τ_{11} and τ_{01} are the PCEs within the strata $(S_1, S_0) = (1, 1)$ and $(0, 1)$ under model 4, respectively. Comparing the two rows of plots in Figure 2(a), we can see that as the sample size increases, the posterior 95% credible intervals of τ_{11}^{m} become narrower and always cover the true value, regardless of the priors. For model 4, the posterior distributions of the PCEs change greatly and the posterior 95% credible intervals do not shrink as those under model 3. When the sample size is 50,000, the posterior distribution of τ_{11} deviates greatly from the true value with the flat prior Beta(1, 1) and is not unimodal with the prior Beta(0.5, 0.5). This is in sharp contrast to standard Bayesian problems in which the Beta(1, 1) and Beta(0.5, 0.5) priors result in small discrepancies. The drastic differences with different sample sizes and priors show strong evidence of the nonidentifiability or weakly identifiability of model 4, which can yield misleading estimates and inferences.

We then use models 3 and 4 to analyze data simulated from DGP 4. The true model 4 is not identifiable, and model 3 is misspecified. Figure 2(b) shows the results for τ_{11}^{m}, τ_{11} and τ_{01}. Although model 3 is not the true model, the result under this model is very stable under different priors. The 95% credible intervals of τ_{11}^{m} cover the true value. This may be due to our choice of small $\pi_{01,1}$ and $\pi_{01,2}$, which makes model 3 only slightly deviates from the true model. In contrast, the result of model 4 changes drastically under different priors even when the sample size is large. The posterior distributions of τ_{01} are multimodal even with a very large sample size. Therefore, using an unidentifiable model may lead to an undesirable result even if it is a true model.

Our simulation demonstrates that identification is important in the Bayesian inference. Otherwise, the results are extremely sensitive to the priors. More importantly, the simulation suggests that when the proposed model is not identifiable, using an identifiable model "close" to it may be a compromising solution.

6. APPLICATION TO THE JOB SEARCH INTERVENTION STUDY

The Job Search Intervention Study was a randomized field experiment investigating the efficacy of a job training intervention on unemployed workers (Vinokur, Price and Schul, 1995, Vinokur and Schul, 1997, Tingley et al., 2014). The program was designed not only to increase reemployment among the unemployed but also to enhance the mental health of the job seekers. In the study, 600 unemployed workers were randomly assigned to the treatment group ($Z = 1$) and 299 were assigned to the control group ($Z = 0$). Those in the treatment group participated in workshops that covered skills for job search and coping with stress. Those in the control group received a booklet describing job-search tips. The intermediate variable S is a measure of job-search self-efficacy ranged from 1 to 5. It measures the participants' confidence in being able to successfully perform six essential job-search activities including completing a job application or resume, using their social network to discover promising job openings, and getting their point across in a job interview. The outcome Y is a measure of depressive symptoms based on the Hopkins Symptom Checklist. It measures how much they had been bothered or distressed in the last two weeks by various depression symptoms such as feeling blue, having thoughts of ending one's life, and crying easily. Let W be the previous occupation, which is a nominal variable with seven categories.

Assume that (S_1, S_0) given W follows (7), where $\rho(w)$ is the correlation coefficient of S_1 and S_0 given $W = w$. Further assume linear models for Y_1 and Y_0,

$$Y_z = \beta_{z0} + \beta_{z1} S_1 + \beta_{z2} S_0 + \epsilon_{Y_z},$$

where

$$\epsilon_{Y_1} \sim N(0, \sigma_{Y_1}^2), \quad \epsilon_{Y_0} \sim N(0, \sigma_{Y_0}^2),$$
$$(\epsilon_{Y_1}, \epsilon_{Y_0}) \perp\!\!\!\perp (S_1, S_0, W).$$

We choose the linear model because of its simplicity for illustration, and acknowledge its limitation and leave the task of building more flexible models for Y_1 and Y_0 to future work. Under this model,

$$\tau_{s_1 s_0} = \beta_{10} - \beta_{00} + (\beta_{11} - \beta_{01}) s_1 + (\beta_{12} - \beta_{02}) s_0.$$

We assume $\rho(w) = \rho$ and treat ρ as the sensitivity parameter within $\{0, 0.2, 0.4, 0.6, 0.8\}$. From Corollary 2, the PCEs are identifiable. We use a Bayesian approach and simulate the posterior distributions of the PCEs. To assess the sensitivity of our results to different priors, we choose two different priors. Let $\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11}, \beta_{12})$, $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01}, \beta_{02})$ and $\boldsymbol{\mu}_w = (\mu_1(w), \mu_0(w))$. For the first prior, we choose multivariate Normal priors for $\boldsymbol{\beta}_z$ and $\boldsymbol{\mu}_w$: $\boldsymbol{\beta}_z \sim \boldsymbol{N}_3(\mathbf{0}, \boldsymbol{\Omega}_z)$, $\boldsymbol{\mu}_w \sim \boldsymbol{N}_2(\mathbf{0}, \boldsymbol{\Omega})$, with

$$\boldsymbol{\Omega}_z = 10^2 \operatorname{diag}(1, 1, 1), \quad \boldsymbol{\Omega} = 10^2 \operatorname{diag}(1, 1)$$

for $z = 0, 1$, and $w = 1, \ldots, 7$. We choose the following noninformative priors for the other parameters: $f(\sigma_{zw}^2) \propto 1/\sigma_{zw}^2$, $f(\sigma_{Y_z}^2) \propto 1/\sigma_{Y_z}^2$, $\{\mathbb{P}(W = 1), \ldots, \mathbb{P}(W = 7)\} \sim$ Dirichlet$(1, \ldots, 1)$ and $\mathbb{P}(Z = 1 \mid W = w) \sim$ Beta$(1, 1)$, where $z = 0, 1$ and $w = 1, \ldots, 7$. For the second prior, we choose

$$\boldsymbol{\Omega}_z = \operatorname{diag}(1, 1, 1), \quad \boldsymbol{\Omega} = \operatorname{diag}(1, 1)$$

and keep other prior distributions unchanged. We will present the results for the first prior in the main text and show the sensitivity check of the results to different priors in Appendix C.2.

Figure 3 shows the posterior medians of $\tau_{s_1 s_0}$ for all (s_1, s_0) under $\rho = 0$. The surface of these posterior medi-

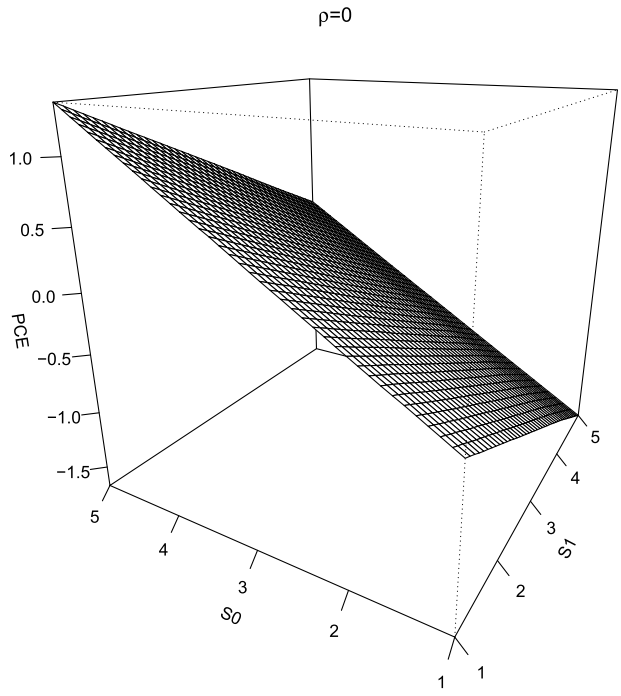

FIG. 3. *Posterior medians of the PCEs with* $\rho = 0$.

ans rises from its lowest point at principal stratum (5, 1) to its highest point at principal stratum (1, 5). In general, the estimated PCE increases as the difference between S_1 and S_0 decreases. That is, for people who gained more for the job-search self-efficacy from the treatment, the treatment lowered the risk of depression to a larger extent. Imai, Keele and Tingley (2010) analyzed this data using a mediation analysis and found that the indirect effect of the treatment through job-search self-efficacy was negative. So the program participation decreased depressive symptoms by increasing the level of job search self-efficacy. Jo et al. (2011) used the principal stratification approach by dichotomizing the job-search self-efficacy and found that the treatment had a negative effect on the depression for people whose job-search self-efficacy was improved by the treatment. Our conclusion corroborates with their findings.

For sensitivity analysis, we focus on five principal strata, consisting of the maximum, minimum, 25%, 50%, and 75% quantiles of S_1 and S_0. Table 3 shows their posterior medians and 95% credible intervals for different values of ρ. The point estimates are not sensitive to the values of ρ, and the interval estimates are not sensitive to small values of ρ. But as ρ grows larger, the intervals tend to become wider, which makes the results not significant.

Appendix C contains more details for the data analysis. Corollary 2 requires W to be continuous but W is categorical in our application. Appendix C.1 gives a formal justification of the identifiability of the PCEs in our model with a discrete W. The Normality assumptions on the outcomes are invoked for convenience in the Bayesian computation. In fact, without Normality, we can use the method of moments to estimate the PCEs. Appendix C.2 presents the results from the method of moments which are similar to those from the Bayesian inference. Including additional covariates can make Assumption 3 more plausible. Appendix C.3 shows an analysis with more covariates.

7. DISCUSSION

7.1 Summary and Extensions

Identification of the PCEs is an important but challenging problem. Although several empirical studies have leveraged auxiliary variables to improve inference for the PCEs, formal identification results have not been established especially for nonbinary intermediate variables. Our results supplement previous empirical studies with theoretical justifications for identification. We give identification results for several models based on Normal distributions, which can be generalized to other commonly used distributions. Appendix B.4 gives identification results for models based on t distributions, which are useful for robust analysis of data with heavy tails.

TABLE 3
Posterior medians and credible intervals of some PCEs. The intervals excluding zero are highlighted in bold

(S_1, S_0)	$\rho = 0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
(1.00, 5.00)	1.363	1.901	1.676	1.790	1.530
	(−0.332, 3.164)	(−0.504, 4.681)	(−0.837, 4.125)	(−1.167, 5.182)	(−1.331, 4.832)
(3.67, 4.50)	0.288	0.392	0.366	0.389	0.318
	(−0.009, 0.613)	(−0.053, 0.962)	(−0.143, 0.876)	(−0.240, 1.107)	(−0.310, 1.047)
(4.17, 4.00)	−0.093	−0.112	−0.100	−0.104	−0.099
	(−0.197, 0.009)	(−0.227, 0.004)	(−0.220, 0.009)	(−0.240, 0.011)	(−0.234, 0.017)
(4.67, 3.58)	−0.439	−0.563	−0.522	−0.550	−0.476
	(−0.815, −0.077)	**(−1.202, −0.030)**	(−1.104, 0.053)	(−1.362, 0.152)	(−1.315, 0.230)
(5.00, 1.67)	−1.386	−1.732	−1.700	−1.773	−1.496
	(−2.451, −0.428)	**(−3.428, −0.338)**	**(−3.451, −0.011)**	(−4.251, 0.368)	(−4.166, 0.717)

Researchers have conducted sensitivity analyses for the principal ignorability and the auxiliary independence. For example, Ding and Lu (2017) proposed the sensitivity analysis for principal ignorability with a binary intermediate variable, and Jiang, Ding and Geng (2016) proposed the sensitivity analysis for auxiliary independence using a random-effects model. However, there is no general setup for the sensitivity analysis of these assumptions, which depends on the specification of the model and types of the outcomes and the intermediate variables. We believe that sensitivity analysis should be routinely conducted in problems with principal stratification, but leave the development and the technical details to future research.

7.2 Comparing Two Strategies

Auxiliary variables play different roles in identifying the PCEs, depending on the underlying assumptions. Under principal ignorability, auxiliary variables can be viewed as "confounders" between the principal stratification variable and the outcome. In contrast, under auxiliary independence, auxiliary variables can be treated as an "instrumental variables" for the relationship between the principal stratification and the outcome. Therefore, the comparison between the principal ignorability and auxiliary independence for identifying the PCEs resembles the comparison between the ignorability assumption (Rosenbaum and Rubin, 1983) and the instrumental variable method (Angrist, Imbens and Rubin, 1996) for identifying the average causal effect. The methods based on principal ignorability are easy to employ because the assumption generally conditions on all baseline variables. However, they bear similar disadvantages as the methods based on ignorability for estimating average causal effect—we do not know whether we have conditioned on sufficient variables (Pearl, 2000, Pearl, 2009). In contrast, the methods based on auxiliary independence may be burdening to analysts and content experts because one needs to carve out a specific baseline variable as a designated auxiliary variable. However, the advantage is that we can intentionally target the variable based on science and experts' knowledge or by design. For example, this assumption can possibly be used in a multicenter trial as in Example 4, and in the augmented design for assessing the effect of vaccination as in Example 3.

Although we restrict the auxiliary variable W to be pretreatment in the paper, the auxiliary independence assumption allows it to be affected by the treatment. It only requires the auxiliary variable to be independent of the outcome conditional on the treatment and principal strata, which can hold even if the auxiliary variable is posttreatment. For example, for a binary S, Mealli and Pacini (2013) identify the PCEs in completely randomized experiments using a secondary outcome as the auxiliary variable. In contrast, the principal ignorability assumption is unlikely to hold with a posttreatment auxiliary variable. The required independence would fail due to the bias induced by conditioning on a posttreatment variable.

7.3 Alternative Identification Strategies

Alternative identification strategies do exist without requiring an auxiliary variable. For a binary intermediate variable, without monotonicity or exclusion restriction, Hirano et al. (2000) suggested using parallel outcome models to improve identifiability where the regression coefficients of the covariates are the same for all types of noncompliers. Mealli, Pacini and Stanghellini (2016) used the concentration graph theory to study the identification of the PCEs. It is of interest to combine these strategies in theory and practice.

The identification of PCEs is closely related to the identification of finite mixture models. For example, with a binary intermediate variable, the observed data with $(Z = 1, S = 1)$ is a mixture of principal strata $(S_1 = 1, S_0 = 1)$ and $(S_1 = 1, S_0 = 0)$, and the observed data with $(Z = 1, S = 0)$ is a mixture of principal strata $(S_1 = 0, S_0 = 0)$ and $(S_1 = 0, S_0 = 1)$. From this perspective, principal ignorability and auxiliary independence help to separate the components in the finite mixture model. Researchers sometimes use parametric finite mixture models for principal stratification problems (Zhang, Rubin and Mealli, 2009, Frumento et al., 2012). However, even though those models are parametrically identifiable, the estimators often have poor finite-sample properties (Frumento et al., 2016, Feller et al., 2019). These findings echo the caveat from Cox and Donnelly (2011), page 96: "If an issue can be addressed nonparametrically then it will often be better to tackle it parametrically; however, if it cannot be resolved nonparametrically then it is usually dangerous to resolve it parametrically." This is an important motivation for us to seek nonparametric and semiparametric identifiability as presented in this paper.

ACKNOWLEDGMENTS

We thank the editor and two reviewers for constructive comments. Dr. Trang Q. Nguyen helped us to correct an error in an early version of the paper. Peng Ding was partially supported by the National Science Foundation (grant # 1945136).

SUPPLEMENTARY MATERIAL

Supplement to "Identification of Causal Effects Within Principal Strata Using Auxiliary Variables" (DOI: 10.1214/20-STS810SUPP; .pdf). The supplementary material includes proofs of the theorems and propositions, additional results for identification, and more details for the data analysis.

REFERENCES

ANGRIST, J. D., IMBENS, G. W. and RUBIN, D. B. (1996). Identification of causal effects using instrumental variables (with discussion). *J. Amer. Statist. Assoc.* **91** 444–455.

BARTOLUCCI, F. and GRILLI, L. (2011). Modeling partial compliance through copulas in a principal stratification framework. *J. Amer. Statist. Assoc.* **106** 469–479. MR2866975 https://doi.org/10.1198/jasa.2011.ap09094

BICKEL, P. J. and DOKSUM, K. A. (2015). *Mathematical Statistics: Basic Ideas and Selected Topics, Vol. I*. CRC Press, Boca Raton, FL.

CHENG, J. and SMALL, D. S. (2006). Bounds on causal effects in three-arm trials with non-compliance. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 815–836. MR2301296 https://doi.org/10.1111/j.1467-9868.2006.00568.x

COCHRAN, W. G. (1957). Analysis of covariance: Its nature and uses. *Biometrics* **13** 261–281. MR0090952 https://doi.org/10.2307/2527916

CONLON, A. S. C., TAYLOR, J. M. G. and ELLIOTT, M. R. (2017). Surrogacy assessment using principal stratification and a Gaussian copula model. *Stat. Methods Med. Res.* **26** 88–107. MR3592714 https://doi.org/10.1177/0962280214539655

COX, D. R. and DONNELLY, C. A. (2011). *Principles of Applied Statistics*. Cambridge Univ. Press, Cambridge. MR2817147 https://doi.org/10.1017/CBO9781139005036

DANIELS, M. J., ROY, J. A., KIM, C., HOGAN, J. W. and PERRI, M. G. (2012). Bayesian inference for the causal effect of mediation. *Biometrics* **68** 1028–1036. MR3040009 https://doi.org/10.1111/j.1541-0420.2012.01781.x

DING, P. and LI, F. (2018). Causal inference: A missing data perspective. *Statist. Sci.* **33** 214–237. MR3797711 https://doi.org/10.1214/18-STS645

DING, P. and LU, J. (2017). Principal stratification analysis using principal scores. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 757–777. MR3641406 https://doi.org/10.1111/rssb.12191

DING, P., GENG, Z., YAN, W. and ZHOU, X.-H. (2011). Identifiability and estimation of causal effects by principal stratification with outcomes truncated by death. *J. Amer. Statist. Assoc.* **106** 1578–1591. MR2896858 https://doi.org/10.1198/jasa.2011.tm10265

EFRON, B. and FELDMAN, D. (1991). Compliance as an explanatory variable in clinical trials (with discussion). *J. Amer. Statist. Assoc.* **86** 9–17.

ELLIOTT, M. R., RAGHUNATHAN, T. E. and LI, Y. (2010). Bayesian inference for causal mediation effects using principal stratification with dichotomous mediators and outcomes. *Biostatistics* **11** 353–372.

EVERITT, B. S. and HAND, D. J. (1981). *Finite Mixture Distributions. Monographs on Applied Probability and Statistics*. CRC Press, London. MR0624267

FELLER, A., GREIF, E., HO, N., MIRATRIX, L. and PILLAI, N. (2019). Weak separation in mixture models and implications for principal stratification. arXiv preprint. Available at arXiv:1602.06595.

FOLLMANN, D. A. (2000). On the effect of treatment among would-be treatment compliers: An analysis of the multiple risk factor intervention trial. *J. Amer. Statist. Assoc.* **95** 1101–1109. MR1821718 https://doi.org/10.2307/2669746

FOLLMANN, D. (2006). Augmented designs to assess immune response in vaccine trials. *Biometrics* **62** 1161–1169. MR2307441 https://doi.org/10.1111/j.1541-0420.2006.00569.x

FORASTIERE, L., MATTEI, A. and DING, P. (2018). Principal ignorability in mediation analysis: Through and beyond sequential ignorability. *Biometrika* **105** 979–986. MR3877878 https://doi.org/10.1093/biomet/asy053

FRANGAKIS, C. E. and RUBIN, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58** 21–29. MR1891039 https://doi.org/10.1111/j.0006-341X.2002.00021.x

FRUMENTO, P., MEALLI, F., PACINI, B. and RUBIN, D. B. (2012). Evaluating the effect of training on wages in the presence of noncompliance, nonemployment, and missing outcome data. *J. Amer. Statist. Assoc.* **107** 450–466. MR2980057 https://doi.org/10.1080/01621459.2011.643719

FRUMENTO, P., MEALLI, F., PACINI, B. and RUBIN, D. B. (2016). The fragility of standard inferential approaches in principal stratification models relative to direct likelihood approaches. *Stat. Anal. Data Min.* **9** 58–70. MR3465093 https://doi.org/10.1002/sam.11299

GABRIEL, E. E. and FOLLMANN, D. (2016). Augmented trial designs for evaluation of principal surrogates. *Biostatistics* **17** 453–467. MR3603947 https://doi.org/10.1093/biostatistics/kxv055

GALLOP, R., SMALL, D. S., LIN, J. Y., ELLIOTT, M. R., JOFFE, M. and TEN HAVE, T. R. (2009). Mediation analysis with principal stratification. *Stat. Med.* **28** 1108–1130. MR2662200 https://doi.org/10.1002/sim.3533

GILBERT, P. B. and HUDGENS, M. G. (2008). Evaluating candidate principal surrogate endpoints. *Biometrics* **64** 1146–1154. MR2522262 https://doi.org/10.1111/j.1541-0420.2008.01014.x

GUSTAFSON, P. (2009). What are the limits of posterior distributions arising from nonidentified models and why should we care? *J. Amer. Statist. Assoc.* **104** 1682–1695. MR2750585 https://doi.org/10.1198/jasa.2009.tm08603

GUSTAFSON, P. (2015). *Bayesian Inference for Partially Identified Models: Exploring the limits of limited data. Monographs on Statistics and Applied Probability* **141**. CRC Press, Boca Raton, FL. MR3642458

HIRANO, K., IMBENS, G. W., RUBIN, D. B. and ZHOU, X.-H. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics* **1** 69–88.

HU, Y. and SHIU, J.-L. (2018). Nonparametric identification using instrumental variables: Sufficient conditions for completeness. *Econometric Theory* **34** 659–693. MR3803171 https://doi.org/10.1017/S0266466617000251

HUANG, Y. and GILBERT, P. B. (2011). Comparing biomarkers as principal surrogate endpoints. *Biometrics* **67** 1442–1451. MR2872395 https://doi.org/10.1111/j.1541-0420.2011.01603.x

HUDGENS, M. G. and GILBERT, P. B. (2009). Assessing vaccine effects in repeated low-dose challenge experiments. *Biometrics* **65** 1223–1232. MR2756510 https://doi.org/10.1111/j.1541-0420.2009.01189.x

IMAI, K. (2008). Sharp bounds on the causal effects in randomized experiments with "truncation-by-death". *Statist. Probab. Lett.* **78** 144–149. MR2382067 https://doi.org/10.1016/j.spl.2007.05.015

IMAI, K., KEELE, L. and TINGLEY, D. (2010). A general approach to causal mediation analysis. *Psychol. Methods* **15** 309–334.

IMBENS, G. W. and RUBIN, D. B. (2015). *Causal Inference—for Statistics, Social, and Biomedical Sciences: An introduction*. Cambridge Univ. Press, New York. MR3309951 https://doi.org/10.1017/CBO9781139025751

JIANG, Z. and DING, P. (2021). Supplement to "Identification of causal effects within principal strata using auxiliary variables." https://doi.org/10.1214/20-STS810SUPP

JIANG, Z., DING, P. and GENG, Z. (2016). Principal causal effect identification and surrogate end point evaluation by multiple trials. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 829–848. MR3534352 https://doi.org/10.1111/rssb.12135

JIN, H. and RUBIN, D. B. (2008). Principal stratification for causal inference with extended partial compliance. *J. Amer. Statist. Assoc.* **103** 101–111. MR2463484 https://doi.org/10.1198/016214507000000347

JO, B. (2008). Causal inference in randomized experiments with mediational processes. *Psychol. Methods* **13** 314–336.

JO, B. and STUART, E. A. (2009). On the use of propensity scores in principal causal effect estimation. *Stat. Med.* **28** 2857–2875. MR2750169 https://doi.org/10.1002/sim.3669

JO, B., STUART, E. A., MACKINNON, D. P. and VINOKUR, A. D. (2011). The use of propensity scores in mediation analysis. *Multivar. Behav. Res.* **46** 425–452.

JOFFE, M. M., SMALL, D. and HSU, C.-Y. (2007). Defining and estimating intervention effects for groups that will develop an auxiliary outcome. *Statist. Sci.* **22** 74–97. MR2408662 https://doi.org/10.1214/088342306000000655

KIM, C., HENNEMAN, L. R. F., CHOIRAT, C. and ZIGLER, C. M. (2020). Health effects of power plant emissions through ambient air quality. *J. Roy. Statist. Soc. Ser. A* **183** 1677–1703. MR4157831

LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. Springer, New York. MR2135927

LITTLE, R. J. and YAU, L. H. (1998). Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using Rubin's causal model. *Psychol. Methods* **3** 147.

MATTEI, A. and MEALLI, F. (2011). Augmented designs to assess principal strata direct effects. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 729–752. MR2867456 https://doi.org/10.1111/j.1467-9868.2011.00780.x

MEALLI, F. and PACINI, B. (2013). Using secondary outcomes to sharpen inference in randomized experiments with noncompliance. *J. Amer. Statist. Assoc.* **108** 1120–1131. MR3174688 https://doi.org/10.1080/01621459.2013.802238

MEALLI, F., PACINI, B. and STANGHELLINI, E. (2016). Identification of principal causal effects using additional outcomes in concentration graphs. *J. Educ. Behav. Stat.* **41** 463–480.

NELSEN, R. B. (2006). *An Introduction to Copulas*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2197664 https://doi.org/10.1007/s11229-005-3715-x

PEARL, J. (2000). *Causality*: *Models*, *reasoning*, *and inference*. Cambridge Univ. Press, Cambridge. MR1744773

PEARL, J. (2009). Letter to the editor: Remarks on the method of propensity score. *Stat. Med.* **28** 1415–1416. MR2724701 https://doi.org/10.1002/sim.3521

QIN, L., GILBERT, P. B., FOLLMANN, D. and LI, D. (2008). Assessing surrogate endpoints in vaccine trials with case-cohort sampling and the Cox model. *Ann. Appl. Stat.* **2** 386–407. MR2415608 https://doi.org/10.1214/07-AOAS132

ROSENBAUM, P. R. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *J. Roy. Statist. Soc. Ser. A* **147** 656–666.

ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974 https://doi.org/10.1093/biomet/70.1.41

ROY, J., HOGAN, J. W. and MARCUS, B. H. (2008). Principal stratification with predictors of compliance for randomized trials with 2 active treatments. *Biostatistics* **9** 277–289.

RUBIN, D. B. (2004). Direct and indirect causal effects via potential outcomes. *Scand. J. Stat.* **31** 161–170. MR2066246 https://doi.org/10.1111/j.1467-9469.2004.02-123.x

RUBIN, D. B. (2006). Causal inference through potential outcomes and principal stratification: Application to studies with "censoring" due to death. *Statist. Sci.* **21** 299–309. MR2339125 https://doi.org/10.1214/088342306000000114

SCHWARTZ, S. L., LI, F. and MEALLI, F. (2011). A Bayesian semiparametric approach to intermediate variables in causal inference. *J. Amer. Statist. Assoc.* **106** 1331–1344. MR2896839 https://doi.org/10.1198/jasa.2011.ap10425

SOMMER, A. and ZEGER, S. L. (1991). On estimating efficacy from clinical trials. *Stat. Med.* **10** 45–52.

STUART, E. A. and JO, B. (2015). Assessing the sensitivity of methods for estimating principal causal effects. *Stat. Methods Med. Res.* **24** 657–674. MR3428422 https://doi.org/10.1177/0962280211421840

TINGLEY, D., YAMAMOTO, T., HIROSE, K., KEELE, L. and IMAI, K. (2014). Mediation: R package for causal mediation analysis. *J. Stat. Softw.* **59** 1–38.

VANDERWEELE, T. J. (2008). Simple relations between principal stratification and direct and indirect effects. *Statist. Probab. Lett.* **78** 2957–2962. MR2516810 https://doi.org/10.1016/j.spl.2008.05.029

VINOKUR, A. D., PRICE, R. H. and SCHUL, Y. (1995). Impact of the JOBS intervention on unemployed workers varying in risk for depression. *Am. J. Community Psychol.* **23** 39–74.

VINOKUR, A. D. and SCHUL, Y. (1997). Mastery and inoculation against setbacks as active ingredients in the jobs intervention for the unemployed. *J. Consult. Clin. Psychol.* **65** 867.

WANG, L., ZHOU, X.-H. and RICHARDSON, T. S. (2017). Identification and estimation of causal effects with outcomes truncated by death. *Biometrika* **104** 597–612. MR3694585 https://doi.org/10.1093/biomet/asx034

YANG, F. and DING, P. (2018). Using survival information in truncation by death problems without the monotonicity assumption. *Biometrics* **74** 1232–1239. MR3908141 https://doi.org/10.1111/biom.12883

YUAN, L.-H., FELLER, A. and MIRATRIX, L. W. (2019). Identifying and estimating principal causal effects in a multi-site trial of early college high schools. *Ann. Appl. Stat.* **13** 1348–1369. MR4019142 https://doi.org/10.1214/18-AOAS1235

ZHANG, J. L. and RUBIN, D. B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *J. Educ. Behav. Stat.* **28** 353–368.

ZHANG, J. L., RUBIN, D. B. and MEALLI, F. (2009). Likelihood-based analysis of causal effects of job-training programs using principal stratification. *J. Amer. Statist. Assoc.* **104** 166–176. MR2663040 https://doi.org/10.1198/jasa.2009.0012

ZIGLER, C. M. and BELIN, T. R. (2012). A Bayesian approach to improved estimation of causal effect predictiveness for a principal surrogate endpoint. *Biometrics* **68** 922–932. MR3055197 https://doi.org/10.1111/j.1541-0420.2011.01736.x