Journal of Machine Learning Research 27 (2026) 1-65 Submitted 3/25; Revised 6/26; Published 7/26

Identifiability of the Instrumental Variable Model with the Treatment and Outcome Missing Not at Random

Shuozhi Zuo shuozhi@umich.edu
Department of Statistics
University of Michigan
Ann Arbor, MI 48109, USA

Peng Ding† pengdingpku@berkeley.edu
Department of Statistics
University of California
Berkeley, CA 94720, USA

Fan Yang† yangfan1987@tsinghua.edu.cn
Yau Mathematical Sciences Center
Tsinghua University
Haidian District, Beijing 100084, China

Yanqi Lake Beijing Institute of Mathematical Sciences and Applications
Huairou District, Beijing 101408, China

†*: Corresponding author*

Editor: Kun Zhang

Abstract

Under the instrumental variable model of Imbens and Angrist (1994) and Angrist et al. (1996), we can identify the local average treatment effect, also known as the complier average causal effect (CACE). In practice, however, the treatment and outcome are often missing. When they are missing not at random (MNAR), the underlying data distribution cannot be recovered, so the CACE is generally not identifiable without further assumptions. We study the conditions under which the CACE remains identifiable when data are MNAR. Searching exhaustively over missingness mechanisms, we characterize those that identify the CACE without auxiliary information in two settings: (1) missing data in the treatment alone or the outcome alone, and (2) missing data in both the treatment and the outcome under prospective data collection. We unify existing results and establish new ones, giving a complete picture of identifiability in each setting. The results have two practical implications. First, before analyzing data under the instrumental variable model, one should check whether the CACE is identifiable under the assumed missingness mechanism. Second, because the true mechanism is usually unknown and untestable, it is more robust to estimate the CACE under several plausible mechanisms and compare the results.

Keywords: causal inference, complier average causal effect, local average treatment effect, nonignorable missing data, nonparametric identification, potential outcome

1. Introduction

Instrumental variable (IV) methods are widely used to study causal effects when the treatment and outcome are confounded by unmeasured variables. An IV is a variable that meets the following conditions: (*i*) it is associated with the treatment, (*ii*) it does not directly affect the outcome except through the treatment, and (*iii*) it is uncorrelated with the unmeasured confounders given the observed covariates. An IV therefore isolates variation in the treatment that is free of unmeasured confounding, and this variation identifies the causal effect of the treatment on the outcome. Imbens and Angrist (1994) and Angrist et al. (1996) added the monotonicity assumption that there are no defiers, that is, no units who take the treatment only when not encouraged by the IV. Under this assumption, the two-stage least squares estimator identifies the complier average causal effect (CACE), where compliers are the units who take the treatment only when encouraged by the IV. A separate line of work bounds effects rather than point-identifying them: Balke and Pearl (1997) established sharp nonparametric bounds on the average treatment effect for the entire population. We follow Imbens and Angrist (1994) and Angrist et al. (1996) and focus on point identification of the CACE.

Missing data in the treatment and outcome are common in empirical IV studies (Mealli et al., 2004; Lui and Chang, 2010; Eren and Ozbeklik, 2013; Frölich and Huber, 2014; von Hinke et al., 2016; Huber, 2021; Rai, 2023). The difficulty of identification depends on the missingness mechanism. When data are missing completely at random (MCAR), so that missingness is independent of all variables, a complete-case analysis identifies the CACE. When data are missing at random (MAR), so that missingness is independent of the unobservables given the observables, standard approaches such as multiple imputation or nonresponse weighting yield valid inference. When data are missing not at random (MNAR), so that missingness depends on the unobservables even given the observables, identifying the CACE generally requires additional assumptions. In the IV setting, MNAR arises in two main ways. First, missingness may depend on the latent compliance status, that is, on how the unit responds to the IV encouragement. For instance, individuals who comply with the IV may be more willing to report their treatment and outcome than those who do not. Second, missingness may depend on the missing values themselves. For instance, individuals may be reluctant to disclose treatment or outcome values that are private or sensitive. Section 3 gives empirical examples of both.

The IV literature has mostly studied identification of the CACE when the outcome is MNAR, and the mechanisms it proposes fall into two groups. The first group lets missingness depend on the latent compliance status. Frangakis and Rubin (1999) established nonparametric identification under one-sided noncompliance using two conditions: latent ignorability, namely that outcome missingness is conditionally independent of the outcome given the latent compliance status, treatment, and IV; and a response exclusion restriction, namely that the IV affects missingness only through the treatment. Zhou and Li (2006) and O'Malley and Normand (2005) extended these results to two-sided noncompliance for binary and continuous outcomes, respectively. Under the same mechanism, Peng et al. (2004) used parametric models to estimate the CACE for continuous and categorical outcomes, Taylor and Zhou (2009) proposed multiple imputation methods, and Frölich and Huber (2014) extended the setting to multiple outcome periods. Departing from Frangakis and

Rubin (1999), Mealli et al. (2004) and Mattei and Mealli (2007) obtained nonparametric identification under one-sided noncompliance and latent ignorability using an alternative response exclusion restriction, namely that for compliers the outcome missingness is unaffected by the IV. Jo et al. (2010) provided parametric estimation strategies for both the Frangakis and Rubin (1999) and Mealli et al. (2004) mechanisms, and Nguyen et al. (2024) and Huber (2021) examined the role and plausibility of latent ignorability in practice.

The second group lets missingness depend on the outcome value itself. Chen et al. (2009a) gave nonparametric identification results for discrete outcomes; Chen et al. (2015) extended these to continuous outcomes whose distribution belongs to an exponential family. By incorporating covariates that affect the outcome but not its missingness, Chen et al. (2009a) further allowed missingness to depend on the latent compliance status and/or the IV in addition to the outcome value. Imai (2009) provided nonparametric identification for a binary outcome whose missingness depends on both the missing outcome value and the treatment.

This paper studies nonparametric identification of the CACE without auxiliary information. When missingness occurs only in the outcome, we unify existing results and provide new findings for the MNAR case. Treatment MNAR has received less attention in the IV literature than outcome MNAR; an exception is Calvi et al. (2022), who studied identification under treatment MNAR using two proxies. For missingness in the treatment alone, as for missingness in the outcome alone, we search exhaustively over all missingness mechanisms. We identify the most general ones that permit nonparametric identification under positivity and conditional dependence conditions, and give counterexamples for the rest. We then extend the analysis to the case where both the treatment and outcome are missing. Because the missingness mechanism is usually unknown and untestable, Huber (2021) recommends comparing alternative mechanisms as a robustness check. We echo this recommendation and supply the theoretical basis for such comparisons.

The rest of the paper is organized as follows. Section 2 introduces the notation and basic concepts for the IV model. Section 3 gives real-world IV examples where the treatment and/or outcome may be MNAR. Sections 4, 5, and 6 present the nonparametric identification results for missingness in the outcome alone, in the treatment alone, and in both. Section 7 illustrates the practical implications through a simulation study, and Section 8 applies the framework to the National Job Corps Study. We conclude with a discussion and present the proofs in the appendix.

2. Review of the IV Model without Missing Data

We focus on the setup with a binary IV and a binary treatment. Let Z denote the IV, with $Z = 1$ encouraging receipt of the treatment and $Z = 0$ otherwise, and let D denote the treatment received, with $D = 1$ and $D = 0$ for the treatment and control conditions. Noncompliance arises when $Z \neq D$ for some units. Let Y denote the outcome of interest. To simplify the presentation, we suppress the pretreatment covariates from the notation and condition on them implicitly, and we assume that Z is randomized within strata of the covariates. All assumptions and identification results below hold conditional on the covariates. We adopt the potential outcomes framework, defining the potential treatments $\{D(1), D(0)\}$ and potential outcomes $\{Y(1), Y(0)\}$ with respect to the IV. The observed values are $D = ZD(1) + (1 - Z)D(0)$ and $Y = ZY(1) + (1 - Z)Y(0)$. We index potential outcomes by the IV as $Y(z)$; the alternative notations $Y(d)$ and $Y(z, d)$ also appear in the literature; see Chapter 21.6 of Ding (2024) for a discussion. Following Imbens and Angrist (1994) and Angrist et al. (1996), we classify units into four latent compliance statuses, denoted U, according to the joint potential treatments $\{D(1), D(0)\}$:

$$U = \begin{cases} a, & \text{if } D(1) = 1 \text{ and } D(0) = 1; \\ c, & \text{if } D(1) = 1 \text{ and } D(0) = 0; \\ d, & \text{if } D(1) = 0 \text{ and } D(0) = 1; \\ n, & \text{if } D(1) = 0 \text{ and } D(0) = 0, \end{cases}$$

where a, c, d, and n denote always-taker, complier, defier, and never-taker, respectively. Two-sided noncompliance refers to the situation in which some units in both the $Z = 1$ and $Z = 0$ groups fail to comply with their IV encouragement. One-sided noncompliance refers to the situation in which units in the $Z = 0$ group have no access to the treatment, so that $D(0) = 0$ and only compliers and never-takers are present. Given U and Z, the treatment $D = ZD(1)+(1-Z)D(0)$ is a deterministic function of Z, so that $\{Y(1), Y(0)\} \perp\!\!\!\perp D \mid U, Z$. Conditioning on U therefore removes all confounding between D and Y, and we use the same U throughout the directed acyclic graphs (DAGs) and in $\mathbb{P}(Z, U, D, Y)$. The causal effect of interest is the CACE,

$$\text{CACE} = \mathbb{E}\{Y(1) - Y(0) \mid U = c\}.$$

We impose the following IV assumptions throughout the paper.

IV Assumptions
(1) *Randomization:* $Z \perp\!\!\!\perp \{D(1), D(0), Y(1), Y(0)\}$*;*
(2) *Monotonicity:* $D(1) \geq D(0)$*, which implies that there are no defiers;*
(3) *Nonzero average causal effect of* Z *on* D*:* $\mathbb{E}\{D(1) - D(0)\} \neq 0$*;*
(4) *Exclusion restriction:* $Y(1) = Y(0)$ *for always-takers* $(U = a)$ *and never-takers* $(U = n)$*.*

Under the IV assumptions, Imbens and Angrist (1994) and Angrist et al. (1996) showed that the CACE is identified by

$$\text{CACE} = \frac{\mathbb{E}(Y \mid Z = 1) - \mathbb{E}(Y \mid Z = 0)}{\mathbb{P}(D = 1 \mid Z = 1) - \mathbb{P}(D = 1 \mid Z = 0)}. \tag{1}$$

The right-hand side of Equation (1) divides the difference in mean outcomes across the two IV groups by the difference in mean treatment received, and both differences are estimable from the sample moments of Y and D.

Equation (1) applies only when no data are missing; the rest of the paper addresses identification when D, Y, or both are missing.

3. Examples of IV Studies with Missing Data

We review four empirical IV studies where the treatment and/or outcome may be MNAR.

Example 1 (MNAR in Y) *Mealli et al. (2004) examined the effect of an enhanced training course on the practice of breast self-examination (BSE) in Faenza, Italy. In their study, 657 women were randomly assigned to either the standard treatment of receiving only mailed information about BSE ($Z = 0$) or the enhanced treatment of an additional training course ($Z = 1$). The binary indicator D records whether a woman received the enhanced treatment, and the outcome Y records whether she practiced BSE one year later. Only 55% of the women assigned to the enhanced treatment adhered to their assignment. Those assigned to the standard treatment had no access to the enhanced treatment, so the noncompliance is one-sided. The outcome is missing for 35% of the women. Mealli et al. (2004) suggested that missingness may depend on the compliance status U and the treatment assignment Z. For example, women who complied with their assigned treatment might be more likely to respond to the survey. Beyond U and Z, Small and Cheng (2009) argued that missingness in Y may also depend on Y itself, since women who practiced BSE might be more willing to respond than those who did not.*

Example 2 (MNAR in D and MAR in Y) *Rai (2023) examined the effect of educational attainment and IQ scores on wages, using a second measure of ability as an IV to address measurement error in the IQ scores. The data come from the 1976 National Longitudinal Survey of Young Men in the United States (Card, 1993). Here Z measures scores on a "Knowledge of the World of Work" test, D is the IQ score, and Y is the hourly wage in 1976. About 58% of subjects have both D and Y observed, 26% have Y but not D, 11% have D but not Y, and 5% have both missing. Rai (2023) argued that missingness in Y stems from survey design rather than intentional nonresponse, so it is reasonable to model Y as MAR. Missingness in D, in contrast, could be MNAR if it arises from individuals or schools unwilling to report low IQ scores.*

Example 3 (MNAR in both D and Y) *von Hinke Kessler Scholder et al. (2014) studied the effect of prenatal alcohol exposure on children's academic achievement for a cohort born in the Avon area of England, using a validated genetic variant as the IV (Zuccolo et al., 2009). Here $Z = 1$ when the mother carried the rare variant and $Z = 0$ otherwise. The treatment is $D = 1$ if the mother reported drinking any amount at any time during pregnancy and $D = 0$ if she reported not drinking in any of the three trimesters. The outcomes (Y) are several measures of children's academic achievement, namely scores on nationally set examinations at different ages, drawn from the National Pupil Database in England. Depending on the outcome measure, the fraction of subjects missing D and/or Y ranges from 40% to 53%. Missingness in Y could arise from problems linking the data sets, from children not sitting some or all of the exams, or from other factors (Jay et al., 2019). Linkage problems might be MAR, but not sitting exams could be associated with lower achievement and thus be MNAR. Similarly, because of the stigma around drinking during pregnancy, mothers who drank might be less likely to report it, which could make D MNAR.*

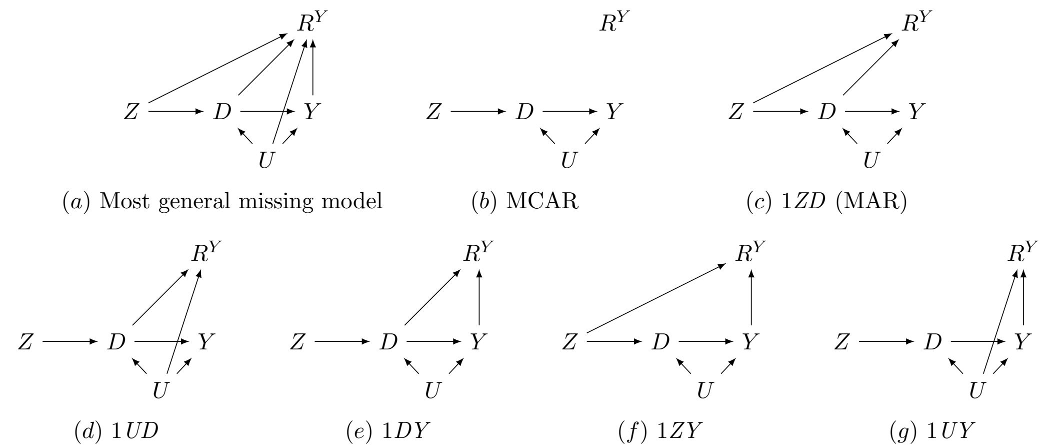
Example 4 (MNAR in both D and Y) *Using the National Education Longitudinal Study of 1988, Eren and Ozbeklik (2013) examined the effect of noncognitive ability on the earnings of young men, using an earlier measure of the same ability as the IV to address measurement error. Here Z is the standardized eighth-grade Rosenberg and Rotter scales, D is the tenth-grade noncognitive ability (the standardized average of the Rosenberg and Rotter scales),* *and Y is log weekly earnings. About 18% of subjects are missing D and/or Y. Because the data were collected by survey, individuals with low noncognitive ability might be less willing to report it, and those with high earnings might be less willing to disclose the amount. Both patterns would make D and Y MNAR.*

These examples show that the plausible missingness DAGs should be chosen using subject-matter knowledge and the study design. The identification results then determine which of these plausible mechanisms permit recovery of the CACE. In Example 1, the concern that response may depend on compliance status and on the unobserved outcome suggests considering mechanisms in which R^Y depends on U and/or Y. In Example 2, reluctance to report low IQ scores suggests considering missingness mechanisms in the treatment in which R^D may depend on the missing treatment value D. In Example 3, the stigma around drinking during pregnancy and the link between low achievement and not sitting exams suggest combined mechanisms in which R^D depends on D and R^Y on Y. In Example 4, reluctance to report low noncognitive ability and high earnings points to the same structures; both fall under the combined mechanisms of Section 6. This motivates the strategy we adopt in the rest of the paper. Because the mechanism cannot be tested from the observed data, we recommend using subject-matter knowledge and the data-collection design to select a set of plausible mechanisms, estimating the CACE under each identifiable one, and comparing the results. Agreement across mechanisms supports the stability of the substantive conclusion, whereas large discrepancies signal sensitivity to the missingness assumptions and should be reported. We illustrate this comparison across identifiable mechanisms in Sections 7 and 8. When the plausible set also includes nonidentifiable mechanisms, the comparison can be further supplemented with a formal sensitivity analysis, such as those of Small and Cheng (2009) and Zuo et al. (2025).

4. Missingness Only in the Outcome

This section considers missingness only in the outcome. Let R^Y be the response indicator for Y, with $R^Y = 1$ if Y is observed and $R^Y = 0$ otherwise. We describe missingness mechanisms with DAGs (Mohan and Pearl, 2021; Fay, 1986; Ma et al., 2003). The most general mechanism lets R^Y depend on all of Z, U, D, and Y, and it does not permit nonparametric identification of the CACE (see Section C.1 of the appendix for counterexamples). Figure 1 displays this most general mechanism along with MCAR, MAR, and four MNAR mechanisms that permit nonparametric identification of the CACE. We label each mechanism by the prefix 1 followed by the variables that R^Y may depend on. For example, Assumption $1ZD$ is the mechanism in which R^Y depends on (Z, D) and is conditionally independent of (U, Y).

When data are MCAR, i.e., $R^Y \perp\!\!\!\perp (Z, U, D, Y)$ and $\mathbb{P}(Z, D, Y) = \mathbb{P}(Z, D, Y \mid R^Y = 1)$, a complete-case analysis gives a consistent estimate of the CACE. Besides MCAR, the CACE is nonparametrically identifiable under the MAR mechanism and under each of the four MNAR mechanisms in Figure 1 (d)–(g). Assumptions $1ZD$ and $1UD$ require only positivity. The self-censoring mechanisms, in which R^Y may depend on the missing outcome value Y itself (Li et al., 2023) (Assumptions $1DY$, $1ZY$, and $1UY$), require a binary outcome; Assumptions $1DY$ and $1ZY$ additionally require two-sided noncompliance and a conditional dependence condition. Table 1 summarizes these conditions. These four



(a) Most general missing model (b) MCAR (c) $1ZD$ (MAR)

(d) $1UD$ (e) $1DY$ (f) $1ZY$ (g) $1UY$

Figure 1: The DAGs in (a)–(g) describe the most general missingness mechanism, MCAR, MAR, and four MNAR mechanisms, respectively, when missingness exists only in the outcome.

Table 1: Summary of the mechanisms for missingness in the outcome, studied in Section 4. Each row identifies the CACE under the listed conditions. A checkmark indicates that a positivity condition is required; the exact positivity condition is stated in Theorem 1. The symbol $*$ means no restriction on that aspect.

Mechanism	Missingness assumption	Positivity	Outcome	Noncompliance	Dependence
$1ZD$	$R^Y \perp\!\!\!\perp (U, Y) \mid (Z, D)$	✓	$*$	$*$	$*$
$1UD$	$R^Y \perp\!\!\!\perp (Z, Y) \mid (U, D)$	✓	$*$	$*$	$*$
$1DY$	$R^Y \perp\!\!\!\perp (Z, U) \mid (D, Y)$	✓	Binary	Two-sided	$Y \not\perp\!\!\!\perp Z \mid D = d$ for $d = 0, 1$
$1ZY$	$R^Y \perp\!\!\!\perp (U, D) \mid (Z, Y)$	✓	Binary	Two-sided	$Y \not\perp\!\!\!\perp D \mid Z = z$ for $z = 0, 1$
$1UY$	$R^Y \perp\!\!\!\perp (Z, D) \mid (U, Y)$	✓	Binary	$*$	$*$

MNAR mechanisms are the most general that permit nonparametric identification: each lets R^Y depend on two variables, and allowing R^Y to depend on any further variable yields a mechanism under which the CACE is no longer identifiable. We provide counterexamples for the nonidentifiable mechanisms in Section C.1 of the appendix, namely the two-variable mechanism $1ZU$ and the three-variable mechanisms. We now state the MAR and four MNAR assumptions.

Assumption $1ZD$ $R^Y \perp\!\!\!\perp (U, Y) \mid (Z, D)$.

Assumption $1ZD$ is the MAR mechanism, since the probability of observing Y depends only on the fully observed IV Z and treatment D.

Assumption $1UD$ $R^Y \perp\!\!\!\perp (Z, Y) \mid (U, D)$.

Assumption $1UD$ allows the missingness in Y to depend on U and D. It corresponds to latent ignorability together with the response exclusion restriction that Z affects R^Y only through D (Frangakis and Rubin, 1999; Zhou and Li, 2006; O'Malley and Normand, 2005).

Assumption $1DY$ $R^Y \perp\!\!\!\perp (Z, U) \mid (D, Y)$.

Assumption $1DY$ is a self-censoring mechanism. It was introduced by Imai (2009), who considered the setting of two-sided noncompliance and a binary outcome, and it also appears as Assumption 9 of Chen et al. (2009b).

Assumption $1ZY$ $R^Y \perp\!\!\!\perp (U, D) \mid (Z, Y)$.

Assumption $1ZY$ allows both the IV and the outcome value to affect the probability of observing Y. Small and Cheng (2009) studied parametric identification under this mechanism, whereas we study its nonparametric identification.

Assumption $1UY$ $R^Y \perp\!\!\!\perp (Z, D) \mid (U, Y)$.

Assumption $1UY$ instead allows the latent compliance status and the outcome value to affect the probability of observing Y. As with Assumption $1ZY$, Small and Cheng (2009) studied parametric identification, whereas we study its nonparametric identification.

The following theorem formalizes the identification results in Table 1.

Theorem 1 *Suppose the IV assumptions hold. The CACE is nonparametrically identifiable in each of the following cases:*
($1ZD$) *under Assumption* $1ZD$, *if* $\mathbb{P}(R^Y = 1 \mid Z = z, D = d) > 0$ *for all* z *and* d;
($1UD$) *under Assumption* $1UD$, *if* $\mathbb{P}(R^Y = 1 \mid U = c, D = d) > 0$ *for* $d = 0, 1$;
($1DY$) *under Assumption* $1DY$, *for a binary* Y *with two-sided noncompliance, if* $\mathbb{P}(R^Y = 1 \mid D = d, Y = y) > 0$ *for all* d *and* y *and* $Y \not\perp\!\!\!\perp Z \mid D = d$ *for* $d = 0, 1$;
($1ZY$) *under Assumption* $1ZY$, *for a binary* Y *with two-sided noncompliance, if* $\mathbb{P}(R^Y = 1 \mid Z = z, Y = y) > 0$ *for all* z *and* y *and* $Y \not\perp\!\!\!\perp D \mid Z = z$ *for* $z = 0, 1$;
($1UY$) *under Assumption* $1UY$, *for a binary* Y, *if* $\mathbb{P}(R^Y = 1 \mid U = c, Y = y) > 0$ *for* $y = 0, 1$.

In Theorem 1, part ($1ZD$) is the MAR case, and part ($1UD$) aligns with Frangakis and Rubin (1999), Zhou and Li (2006), and O'Malley and Normand (2005). Part ($1DY$) corresponds to Imai (2009) and Chen et al. (2009b), where we additionally highlight the role of two-sided noncompliance. Parts ($1ZY$) and ($1UY$) are new nonparametric identification results under outcome MNAR.

All five parts of Theorem 1 include a positivity assumption, which ensures that observed data on Y are available in all relevant strata. Identification of the CACE under MAR is standard, and identification under Assumption $1UD$ has been discussed in the literature, so we refer readers to the appendix for details. Below, we focus on parts ($1DY$) through ($1UY$) and give the intuition behind the conditions required by the three mechanisms that involve Y.

One-sided noncompliance is a special case of two-sided noncompliance with no always-takers, so one might expect that identification under the latter implies identification under the former. Parts ($1DY$) and ($1ZY$) of Theorem 1 show otherwise: in both, the CACE is identifiable under two-sided but not under one-sided noncompliance. We illustrate the reason using $(1DY)$. Under Assumption $1DY$,

$$\mathbb{P}(D = d, Y = y \mid Z = z) = \frac{\mathbb{P}(D = d, Y = y, R^Y = 1 \mid Z = z)}{\mathbb{P}(R^Y = 1 \mid D = d, Y = y)}.$$

Therefore, the conditional distribution $\mathbb{P}(D, Y \mid Z)$ is identifiable, and so is the CACE if $\mathbb{P}(R^Y \mid D, Y)$ is identifiable. Define

$$\eta_d(y) = \frac{\mathbb{P}(R^Y = 0 \mid D = d, Y = y)}{\mathbb{P}(R^Y = 1 \mid D = d, Y = y)}$$

for all d and y. For each $d = 0, 1$, we have the following system of linear equations with $\{\eta_d(y) : y \in \mathcal{Y}\}$ as the unknowns:

$$\begin{aligned}\mathbb{P}(D = d, R^Y = 0 \mid Z = z) &= \sum_{y \in \mathcal{Y}} \mathbb{P}(D = d, Y = y, R^Y = 0 \mid Z = z) \\ &= \sum_{y \in \mathcal{Y}} \mathbb{P}(D = d, Y = y, R^Y = 1 \mid Z = z)\eta_d(y)\end{aligned}$$

for $z = 0, 1$. The solutions $\eta_d(y)$ are unique if Y is binary and $Y \not\perp\!\!\!\perp Z \mid D = d$ for $d = 0, 1$. Under one-sided noncompliance, however, $D = 1$ occurs only when $Z = 1$, so for $d = 1$ we are left with the single equation

$$\mathbb{P}(D = 1, R^Y = 0 \mid Z = 1) = \sum_{y \in \mathcal{Y}} \mathbb{P}(D = 1, Y = y, R^Y = 1 \mid Z = 1)\eta_1(y).$$

With one equation and two unknowns $\eta_1(0)$ and $\eta_1(1)$, $\eta_1(y)$ is not identifiable without further assumptions. In essence, under one-sided noncompliance there is no variation in Z within the stratum $D = 1$, and the system needs that variation to pin down $\eta_1(y)$. The same reasoning applies to Theorem 1 $(1ZY)$. Equivalently, the conditional dependence conditions themselves fail. Under one-sided noncompliance, $D = 1$ forces $Z = 1$, so $Y \perp\!\!\!\perp Z \mid D = 1$ and the condition in part $(1DY)$ fails. Likewise, $Z = 0$ forces $D = 0$, so $Y \perp\!\!\!\perp D \mid Z = 0$ and the condition in part $(1ZY)$ fails. Section C.1 of the appendix gives explicit one-sided counterexamples for both Assumptions $1DY$ and $1ZY$.

If Assumptions $1DY$ and $1ZY$ are strengthened to Assumption $1Y$ of Chen et al. (2009a), namely $R^Y \perp\!\!\!\perp (Z, U, D) \mid Y$, the CACE becomes identifiable for a binary Y under both one-sided and two-sided noncompliance. Identification is then possible for a discrete Y with up to three categories under one-sided noncompliance and up to four under two-sided noncompliance; see Section B.1.6 of the appendix.

Theorem 1 $(1UY)$ assumes a binary Y. We now explain where this condition is used. Under Assumption $1UY$,

$$\mathbb{P}(U = n, D = 0, Y = y, R^Y = 1 \mid Z = 0) = \mathbb{P}(U = n, D = 0, Y = y, R^Y = 1 \mid Z = 1),$$

which lets us recover the complier quantities at $z = d = 0$ from the never-taker quantities. Specifically,

$$\mathbb{P}(U = c, D = 0, Y = y, R^Y = 1 \mid Z = 0)$$

$$= \mathbb{P}(D=0, Y=y, R^Y=1 \mid Z=0) - \mathbb{P}(U=n, D=0, Y=y, R^Y=1 \mid Z=0).$$

Similarly, we identify $\mathbb{P}(U=c, D=1, Y=y, R^Y=1 \mid Z=1)$ from the always-taker quantities. We can therefore identify the ratio

$$\frac{\mathbb{P}(Y=y \mid U=c, D=0)}{\mathbb{P}(Y=y \mid U=c, D=1)} = \frac{\mathbb{P}(U=c, D=0, Y=y, R^Y=1 \mid Z=0)}{\mathbb{P}(U=c, D=1, Y=y, R^Y=1 \mid Z=1)}$$

for all y. In the special case where this ratio equals 1 for all y, the CACE equals 0 and is identified without recovering the individual probabilities $\mathbb{P}(Y=y \mid U=c, D=d)$. In general, however, identifying the CACE requires $\mathbb{P}(Y=y \mid U=c, D=d)$ for all y and d, which we can recover only when Y is binary.

Summary. We searched exhaustively over all missingness mechanisms in the outcome. Theorem 1 gives the most general ones that identify the CACE, collected in Table 1, and the counterexamples in Section C.1 of the appendix show that the rest are not identifiable. Therefore, the classification in Theorem 1 is complete. Beyond the CACE, the proof of Theorem 1 also settles the identifiability of the joint distribution $\mathbb{P}(Z, U, D, Y)$. It is identifiable under $(1ZD)$, $(1DY)$, and $(1ZY)$, identifiable under $(1UD)$ with additional positivity conditions, and not generally identifiable under $(1UY)$.

5. Missingness Only in the Treatment

This section considers missingness only in the treatment. Let R^D be the response indicator for D, with $R^D=1$ if D is observed and $R^D=0$ otherwise. The most general mechanism lets R^D depend on all of Z, U, D, and Y, and it does not permit nonparametric identification of the CACE (see Section C.2 of the appendix for counterexamples). Figure 2 displays this most general mechanism along with MCAR, MAR, and five MNAR mechanisms that permit nonparametric identification of the CACE. We label each mechanism by the prefix 2 followed by the variables that R^D may depend on. For example, Assumption $2ZY$ is the mechanism in which R^D depends on (Z, Y) and is conditionally independent of (U, D).

When data are MCAR, i.e., $R^D \perp\!\!\!\perp (Z, U, D, Y)$ and $\mathbb{P}(Z, D, Y) = \mathbb{P}(Z, D, Y \mid R^D = 1)$, a complete-case analysis gives a consistent estimate of the CACE. Besides MCAR, the CACE is nonparametrically identifiable under the MAR mechanism and under each of the five MNAR mechanisms in Figure 2 (d)–(h). Assumptions $2ZY$ and $2UD$ require only positivity. Beyond positivity, Assumption $2UY$ requires a binary outcome, Assumptions $2DY$ and $2ZD$ require a conditional dependence condition (for $2DY$, only under two-sided noncompliance), and Assumption $2ZU$ requires one-sided noncompliance. Table 2 summarizes these conditions. These five MNAR mechanisms are the most general that permit nonparametric identification: each lets R^D depend on two variables, and allowing R^D to depend on any further variable yields a mechanism under which the CACE is no longer identifiable. We provide counterexamples for the nonidentifiable mechanisms in Section C.2 of the appendix, namely the three-variable mechanisms. We now state the MAR and five MNAR assumptions.

Assumption $2ZY$ $R^D \perp\!\!\!\perp (U, D) \mid (Z, Y)$.

Assumption $2ZY$ is the MAR mechanism, since the probability of observing D depends only on the fully observed IV Z and outcome Y. In retrospective studies, R^D may depend

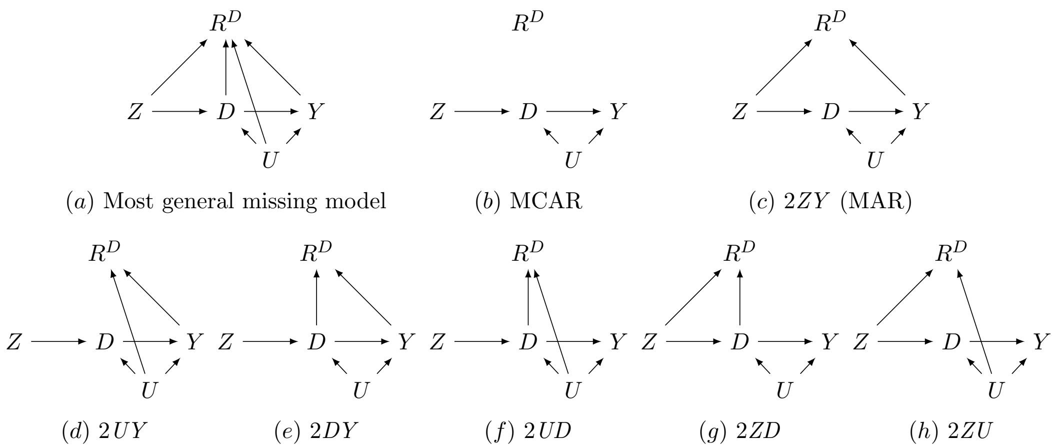


Figure 2: The DAGs in (a)–(h) describe the most general missingness mechanism, MCAR, MAR, and five MNAR mechanisms, respectively, when missingness exists only in the treatment.

Table 2: Summary of the mechanisms for missingness in the treatment, studied in Section 5. Each row identifies the CACE under the listed conditions. A checkmark indicates that a positivity condition is required; the exact positivity condition is stated in Theorem 2. The symbol $*$ means no restriction on that aspect.

Mechanism	Missingness assumption	Positivity	Outcome	Noncompliance	Dependence
$2ZY$	$R^D \perp\!\!\!\perp (U, D) \mid (Z, Y)$	✓	$*$	$*$	$*$
$2UY$	$R^D \perp\!\!\!\perp (Z, D) \mid (U, Y)$	✓	Binary	$*$	$*$
$2DY$	$R^D \perp\!\!\!\perp (Z, U) \mid (D, Y)$	✓	$*$	$*$	One-sided: $*$ Two-sided: $D \not\perp\!\!\!\perp Z \mid Y = y$ for all y
$2UD$	$R^D \perp\!\!\!\perp (Z, Y) \mid (U, D)$	✓	$*$	$*$	$*$
$2ZD$	$R^D \perp\!\!\!\perp (U, Y) \mid (Z, D)$	✓	$*$	$*$	One-sided: $Y \not\perp\!\!\!\perp D \mid Z = 1$ Two-sided: $Y \not\perp\!\!\!\perp D \mid Z = z$ for $z = 0, 1$
$2ZU$	$R^D \perp\!\!\!\perp (D, Y) \mid (Z, U)$	✓	$*$	One-sided	$*$

on Y when D is collected after Y is measured. For example, researchers may ask patients about their smoking history to study how smoking affects a health outcome of interest. If patients in poorer health are less likely to participate in a survey than those in good health, then R^D depends on Y.

Assumption $2UY$ $R^D \perp\!\!\!\perp (Z, D) \mid (U, Y)$.

Assumption $2UY$ allows the treatment response indicator to depend on the latent compliance status U and the fully observed outcome Y.

Assumption $2DY$ $R^D \perp\!\!\!\perp (Z, U) \mid (D, Y)$.

Assumption $2DY$ allows the treatment response indicator to depend on the missing treatment value D and the fully observed outcome Y.

Assumption $2UD$ $R^D \perp\!\!\!\perp (Z, Y) \mid (U, D)$.

Assumption $2UD$ allows the treatment response indicator to depend on the latent compliance status U and the missing treatment value D.

Assumption $2ZD$ $R^D \perp\!\!\!\perp (U, Y) \mid (Z, D)$.

Assumption $2ZD$ allows the treatment response indicator to depend on the fully observed IV Z and the missing treatment value D. It is analogous to Assumption 1 of Zuo et al. (2025) in the mediation analysis setting, where the mediator is MNAR and its missingness may depend on the fully observed treatment and on the incompletely observed mediator value. Under that mapping, our Z is their treatment, our D their mediator, and our R^D their mediator response indicator.

Assumption $2ZU$ $R^D \perp\!\!\!\perp (D, Y) \mid (Z, U)$.

Assumption $2ZU$ allows the treatment response indicator to depend on the fully observed IV Z and the latent compliance status U. Because Z and U jointly determine D, this mechanism is more general than Assumptions $2UD$ and $2ZD$.

The following theorem formalizes the identification results in Table 2.

Theorem 2 *Suppose the IV assumptions hold. The CACE is nonparametrically identifiable in each of the following cases:*
($2ZY$) *under Assumption* $2ZY$*, if* $\mathbb{P}(R^D = 1 \mid Z = z, Y = y) > 0$ *for all* z *and* y*;*
($2UY$) *under Assumption* $2UY$*, for a binary* Y*, if* $\mathbb{P}(R^D = 1 \mid U = c, Y = y) > 0$ *for* $y = 0, 1$*;*
($2DY$) *under Assumption* $2DY$*, if* $\mathbb{P}(R^D = 1 \mid D = d, Y = y) > 0$ *for all* d *and* y*, and either (i) one-sided noncompliance, or (ii) two-sided noncompliance with* $D \not\perp\!\!\!\perp Z \mid Y = y$ *for all* y*;*
($2UD$) *under Assumption* $2UD$*, if* $\mathbb{P}(R^D = 1 \mid U = c, D = d) > 0$ *for* $d = 0, 1$*;*
($2ZD$) *under Assumption* $2ZD$*, if* $\mathbb{P}(R^D = 1 \mid Z = z, D = d) > 0$ *for all* z *and* d*, and either (i) one-sided noncompliance with* $Y \not\perp\!\!\!\perp D \mid Z = 1$*, or (ii) two-sided noncompliance with* $Y \not\perp\!\!\!\perp D \mid Z = z$ *for* $z = 0, 1$*;*
($2ZU$) *under Assumption* $2ZU$*, with one-sided noncompliance, if* $\mathbb{P}(R^D = 1 \mid Z = 1, U = u) > 0$ *for* $u = n, c$*.*

In Theorem 2, part ($2ZY$) is the MAR case, and its identification is standard. Parts ($2UY$) through ($2ZU$) are new nonparametric identification results under treatment MNAR. We now give the intuition behind the conditions these MNAR mechanisms require.

In Theorem 2 ($2UY$), the binary-Y condition parallels that in Theorem 1 ($1UY$). Under Assumption $2UY$, we can identify the ratio

$$\frac{\mathbb{P}(Y = y \mid U = c, D = 0)}{\mathbb{P}(Y = y \mid U = c, D = 1)}$$

for all y. When this ratio equals 1 for all y, the CACE equals 0. Otherwise, identifying the CACE requires the probabilities $\mathbb{P}(Y = y \mid U = c, D = d)$, which we can recover when Y is binary.

In Theorem 2 ($2DY$), we require $D \not\perp\!\!\!\perp Z \mid Y = y$ for all y under two-sided noncompliance but not under one-sided noncompliance. To see this, note that under Assumption $2DY$,

$$\mathbb{P}(D = d, Y = y \mid Z = z) = \frac{\mathbb{P}(D = d, Y = y, R^D = 1 \mid Z = z)}{\mathbb{P}(R^D = 1 \mid D = d, Y = y)}.$$

Therefore, the conditional distribution $\mathbb{P}(D, Y \mid Z)$ is identifiable, and so is the CACE if $\mathbb{P}(R^D \mid D, Y)$ is identifiable. Define

$$\zeta_y(d) = \frac{\mathbb{P}(R^D = 0 \mid D = d, Y = y)}{\mathbb{P}(R^D = 1 \mid D = d, Y = y)}$$

for all d and y. For each $y \in \mathcal{Y}$, we have the following system of linear equations with $\{\zeta_y(d) : d = 0, 1\}$ as the unknowns:

$$\begin{aligned}\mathbb{P}(Y = y, R^D = 0 \mid Z = z) &= \sum_{d=0}^{1} \mathbb{P}(D = d, Y = y, R^D = 0 \mid Z = z) \\ &= \sum_{d=0}^{1} \mathbb{P}(D = d, Y = y, R^D = 1 \mid Z = z)\zeta_y(d)\end{aligned}$$

for $z = 0, 1$. Under two-sided noncompliance, the solutions $\zeta_y(d)$ are unique if $D \not\perp\!\!\!\perp Z \mid Y = y$ for all y. This condition is not needed under one-sided noncompliance: when $D(0) = 0$, we identify $\zeta_y(0)$ as

$$\zeta_y(0) = \frac{\mathbb{P}(Y = y, R^D = 0 \mid Z = 0)}{\mathbb{P}(D = 0, Y = y, R^D = 1 \mid Z = 0)},$$

and then $\zeta_y(1)$ as

$$\zeta_y(1) = \frac{\mathbb{P}(Y = y, R^D = 0 \mid Z = 1) - \mathbb{P}(D = 0, Y = y, R^D = 1 \mid Z = 1)\zeta_y(0)}{\mathbb{P}(D = 1, Y = y, R^D = 1 \mid Z = 1)}.$$

Theorem 2 ($2UD$) requires no condition beyond positivity, because under Assumption $2UD$,

$$\mathbb{P}(U = u, D = d, Y = y, R^D = 1 \mid Z = 0) = \mathbb{P}(U = u, D = d, Y = y, R^D = 1 \mid Z = 1)$$

for $(u, d) = (n, 0), (a, 1)$, so we can recover the complier quantities at $z = 0$ and $z = 1$ from the never-taker quantities at $z = 1$ and the always-taker quantities at $z = 0$, respectively. Specifically,

$$\mathbb{P}(U = c, D = 0, Y = y, R^D = 1 \mid Z = 0)$$

$$= \mathbb{P}(D = 0, Y = y, R^D = 1 \mid Z = 0) - \mathbb{P}(U = n, D = 0, Y = y, R^D = 1 \mid Z = 1).$$

Therefore, we can identify $\mathbb{P}(Y = y \mid U = c, D = 0)$ by

$$\mathbb{P}(Y = y \mid U = c, D = 0) = \frac{\mathbb{P}(U = c, D = 0, Y = y, R^D = 1 \mid Z = 0)}{\mathbb{P}(U = c, D = 0, R^D = 1 \mid Z = 0)}.$$

Similarly, we can identify $\mathbb{P}(Y = y \mid U = c, D = 1)$ and therefore the CACE.

In Theorem 2 ($2ZD$), we require $Y \not\perp\!\!\!\perp D \mid Z = z$ for $z = 0, 1$ under two-sided noncompliance, which reduces to $Y \not\perp\!\!\!\perp D \mid Z = 1$ under one-sided noncompliance. To see this, note that under Assumption $2ZD$,

$$\mathbb{P}(D = d, Y = y \mid Z = z) = \frac{\mathbb{P}(D = d, Y = y, R^D = 1 \mid Z = z)}{\mathbb{P}(R^D = 1 \mid Z = z, D = d)}.$$

Therefore, the conditional distribution $\mathbb{P}(D, Y \mid Z)$ is identifiable, and so is the CACE if $\mathbb{P}(R^D \mid Z, D)$ is identifiable. Define

$$\zeta_z(d) = \frac{\mathbb{P}(R^D = 0 \mid Z = z, D = d)}{\mathbb{P}(R^D = 1 \mid Z = z, D = d)}$$

for all z and d. We then have the following system of linear equations with $\{\zeta_z(d) : d = 0, 1\}$ as the unknowns:

$$\mathbb{P}(Y = y, R^D = 0 \mid Z = z) = \sum_{d=0}^{1} \mathbb{P}(D = d, Y = y, R^D = 1 \mid Z = z)\zeta_z(d)$$

for each $y \in \mathcal{Y}$. Under two-sided noncompliance, the solutions $\{\zeta_z(d) : z = 0, 1, d = 0, 1\}$ are unique if $Y \not\perp\!\!\!\perp D \mid Z = z$ for $z = 0, 1$. Under one-sided noncompliance, the solutions $\{\zeta_1(d) : d = 0, 1\}$ are unique if $Y \not\perp\!\!\!\perp D \mid Z = 1$. For $z = 0$, we identify $\zeta_0(0)$ as

$$\zeta_0(0) = \frac{\mathbb{P}(Y = y, R^D = 0 \mid Z = 0)}{\mathbb{P}(D = 0, Y = y, R^D = 1 \mid Z = 0)}$$

for any y.

In Theorem 2 ($2ZU$), identification under one-sided noncompliance relies on two facts: Y is conditionally independent of R^D given Z and U, and a unit's compliance status in the $Z = 1$ group is known once its D is observed. We first identify $\mathbb{P}(Y = y \mid U = n, D = 0)$ and $\mathbb{P}(Y = y \mid U = c, D = 1)$:

$$\mathbb{P}(Y = y \mid U = n, D = 0) = \mathbb{P}(Y = y \mid Z = 1, U = n, D = 0, R^D = 1),$$
$$\mathbb{P}(Y = y \mid U = c, D = 1) = \mathbb{P}(Y = y \mid Z = 1, U = c, D = 1, R^D = 1).$$

Since

$$\begin{aligned}\mathbb{P}(Y = y \mid Z = 1) = {} & \mathbb{P}(Y = y \mid U = n, D = 0)\mathbb{P}(U = n, D = 0 \mid Z = 1) \\ & + \mathbb{P}(Y = y \mid U = c, D = 1)\{1 - \mathbb{P}(U = n, D = 0 \mid Z = 1)\},\end{aligned}$$

we can identify $\mathbb{P}(U = n, D = 0 \mid Z = 1)$. Since

$$\mathbb{P}(Y = y \mid Z = 0) = \mathbb{P}(Y = y \mid U = n, D = 0)\mathbb{P}(U = n, D = 0 \mid Z = 1) + \mathbb{P}(Y = y \mid U = c, D = 0)\{1 - \mathbb{P}(U = n, D = 0 \mid Z = 1)\},$$

we can identify $\mathbb{P}(Y = y \mid U = c, D = 0)$, and hence the CACE. Under two-sided noncompliance, the CACE is not identifiable without further assumptions. Section C.2 of the appendix gives a counterexample.

Summary. We searched exhaustively over all missingness mechanisms in the treatment. Theorem 2 gives the most general ones that identify the CACE, collected in Table 2, and the counterexamples in Section C.2 of the appendix show that the rest are not identifiable. Therefore, the classification in Theorem 2 is complete. Beyond the CACE, the proof of Theorem 2 also settles the identifiability of the joint distribution $\mathbb{P}(Z, U, D, Y)$. It is identifiable under $(2ZY)$, $(2DY)$, $(2ZD)$, and $(2ZU)$. Under $(2UD)$ and $(2UY)$, it is identifiable if a linear system pins down the odds of nonresponse in D uniquely and additional positivity conditions hold, with $(2UY)$ also requiring one-sided noncompliance. Sections B.2.4 and B.2.2 of the appendix give the exact conditions.

6. Missingness in Both the Treatment and Outcome

This section considers missingness in both the treatment and outcome. We focus on prospective studies in which D is collected before Y, so the future variables Y and R^Y do not affect R^D. Each combined mechanism pairs one missingness mechanism in the outcome from Section 4 with one of Assumptions $2UD$, $2ZD$, and $2ZU$ from Section 5. These three are the missingness mechanisms in the treatment compatible with the prospective design. The key distinction is whether the direct path $R^D \to R^Y$ may be included. It may be included under $2UD$, because R^D remains conditionally independent of Z given (U, D). Under $2ZD$ or $2ZU$, this path breaks identification; see Section C.3 of the appendix for counterexamples. For the mechanisms that exclude the path, Section A of the appendix shows that removing one of the other direct paths into R^D or R^Y restores identification.

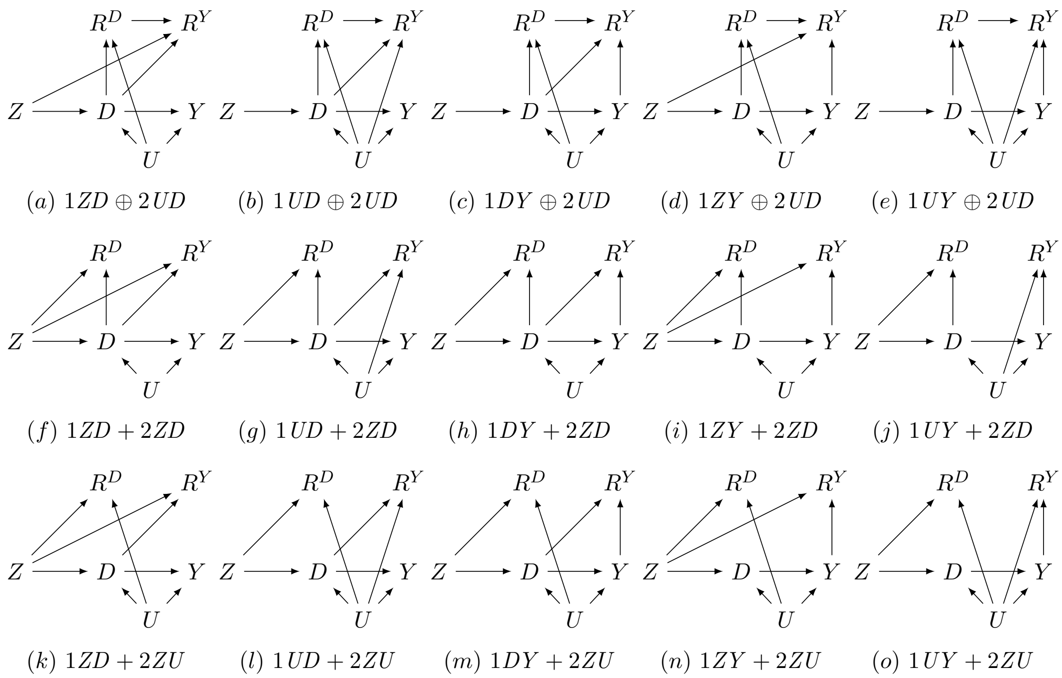
We label the combined mechanisms with both prefixes: the 1 label lists the variables that R^Y may depend on, other than R^D, and the 2 label lists the variables that R^D may depend on. We write $\oplus$ when the direct path $R^D \to R^Y$ is allowed and $+$ when it is not. We write $1\mathcal{M}$ for any of the five missingness mechanisms in the outcome $1ZD$, $1UD$, $1DY$, $1ZY$, and $1UY$. For example, Assumption $1ZD \oplus 2UD$ allows R^Y to depend on (Z, D, R^D) and R^D on (U, D), whereas Assumption $1ZD + 2ZD$ excludes the path $R^D \to R^Y$ and allows both response indicators to depend on (Z, D).

Figure 3 shows the DAGs for the combined mechanisms, and Table 3 summarizes the three classes studied in this section. The subsections below give the formal assumptions and theorems, grouped by the mechanism for missingness in the treatment.

6.1 The Missing Outcome Models Combined with Assumption $2UD$

The first class pairs $1\mathcal{M}$ with Assumption $2UD$. It is the only class that allows the direct path $R^D \to R^Y$ without removing any other direct path.

Assumption $1ZD \oplus 2UD$ *$R^D \perp\!\!\!\perp (Z, Y) \mid (U, D)$ and $R^Y \perp\!\!\!\perp (U, Y) \mid (Z, D, R^D)$.*



(a) $1ZD \oplus 2UD$ (b) $1UD \oplus 2UD$ (c) $1DY \oplus 2UD$ (d) $1ZY \oplus 2UD$ (e) $1UY \oplus 2UD$

(f) $1ZD + 2ZD$ (g) $1UD + 2ZD$ (h) $1DY + 2ZD$ (i) $1ZY + 2ZD$ (j) $1UY + 2ZD$

(k) $1ZD + 2ZU$ (l) $1UD + 2ZU$ (m) $1DY + 2ZU$ (n) $1ZY + 2ZU$ (o) $1UY + 2ZU$

Figure 3: The DAGs illustrate the mechanisms obtained by combining Assumptions $1ZD$, $1UD$, $1DY$, $1ZY$, and $1UY$ with Assumptions $2UD$, $2ZD$, and $2ZU$. Panels (m) and (n), corresponding to $1DY + 2ZU$ and $1ZY + 2ZU$, are not included in the identification results because the $1DY$ and $1ZY$ mechanisms require two-sided noncompliance, whereas $2ZU$ requires one-sided.

Table 3: Summary of the combined mechanisms studied in Section 6. Here $1\mathcal{M}$ denotes any of $1ZD$, $1UD$, $1DY$, $1ZY$, and $1UY$. A checkmark indicates that the direct path $R^D \to R^Y$ is allowed, and a $\times$ indicates that it is excluded. Theorems 3–5 state the exact conditions.

Mechanism	Missingness assumption	$R^D \to R^Y$	Conditions
$1\mathcal{M} \oplus 2UD$	$R^D \perp\!\!\!\perp (Z,Y) \mid (U,D)$; R^Y follows $1\mathcal{M}$ conditional on R^D	✓	$1\mathcal{M}$ condition + $2UD$ condition.
$1\mathcal{M} + 2ZD$	$R^D \perp\!\!\!\perp (U,Y) \mid (Z,D)$; R^Y follows $1\mathcal{M}$	$\times$	$1\mathcal{M}$ condition + $2ZD$ condition.
$1\mathcal{M} + 2ZU$	$R^D \perp\!\!\!\perp (D,Y) \mid (Z,U)$; R^Y follows $1\mathcal{M}$	$\times$	$1\mathcal{M}$ condition + $2ZU$ condition. For $1DY$ and $1ZY$, the outcome and treatment mechanisms impose conflicting noncompliance conditions, so we obtain no identification result.

Assumption $1UD \oplus 2UD$ *$R^D \perp\!\!\!\perp (Z,Y) \mid (U,D)$ and $R^Y \perp\!\!\!\perp (Z,Y) \mid (U,D,R^D)$.*

Assumption $1DY \oplus 2UD$ *$R^D \perp\!\!\!\perp (Z, Y) \mid (U, D)$ and $R^Y \perp\!\!\!\perp (Z, U) \mid (D, Y, R^D)$.*

Assumption $1ZY \oplus 2UD$ *$R^D \perp\!\!\!\perp (Z, Y) \mid (U, D)$ and $R^Y \perp\!\!\!\perp (U, D) \mid (Z, Y, R^D)$.*

Assumption $1UY \oplus 2UD$ *$R^D \perp\!\!\!\perp (Z, Y) \mid (U, D)$ and $R^Y \perp\!\!\!\perp (Z, D) \mid (U, Y, R^D)$.*

The following theorem formalizes the identification results for this first class of mechanisms.

Theorem 3 *Suppose the IV assumptions hold. The CACE is nonparametrically identifiable in each of the following cases:*
($1ZD \oplus 2UD$) under Assumption $1ZD \oplus 2UD$, if $\mathbb{P}(R^D = 1 \mid U = c, D = d) > 0$ for $d = 0, 1$, and $\mathbb{P}(R^Y = 1 \mid Z = z, D = d, R^D = 1) > 0$ for all z and d;
($1UD \oplus 2UD$) under Assumption $1UD \oplus 2UD$, if $\mathbb{P}(R^D = 1 \mid U = c, D = d) > 0$ and $\mathbb{P}(R^Y = 1 \mid U = c, D = d, R^D = 1) > 0$ for $d = 0, 1$;
($1DY \oplus 2UD$) under Assumption $1DY \oplus 2UD$, for a binary Y with two-sided noncompliance, if $\mathbb{P}(R^D = 1 \mid U = c, D = d) > 0$ for $d = 0, 1$, $\mathbb{P}(R^Y = 1 \mid D = d, Y = y, R^D = 1) > 0$ for all d and y, and $Y \not\perp\!\!\!\perp Z \mid (D = d, R^D = 1)$ for $d = 0, 1$;
($1ZY \oplus 2UD$) under Assumption $1ZY \oplus 2UD$, for a binary Y with two-sided noncompliance, if $\mathbb{P}(R^D = 1 \mid U = c, D = d) > 0$ for $d = 0, 1$, $\mathbb{P}(R^Y = 1 \mid Z = z, Y = y, R^D = 1) > 0$ for all z and y, and $Y \not\perp\!\!\!\perp D \mid (Z = z, R^D = 1)$ for $z = 0, 1$;
($1UY \oplus 2UD$) under Assumption $1UY \oplus 2UD$, for a binary Y, if $\mathbb{P}(R^D = 1 \mid U = c, D = d) > 0$ for $d = 0, 1$, and $\mathbb{P}(R^Y = 1 \mid U = c, Y = y, R^D = 1) > 0$ for $y = 0, 1$.

The conditions in Theorem 3 inherit the corresponding conditions for missingness in the outcome from Theorem 1, conditioning on $R^D = 1$ wherever R^Y depends on R^D, and add the $2UD$ positivity condition from Theorem 2. Because $R^D \perp\!\!\!\perp Z \mid (U, D)$, the direct path $R^D \to R^Y$ does not disrupt the cross-Z comparisons used to isolate compliers. The case-specific recovery formulas mirror those in Section 4; see the proof for details.

6.2 The Missing Outcome Models Combined with Assumption $2ZD$

The second class pairs $1\mathcal{M}$ with Assumption $2ZD$. Here, the identification results do not allow the direct path $R^D \to R^Y$. Section C.3 of the appendix gives counterexamples.

Assumption $1ZD + 2ZD$ *$R^D \perp\!\!\!\perp (U, Y) \mid (Z, D)$ and $R^Y \perp\!\!\!\perp (U, Y, R^D) \mid (Z, D)$.*

Assumption $1UD + 2ZD$ *$R^D \perp\!\!\!\perp (U, Y) \mid (Z, D)$ and $R^Y \perp\!\!\!\perp (Z, Y, R^D) \mid (U, D)$.*

Assumption $1DY + 2ZD$ *$R^D \perp\!\!\!\perp (U, Y) \mid (Z, D)$ and $R^Y \perp\!\!\!\perp (Z, U, R^D) \mid (D, Y)$.*

Assumption $1ZY + 2ZD$ *$R^D \perp\!\!\!\perp (U, Y) \mid (Z, D)$ and $R^Y \perp\!\!\!\perp (U, D, R^D) \mid (Z, Y)$.*

Assumption $1UY + 2ZD$ *$R^D \perp\!\!\!\perp (U, Y) \mid (Z, D)$ and $R^Y \perp\!\!\!\perp (Z, D, R^D) \mid (U, Y)$.*

Define the random vector $Y^\dagger = (Y \cdot R^Y, R^Y)$, so that $\mathbb{P}\{Y^\dagger = (y, 1)\} = \mathbb{P}(Y = y, R^Y = 1)$ for all $y \in \mathcal{Y}$ and $\mathbb{P}\{Y^\dagger = (0, 0)\} = \mathbb{P}(R^Y = 0)$. In words, $Y^\dagger$ records the outcome value when Y is observed and records only the missingness otherwise. The following theorem formalizes the identification results for this second class of mechanisms.

Theorem 4 *Suppose the IV assumptions hold. The CACE is nonparametrically identifiable in each of the following cases:*
($1ZD + 2ZD$) *under Assumption* $1ZD + 2ZD$, *if* $\mathbb{P}(R^D = 1 \mid Z = z, D = d) > 0$ *and* $\mathbb{P}(R^Y = 1 \mid Z = z, D = d) > 0$ *for all* z *and* d, *and either (i) one-sided noncompliance with* $Y^\dagger \not\perp\!\!\!\perp D \mid Z = 1$, *or (ii) two-sided noncompliance with* $Y^\dagger \not\perp\!\!\!\perp D \mid Z = z$ *for* $z = 0, 1$;
($1UD+2ZD$) *under Assumption* $1UD+2ZD$, *if* $\mathbb{P}(R^D = 1 \mid Z = z, D = d) > 0$ *for all* z *and* d *and* $\mathbb{P}(R^Y = 1 \mid U = c, D = d) > 0$ *for* $d = 0, 1$, *and either (i) one-sided noncompliance with* $Y^\dagger \not\perp\!\!\!\perp D \mid Z = 1$, *or (ii) two-sided noncompliance with* $Y^\dagger \not\perp\!\!\!\perp D \mid Z = z$ *for* $z = 0, 1$;
($1DY+2ZD$) *under Assumption* $1DY+2ZD$, *for a binary* Y *with two-sided noncompliance, if* $\mathbb{P}(R^D = 1 \mid Z = z, D = d) > 0$ *for all* z *and* d, $\mathbb{P}(R^Y = 1 \mid D = d, Y = y) > 0$ *for all* d *and* y, $Y^\dagger \not\perp\!\!\!\perp D \mid Z = z$ *for* $z = 0, 1$, *and* $Y \not\perp\!\!\!\perp Z \mid D = d$ *for* $d = 0, 1$;
($1ZY+2ZD$) *under Assumption* $1ZY+2ZD$, *for a binary* Y *with two-sided noncompliance, if* $\mathbb{P}(R^D = 1 \mid Z = z, D = d) > 0$ *for all* z *and* d, $\mathbb{P}(R^Y = 1 \mid Z = z, Y = y) > 0$ *for all* z *and* y, *and* $Y \not\perp\!\!\!\perp D \mid Z = z$ *for* $z = 0, 1$;
($1UY + 2ZD$) *under Assumption* $1UY + 2ZD$, *for a binary* Y, *if* $\mathbb{P}(R^D = 1 \mid Z = z, D = d) > 0$ *for all* z *and* d *and* $\mathbb{P}(R^Y = 1 \mid U = c, Y = y) > 0$ *for* $y = 0, 1$, *and either (i) one-sided noncompliance with* $Y^\dagger \not\perp\!\!\!\perp D \mid Z = 1$, *or (ii) two-sided noncompliance with* $Y^\dagger \not\perp\!\!\!\perp D \mid Z = z$ *for* $z = 0, 1$.

The conditions in Theorem 4 combine the corresponding conditions for missingness in the outcome from Theorem 1 with the $2ZD$ condition from Theorem 2. Where stated, the dependence condition from Theorem 2 changes from $Y \not\perp\!\!\!\perp D \mid Z = z$ to $Y^\dagger \not\perp\!\!\!\perp D \mid Z = z$, so that dependence between D and either the observed outcome value or the outcome response indicator can help recover $\mathbb{P}(R^D = 1 \mid Z = z, D = d)$. Adding the direct path $R^D \to R^Y$ makes this recovery non-unique; see Section C.3 of the appendix for counterexamples.

6.3 The Missing Outcome Models Combined with Assumption $2ZU$

The third class pairs $1\mathcal{M}$ with Assumption $2ZU$. Only Assumptions $1ZD + 2ZU$, $1UD + 2ZU$, and $1UY + 2ZU$ yield identification. Assumptions $1DY$ and $1ZY$ are excluded because they require two-sided noncompliance, whereas $2ZU$ identifies the CACE only under one-sided noncompliance. As in Section 6.2, the direct path $R^D \to R^Y$ is not allowed.

Assumption $1ZD + 2ZU$ $R^D \perp\!\!\!\perp (D, Y) \mid (Z, U)$ *and* $R^Y \perp\!\!\!\perp (U, Y, R^D) \mid (Z, D)$.

Assumption $1UD + 2ZU$ $R^D \perp\!\!\!\perp (D, Y) \mid (Z, U)$ *and* $R^Y \perp\!\!\!\perp (Z, Y, R^D) \mid (U, D)$.

Assumption $1DY + 2ZU$ $R^D \perp\!\!\!\perp (D, Y) \mid (Z, U)$ *and* $R^Y \perp\!\!\!\perp (Z, U, R^D) \mid (D, Y)$.

Assumption $1ZY + 2ZU$ $R^D \perp\!\!\!\perp (D, Y) \mid (Z, U)$ *and* $R^Y \perp\!\!\!\perp (U, D, R^D) \mid (Z, Y)$.

Assumption $1UY + 2ZU$ $R^D \perp\!\!\!\perp (D, Y) \mid (Z, U)$ *and* $R^Y \perp\!\!\!\perp (Z, D, R^D) \mid (U, Y)$.

The following theorem formalizes the identification results for this third class of mechanisms.

Theorem 5 *Suppose the IV assumptions hold. The CACE is nonparametrically identifiable in each of the following cases:*
($1ZD + 2ZU$) *under Assumption* $1ZD + 2ZU$, *with one-sided noncompliance, if* $\mathbb{P}(R^D = 1 \mid$

$Z = 1, U = u) > 0$ for $u = n, c$, and $\mathbb{P}(R^Y = 1 \mid Z = z, D = d) > 0$ for all z and d;
$(1UD + 2ZU)$ under Assumption $1UD + 2ZU$, with one-sided noncompliance, if $\mathbb{P}(R^D = 1 \mid Z = 1, U = u) > 0$ for $u = n, c$, and $\mathbb{P}(R^Y = 1 \mid U = c, D = d) > 0$ for $d = 0, 1$;
$(1UY + 2ZU)$ under Assumption $1UY + 2ZU$, for a binary Y with one-sided noncompliance, if $\mathbb{P}(R^D = 1 \mid Z = 1, U = u) > 0$ for $u = n, c$, and $\mathbb{P}(R^Y = 1 \mid U = c, Y = y) > 0$ for $y = 0, 1$.

The conditions in Theorem 5 combine the $2ZU$ condition from Theorem 2 with the conditions for missingness in the outcome from Theorem 1 that do not require two-sided noncompliance. The CACE therefore remains identifiable under $1ZD$, $1UD$, and $1UY$ with one-sided noncompliance, but not under $1DY$ and $1ZY$. We refer readers to the proofs for details.

7. Numerical Study

We illustrate the identification results for missingness in the outcome alone. The study makes three points. First, an estimator that assumes the correct identifiable mechanism recovers the CACE. Second, a nonidentifiable mechanism leaves an interval of CACE values consistent with the observed data, so no estimator can recover the CACE. Third, an identifiable but incorrect mechanism can produce substantial bias. The first and third points concern estimation, and we study them with simulated samples. The second concerns identification, which is a property of the observed-data distribution and not of any one sample. We therefore study it with that distribution itself, computed from the data-generating process rather than estimated from a sample.

7.1 Setup

The complete data follow the IV model of Section 2 with two-sided noncompliance and a binary outcome:

$$Z \sim \text{Bernoulli}(0.5), \qquad \big(\mathbb{P}(U = a), \mathbb{P}(U = c), \mathbb{P}(U = n)\big) = (0.25, 0.50, 0.25),$$

with $D = 1$ for always-takers, $D = 0$ for never-takers, and $D = Z$ for compliers. Write $q_{ud} = \mathbb{P}(Y = 1 \mid U = u, D = d)$ for the outcome probabilities. We set $q_{c0} = 0.10$, $q_{c1} = 0.65$, $q_{a1} = 0.05$, and $q_{n0} = 0.65$, so the true CACE is $q_{c1} - q_{c0} = 0.55$. We then generate the outcome response indicator R^Y under each of nine mechanisms in turn. Under a given mechanism, the response probability follows a logistic model in the variables that mechanism allows, calibrated to an average response rate of about 65% and bounded away from zero and one. By Theorem 1, the CACE is identifiable under five of the nine, $1ZD$, $1UD$, $1DY$, $1ZY$, and $1UY$, and not under the remaining four, $1ZU$, $1ZDY$, $1UDY$, and $1ZUY$.

Each mechanism defines a model with three sets of parameters: the distribution of the compliance status U, the outcome probabilities q_{ud}, and one response probability for each combination of the variables the mechanism allows. The model imposes no restrictions beyond the conditional independences the mechanism encodes. For the first and third points, we estimate the CACE with the closed-form estimator that the proof of Theorem 1 supplies for each identifiable mechanism, drawing 1000 Monte Carlo replications of size $n = 5000$.

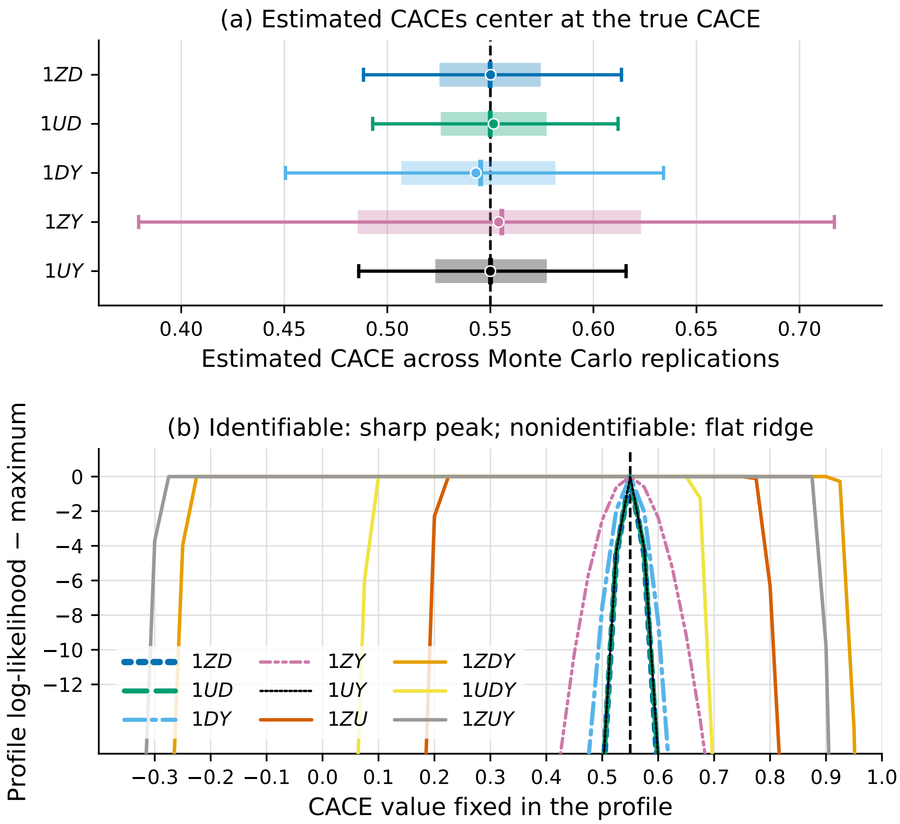


Figure 4: Identifiability determines whether the CACE can be recovered. The complete data are the same throughout, and the mechanisms differ only in how R^Y is generated, so the true CACE is 0.55 in both panels (dashed line). (a) Sampling distributions of the estimator that assumes the true mechanism, under each of the five identifiable mechanisms (1000 replications, sample size $n = 5000$; bars: interquartile range and 5th–95th percentiles, median and mean marked); all five center at the true CACE. (b) Profile log-likelihoods computed from the true observed-data distribution rather than from a sample, each shifted to peak at zero. A curve touches zero at every CACE value consistent with that distribution: the five identifiable mechanisms (dashed and dotted) touch zero only at the true CACE, and the four nonidentifiable ones (solid) stay at zero across an interval.

7.2 Identifiable Mechanisms Recover the CACE

Panel (a) of Figure 4 shows the sampling distribution of the estimator that assumes the true mechanism, under each of the five identifiable mechanisms. All five center at the true CACE, as Theorem 1 implies, but they differ in spread. Assumptions $1ZD$, $1UD$, and $1UY$ require no dependence condition and give the most precise estimators. Assumptions $1DY$ and $1ZY$ require a conditional dependence condition and give less precise ones: their estimators solve a linear system whose determinant measures that dependence, so the weaker the dependence, the more variable the estimate. Of the five, $1ZY$ is the most variable.

7.3 Which CACE Values the Observed Data Allow

The spread in panel (a) reflects sampling noise. We now set that aside and ask what the observed-data distribution determines on its own. For each mechanism, we compute the exact probabilities that the data-generating process assigns to the cells of (Z, D, R^Y, Y), with Y recorded only when $R^Y = 1$. We then collect every parameter configuration that reproduces these probabilities and record the CACE each one gives. The mechanism identifies the CACE if all of these values agree, and does not if any two disagree. The counterexamples in Section C.1 of the appendix construct disagreeing pairs algebraically.

We carry this out with a profile likelihood. A sample log-likelihood weights each cell by the number of observations that fall in it. Here we weight each cell by its exact probability, so no sampling enters. For a candidate τ_0, we impose $q_{c1} - q_{c0} = \tau_0$ and maximize this log-likelihood over the remaining parameters. It attains its largest possible value exactly when the parameters reproduce the observed-data distribution, so the constrained maximum reaches that value if and only if τ_0 is consistent with the distribution. Panel (b) of Figure 4 plots the constrained maximum against τ_0, shifted so that the curve peaks at zero. The CACE values consistent with the distribution are those where the curve touches zero.

Under the five identifiable mechanisms, the curve touches zero only at the true CACE 0.55, giving a single sharp peak. Under the four nonidentifiable mechanisms, it stays at zero across a broad interval, and no amount of data can distinguish the values in that interval.

Table 4: Maximizing the log-likelihood of Section 7.3 without the constraint on the CACE, from 100 random starting values. The last column gives the smallest and largest CACE among the solutions that reach the maximum: a single value, the true CACE 0.55, under the identifiable mechanisms, and a range of distinct values under the nonidentifiable ones.

Mechanism	Identifiable	Range of CACE estimates
$1ZD$	Yes	(0.55, 0.55)
$1UD$	Yes	(0.55, 0.55)
$1DY$	Yes	(0.55, 0.55)
$1ZY$	Yes	(0.55, 0.55)
$1UY$	Yes	(0.55, 0.55)
$1ZU$	No	(0.28, 0.71)
$1ZDY$	No	(0.23, 0.86)
$1UDY$	No	(0.30, 0.62)
$1ZUY$	No	(0.23, 0.75)

Table 4 makes the same point without the profile: maximizing that log-likelihood over all parameters, from 100 random starting values, returns the true CACE every time under the identifiable mechanisms, and a range of values that all attain the maximum under the nonidentifiable ones.

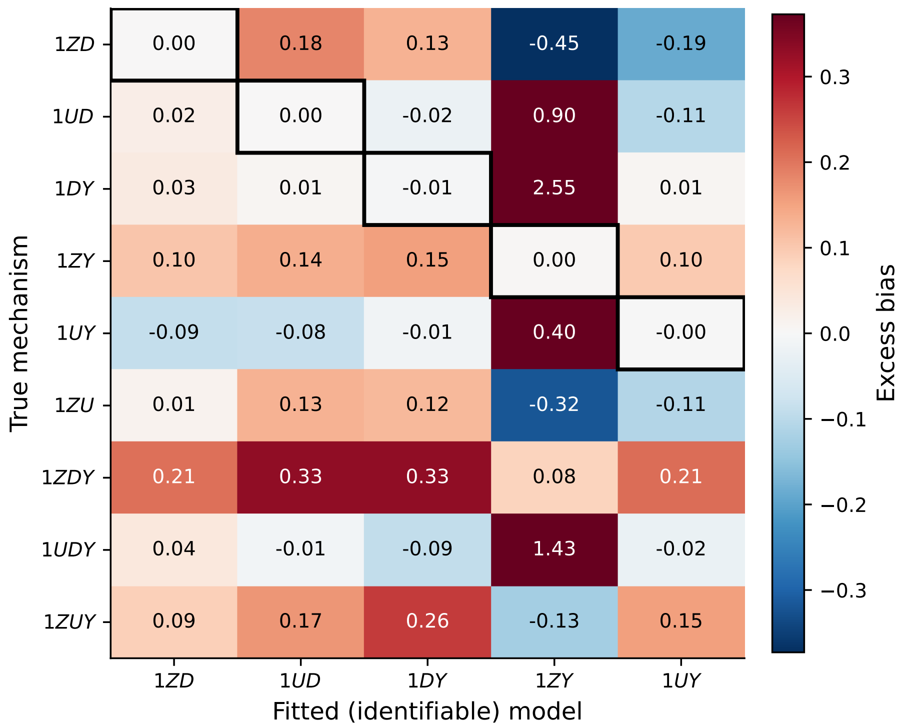


Figure 5: Excess bias, the average difference between the estimated CACE and the estimate the same sample would give with no missing data, over 1000 replications of size $n = 5000$, by true mechanism (rows) and fitted mechanism (columns). In the outlined cells, the fitted mechanism is the true one.

7.4 Fitting the Wrong Mechanism Biases the CACE

The observed data cannot test the mechanism, so the one an analyst assumes may differ from the truth. Figure 5 reports the cost of that mismatch, pairing each true mechanism (row) with each fitted identifiable mechanism (column). Each cell gives the excess bias, measured against the estimate the same sample would give with no missing data. In the outlined cells, the two agree, and the excess bias is close to zero. Elsewhere, the excess bias can exceed the true CACE of 0.55 several times over, reaching 2.55 when $1ZY$ is fitted to data generated under $1DY$. The last four rows are the nonidentifiable mechanisms, for which no fitted model is correct; a small entry in one of these rows reflects the particular design rather than recovery of the CACE. This is why we recommend the strategy of Section 3: estimate the CACE under every plausible identifiable mechanism and compare the results.

8. Empirical Illustration: The National Job Corps Study

We illustrate the theory using the National Job Corps Study (Schochet et al., 2001, 2006), a randomized evaluation of the Job Corps training program. Random assignment to Job Corps is the instrument Z, actual enrollment status is the treatment D, and weekly earnings in the fourth year after assignment are the outcome Y. We adjust for the baseline covariates X, namely gender, age, race, education, baseline earnings, child status, and arrest history, through parametric models, with details in the replication materials.[1] The study has one-sided noncompliance because those assigned to control had no access to the program, so $D = 0$ whenever $Z = 0$. The full sample has 8,707 subjects, of whom 5,084 were assigned to Job Corps. Among those assigned, 19.7% are known not to have enrolled, and enrollment is missing for a further 27.8%. The outcome is missing for 21.4% of the full sample.

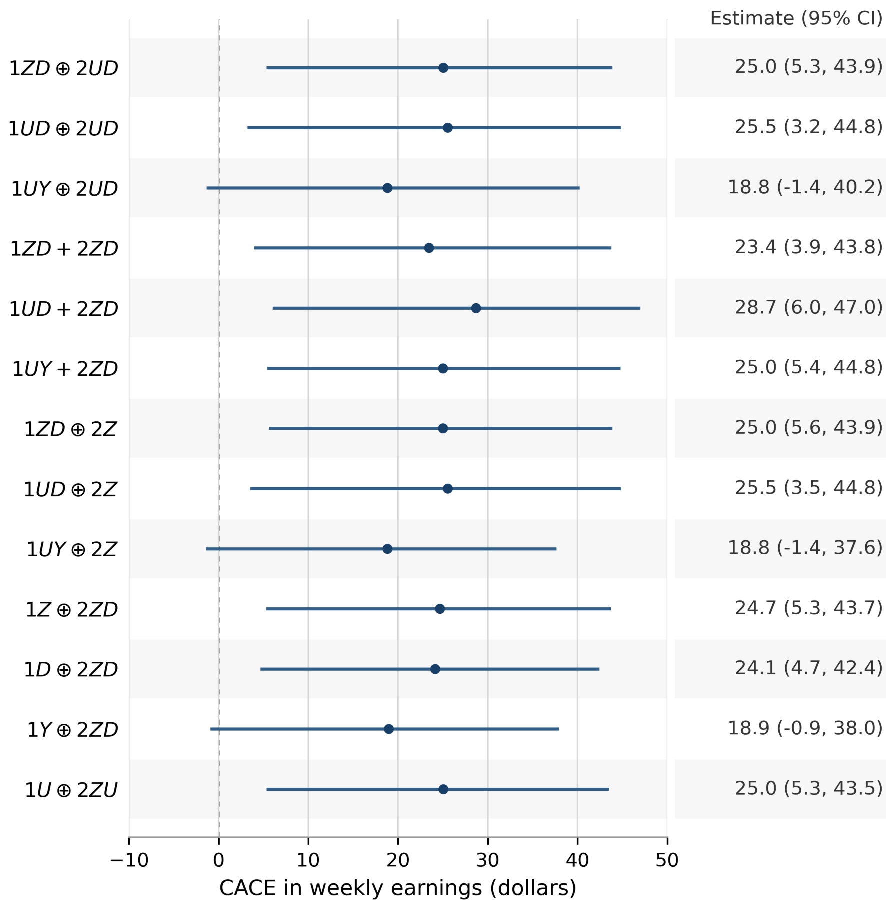


Figure 6: Estimated CACE of Job Corps enrollment on weekly earnings, under each of the 13 identified missingness mechanisms in Section 6 and Appendix A. Points are point estimates and bars are 95% percentile bootstrap confidence intervals.

1. Replication code and data for both the simulation study and this empirical application are available at `https://github.com/sushi133/IV_MNAR`.

We estimate the CACE under every mechanism in Section 6 and Appendix A that identifies the effect when both D and Y are missing. We exclude two groups. First, the mechanisms that require two-sided noncompliance, since the study is one-sided. Second, the $2ZU$ mechanisms that duplicate a $2ZD$ counterpart under one-sided noncompliance, since D determines U when $Z = 1$ and D is known by design when $Z = 0$. This leaves 13 distinct mechanisms. Several of them require a binary outcome. For these, we assume that any outcome-dependent missingness acts only through the indicator of positive earnings $H = \mathbf{1}(Y > 0)$, a natural dichotomization here since 18.9% of observed outcomes are zero. Figure 6 reports the resulting estimates and 95% percentile bootstrap confidence intervals based on 500 resamples.

The estimated effect is positive under all 13 mechanisms, ranging from \$19 to \$29 per week. Ten of the 13 confidence intervals exclude zero, and the remaining three extend slightly below it. The sign of the effect is therefore stable across the identified mechanisms, whereas its statistical significance depends on the mechanism.

9. Discussion

For missingness in the outcome alone and in the treatment alone, we searched exhaustively over the missingness mechanisms and characterized the most general ones that permit nonparametric identification of the CACE. The numerical study illustrates the practical value of this classification. We then extended the results to the case where both the treatment and outcome are missing, focusing on prospective studies in which the future variables Y and R^Y do not affect R^D.

Our central practical recommendation, developed in Section 3, is to estimate the CACE under all plausible identifiable mechanisms and report how much it varies, rather than commit to a single untestable mechanism. When nonidentifiable mechanisms are also plausible, this comparison can be supplemented with a formal sensitivity analysis.

The instrument Z may itself be missing. Let R^Z be the response indicator for Z, with $R^Z = 1$ if Z is observed and $R^Z = 0$ otherwise. Our analysis never uses the marginal distribution of Z. Each identification result relies on only one of the two conditional laws $\mathbb{P}(Y, D \mid Z)$ or $\mathbb{P}(Y \mid U = c, Z)$. If $R^Z \perp\!\!\!\perp (U, D, Y) \mid Z$ for the results based on $\mathbb{P}(Y, D \mid Z)$, or $R^Z \perp\!\!\!\perp Y \mid (U, D, Z)$ for those based on $\mathbb{P}(Y \mid U = c, Z)$, then the relevant law is unchanged on the units with Z observed. It can therefore be recovered from those units, and the CACE remains identifiable. When these conditions fail, restricting to the units with Z observed distorts the relevant law, and identification would require modeling the missingness in Z jointly with that in D and Y. The same dichotomy applies to a missing pretreatment covariate X when the target is the CACE within levels of X. A full treatment of missingness in Z or X is left to future work.

Acknowledgments

We thank the action editor and reviewers for helpful comments. Peng Ding was supported by the U.S. National Science Foundation (grants # 1945136, # 2514234). The authors declare no competing interests.

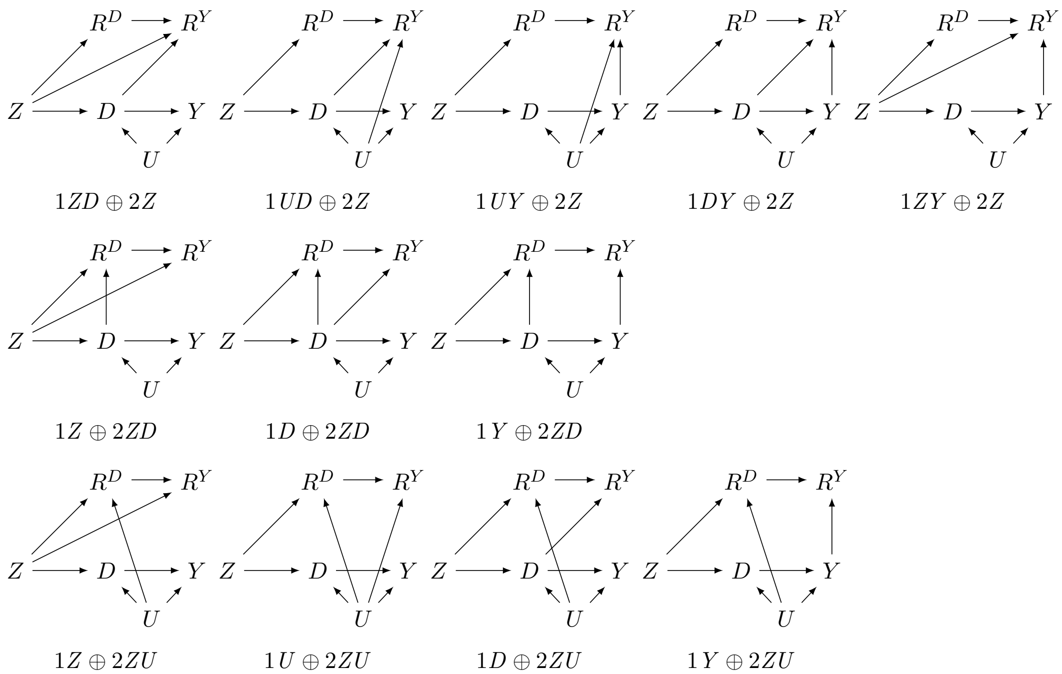


Figure 7: The DAGs illustrate the alternative combined mechanisms that retain identification of the CACE while allowing the direct path $R^D \to R^Y$.

Appendix A. Alternative Missingness Mechanisms

Section 6 allows the direct path $R^D \to R^Y$ only under Assumption $2UD$. This appendix shows that the path can also be accommodated under the other missingness mechanisms in the treatment, provided one of the remaining direct paths into R^D or R^Y is removed in exchange. Applying this trade-off to the eight combined mechanisms of Sections 6.2 and 6.3 yields eighteen alternative mechanisms that retain identification. Five of these reduce to the more general mechanisms of Section 6.1 and are omitted. A further reduction comes from noncompliance. Assumptions $1U \oplus 2ZU$ and $1U \oplus 2ZD$ both identify the CACE only under one-sided noncompliance, so we keep the more general $1U \oplus 2ZU$ and omit $1U \oplus 2ZD$. Section C.3 below gives a counterexample for $1U \oplus 2ZD$ under two-sided noncompliance, confirming that nothing is lost. The twelve remaining mechanisms are shown in Figure 7 and summarized in Table 5. They fall into two groups according to what R^D may depend on, covered in Sections A.1 and A.2.

A.1 The Outcome Models Combined with Treatment Missingness Depending on Z

In the first group, R^D depends only on the instrument Z, and R^Y follows one of the mechanisms of Section 4 and may also depend on R^D.

Assumption $1ZD \oplus 2Z$ *$R^D \perp\!\!\!\perp (U, D, Y) \mid Z$ and $R^Y \perp\!\!\!\perp (U, Y) \mid (Z, D, R^D)$.*

Table 5: Summary of the alternative combined mechanisms in Appendix A that allow the direct path $R^D \to R^Y$. Each row identifies the CACE under the listed conditions. A checkmark in the positivity or dependence column indicates that the corresponding condition is required; the exact conditions are stated in Theorems 6 and 7. The symbol $*$ means no restriction on that aspect.

Mechanism	R^D depends on	R^Y depends on	Positivity	Outcome	Noncompliance	Dependence
Group 1: R^D depends on Z only (Theorem 6)						
$1ZD \oplus 2Z$	Z	Z, D, R^D	✓	$*$	$*$	$*$
$1UD \oplus 2Z$	Z	U, D, R^D	✓	$*$	$*$	$*$
$1UY \oplus 2Z$	Z	U, Y, R^D	✓	Binary	$*$	$*$
$1DY \oplus 2Z$	Z	D, Y, R^D	✓	Binary	Two-sided	✓
$1ZY \oplus 2Z$	Z	Z, Y, R^D	✓	Binary	Two-sided	✓
Group 2: R^D depends on (Z, D) or (Z, U) (Theorem 7)						
$1Z \oplus 2ZD$	Z, D	Z, R^D	✓	$*$	$*$	✓
$1D \oplus 2ZD$	Z, D	D, R^D	✓	$*$	$*$	✓
$1Y \oplus 2ZD$	Z, D	Y, R^D	✓	Binary	$*$	✓
$1Z \oplus 2ZU$	Z, U	Z, R^D	✓	$*$	One-sided	$*$
$1U \oplus 2ZU$	Z, U	U, R^D	✓	$*$	One-sided	✓
$1D \oplus 2ZU$	Z, U	D, R^D	✓	$*$	One-sided	✓
$1Y \oplus 2ZU$	Z, U	Y, R^D	✓	Binary	One-sided	✓

Assumption $1UD \oplus 2Z$ *$R^D \perp\!\!\!\perp (U, D, Y) \mid Z$ and $R^Y \perp\!\!\!\perp (Z, Y) \mid (U, D, R^D)$.*

Assumption $1UY \oplus 2Z$ *$R^D \perp\!\!\!\perp (U, D, Y) \mid Z$ and $R^Y \perp\!\!\!\perp (Z, D) \mid (U, Y, R^D)$.*

Assumption $1DY \oplus 2Z$ *$R^D \perp\!\!\!\perp (U, D, Y) \mid Z$ and $R^Y \perp\!\!\!\perp (Z, U) \mid (D, Y, R^D)$.*

Assumption $1ZY \oplus 2Z$ *$R^D \perp\!\!\!\perp (U, D, Y) \mid Z$ and $R^Y \perp\!\!\!\perp (U, D) \mid (Z, Y, R^D)$.*

The following theorem formalizes the identification results for this group.

Theorem 6 *Suppose the IV assumptions hold. The CACE is nonparametrically identifiable in each of the following cases:*
($1ZD \oplus 2Z$) under Assumption $1ZD \oplus 2Z$, if $\mathbb{P}(R^D = 1 \mid Z = z) > 0$ for $z = 0, 1$ and $\mathbb{P}(R^Y = 1 \mid Z = z, D = d, R^D = 1) > 0$ for all z and d;
($1UD \oplus 2Z$) under Assumption $1UD \oplus 2Z$, if $\mathbb{P}(R^D = 1 \mid Z = z) > 0$ for $z = 0, 1$ and $\mathbb{P}(R^Y = 1 \mid U = c, D = d, R^D = 1) > 0$ for $d = 0, 1$;
($1UY \oplus 2Z$) under Assumption $1UY \oplus 2Z$, for a binary Y, if $\mathbb{P}(R^D = 1 \mid Z = z) > 0$ for $z = 0, 1$ and $\mathbb{P}(R^Y = 1 \mid U = c, Y = y, R^D = 1) > 0$ for $y = 0, 1$;
($1DY \oplus 2Z$) under Assumption $1DY \oplus 2Z$, for a binary Y with two-sided noncompliance, if $\mathbb{P}(R^D = 1 \mid Z = z) > 0$ for $z = 0, 1$, $\mathbb{P}(R^Y = 1 \mid D = d, Y = y, R^D = 1) > 0$ for all d and y, and $Y \not\perp\!\!\!\perp Z \mid (D = d, R^D = 1)$ for $d = 0, 1$;
($1ZY \oplus 2Z$) under Assumption $1ZY \oplus 2Z$, for a binary Y with two-sided noncompliance, if $\mathbb{P}(R^D = 1 \mid Z = z) > 0$ for $z = 0, 1$, $\mathbb{P}(R^Y = 1 \mid Z = z, Y = y, R^D = 1) > 0$ for all z and y, and $Y \not\perp\!\!\!\perp D \mid Z = z$ for $z = 0, 1$.

A.2 The Outcome Models Combined with Treatment Missingness Depending on (Z, D) or (Z, U)

In the second group, R^D depends on (Z, D) or on (Z, U), and R^Y depends on one variable and may also depend on R^D.

Assumption $1Z \oplus 2ZD$ *$R^D \perp\!\!\!\perp (U, Y) \mid (Z, D)$ and $R^Y \perp\!\!\!\perp (U, D, Y) \mid (Z, R^D)$.*

Assumption $1D \oplus 2ZD$ *$R^D \perp\!\!\!\perp (U, Y) \mid (Z, D)$ and $R^Y \perp\!\!\!\perp (Z, U, Y) \mid (D, R^D)$.*

Assumption $1Y \oplus 2ZD$ *$R^D \perp\!\!\!\perp (U, Y) \mid (Z, D)$ and $R^Y \perp\!\!\!\perp (Z, U, D) \mid (Y, R^D)$.*

Assumption $1Z \oplus 2ZU$ *$R^D \perp\!\!\!\perp (D, Y) \mid (Z, U)$ and $R^Y \perp\!\!\!\perp (U, D, Y) \mid (Z, R^D)$.*

Assumption $1U \oplus 2ZU$ *$R^D \perp\!\!\!\perp (D, Y) \mid (Z, U)$ and $R^Y \perp\!\!\!\perp (Z, D, Y) \mid (U, R^D)$.*

Assumption $1D \oplus 2ZU$ *$R^D \perp\!\!\!\perp (D, Y) \mid (Z, U)$ and $R^Y \perp\!\!\!\perp (Z, U, Y) \mid (D, R^D)$.*

Assumption $1Y \oplus 2ZU$ *$R^D \perp\!\!\!\perp (D, Y) \mid (Z, U)$ and $R^Y \perp\!\!\!\perp (Z, U, D) \mid (Y, R^D)$.*

The following theorem formalizes the identification results for this group.

Theorem 7 *Suppose the IV assumptions hold. The CACE is nonparametrically identifiable in each of the following cases:*
($1Z \oplus 2ZD$) under Assumption $1Z \oplus 2ZD$, if $\mathbb{P}(R^D = 1 \mid Z = z, D = d) > 0$ for all z and d and $\mathbb{P}(R^Y = 1 \mid Z = z, R^D = r^D) > 0$ for all z and r^D, and either (i) one-sided noncompliance with $Y \not\perp\!\!\!\perp D \mid Z = 1$, or (ii) two-sided noncompliance with $Y \not\perp\!\!\!\perp D \mid Z = z$ for $z = 0, 1$;
($1D \oplus 2ZD$) under Assumption $1D \oplus 2ZD$, if $\mathbb{P}(R^D = 1 \mid Z = z, D = d) > 0$ for all z and d, $\mathbb{P}(R^Y = 1 \mid D = d, R^D = r^D) > 0$ for all d and r^D, and $D \not\perp\!\!\!\perp Z$, and either (i) one-sided noncompliance with $Y \not\perp\!\!\!\perp D \mid Z = 1$, or (ii) two-sided noncompliance with $Y \not\perp\!\!\!\perp D \mid Z = z$ for $z = 0, 1$;
($1Y \oplus 2ZD$) under Assumption $1Y \oplus 2ZD$, for a binary Y, if $\mathbb{P}(R^D = 1 \mid Z = z, D = d) > 0$ for all z and d, $\mathbb{P}(R^Y = 1 \mid Y = y, R^D = r^D) > 0$ for all y and r^D, $Y \not\perp\!\!\!\perp (Z, D) \mid (R^D = 1)$, and $Y \not\perp\!\!\!\perp Z \mid (R^D = 0)$, and either (i) one-sided noncompliance with $Y \not\perp\!\!\!\perp D \mid Z = 1$, or (ii) two-sided noncompliance with $Y \not\perp\!\!\!\perp D \mid Z = z$ for $z = 0, 1$;
($1Z \oplus 2ZU$) under Assumption $1Z \oplus 2ZU$, with one-sided noncompliance, if $\mathbb{P}(R^D = 1 \mid Z = 1, U = u) > 0$ for $u = n, c$ and $\mathbb{P}(R^Y = 1 \mid Z = z, R^D = r^D) > 0$ for all z and r^D;
($1U \oplus 2ZU$) under Assumption $1U \oplus 2ZU$, with one-sided noncompliance, if $\mathbb{P}(R^D = 1 \mid Z = 1, U = n) > 0$, $\mathbb{P}(R^D = 1 \mid Z = z, U = c) > 0$ for $z = 0, 1$, $\mathbb{P}(R^Y = 1 \mid U = c, R^D = 1) > 0$, and $R^Y \not\perp\!\!\!\perp U \mid (R^D = 1)$;
($1D \oplus 2ZU$) under Assumption $1D \oplus 2ZU$, with one-sided noncompliance, if $\mathbb{P}(R^D = 1 \mid Z = 1, U = u) > 0$ for $u = n, c$, $\mathbb{P}(R^Y = 1 \mid D = d, R^D = 1) > 0$ for $d = 0, 1$, $\mathbb{P}(R^Y = 1 \mid D = 0, R^D = 0) > 0$, and $Y \not\perp\!\!\!\perp U \mid (Z = 1)$;
($1Y \oplus 2ZU$) under Assumption $1Y \oplus 2ZU$, for a binary Y with one-sided noncompliance, if $\mathbb{P}(R^D = 1 \mid Z = 1, U = u) > 0$ for $u = n, c$, $\mathbb{P}(R^Y = 1 \mid Y = y, R^D = r^D) > 0$ for all y and r^D, $Y \not\perp\!\!\!\perp (Z, D) \mid (R^D = 1)$, and $Y \not\perp\!\!\!\perp Z \mid (R^D = 0)$.

Appendix B. Proofs

The main paper focuses on binary Z and D. In this appendix, some arguments are stated first for more general supports using a completeness condition. Let $f(a \mid b)$ denote the conditional density or probability mass function of A given $B = b$. We say that $f(a \mid b)$ is complete in B if

$$\int g(a) f(a \mid b) \mathrm{d}\nu(a) = 0 \quad \text{for all } b$$

implies $g(a) = 0$ almost surely for every square-integrable function g (Newey and Powell, 2003), where $\nu(\cdot)$ is the Lebesgue measure for a continuous variable and the counting measure for a discrete variable.

For the binary variables in the main theorems, this completeness requirement reduces to a finite-dimensional rank condition. To see this, suppose A is binary and write $p_b = \mathbb{P}(A = 1 \mid B = b)$. The display above then becomes

$$(1 - p_b)g(0) + p_b g(1) = 0 \quad \text{for all } b.$$

If B takes values $b_1, \ldots, b_m$, these equations can be written as

$$\begin{pmatrix} 1 - p_{b_1} & p_{b_1} \\ \vdots & \vdots \\ 1 - p_{b_m} & p_{b_m} \end{pmatrix} \begin{pmatrix} g(0) \\ g(1) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

The only solution is $g(0) = g(1) = 0$ exactly when this matrix has rank two. Since the two columns are $(1 - p_b)$ and p_b, this holds if and only if p_b is not constant in b. Thus, for binary B, completeness is equivalent to

$$\mathbb{P}(A = 1 \mid B = 0) \neq \mathbb{P}(A = 1 \mid B = 1),$$

that is, $A \not\perp\!\!\!\perp B$. When B has more than two support points, the analogous condition is that $\mathbb{P}(A = 1 \mid B = b)$ varies with b; and if A has K categories, completeness reduces to the requirement that the matrix with entries $\mathbb{P}(A = a \mid B = b)$ has column rank K. The nonindependence conditions in the main text should therefore be read as the finite-support rank conditions arising in the binary cases, not as generic completeness conditions for arbitrary continuous variables.

B.1 Proof of Theorem 1

We prove each case of Theorem 1 separately. We then prove the result for Assumption 1Y of Chen et al. (2009a), discussed in Section 4.

B.1.1 Assumption 1ZD

The identification of $\mathbb{P}(D = d, Y = y \mid Z = z)$ follows from

$$\mathbb{P}(D = d, Y = y \mid Z = z) = \frac{\mathbb{P}(D = d, Y = y, R^Y = 1 \mid Z = z)}{\mathbb{P}(R^Y = 1 \mid Z = z, D = d)}.$$

In addition, since

$$
\begin{aligned}
&\mathbb{P}(U=n, D=0, Y=y \mid Z=0)=\mathbb{P}(U=n, D=0, Y=y \mid Z=1),\\
&\mathbb{P}(U=a, D=1, Y=y \mid Z=1)=\mathbb{P}(U=a, D=1, Y=y \mid Z=0),\\
&\mathbb{P}(U=c, D=0, Y=y \mid Z=0)=\mathbb{P}(D=0, Y=y \mid Z=0)-\mathbb{P}(U=n, D=0, Y=y \mid Z=0),\\
&\mathbb{P}(U=c, D=1, Y=y \mid Z=1)=\mathbb{P}(D=1, Y=y \mid Z=1)-\mathbb{P}(U=a, D=1, Y=y \mid Z=1),
\end{aligned}
$$

we can identify $\mathbb{P}(Z=z, U=u, D=d, Y=y)$. Therefore, identifying $\mathbb{P}(D, Y \mid Z)$ implies identifying $\mathbb{P}(Z, U, D, Y)$.

B.1.2 Assumption $1UD$

We focus on the identification of $\mathbb{P}(Y=y \mid U=c, D=d)$ to identify the CACE. Note that

$$
\begin{aligned}
&\mathbb{P}(Y=y \mid U=c, D=0)\\
&=\mathbb{P}(Y=y \mid Z=0, U=c, D=0, R^Y=1)\\
&=\frac{\mathbb{P}(U=c, D=0, Y=y, R^Y=1 \mid Z=0)}{\mathbb{P}(U=c, D=0, R^Y=1 \mid Z=0)}\\
&=\frac{\mathbb{P}(D=0, Y=y, R^Y=1 \mid Z=0)-\mathbb{P}(U=n, D=0, Y=y, R^Y=1 \mid Z=0)}{\mathbb{P}(D=0, R^Y=1 \mid Z=0)-\mathbb{P}(U=n, D=0, R^Y=1 \mid Z=0)}.
\end{aligned}
$$

Since

$$
\begin{aligned}
&\mathbb{P}(U=n, D=0, Y=y, R^Y=1 \mid Z=0)\\
&=\mathbb{P}(Y=y, R^Y=1 \mid Z=0, U=n, D=0)\mathbb{P}(U=n, D=0 \mid Z=0)\\
&=\mathbb{P}(Y=y, R^Y=1 \mid Z=1, U=n, D=0)\mathbb{P}(U=n, D=0 \mid Z=1)\\
&=\mathbb{P}(U=n, D=0, Y=y, R^Y=1 \mid Z=1),
\end{aligned}
$$

we can identify both $\mathbb{P}(U=n, D=0, Y=y, R^Y=1 \mid Z=0)$ and $\mathbb{P}(U=n, D=0, R^Y=1 \mid Z=0)=\int_{y\in\mathcal{Y}}\mathbb{P}(U=n, D=0, Y=y, R^Y=1 \mid Z=0)\mathrm{d}y$. Therefore, $\mathbb{P}(Y=y \mid U=c, D=0)$ is identifiable. Similarly, we can identify $\mathbb{P}(Y=y \mid U=c, D=1)$. Then, we identify $\text{CACE}=\mathbb{E}(Y \mid U=c, D=1)-\mathbb{E}(Y \mid U=c, D=0)$.

In addition, we can identify $\mathbb{P}(Z, U, D, Y)$ under additional positivity conditions. For z that is consistent with the (u, d) combination,

$$
\begin{aligned}
\mathbb{P}(R^Y=1 \mid U=u, D=d)&=\mathbb{P}(R^Y=1 \mid Z=z, U=u, D=d)\\
&=\frac{\mathbb{P}(U=u, D=d, R^Y=1 \mid Z=z)}{\mathbb{P}(U=u, D=d \mid Z=z)}.
\end{aligned}
$$

If $\mathbb{P}(R^Y=1 \mid U=u, D=d)>0$ for $(u, d)=(a,1),(n,0),(c,1),(c,0)$, the identification of $\mathbb{P}(Z=z, U=u, D=d, Y=y)$ follows from

$$
\mathbb{P}(Z=z, U=u, D=d, Y=y)=\frac{\mathbb{P}(Z=z, U=u, D=d, Y=y, R^Y=1)}{\mathbb{P}(R^Y=1 \mid U=u, D=d)}.
$$

B.1.3 Assumption $1DY$

We focus on the identification of $\mathbb{P}(D = d, Y = y \mid Z = z)$ to identify the CACE. Define $\mathbb{P}_{dy1|z} = \mathbb{P}(D = d, Y = y, R^Y = 1 \mid Z = z)$, $\mathbb{P}_{d+0|z} = \mathbb{P}(D = d, R^Y = 0 \mid Z = z)$, $\eta_d(y) = \frac{\mathbb{P}(R^Y=0|D=d,Y=y)}{\mathbb{P}(R^Y=1|D=d,Y=y)}$. Since

$$\mathbb{P}_{dy1|z} = \mathbb{P}(D = d, Y = y \mid Z = z)\mathbb{P}(R^Y = 1 \mid D = d, Y = y),$$

we have

$$\mathbb{P}_{d+0|z} = \int_{y\in\mathcal{Y}} \mathbb{P}(D = d, Y = y, R^Y = 0 \mid Z = z)\mathrm{d}y = \int_{y\in\mathcal{Y}} \mathbb{P}_{dy1|z}\eta_d(y)\mathrm{d}y$$

for each $z \in \mathcal{Z}$. The solutions $\eta_d(y)$ are unique if $\mathbb{P}(D = d, Y, R^Y = 1 \mid Z)$ is complete in Z for all d. For binary Z and D, the solutions $\eta_d(y)$ are unique if Y is binary and $Y \not\perp\!\!\!\perp Z \mid D = d$ for $d = 0, 1$ under two-sided noncompliance.

We can identify $\mathbb{P}(R^Y = 1 \mid D = d, Y = y)$ once $\eta_d(y)$ is identified. Then, the identification of $\mathbb{P}(D = d, Y = y \mid Z = z)$ follows from

$$\mathbb{P}(D = d, Y = y \mid Z = z) = \frac{\mathbb{P}_{dy1|z}}{\mathbb{P}(R^Y = 1 \mid D = d, Y = y)}.$$

B.1.4 Assumption $1ZY$

We focus on the identification of $\mathbb{P}(D = d, Y = y \mid Z = z)$ to identify the CACE. Define $\mathbb{P}_{dy1|z} = \mathbb{P}(D = d, Y = y, R^Y = 1 \mid Z = z)$, $\mathbb{P}_{d+0|z} = \mathbb{P}(D = d, R^Y = 0 \mid Z = z)$, $\eta_z(y) = \frac{\mathbb{P}(R^Y=0|Z=z,Y=y)}{\mathbb{P}(R^Y=1|Z=z,Y=y)}$. Since

$$\mathbb{P}_{dy1|z} = \mathbb{P}(D = d, Y = y \mid Z = z)\mathbb{P}(R^Y = 1 \mid Z = z, Y = y),$$

we have

$$\mathbb{P}_{d+0|z} = \int_{y\in\mathcal{Y}} \mathbb{P}(D = d, Y = y, R^Y = 0 \mid Z = z)\mathrm{d}y = \int_{y\in\mathcal{Y}} \mathbb{P}_{dy1|z}\eta_z(y)\mathrm{d}y$$

for each $d \in \mathcal{D}$. The solutions $\eta_z(y)$ are unique if $\mathbb{P}(D, Y, R^Y = 1 \mid Z = z)$ is complete in D for all z. For binary Z and D, the solutions $\eta_z(y)$ are unique if Y is binary and $Y \not\perp\!\!\!\perp D \mid Z = z$ for $z = 0, 1$ under two-sided noncompliance.

We can identify $\mathbb{P}(R^Y = 1 \mid Z = z, Y = y)$ once $\eta_z(y)$ is identified. Then, the identification of $\mathbb{P}(D = d, Y = y \mid Z = z)$ follows from

$$\mathbb{P}(D = d, Y = y \mid Z = z) = \frac{\mathbb{P}_{dy1|z}}{\mathbb{P}(R^Y = 1 \mid Z = z, Y = y)}.$$

B.1.5 Assumption $1UY$

We focus on the identification of the CACE. Note that

$$\mathbb{P}(U = c, D = 0, Y = y, R^Y = 1 \mid Z = 0)$$

$$= \mathbb{P}(D = 0, Y = y, R^Y = 1 \mid Z = 0) - \mathbb{P}(U = n, D = 0, Y = y, R^Y = 1 \mid Z = 0).$$

Since

$$\begin{aligned}
&\mathbb{P}(U = n, D = 0, Y = y, R^Y = 1 \mid Z = 0) \\
&= \mathbb{P}(Y = y, R^Y = 1 \mid Z = 0, U = n, D = 0)\mathbb{P}(U = n, D = 0 \mid Z = 0) \\
&= \mathbb{P}(Y = y, R^Y = 1 \mid Z = 1, U = n, D = 0)\mathbb{P}(U = n, D = 0 \mid Z = 1) \\
&= \mathbb{P}(U = n, D = 0, Y = y, R^Y = 1 \mid Z = 1),
\end{aligned}$$

we can identify $\mathbb{P}(U = c, D = 0, Y = y, R^Y = 1 \mid Z = 0)$. Similarly, we can identify $\mathbb{P}(U = c, D = 1, Y = y, R^Y = 1 \mid Z = 1)$. Since

$$\begin{aligned}
\frac{\mathbb{P}(Y = 1 \mid U = c, D = 0)}{\mathbb{P}(Y = 1 \mid U = c, D = 1)} &= \frac{\mathbb{P}(U = c, D = 0, Y = 1, R^Y = 1 \mid Z = 0)}{\mathbb{P}(U = c, D = 1, Y = 1, R^Y = 1 \mid Z = 1)}, \\
\frac{1 - \mathbb{P}(Y = 1 \mid U = c, D = 0)}{1 - \mathbb{P}(Y = 1 \mid U = c, D = 1)} &= \frac{\mathbb{P}(U = c, D = 0, Y = 0, R^Y = 1 \mid Z = 0)}{\mathbb{P}(U = c, D = 1, Y = 0, R^Y = 1 \mid Z = 1)},
\end{aligned}$$

we can identify the CACE.

However, $\mathbb{P}(Z, U, D, Y)$ is not identifiable under either two-sided or one-sided noncompliance, because it has more unknown parameters than the observable data frequencies have degrees of freedom. Under two-sided noncompliance, the unknown parameters of $\mathbb{P}(Z, U, D, Y)$ under Assumption $1UY$ are: $\mathbb{P}(Z = 1)$; $\mathbb{P}(U = u)$ for $u = a, n$; $\mathbb{P}(Y = 1 \mid U = u, D = d)$ for $(u, d) = (a, 1), (n, 0), (c, 1), (c, 0)$; and $\mathbb{P}(R^Y = 1 \mid U = u, Y = y)$ for $u = a, n, c$ and $y = 0, 1$. In total, there are 13 unknown parameters. In contrast, we have the following observable data frequencies: $\mathbb{P}(Z = z, D = d, Y = y, R^Y = 1)$ for $z = 0, 1$, $d = 0, 1$, and $y = 0, 1$; and $\mathbb{P}(Z = z, D = d, R^Y = 0)$ for $z = 0, 1$ and $d = 0, 1$. There are 12 observable data frequencies. Because they sum to 1, the degrees of freedom equal 11. Since $11 < 13$, $\mathbb{P}(Z, U, D, Y)$ is not identifiable. Under one-sided noncompliance, there are 9 unknown parameters and 8 degrees of freedom, so $\mathbb{P}(Z, U, D, Y)$ is again not identifiable.

B.1.6 Assumption $1Y$

Under Assumption $1Y$, we have

$$\mathbb{P}(D = d, Y = y \mid Z = z) = \frac{\mathbb{P}(D = d, Y = y, R^Y = 1 \mid Z = z)}{\mathbb{P}(R^Y = 1 \mid Y = y)}.$$

Therefore, the conditional distribution $\mathbb{P}(D, Y \mid Z)$ is identifiable, and so is the CACE if $\mathbb{P}(R^Y \mid Y)$ is identifiable. Define

$$\eta(y) = \frac{\mathbb{P}(R^Y = 0 \mid Y = y)}{\mathbb{P}(R^Y = 1 \mid Y = y)}$$

for all y. We then have the following system of linear equations with $\{\eta(y) : y \in \mathcal{Y}\}$ as the unknowns:

$$\mathbb{P}(D = d, R^Y = 0 \mid Z = z) = \sum_{y \in \mathcal{Y}} \mathbb{P}(D = d, Y = y, R^Y = 0 \mid Z = z)$$

$$= \sum_{y \in \mathcal{Y}} \mathbb{P}(D = d, Y = y, R^Y = 1 \mid Z = z)\eta(y)$$

for all z and d. For a binary Y, the solutions $\eta(y)$ are unique if $Y \not\perp\!\!\!\perp (Z, D)$. For a discrete Y, let $|\mathcal{Y}|$ denote the number of unique values that Y can take. With two-sided noncompliance, we have $(z, d) = (0, 0), (0, 1), (1, 0), (1, 1)$. The solutions $\eta(y)$ are unique if the rank of the $4 \times |\mathcal{Y}|$ matrix with $\mathbb{P}(D = d, Y = y, R^Y = 1 \mid Z = z)$ as the entries equals $|\mathcal{Y}|$, which means that $|\mathcal{Y}| \leq 4$. With one-sided noncompliance, we have $(z, d) = (0, 0), (1, 0), (1, 1)$. Here, the solutions $\eta(y)$ are unique if the rank of the $3 \times |\mathcal{Y}|$ matrix with $\mathbb{P}(D = d, Y = y, R^Y = 1 \mid Z = z)$ as the entries equals $|\mathcal{Y}|$, which means that $|\mathcal{Y}| \leq 3$. ∎

B.2 Proof of Theorem 2

We prove each case of Theorem 2 separately.

B.2.1 Assumption 2ZY

The identification of $\mathbb{P}(D = d, Y = y \mid Z = z)$ follows from

$$\mathbb{P}(D = d, Y = y \mid Z = z) = \frac{\mathbb{P}(D = d, Y = y, R^D = 1 \mid Z = z)}{\mathbb{P}(R^D = 1 \mid Z = z, Y = y)}.$$

B.2.2 Assumption 2UY

We focus on the identification of the CACE. Note that

$$\begin{aligned}
&\mathbb{P}(U = c, D = 0, Y = y, R^D = 1 \mid Z = 0) \\
&= \mathbb{P}(D = 0, Y = y, R^D = 1 \mid Z = 0) - \mathbb{P}(U = n, D = 0, Y = y, R^D = 1 \mid Z = 0).
\end{aligned}$$

Since

$$\mathbb{P}(U = n, D = 0, Y = y, R^D = 1 \mid Z = 0) = \mathbb{P}(U = n, D = 0, Y = y, R^D = 1 \mid Z = 1),$$

we can identify $\mathbb{P}(U = c, D = 0, Y = y, R^D = 1 \mid Z = 0)$. Similarly, we can identify $\mathbb{P}(U = c, D = 1, Y = y, R^D = 1 \mid Z = 1)$. Since

$$\begin{aligned}
\frac{\mathbb{P}(Y = 1 \mid U = c, D = 0)}{\mathbb{P}(Y = 1 \mid U = c, D = 1)} &= \frac{\mathbb{P}(U = c, D = 0, Y = 1, R^D = 1 \mid Z = 0)}{\mathbb{P}(U = c, D = 1, Y = 1, R^D = 1 \mid Z = 1)}, \\
\frac{1 - \mathbb{P}(Y = 1 \mid U = c, D = 0)}{1 - \mathbb{P}(Y = 1 \mid U = c, D = 1)} &= \frac{\mathbb{P}(U = c, D = 0, Y = 0, R^D = 1 \mid Z = 0)}{\mathbb{P}(U = c, D = 1, Y = 0, R^D = 1 \mid Z = 1)},
\end{aligned}$$

we can identify the CACE.

Assumption 2UY differs from Assumption 1UY here. Under Assumption 2UY, $\mathbb{P}(Z, U, D, Y)$ is not identifiable with two-sided noncompliance, but it is identifiable with one-sided noncompliance under a full rank condition and additional positivity conditions. Under two-sided noncompliance, the counting argument of Section B.1.5 gives 13 unknown parameters and 11 degrees of freedom, so $\mathbb{P}(Z, U, D, Y)$ is not identifiable. Under one-sided noncompliance, there are 9 unknown parameters and 9 degrees of freedom, which leaves identification possible.

We now show the identifiability of $\mathbb{P}(Z, U, D, Y)$ with one-sided noncompliance. Define $\mathbb{P}_{zudy1} = \mathbb{P}(Z = z, U = u, D = d, Y = y, R^D = 1)$, $\mathbb{P}_{zy+0} = \mathbb{P}(Z = z, Y = y, R^D = 0)$, $\zeta_y(u) = \frac{\mathbb{P}(R^D=0|U=u,Y=y)}{\mathbb{P}(R^D=1|U=u,Y=y)}$. Since

$$\mathbb{P}_{zudy1} = \mathbb{P}(Z = z, U = u, D = d, Y = y)\mathbb{P}(R^D = 1 \mid U = u, Y = y),$$

we have

$$\begin{aligned} \mathbb{P}_{1y+0} &= \mathbb{P}_{1n0y1}\zeta_y(n) + \mathbb{P}_{1c1y1}\zeta_y(c), \\ \mathbb{P}_{0y+0} &= \mathbb{P}_{0n0y1}\zeta_y(n) + \mathbb{P}_{0c0y1}\zeta_y(c), \end{aligned}$$

for $y = 0, 1$. The solutions $\zeta_y(u)$ are unique if $\mathbb{P}(Y = 1 \mid U = c, D = 0) \neq \mathbb{P}(Y = 1 \mid U = c, D = 1)$.

We can identify $\mathbb{P}(R^D = 1 \mid U = u, Y = y)$ once $\zeta_y(u)$ is identified. Then, if $\mathbb{P}(R^D = 1 \mid U = u, Y = y) > 0$ for $u = n, c$ and $y = 0, 1$, the identification of $\mathbb{P}(Z = z, U = u, D = d, Y = y)$ follows from

$$\mathbb{P}(Z = z, U = u, D = d, Y = y) = \frac{\mathbb{P}_{zudy1}}{\mathbb{P}(R^D = 1 \mid U = u, Y = y)}.$$

B.2.3 Assumption $2DY$

We focus on the identification of $\mathbb{P}(D = d, Y = y \mid Z = z)$ to identify the CACE. Define $\mathbb{P}_{dy1|z} = \mathbb{P}(D = d, Y = y, R^D = 1 \mid Z = z)$, $\mathbb{P}_{y+0|z} = \mathbb{P}(Y = y, R^D = 0 \mid Z = z)$, $\zeta_y(d) = \frac{\mathbb{P}(R^D=0|D=d,Y=y)}{\mathbb{P}(R^D=1|D=d,Y=y)}$. Since

$$\mathbb{P}_{dy1|z} = \mathbb{P}(D = d, Y = y \mid Z = z)\mathbb{P}(R^D = 1 \mid D = d, Y = y),$$

we have

$$\mathbb{P}_{y+0|z} = \int_{d\in\mathcal{D}} \mathbb{P}(D = d, Y = y, R^D = 0 \mid Z = z)\mathrm{d}d = \int_{d\in\mathcal{D}} \mathbb{P}_{dy1|z}\zeta_y(d)\mathrm{d}d$$

for each $z \in \mathcal{Z}$. The solutions $\zeta_y(d)$ are unique if $\mathbb{P}(D, Y = y, R^D = 1 \mid Z)$ is complete in Z for all y. For binary Z and D, we can identify $\zeta_y(d)$ directly under one-sided noncompliance, and under two-sided noncompliance the solutions $\zeta_y(d)$ are unique if $D \not\perp\!\!\!\perp Z \mid Y = y$ for all y.

We can identify $\mathbb{P}(R^D = 1 \mid D = d, Y = y)$ once $\zeta_y(d)$ is identified. Then, the identification of $\mathbb{P}(D = d, Y = y \mid Z = z)$ follows from

$$\mathbb{P}(D = d, Y = y \mid Z = z) = \frac{\mathbb{P}_{dy1|z}}{\mathbb{P}(R^D = 1 \mid D = d, Y = y)}.$$

B.2.4 Assumption $2UD$

We focus on the identification of $\mathbb{P}(Y = y \mid U = c, D = d)$ to identify the CACE. Note that

$$\mathbb{P}(Y = y \mid U = c, D = 0)$$

$$= \frac{\mathbb{P}(U=c, D=0, Y=y, R^D=1 \mid Z=0)}{\mathbb{P}(U=c, D=0, R^D=1 \mid Z=0)}$$
$$= \frac{\mathbb{P}(D=0, Y=y, R^D=1 \mid Z=0) - \mathbb{P}(U=n, D=0, Y=y, R^D=1 \mid Z=0)}{\mathbb{P}(D=0, R^D=1 \mid Z=0) - \mathbb{P}(U=n, D=0, R^D=1 \mid Z=0)}.$$

Since

$$\mathbb{P}(U=n, D=0, Y=y, R^D=1 \mid Z=0) = \mathbb{P}(U=n, D=0, Y=y, R^D=1 \mid Z=1),$$

we can identify $\mathbb{P}(Y=y \mid U=c, D=0)$. Similarly, we can identify $\mathbb{P}(Y=y \mid U=c, D=1)$, and therefore, the CACE.

In addition, we can identify $\mathbb{P}(Z,U,D,Y)$ under a full-rank condition and additional positivity conditions. Define $\mathbb{P}_{zudy1} = \mathbb{P}(Z=z, U=u, D=d, Y=y, R^D=1)$, $\mathbb{P}_{zy+0} = \mathbb{P}(Z=z, Y=y, R^D=0)$, $\zeta(u,d) = \frac{\mathbb{P}(R^D=0 \mid U=u, D=d)}{\mathbb{P}(R^D=1 \mid U=u, D=d)}$. Since

$$\mathbb{P}_{zudy1} = \mathbb{P}(Z=z, U=u, D=d, Y=y)\mathbb{P}(R^D=1 \mid U=u, D=d),$$

we have

$$\mathbb{P}_{0y+0} = \mathbb{P}_{0n0y1}\zeta(n,0) + \mathbb{P}_{0a1y1}\zeta(a,1) + \mathbb{P}_{0c0y1}\zeta(c,0),$$
$$\mathbb{P}_{1y+0} = \mathbb{P}_{1n0y1}\zeta(n,0) + \mathbb{P}_{1a1y1}\zeta(a,1) + \mathbb{P}_{1c1y1}\zeta(c,1),$$

for each $y \in \mathcal{Y}$. The solutions $\zeta(u,d)$ are unique if the above system of linear equations has full rank.

We can identify $\mathbb{P}(R^D=1 \mid U=u, D=d)$ once $\zeta(u,d)$ is identified. Then, if $\mathbb{P}(R^D=1 \mid U=u, D=d) > 0$ for $(u,d) = (a,1), (n,0), (c,1), (c,0)$, the identification of $\mathbb{P}(Z=z, U=u, D=d, Y=y)$ follows from

$$\mathbb{P}(Z=z, U=u, D=d, Y=y) = \frac{\mathbb{P}_{zudy1}}{\mathbb{P}(R^D=1 \mid U=u, D=d)}.$$

B.2.5 Assumption $2ZD$

We focus on the identification of $\mathbb{P}(D=d, Y=y \mid Z=z)$ to identify the CACE. Define $\mathbb{P}_{dy1|z} = \mathbb{P}(D=d, Y=y, R^D=1 \mid Z=z)$, $\mathbb{P}_{y+0|z} = \mathbb{P}(Y=y, R^D=0 \mid Z=z)$, $\zeta_z(d) = \frac{\mathbb{P}(R^D=0 \mid Z=z, D=d)}{\mathbb{P}(R^D=1 \mid Z=z, D=d)}$. Since

$$\mathbb{P}_{dy1|z} = \mathbb{P}(D=d, Y=y \mid Z=z)\mathbb{P}(R^D=1 \mid Z=z, D=d),$$

we have

$$\mathbb{P}_{y+0|z} = \int_{d \in \mathcal{D}} \mathbb{P}(D=d, Y=y, R^D=0 \mid Z=z)\mathrm{d}d = \int_{d \in \mathcal{D}} \mathbb{P}_{dy1|z}\zeta_z(d)\mathrm{d}d$$

for each $y \in \mathcal{Y}$. The solutions $\zeta_z(d)$ are unique if $\mathbb{P}(D, Y, R^D=1 \mid Z=z)$ is complete in Y for all z. For binary Z and D, the solutions $\zeta_z(d)$ are unique if $Y \not\perp\!\!\!\perp D \mid Z=1$ under one-sided noncompliance, and if $Y \not\perp\!\!\!\perp D \mid Z=z$ for $z=0,1$ under two-sided noncompliance.

We can identify $\mathbb{P}(R^D=1 \mid Z=z, D=d)$ once $\zeta_z(d)$ is identified. Then, the identification of $\mathbb{P}(D=d, Y=y \mid Z=z)$ follows from

$$\mathbb{P}(D=d, Y=y \mid Z=z) = \frac{\mathbb{P}_{dy1|z}}{\mathbb{P}(R^D=1 \mid Z=z, D=d)}.$$

B.2.6 Assumption $2ZU$

The identification of $\mathbb{P}(Y = y \mid U = n, D = 0)$ and $\mathbb{P}(Y = y \mid U = c, D = 1)$ follows from

$$\mathbb{P}(Y = y \mid U = n, D = 0) = \mathbb{P}(Y = y \mid Z = 1, U = n, D = 0, R^D = 1),$$
$$\mathbb{P}(Y = y \mid U = c, D = 1) = \mathbb{P}(Y = y \mid Z = 1, U = c, D = 1, R^D = 1).$$

Since

$$\begin{aligned}\mathbb{P}(Y = y \mid Z = 1) &= \mathbb{P}(U = n, D = 0, Y = y \mid Z = 1) + \mathbb{P}(U = c, D = 1, Y = y \mid Z = 1)\\ &= \mathbb{P}(Y = y \mid U = n, D = 0)\mathbb{P}(U = n, D = 0 \mid Z = 1)\\ &\quad + \mathbb{P}(Y = y \mid U = c, D = 1)\{1 - \mathbb{P}(U = n, D = 0 \mid Z = 1)\},\end{aligned}$$

we can identify $\mathbb{P}(U = n, D = 0 \mid Z = 1)$. Since

$$\begin{aligned}\mathbb{P}(Y = y \mid Z = 0) &= \mathbb{P}(U = n, D = 0, Y = y \mid Z = 0) + \mathbb{P}(U = c, D = 0, Y = y \mid Z = 0)\\ &= \mathbb{P}(Y = y \mid U = n, D = 0)\mathbb{P}(U = n, D = 0 \mid Z = 1)\\ &\quad + \mathbb{P}(Y = y \mid U = c, D = 0)\{1 - \mathbb{P}(U = n, D = 0 \mid Z = 1)\},\end{aligned}$$

we can identify $\mathbb{P}(Y = y \mid U = c, D = 0)$. In addition, we can identify $\mathbb{P}(Z = z, U = u, D = d, Y = y)$. ■

B.3 Proof of Theorem 3

We prove each case of Theorem 3 separately.

B.3.1 Assumption $1ZD \oplus 2UD$

Note that

$$\begin{aligned}&\mathbb{P}(Y = y \mid U = c, D = 0)\\ &= \frac{\mathbb{P}(U = c, D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 0)}{\mathbb{P}(U = c, D = 0, R^D = 1, R^Y = 1 \mid Z = 0)}\\ &= \frac{\mathbb{P}(D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 0) - \mathbb{P}(U = n, D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 0)}{\mathbb{P}(D = 0, R^D = 1, R^Y = 1 \mid Z = 0) - \mathbb{P}(U = n, D = 0, R^D = 1, R^Y = 1 \mid Z = 0)}.\end{aligned}$$

Since

$$\begin{aligned}&\mathbb{P}(U = n, D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 0)\\ &= \mathbb{P}(U = n, D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 1)\frac{\mathbb{P}(R^Y = 1 \mid Z = 0, D = 0, R^D = 1)}{\mathbb{P}(R^Y = 1 \mid Z = 1, D = 0, R^D = 1)},\end{aligned}$$

we can identify $\mathbb{P}(Y = y \mid U = c, D = 0)$, and similarly, $\mathbb{P}(Y = y \mid U = c, D = 1)$.

B.3.2 Assumption $1UD \oplus 2UD$

Note that

$$\mathbb{P}(Y = y \mid U = c, D = 0)$$

$$
\begin{aligned}
&= \frac{\mathbb{P}(U=c, D=0, Y=y, R^D=1, R^Y=1 \mid Z=0)}{\mathbb{P}(U=c, D=0, R^D=1, R^Y=1 \mid Z=0)} \\
&= \frac{\mathbb{P}(D=0, Y=y, R^D=1, R^Y=1 \mid Z=0) - \mathbb{P}(U=n, D=0, Y=y, R^D=1, R^Y=1 \mid Z=0)}{\mathbb{P}(D=0, R^D=1, R^Y=1 \mid Z=0) - \mathbb{P}(U=n, D=0, R^D=1, R^Y=1 \mid Z=0)}.
\end{aligned}
$$

Since

$$
\mathbb{P}(U=n, D=0, Y=y, R^D=1, R^Y=1 \mid Z=0) = \mathbb{P}(U=n, D=0, Y=y, R^D=1, R^Y=1 \mid Z=1),
$$

we can identify $\mathbb{P}(Y=y \mid U=c, D=0)$, and similarly, $\mathbb{P}(Y=y \mid U=c, D=1)$.

B.3.3 Assumption $1DY \oplus 2UD$

Note that

$$
\begin{aligned}
&\mathbb{P}(Y=y, R^Y=1 \mid U=c, D=0, R^D=1) \\
&= \frac{\mathbb{P}(U=c, D=0, Y=y, R^D=1, R^Y=1 \mid Z=0)}{\mathbb{P}(U=c, D=0, R^D=1 \mid Z=0)} \\
&= \frac{\mathbb{P}(D=0, Y=y, R^D=1, R^Y=1 \mid Z=0) - \mathbb{P}(U=n, D=0, Y=y, R^D=1, R^Y=1 \mid Z=0)}{\mathbb{P}(D=0, R^D=1 \mid Z=0) - \mathbb{P}(U=n, D=0, R^D=1 \mid Z=0)}.
\end{aligned}
$$

Since

$$
\mathbb{P}(U=n, D=0, Y=y, R^D=1, R^Y=1 \mid Z=0) = \mathbb{P}(U=n, D=0, Y=y, R^D=1, R^Y=1 \mid Z=1),
$$
$$
\mathbb{P}(U=n, D=0, R^D=1 \mid Z=0) = \mathbb{P}(U=n, D=0, R^D=1 \mid Z=1),
$$

we can identify $\mathbb{P}(Y=y, R^Y=1 \mid U=c, D=0, R^D=1)$, and similarly, $\mathbb{P}(Y=y, R^Y=1 \mid U=c, D=1, R^D=1)$. The identification of $\mathbb{P}(R^Y=1 \mid D=d, Y=y, R^D=1)$ follows the same logic as in Assumption $1DY \oplus 2Z$, as detailed in Section B.6.4. The identification of $\mathbb{P}(Y=y \mid U=c, D=d)$ follows from

$$
\mathbb{P}(Y=y \mid U=c, D=d) = \frac{\mathbb{P}(Y=y, R^Y=1 \mid U=c, D=d, R^D=1)}{\mathbb{P}(R^Y=1 \mid D=d, Y=y, R^D=1)}.
$$

B.3.4 Assumption $1ZY \oplus 2UD$

Note that

$$
\begin{aligned}
&\mathbb{P}(Y=y, R^Y=1 \mid Z=0, U=c, D=0, R^D=1) \\
&= \frac{\mathbb{P}(U=c, D=0, Y=y, R^D=1, R^Y=1 \mid Z=0)}{\mathbb{P}(U=c, D=0, R^D=1 \mid Z=0)} \\
&= \frac{\mathbb{P}(D=0, Y=y, R^D=1, R^Y=1 \mid Z=0) - \mathbb{P}(U=n, D=0, Y=y, R^D=1, R^Y=1 \mid Z=0)}{\mathbb{P}(D=0, R^D=1 \mid Z=0) - \mathbb{P}(U=n, D=0, R^D=1 \mid Z=0)}.
\end{aligned}
$$

Since

$$
\mathbb{P}(U=n, D=0, Y=y, R^D=1, R^Y=1 \mid Z=0)
$$

$$= \mathbb{P}(U = n, D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 1)\frac{\mathbb{P}(R^Y = 1 \mid Z = 0, Y = y, R^D = 1)}{\mathbb{P}(R^Y = 1 \mid Z = 1, Y = y, R^D = 1)},$$
$$\mathbb{P}(U = n, D = 0, R^D = 1 \mid Z = 0) = \mathbb{P}(U = n, D = 0, R^D = 1 \mid Z = 1),$$

we can identify $\mathbb{P}(Y = y, R^Y = 1 \mid Z = 0, U = c, D = 0, R^D = 1)$, and similarly, $\mathbb{P}(Y = y, R^Y = 1 \mid Z = 1, U = c, D = 1, R^D = 1)$. The identification of $\mathbb{P}(R^Y = 1 \mid Z = z, Y = y, R^D = 1)$ follows the same logic as in Assumption $1ZY \oplus 2Z$, as detailed in Section B.6.5, with a slightly modified conditional dependence condition, $Y \not\perp\!\!\!\perp D \mid (Z = z, R^D = 1)$ for $z = 0, 1$, where Z and D are binary. The identification of $\mathbb{P}(Y = y \mid U = c, D = d)$ follows from

$$\mathbb{P}(Y = y \mid U = c, D = d) = \frac{\mathbb{P}(Y = y, R^Y = 1 \mid Z = z, U = c, D = d, R^D = 1)}{\mathbb{P}(R^Y = 1 \mid Z = z, Y = y, R^D = 1)}.$$

B.3.5 Assumption $1UY \oplus 2UD$

Note that

$$\mathbb{P}(U = c, D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 0)$$
$$= \mathbb{P}(D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 0) - \mathbb{P}(U = n, D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 0),$$
$$\mathbb{P}(U = c, D = 0, R^D = 1 \mid Z = 0)$$
$$= \mathbb{P}(D = 0, R^D = 1 \mid Z = 0) - \mathbb{P}(U = n, D = 0, R^D = 1 \mid Z = 0).$$

Since

$$\mathbb{P}(U = n, D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 0) = \mathbb{P}(U = n, D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 1),$$
$$\mathbb{P}(U = n, D = 0, R^D = 1 \mid Z = 0) = \mathbb{P}(U = n, D = 0, R^D = 1 \mid Z = 1),$$

we can identify $\mathbb{P}(U = c, D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 0)$ and $\mathbb{P}(U = c, D = 0, R^D = 1 \mid Z = 0)$, and similarly, $\mathbb{P}(U = c, D = 1, Y = y, R^D = 1, R^Y = 1 \mid Z = 1)$ and $\mathbb{P}(U = c, D = 1, R^D = 1 \mid Z = 1)$. Since

$$\frac{\mathbb{P}(Y = 1 \mid U = c, D = 0)}{\mathbb{P}(Y = 1 \mid U = c, D = 1)} = \frac{\mathbb{P}(U = c, D = 0, Y = 1, R^D = 1, R^Y = 1 \mid Z = 0)}{\mathbb{P}(U = c, D = 1, Y = 1, R^D = 1, R^Y = 1 \mid Z = 1)} \cdot \frac{\mathbb{P}(U = c, D = 1, R^D = 1 \mid Z = 1)}{\mathbb{P}(U = c, D = 0, R^D = 1 \mid Z = 0)},$$
$$\frac{1 - \mathbb{P}(Y = 1 \mid U = c, D = 0)}{1 - \mathbb{P}(Y = 1 \mid U = c, D = 1)} = \frac{\mathbb{P}(U = c, D = 0, Y = 0, R^D = 1, R^Y = 1 \mid Z = 0)}{\mathbb{P}(U = c, D = 1, Y = 0, R^D = 1, R^Y = 1 \mid Z = 1)} \cdot \frac{\mathbb{P}(U = c, D = 1, R^D = 1 \mid Z = 1)}{\mathbb{P}(U = c, D = 0, R^D = 1 \mid Z = 0)},$$

we can identify the CACE. ■

B.4 Proof of Theorem 4

We prove each case of Theorem 4 separately.

B.4.1 Assumption $1ZD + 2ZD$

Define $\mathbb{P}_{dy11|z} = \mathbb{P}(D = d, Y = y, R^D = 1, R^Y = 1 \mid Z = z)$, $\mathbb{P}_{y+01|z} = \mathbb{P}(Y = y, R^D = 0, R^Y = 1 \mid Z = z)$, $\mathbb{P}_{d1+0|z} = \mathbb{P}(D = d, R^D = 1, R^Y = 0 \mid Z = z)$, $\mathbb{P}_{+0+0|z} = \mathbb{P}(R^D = 0, R^Y = 0 \mid Z = z)$, $\zeta_z(d) = \frac{\mathbb{P}(R^D=0|Z=z,D=d)}{\mathbb{P}(R^D=1|Z=z,D=d)}$. Since

$$\begin{aligned}
\mathbb{P}_{dy11|z} &= \mathbb{P}(D = d, Y = y, R^Y = 1 \mid Z = z)\mathbb{P}(R^D = 1 \mid Z = z, D = d),\\
\mathbb{P}_{d1+0|z} &= \mathbb{P}(D = d, R^Y = 0 \mid Z = z)\mathbb{P}(R^D = 1 \mid Z = z, D = d),
\end{aligned}$$

we have

$$\mathbb{P}_{y+01|z} = \int_{d\in\mathcal{D}} \mathbb{P}(D = d, Y = y, R^D = 0, R^Y = 1 \mid Z = z)\mathrm{d}d = \int_{d\in\mathcal{D}} \mathbb{P}_{dy11|z}\zeta_z(d)\mathrm{d}d$$

for each $y \in \mathcal{Y}$, and

$$\mathbb{P}_{+0+0|z} = \int_{d\in\mathcal{D}} \mathbb{P}(D = d, R^D = 0, R^Y = 0 \mid Z = z)\mathrm{d}d = \int_{d\in\mathcal{D}} \mathbb{P}_{d1+0|z}\zeta_z(d)\mathrm{d}d.$$

The solutions $\zeta_z(d)$ are unique if $\mathbb{P}(D, Y^\dagger, R^D = 1 \mid Z = z)$ is complete in $Y^\dagger$ for all z. For binary Z and D, the solutions $\zeta_z(d)$ are unique if $Y^\dagger \not\perp\!\!\!\perp D \mid Z = 1$ under one-sided noncompliance, and if $Y^\dagger \not\perp\!\!\!\perp D \mid Z = z$ for $z = 0, 1$ under two-sided noncompliance.

We can identify $\mathbb{P}(R^D = 1 \mid Z = z, D = d)$ once $\zeta_z(d)$ is identified. The identification of $\mathbb{P}(R^Y = 1 \mid Z = z, D = d)$ follows from

$$\mathbb{P}(R^Y = 1 \mid Z = z, D = d) = \mathbb{P}(R^Y = 1 \mid Z = z, D = d, R^D = 1).$$

The identification of $\mathbb{P}(D = d, Y = y \mid Z = z)$ follows from

$$\mathbb{P}(D = d, Y = y \mid Z = z) = \frac{\mathbb{P}_{dy11|z}}{\mathbb{P}(R^D = 1 \mid Z = z, D = d)\mathbb{P}(R^Y = 1 \mid Z = z, D = d)}.$$

B.4.2 Assumption $1UD + 2ZD$

The identification of $\mathbb{P}(R^D = 1 \mid Z = z, D = d)$ follows the same logic as in Assumption $1ZD + 2ZD$, as detailed in Section B.4.1. Note that

$$\begin{aligned}
&\mathbb{P}(Y = y \mid U = c, D = 0)\\
&= \frac{\mathbb{P}(U = c, D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 0)}{\mathbb{P}(U = c, D = 0, R^D = 1, R^Y = 1 \mid Z = 0)}\\
&= \frac{\mathbb{P}(D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 0) - \mathbb{P}(U = n, D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 0)}{\mathbb{P}(D = 0, R^D = 1, R^Y = 1 \mid Z = 0) - \mathbb{P}(U = n, D = 0, R^D = 1, R^Y = 1 \mid Z = 0)}.
\end{aligned}$$

Since

$$\begin{aligned}
&\mathbb{P}(U = n, D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 0)\\
&= \mathbb{P}(U = n, D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 1)\frac{\mathbb{P}(R^D = 1 \mid Z = 0, D = 0)}{\mathbb{P}(R^D = 1 \mid Z = 1, D = 0)},
\end{aligned}$$

we can identify $\mathbb{P}(Y = y \mid U = c, D = 0)$, and similarly, $\mathbb{P}(Y = y \mid U = c, D = 1)$.

B.4.3 Assumption $1DY + 2ZD$

The identification of $\mathbb{P}(R^D = 1 \mid Z = z, D = d)$ follows the same logic as in Assumption $1ZD + 2ZD$, as detailed in Section B.4.1. Define $\mathbb{P}_{dy11|z} = \mathbb{P}(D = d, Y = y, R^D = 1, R^Y = 1 \mid Z = z)$, $\mathbb{P}_{d1+0|z} = \mathbb{P}(D = d, R^D = 1, R^Y = 0 \mid Z = z)$, $\eta_d(y) = \frac{\mathbb{P}(R^Y=0|D=d,Y=y)}{\mathbb{P}(R^Y=1|D=d,Y=y)}$. Since

$$\mathbb{P}_{dy11|z} = \mathbb{P}(D = d, Y = y, R^D = 1 \mid Z = z)\mathbb{P}(R^Y = 1 \mid D = d, Y = y),$$

we have

$$\mathbb{P}_{d1+0|z} = \int_{y\in\mathcal{Y}} \mathbb{P}(D = d, Y = y, R^D = 1, R^Y = 0 \mid Z = z)\mathrm{d}y = \int_{y\in\mathcal{Y}} \mathbb{P}_{dy11|z}\eta_d(y)\mathrm{d}y$$

for each $z \in \mathcal{Z}$. The solutions $\eta_d(y)$ are unique if $\mathbb{P}(D = d, Y, R^D = 1, R^Y = 1 \mid Z)$ is complete in Z for all d. For binary Z and D, the solutions $\eta_d(y)$ are unique if Y is binary and $Y \not\perp\!\!\!\perp Z \mid D = d$ for $d = 0, 1$ under two-sided noncompliance.

We can identify $\mathbb{P}(R^Y = 1 \mid D = d, Y = y)$ once $\eta_d(y)$ is identified. The identification of $\mathbb{P}(D = d, Y = y \mid Z = z)$ follows from

$$\mathbb{P}(D = d, Y = y \mid Z = z) = \frac{\mathbb{P}_{dy11|z}}{\mathbb{P}(R^D = 1 \mid Z = z, D = d)\mathbb{P}(R^Y = 1 \mid D = d, Y = y)}.$$

B.4.4 Assumption $1ZY + 2ZD$

Define $\mathbb{P}_{dy11|z} = \mathbb{P}(D = d, Y = y, R^D = 1, R^Y = 1 \mid Z = z)$, $\mathbb{P}_{y+01|z} = \mathbb{P}(Y = y, R^D = 0, R^Y = 1 \mid Z = z)$, $\mathbb{P}_{d1+0|z} = \mathbb{P}(D = d, R^D = 1, R^Y = 0 \mid Z = z)$, $\zeta_z(d) = \frac{\mathbb{P}(R^D=0|Z=z,D=d)}{\mathbb{P}(R^D=1|Z=z,D=d)}$, $\eta_z(y) = \frac{\mathbb{P}(R^Y=0|Z=z,Y=y)}{\mathbb{P}(R^Y=1|Z=z,Y=y)}$. Since

$$\mathbb{P}_{dy11|z} = \mathbb{P}(D = d, Y = y \mid Z = z)\mathbb{P}(R^Y = 1 \mid Z = z, Y = y)\mathbb{P}(R^D = 1 \mid Z = z, D = d),$$

we have

$$\mathbb{P}_{y+01|z} = \int_{d\in\mathcal{D}} \mathbb{P}(D = d, Y = y, R^D = 0, R^Y = 1 \mid Z = z)\mathrm{d}d = \int_{d\in\mathcal{D}} \mathbb{P}_{dy11|z}\zeta_z(d)\mathrm{d}d$$

for each $y \in \mathcal{Y}$, and

$$\mathbb{P}_{d1+0|z} = \int_{y\in\mathcal{Y}} \mathbb{P}(D = d, Y = y, R^D = 1, R^Y = 0 \mid Z = z)\mathrm{d}y = \int_{y\in\mathcal{Y}} \mathbb{P}_{dy11|z}\eta_z(y)\mathrm{d}y$$

for each $d \in \mathcal{D}$. The solutions $\zeta_z(d)$ are unique if $\mathbb{P}(D, Y, R^D = 1, R^Y = 1 \mid Z = z)$ is complete in Y for all z, and the solutions $\eta_z(y)$ are unique if $\mathbb{P}(D, Y, R^D = 1, R^Y = 1 \mid Z = z)$ is complete in D for all z. For binary Z and D, the solutions $\zeta_z(d)$ and $\eta_z(y)$ are unique if Y is binary and $Y \not\perp\!\!\!\perp D \mid Z = z$ for $z = 0, 1$ under two-sided noncompliance.

We can identify $\mathbb{P}(R^D = 1 \mid Z = z, D = d)$ and $\mathbb{P}(R^Y = 1 \mid Z = z, Y = y)$ once $\zeta_z(d)$ and $\eta_z(y)$ are identified. The identification of $\mathbb{P}(D = d, Y = y \mid Z = z)$ follows from

$$\mathbb{P}(D = d, Y = y \mid Z = z) = \frac{\mathbb{P}_{dy11|z}}{\mathbb{P}(R^D = 1 \mid Z = z, D = d)\mathbb{P}(R^Y = 1 \mid Z = z, Y = y)}.$$

B.4.5 Assumption $1UY+2ZD$

The identification of $\mathbb{P}(R^D=1 \mid Z=z, D=d)$ follows the same logic as in Assumption $1ZD+2ZD$, as detailed in Section B.4.1. Note that

$$\begin{aligned}&\mathbb{P}(U=c, D=0, Y=y, R^D=1, R^Y=1 \mid Z=0)\\&=\mathbb{P}(D=0, Y=y, R^D=1, R^Y=1 \mid Z=0)-\mathbb{P}(U=n, D=0, Y=y, R^D=1, R^Y=1 \mid Z=0).\end{aligned}$$

Since

$$\begin{aligned}&\mathbb{P}(U=n, D=0, Y=y, R^D=1, R^Y=1 \mid Z=0)\\&=\mathbb{P}(U=n, D=0, Y=y, R^D=1, R^Y=1 \mid Z=1)\frac{\mathbb{P}(R^D=1 \mid Z=0, D=0)}{\mathbb{P}(R^D=1 \mid Z=1, D=0)},\end{aligned}$$

we can identify $\mathbb{P}(U=c, D=0, Y=y, R^D=1, R^Y=1 \mid Z=0)$, and similarly, $\mathbb{P}(U=c, D=1, Y=y, R^D=1, R^Y=1 \mid Z=1)$. Since

$$\begin{aligned}\frac{\mathbb{P}(Y=1 \mid U=c, D=0)}{\mathbb{P}(Y=1 \mid U=c, D=1)}&=\frac{\mathbb{P}(U=c, D=0, Y=1, R^D=1, R^Y=1 \mid Z=0)}{\mathbb{P}(U=c, D=1, Y=1, R^D=1, R^Y=1 \mid Z=1)}\\&\quad\cdot\frac{\mathbb{P}(R^D=1 \mid Z=1, D=1)}{\mathbb{P}(R^D=1 \mid Z=0, D=0)},\\\frac{1-\mathbb{P}(Y=1 \mid U=c, D=0)}{1-\mathbb{P}(Y=1 \mid U=c, D=1)}&=\frac{\mathbb{P}(U=c, D=0, Y=0, R^D=1, R^Y=1 \mid Z=0)}{\mathbb{P}(U=c, D=1, Y=0, R^D=1, R^Y=1 \mid Z=1)}\\&\quad\cdot\frac{\mathbb{P}(R^D=1 \mid Z=1, D=1)}{\mathbb{P}(R^D=1 \mid Z=0, D=0)},\end{aligned}$$

we can identify the CACE. ∎

B.5 Proof of Theorem 5

We prove each case of Theorem 5 separately.

B.5.1 Assumption $1ZD+2ZU$

The identification of $\mathbb{P}(Y=y \mid U=n, D=0)$ and $\mathbb{P}(Y=y \mid U=c, D=1)$ follows from

$$\begin{aligned}\mathbb{P}(Y=y \mid U=n, D=0)&=\mathbb{P}(Y=y \mid Z=1, U=n, D=0, R^D=1, R^Y=1),\\\mathbb{P}(Y=y \mid U=c, D=1)&=\mathbb{P}(Y=y \mid Z=1, U=c, D=1, R^D=1, R^Y=1).\end{aligned}$$

The identification of $\mathbb{P}(R^Y=1 \mid Z=z, D=d)$ follows from

$$\mathbb{P}(R^Y=1 \mid Z=z, D=d)=\mathbb{P}(R^Y=1 \mid Z=z, D=d, R^D=1).$$

Since

$$\begin{aligned}&\mathbb{P}(Y=y, R^Y=1 \mid Z=1)\\&=\mathbb{P}(U=n, D=0, Y=y, R^Y=1 \mid Z=1)+\mathbb{P}(U=c, D=1, Y=y, R^Y=1 \mid Z=1)\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{P}(Y = y \mid U = n, D = 0)\mathbb{P}(U = n, D = 0 \mid Z = 1)\mathbb{P}(R^Y = 1 \mid Z = 1, D = 0) \\
&\quad + \mathbb{P}(Y = y \mid U = c, D = 1)\{1 - \mathbb{P}(U = n, D = 0 \mid Z = 1)\}\mathbb{P}(R^Y = 1 \mid Z = 1, D = 1),
\end{aligned}
$$

we can identify $\mathbb{P}(U = n, D = 0 \mid Z = 1)$. Since

$$
\begin{aligned}
&\mathbb{P}(Y = y, R^Y = 1 \mid Z = 0) \\
&= \mathbb{P}(U = n, D = 0, Y = y, R^Y = 1 \mid Z = 0) + \mathbb{P}(U = c, D = 0, Y = y, R^Y = 1 \mid Z = 0) \\
&= \mathbb{P}(Y = y \mid U = n, D = 0)\mathbb{P}(U = n, D = 0 \mid Z = 1)\mathbb{P}(R^Y = 1 \mid Z = 0, D = 0) \\
&\quad + \mathbb{P}(Y = y \mid U = c, D = 0)\{1 - \mathbb{P}(U = n, D = 0 \mid Z = 1)\}\mathbb{P}(R^Y = 1 \mid Z = 0, D = 0),
\end{aligned}
$$

we can identify $\mathbb{P}(Y = y \mid U = c, D = 0)$.

B.5.2 Assumption $1UD + 2ZU$

The identification of $\mathbb{P}(Y = y, R^Y = 1 \mid U = n, D = 0)$ and $\mathbb{P}(Y = y, R^Y = 1 \mid U = c, D = 1)$ follows from

$$
\begin{aligned}
\mathbb{P}(Y = y, R^Y = 1 \mid U = n, D = 0) &= \mathbb{P}(Y = y, R^Y = 1 \mid Z = 1, U = n, D = 0, R^D = 1), \\
\mathbb{P}(Y = y, R^Y = 1 \mid U = c, D = 1) &= \mathbb{P}(Y = y, R^Y = 1 \mid Z = 1, U = c, D = 1, R^D = 1).
\end{aligned}
$$

Since

$$
\begin{aligned}
&\mathbb{P}(Y = y, R^Y = 1 \mid Z = 1) \\
&= \mathbb{P}(U = n, D = 0, Y = y, R^Y = 1 \mid Z = 1) + \mathbb{P}(U = c, D = 1, Y = y, R^Y = 1 \mid Z = 1) \\
&= \mathbb{P}(Y = y, R^Y = 1 \mid U = n, D = 0)\mathbb{P}(U = n, D = 0 \mid Z = 1) \\
&\quad + \mathbb{P}(Y = y, R^Y = 1 \mid U = c, D = 1)\{1 - \mathbb{P}(U = n, D = 0 \mid Z = 1)\},
\end{aligned}
$$

we can identify $\mathbb{P}(U = n, D = 0 \mid Z = 1)$. Since

$$
\begin{aligned}
&\mathbb{P}(Y = y, R^Y = 1 \mid Z = 0) \\
&= \mathbb{P}(U = n, D = 0, Y = y, R^Y = 1 \mid Z = 0) + \mathbb{P}(U = c, D = 0, Y = y, R^Y = 1 \mid Z = 0) \\
&= \mathbb{P}(Y = y, R^Y = 1 \mid U = n, D = 0)\mathbb{P}(U = n, D = 0 \mid Z = 1) \\
&\quad + \mathbb{P}(Y = y, R^Y = 1 \mid U = c, D = 0)\{1 - \mathbb{P}(U = n, D = 0 \mid Z = 1)\},
\end{aligned}
$$

we can identify $\mathbb{P}(Y = y, R^Y = 1 \mid U = c, D = 0)$. The identification of $\mathbb{P}(Y = y \mid U = c, D = d)$ follows from

$$
\mathbb{P}(Y = y \mid U = c, D = d) = \frac{\mathbb{P}(Y = y, R^Y = 1 \mid U = c, D = d)}{\mathbb{P}(R^Y = 1 \mid U = c, D = d)}.
$$

B.5.3 Assumption $1UY + 2ZU$

The identification of $\mathbb{P}(Y = y, R^Y = 1 \mid U = c, D = d)$ follows the same logic as in Assumption $1UD + 2ZU$, as detailed in Section B.5.2. Since

$$
\frac{\mathbb{P}(Y = 1 \mid U = c, D = 0)}{\mathbb{P}(Y = 1 \mid U = c, D = 1)} = \frac{\mathbb{P}(Y = 1, R^Y = 1 \mid U = c, D = 0)}{\mathbb{P}(Y = 1, R^Y = 1 \mid U = c, D = 1)},
$$

$$\frac{1-\mathbb{P}(Y=1 \mid U=c, D=0)}{1-\mathbb{P}(Y=1 \mid U=c, D=1)}=\frac{\mathbb{P}(Y=0, R^Y=1 \mid U=c, D=0)}{\mathbb{P}(Y=0, R^Y=1 \mid U=c, D=1)},$$

we can identify the CACE. ■

B.6 Proof of Theorem 6

We prove each case of Theorem 6 separately.

B.6.1 Assumption $1ZD \oplus 2Z$

The identification of $\mathbb{P}(D=d, Y=y \mid Z=z)$ follows from

$$\mathbb{P}(D=d, Y=y \mid Z=z)=\frac{\mathbb{P}(D=d, Y=y, R^D=1, R^Y=1 \mid Z=z)}{\mathbb{P}(R^D=1 \mid Z=z)\mathbb{P}(R^Y=1 \mid Z=z, D=d, R^D=1)}.$$

B.6.2 Assumption $1UD \oplus 2Z$

Note that

$$\begin{aligned}
&\mathbb{P}(Y=y \mid U=c, D=0)\\
&=\frac{\mathbb{P}(U=c, D=0, Y=y, R^D=1, R^Y=1 \mid Z=0)}{\mathbb{P}(U=c, D=0, R^D=1, R^Y=1 \mid Z=0)}\\
&=\frac{\mathbb{P}(D=0, Y=y, R^D=1, R^Y=1 \mid Z=0)-\mathbb{P}(U=n, D=0, Y=y, R^D=1, R^Y=1 \mid Z=0)}{\mathbb{P}(D=0, R^D=1, R^Y=1 \mid Z=0)-\mathbb{P}(U=n, D=0, R^D=1, R^Y=1 \mid Z=0)}.
\end{aligned}$$

Since

$$\begin{aligned}
&\mathbb{P}(U=n, D=0, Y=y, R^D=1, R^Y=1 \mid Z=0)\\
&=\mathbb{P}(U=n, D=0, Y=y, R^D=1, R^Y=1 \mid Z=1)\frac{\mathbb{P}(R^D=1 \mid Z=0)}{\mathbb{P}(R^D=1 \mid Z=1)},
\end{aligned}$$

we can identify $\mathbb{P}(Y=y \mid U=c, D=0)$, and similarly, $\mathbb{P}(Y=y \mid U=c, D=1)$.

B.6.3 Assumption $1UY \oplus 2Z$

Note that

$$\begin{aligned}
&\mathbb{P}(U=c, D=0, Y=y, R^D=1, R^Y=1 \mid Z=0)\\
&=\mathbb{P}(D=0, Y=y, R^D=1, R^Y=1 \mid Z=0)-\mathbb{P}(U=n, D=0, Y=y, R^D=1, R^Y=1 \mid Z=0).
\end{aligned}$$

Since

$$\begin{aligned}
&\mathbb{P}(U=n, D=0, Y=y, R^D=1, R^Y=1 \mid Z=0)\\
&=\mathbb{P}(U=n, D=0, Y=y, R^D=1, R^Y=1 \mid Z=1)\frac{\mathbb{P}(R^D=1 \mid Z=0)}{\mathbb{P}(R^D=1 \mid Z=1)},
\end{aligned}$$

we can identify $\mathbb{P}(U=c, D=0, Y=y, R^D=1, R^Y=1 \mid Z=0)$, and similarly, $\mathbb{P}(U=c, D=1, Y=y, R^D=1, R^Y=1 \mid Z=1)$. Since

$$\frac{\mathbb{P}(Y=1 \mid U=c, D=0)}{\mathbb{P}(Y=1 \mid U=c, D=1)}=\frac{\mathbb{P}(U=c, D=0, Y=1, R^D=1, R^Y=1 \mid Z=0)}{\mathbb{P}(U=c, D=1, Y=1, R^D=1, R^Y=1 \mid Z=1)}\frac{\mathbb{P}(R^D=1 \mid Z=1)}{\mathbb{P}(R^D=1 \mid Z=0)},$$

$$\frac{1-\mathbb{P}(Y=1 \mid U=c, D=0)}{1-\mathbb{P}(Y=1 \mid U=c, D=1)}=\frac{\mathbb{P}(U=c, D=0, Y=0, R^D=1, R^Y=1 \mid Z=0)}{\mathbb{P}(U=c, D=1, Y=0, R^D=1, R^Y=1 \mid Z=1)} \frac{\mathbb{P}(R^D=1 \mid Z=1)}{\mathbb{P}(R^D=1 \mid Z=0)},$$

we can identify the CACE.

B.6.4 Assumption $1DY \oplus 2Z$

Define $\mathbb{P}_{dy11|z}=\mathbb{P}(D=d, Y=y, R^D=1, R^Y=1 \mid Z=z)$, $\mathbb{P}_{d1+0|z}=\mathbb{P}(D=d, R^D=1, R^Y=0 \mid Z=z)$, $\eta_d(y)=\frac{\mathbb{P}(R^Y=0 \mid D=d, Y=y, R^D=1)}{\mathbb{P}(R^Y=1 \mid D=d, Y=y, R^D=1)}$. Since

$$\mathbb{P}_{dy11|z}=\mathbb{P}(D=d, Y=y, R^D=1 \mid Z=z) \mathbb{P}(R^Y=1 \mid D=d, Y=y, R^D=1),$$

we have

$$\mathbb{P}_{d1+0|z}=\int_{y \in \mathcal{Y}} \mathbb{P}(D=d, Y=y, R^D=1, R^Y=0 \mid Z=z) \mathrm{d} y=\int_{y \in \mathcal{Y}} \mathbb{P}_{dy11|z} \eta_d(y) \mathrm{d} y$$

for each $z \in \mathcal{Z}$. The solutions $\eta_d(y)$ are unique if $\mathbb{P}(D=d, Y, R^D=1, R^Y=1 \mid Z)$ is complete in Z for all d. For binary Z and D, the solutions $\eta_d(y)$ are unique if Y is binary and $Y \not\perp\!\!\!\perp Z \mid (D=d, R^D=1)$ for $d=0,1$ under two-sided noncompliance.

We can identify $\mathbb{P}(R^Y=1 \mid D=d, Y=y, R^D=1)$ once $\eta_d(y)$ is identified. The identification of $\mathbb{P}(D=d, Y=y \mid Z=z)$ follows from

$$\mathbb{P}(D=d, Y=y \mid Z=z)=\frac{\mathbb{P}_{dy11|z}}{\mathbb{P}(R^D=1 \mid Z=z) \mathbb{P}(R^Y=1 \mid D=d, Y=y, R^D=1)}.$$

B.6.5 Assumption $1ZY \oplus 2Z$

Define $\mathbb{P}_{dy11|z}=\mathbb{P}(D=d, Y=y, R^D=1, R^Y=1 \mid Z=z)$, $\mathbb{P}_{d1+0|z}=\mathbb{P}(D=d, R^D=1, R^Y=0 \mid Z=z)$, $\eta_z(y)=\frac{\mathbb{P}(R^Y=0 \mid Z=z, Y=y, R^D=1)}{\mathbb{P}(R^Y=1 \mid Z=z, Y=y, R^D=1)}$. Since

$$\mathbb{P}_{dy11|z}=\mathbb{P}(D=d, Y=y, R^D=1 \mid Z=z) \mathbb{P}(R^Y=1 \mid Z=z, Y=y, R^D=1),$$

we have

$$\mathbb{P}_{d1+0|z}=\int_{y \in \mathcal{Y}} \mathbb{P}(D=d, Y=y, R^D=1, R^Y=0 \mid Z=z) \mathrm{d} y=\int_{y \in \mathcal{Y}} \mathbb{P}_{dy11|z} \eta_z(y) \mathrm{d} y$$

for each $d \in \mathcal{D}$. The solutions $\eta_z(y)$ are unique if $\mathbb{P}(D, Y, R^D=1, R^Y=1 \mid Z=z)$ is complete in D for all z. For binary Z and D, the solutions $\eta_z(y)$ are unique if Y is binary and $Y \not\perp\!\!\!\perp D \mid Z=z$ for $z=0,1$ under two-sided noncompliance.

We can identify $\mathbb{P}(R^Y=1 \mid Z=z, Y=y, R^D=1)$ once $\eta_z(y)$ is identified. The identification of $\mathbb{P}(D=d, Y=y \mid Z=z)$ follows from

$$\mathbb{P}(D=d, Y=y \mid Z=z)=\frac{\mathbb{P}_{dy11|z}}{\mathbb{P}(R^D=1 \mid Z=z) \mathbb{P}(R^Y=1 \mid Z=z, Y=y, R^D=1)}.$$

■

B.7 Proof of Theorem 7

We prove each case of Theorem 7 separately.

B.7.1 Assumption $1Z \oplus 2ZD$

Define $\mathbb{P}_{dy11|z} = \mathbb{P}(D = d, Y = y, R^D = 1, R^Y = 1 \mid Z = z)$, $\mathbb{P}_{y+01|z} = \mathbb{P}(Y = y, R^D = 0, R^Y = 1 \mid Z = z)$, $\zeta_z(d) = \frac{\mathbb{P}(R^D=0|Z=z,D=d)}{\mathbb{P}(R^D=1|Z=z,D=d)}$. Since

$$\mathbb{P}_{dy11|z} = \mathbb{P}(D = d, Y = y \mid Z = z)\mathbb{P}(R^D = 1 \mid Z = z, D = d)\mathbb{P}(R^Y = 1 \mid Z = z, R^D = 1),$$

we have

$$\begin{aligned}\mathbb{P}_{y+01|z} &= \int_{d\in\mathcal{D}} \mathbb{P}(D = d, Y = y, R^D = 0, R^Y = 1 \mid Z = z)\mathrm{d}d \\ &= \int_{d\in\mathcal{D}} \mathbb{P}_{dy11|z}\zeta_z(d)\frac{\mathbb{P}(R^Y = 1 \mid Z = z, R^D = 0)}{\mathbb{P}(R^Y = 1 \mid Z = z, R^D = 1)}\mathrm{d}d\end{aligned}$$

for each $y \in \mathcal{Y}$. The solutions $\zeta_z(d)$ are unique if $\mathbb{P}(D, Y, R^D = 1, R^Y = 1 \mid Z = z)$ is complete in Y for all z. For binary Z and D, the solutions $\zeta_z(d)$ are unique if $Y \not\perp\!\!\!\perp D \mid Z = 1$ under one-sided noncompliance, and if $Y \not\perp\!\!\!\perp D \mid Z = z$ for $z = 0, 1$ under two-sided noncompliance.

We can identify $\mathbb{P}(R^D = 1 \mid Z = z, D = d)$ once $\zeta_z(d)$ is identified. The identification of $\mathbb{P}(D = d, Y = y \mid Z = z)$ follows from

$$\mathbb{P}(D = d, Y = y \mid Z = z) = \frac{\mathbb{P}_{dy11|z}}{\mathbb{P}(R^D = 1 \mid Z = z, D = d)\mathbb{P}(R^Y = 1 \mid Z = z, R^D = 1)}.$$

B.7.2 Assumption $1D \oplus 2ZD$

Define $\mathbb{P}_{dy11|z} = \mathbb{P}(D = d, Y = y, R^D = 1, R^Y = 1 \mid Z = z)$, $\mathbb{P}_{y+01|z} = \mathbb{P}(Y = y, R^D = 0, R^Y = 1 \mid Z = z)$, $\zeta_z(d) = \frac{\mathbb{P}(R^D=0|Z=z,D=d)}{\mathbb{P}(R^D=1|Z=z,D=d)}$. Since

$$\mathbb{P}_{dy11|z} = \mathbb{P}(D = d, Y = y \mid Z = z)\mathbb{P}(R^D = 1 \mid Z = z, D = d)\mathbb{P}(R^Y = 1 \mid D = d, R^D = 1),$$

we have

$$\begin{aligned}\mathbb{P}_{y+01|z} &= \int_{d\in\mathcal{D}} \mathbb{P}(D = d, Y = y, R^D = 0, R^Y = 1 \mid Z = z)\mathrm{d}d \\ &= \int_{d\in\mathcal{D}} \mathbb{P}_{dy11|z}\zeta_z(d)\frac{\mathbb{P}(R^Y = 1 \mid D = d, R^D = 0)}{\mathbb{P}(R^Y = 1 \mid D = d, R^D = 1)}\mathrm{d}d\end{aligned}$$

for each $y \in \mathcal{Y}$, and

$$\mathbb{P}(R^D = 0 \mid Z = z) = \int_{d\in\mathcal{D}} \frac{\{\zeta_z(d)\mathbb{P}(R^Y = 1 \mid D = d, R^D = 0)\}\mathbb{P}(D = d, R^D = 1 \mid Z = z)}{\mathbb{P}(R^Y = 1 \mid D = d, R^D = 0)}\mathrm{d}d$$

for each $z \in \mathcal{Z}$. The solutions $\{\zeta_z(d)\mathbb{P}(R^Y = 1 \mid D = d, R^D = 0)\}$ are unique if $\mathbb{P}(D, Y, R^D = 1, R^Y = 1 \mid Z = z)$ is complete in Y for all z, and the solutions $\mathbb{P}(R^Y = 1 \mid D = d, R^D = 0)$

are unique if $\mathbb{P}(D, R^D = 1 \mid Z)$ is complete in Z. For binary Z and D, the solutions $\{\zeta_z(d)\mathbb{P}(R^Y = 1 \mid D = d, R^D = 0)\}$ are unique if $Y \not\perp\!\!\!\perp D \mid Z = 1$ under one-sided noncompliance, and if $Y \not\perp\!\!\!\perp D \mid Z = z$ for $z = 0, 1$ under two-sided noncompliance. The solutions $\mathbb{P}(R^Y = 1 \mid D = d, R^D = 0)$ are unique if $D \not\perp\!\!\!\perp Z$.

We can identify $\mathbb{P}(R^D = 1 \mid Z = z, D = d)$ once $\mathbb{P}(R^Y = 1 \mid D = d, R^D = 0)$ is identified. The identification of $\mathbb{P}(D = d, Y = y \mid Z = z)$ follows from

$$\mathbb{P}(D = d, Y = y \mid Z = z) = \frac{\mathbb{P}_{dy11|z}}{\mathbb{P}(R^D = 1 \mid Z = z, D = d)\mathbb{P}(R^Y = 1 \mid D = d, R^D = 1)}.$$

B.7.3 Assumption $1Y \oplus 2ZD$

Define $\mathbb{P}_{dy11|z} = \mathbb{P}(D = d, Y = y, R^D = 1, R^Y = 1 \mid Z = z)$, $\mathbb{P}_{y+01|z} = \mathbb{P}(Y = y, R^D = 0, R^Y = 1 \mid Z = z)$, $\mathbb{P}_{d1+0|z} = \mathbb{P}(D = d, R^D = 1, R^Y = 0 \mid Z = z)$, $\mathbb{P}_{+0+0|z} = \mathbb{P}(R^D = 0, R^Y = 0 \mid Z = z)$, $\eta(y) = \frac{\mathbb{P}(R^Y=0|Y=y,R^D=1)}{\mathbb{P}(R^Y=1|Y=y,R^D=1)}$, $\xi(y) = \frac{\mathbb{P}(R^Y=0|Y=y,R^D=0)}{\mathbb{P}(R^Y=1|Y=y,R^D=0)}$, $\zeta_z(d) = \frac{\mathbb{P}(R^D=0|Z=z,D=d)}{\mathbb{P}(R^D=1|Z=z,D=d)}$. Since

$$\begin{aligned}
\mathbb{P}_{dy11|z} &= \mathbb{P}(D = d, Y = y \mid Z = z)\mathbb{P}(R^D = 1 \mid Z = z, D = d)\mathbb{P}(R^Y = 1 \mid Y = y, R^D = 1), \\
\mathbb{P}_{y+01|z} &= \mathbb{P}(Y = y, R^D = 0 \mid Z = z)\mathbb{P}(R^Y = 1 \mid Y = y, R^D = 0),
\end{aligned}$$

we have

$$\mathbb{P}_{d1+0|z} = \int_{y\in\mathcal{Y}} \mathbb{P}(D = d, Y = y, R^D = 1, R^Y = 0 \mid Z = z)\mathrm{d}y = \int_{y\in\mathcal{Y}} \mathbb{P}_{dy11|z}\eta(y)\mathrm{d}y$$

for each $(z, d) \in (\mathcal{Z}, \mathcal{D})$,

$$\mathbb{P}_{+0+0|z} = \int_{y\in\mathcal{Y}} \mathbb{P}(Y = y, R^D = 0, R^Y = 0 \mid Z = z)\mathrm{d}y = \int_{y\in\mathcal{Y}} \mathbb{P}_{y+01|z}\xi(y)\mathrm{d}y$$

for each $z \in \mathcal{Z}$, and

$$\begin{aligned}
\mathbb{P}_{y+01|z} &= \int_{d\in\mathcal{D}} \mathbb{P}(D = d, Y = y, R^D = 0, R^Y = 1 \mid Z = z)\mathrm{d}d \\
&= \frac{\mathbb{P}(R^Y = 1 \mid Y = y, R^D = 0)}{\mathbb{P}(R^Y = 1 \mid Y = y, R^D = 1)} \int_{d\in\mathcal{D}} \mathbb{P}_{dy11|z}\zeta_z(d)\mathrm{d}d
\end{aligned}$$

for each $y \in \mathcal{Y}$. The solutions $\eta(y)$ are unique if $\mathbb{P}(D, Y, R^D = 1, R^Y = 1 \mid Z)$ is complete in (Z, D), the solutions $\xi(y)$ are unique if $\mathbb{P}(Y, R^D = 0, R^Y = 1 \mid Z)$ is complete in Z, and the solutions $\zeta_z(d)$ are unique if $\mathbb{P}(D, Y, R^D = 1, R^Y = 1 \mid Z = z)$ is complete in Y for all z. Now take Z and D to be binary. The solutions $\eta(y)$ are unique if $Y \not\perp\!\!\!\perp (Z, D) \mid (R^D = 1)$, and the solutions $\xi(y)$ are unique if Y is binary and $Y \not\perp\!\!\!\perp Z \mid (R^D = 0)$. The solutions $\zeta_z(d)$ are unique if $Y \not\perp\!\!\!\perp D \mid Z = 1$ under one-sided noncompliance, and if $Y \not\perp\!\!\!\perp D \mid Z = z$ for $z = 0, 1$ under two-sided noncompliance.

We can identify $\mathbb{P}(R^Y = 1 \mid Y = y, R^D = 1)$, $\mathbb{P}(R^Y = 1 \mid Y = y, R^D = 0)$, and $\mathbb{P}(R^D = 1 \mid Z = z, D = d)$ once $\eta(y)$, $\xi(y)$, and $\zeta_z(d)$ are identified. The identification of $\mathbb{P}(D = d, Y = y \mid Z = z)$ follows from

$$\mathbb{P}(D = d, Y = y \mid Z = z) = \frac{\mathbb{P}_{dy11|z}}{\mathbb{P}(R^D = 1 \mid Z = z, D = d)\mathbb{P}(R^Y = 1 \mid Y = y, R^D = 1)}.$$

B.7.4 Assumption $1Z \oplus 2ZU$

The identification of $\mathbb{P}(Y = y \mid U = n, D = 0)$ and $\mathbb{P}(Y = y \mid U = c, D = 1)$ follows from

$$\mathbb{P}(Y = y \mid U = n, D = 0) = \mathbb{P}(Y = y \mid Z = 1, U = n, D = 0, R^D = 1, R^Y = 1),$$
$$\mathbb{P}(Y = y \mid U = c, D = 1) = \mathbb{P}(Y = y \mid Z = 1, U = c, D = 1, R^D = 1, R^Y = 1).$$

The identification of $\mathbb{P}(Y = y \mid Z = z)$ follows from

$$\mathbb{P}(Y = y \mid Z = z) = \frac{\mathbb{P}(Y = y, R^D = 0, R^Y = 1 \mid Z = z)}{\mathbb{P}(R^Y = 1 \mid Z = z, R^D = 0)} + \frac{\mathbb{P}(Y = y, R^D = 1, R^Y = 1 \mid Z = z)}{\mathbb{P}(R^Y = 1 \mid Z = z, R^D = 1)}.$$

The identification of $\mathbb{P}(Y = y \mid U = c, D = 0)$ follows the same logic as in Assumption $2ZU$, as detailed in Section B.2.6.

B.7.5 Assumption $1U \oplus 2ZU$

The identification of $\mathbb{P}(Y = y \mid U = c, D = 1)$ follows from

$$\mathbb{P}(Y = y \mid U = c, D = 1) = \mathbb{P}(Y = y \mid Z = 1, U = c, D = 1, R^D = 1, R^Y = 1).$$

Define $\mathbb{P}_{ud1r^Y|z} = \mathbb{P}(U = u, D = d, R^D = 1, R^Y = r^Y \mid Z = z)$, $\mathbb{P}_{01r^Y|0} = \mathbb{P}(D = 0, R^D = 1, R^Y = r^Y \mid Z = 0)$, $\zeta(u) = \frac{\mathbb{P}(R^D=1|Z=0,U=u)}{\mathbb{P}(R^D=1|Z=1,U=u)}$. We have

$$\mathbb{P}_{01r^Y|0} = \mathbb{P}_{n01r^Y|0} + \mathbb{P}_{c01r^Y|0} = \mathbb{P}_{n01r^Y|1}\zeta(n) + \mathbb{P}_{c11r^Y|1}\zeta(c)$$

for $r^Y = 0, 1$. The solutions $\zeta(u)$ are unique if $R^Y \not\perp\!\!\!\perp U \mid (R^D = 1)$. Note that

$$\mathbb{P}(U = c, D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 0)$$
$$= \mathbb{P}(D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 0) - \mathbb{P}(U = n, D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 0).$$

Since

$$\mathbb{P}(U = n, D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 0)$$
$$= \mathbb{P}(U = n, D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 1)\zeta(n),$$

we can identify $\mathbb{P}(U = c, D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 0)$. The identification of $\mathbb{P}(Y = y \mid U = c, D = 0)$ follows from

$$\mathbb{P}(Y = y \mid U = c, D = 0) = \frac{\mathbb{P}(U = c, D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 0)}{\mathbb{P}(U = c, D = 0, R^D = 1, R^Y = 1 \mid Z = 0)}.$$

B.7.6 Assumption $1D \oplus 2ZU$

The identification of $\mathbb{P}(Y = y \mid U = n, D = 0)$ and $\mathbb{P}(Y = y \mid U = c, D = 1)$ follows from

$$\mathbb{P}(Y = y \mid U = n, D = 0) = \mathbb{P}(Y = y \mid Z = 1, U = n, D = 0, R^D = 1, R^Y = 1),$$
$$\mathbb{P}(Y = y \mid U = c, D = 1) = \mathbb{P}(Y = y \mid Z = 1, U = c, D = 1, R^D = 1, R^Y = 1).$$

Define $\mathbb{P}_{udy11|1} = \mathbb{P}(U = u, D = d, Y = y, R^D = 1, R^Y = 1 \mid Z = 1)$, $\mathbb{P}_{y+01|1} = \mathbb{P}(Y = y, R^D = 0, R^Y = 1 \mid Z = 1)$, $\eta(u) = \frac{\mathbb{P}(R^D=0|Z=1,U=u)}{\mathbb{P}(R^D=1|Z=1,U=u)}$. We have

$$\mathbb{P}_{y+01|1} = \mathbb{P}_{n0y11|1}\eta(n)\frac{\mathbb{P}(R^Y = 1 \mid D = 0, R^D = 0)}{\mathbb{P}(R^Y = 1 \mid D = 0, R^D = 1)} + \mathbb{P}_{c1y11|1}\eta(c)\frac{\mathbb{P}(R^Y = 1 \mid D = 1, R^D = 0)}{\mathbb{P}(R^Y = 1 \mid D = 1, R^D = 1)}$$

for $y = 0, 1$. The solutions $\{\eta(u)\mathbb{P}(R^Y = 1 \mid D = d, R^D = 0)\}$ are unique if $Y \not\perp\!\!\!\perp U \mid (Z = 1)$. Since

$$\mathbb{P}(R^Y = 1 \mid D = 0, R^D = 0) = \frac{\mathbb{P}(D = 0, R^D = 0, R^Y = 1 \mid Z = 0)}{\mathbb{P}(D = 0, R^D = 0 \mid Z = 0)},$$

we can identify $\mathbb{P}(R^D = 1 \mid Z = 1, U = n)$. The identification of $\mathbb{P}(U = n, D = 0 \mid Z = 1)$ follows from

$$\mathbb{P}(U = n, D = 0 \mid Z = 1) = \frac{\mathbb{P}(U = n, D = 0, R^D = 1 \mid Z = 1)}{\mathbb{P}(R^D = 1 \mid Z = 1, U = n)}.$$

The identification of $\mathbb{P}(Y = y \mid Z = 0)$ follows from

$$\begin{aligned}
&\mathbb{P}(Y = y \mid Z = 0)\\
&= \mathbb{P}(D = 0, Y = y, R^D = 1 \mid Z = 0) + \mathbb{P}(D = 0, Y = y, R^D = 0 \mid Z = 0)\\
&= \frac{\mathbb{P}(D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 0)}{\mathbb{P}(R^Y = 1 \mid D = 0, R^D = 1)} + \frac{\mathbb{P}(D = 0, Y = y, R^D = 0, R^Y = 1 \mid Z = 0)}{\mathbb{P}(R^Y = 1 \mid D = 0, R^D = 0)}.
\end{aligned}$$

Since

$$\begin{aligned}
\mathbb{P}(Y = y \mid Z = 0) &= \mathbb{P}(U = n, D = 0, Y = y \mid Z = 0) + \mathbb{P}(U = c, D = 0, Y = y \mid Z = 0)\\
&= \mathbb{P}(Y = y \mid U = n, D = 0)\mathbb{P}(U = n, D = 0 \mid Z = 1)\\
&\quad + \mathbb{P}(Y = y \mid U = c, D = 0)\{1 - \mathbb{P}(U = n, D = 0 \mid Z = 1)\},
\end{aligned}$$

we can identify $\mathbb{P}(Y = y \mid U = c, D = 0)$.

B.7.7 Assumption $1Y \oplus 2ZU$

The identification of $\mathbb{P}(R^Y = 1 \mid Y = y, R^D = r^D)$ follows the same logic as in Assumption $1Y \oplus 2ZD$, as detailed in Section B.7.3. The identification of $\mathbb{P}(Y = y \mid U = n, D = 0)$ and $\mathbb{P}(Y = y \mid U = c, D = 1)$ follows from

$$\begin{aligned}
\mathbb{P}(Y = y \mid U = n, D = 0) &= \frac{\mathbb{P}(U = n, D = 0, Y = y, R^D = 1, R^Y = 1 \mid Z = 1)}{\mathbb{P}(U = n, D = 0, R^D = 1 \mid Z = 1)\mathbb{P}(R^Y = 1 \mid Y = y, R^D = 1)},\\
\mathbb{P}(Y = y \mid U = c, D = 1) &= \frac{\mathbb{P}(U = c, D = 1, Y = y, R^D = 1, R^Y = 1 \mid Z = 1)}{\mathbb{P}(U = c, D = 1, R^D = 1 \mid Z = 1)\mathbb{P}(R^Y = 1 \mid Y = y, R^D = 1)}.
\end{aligned}$$

The identification of $\mathbb{P}(Y = y \mid Z = z)$ follows from

$$\mathbb{P}(Y = y \mid Z = z) = \frac{\mathbb{P}(Y = y, R^D = 0, R^Y = 1 \mid Z = z)}{\mathbb{P}(R^Y = 1 \mid Y = y, R^D = 0)} + \frac{\mathbb{P}(Y = y, R^D = 1, R^Y = 1 \mid Z = z)}{\mathbb{P}(R^Y = 1 \mid Y = y, R^D = 1)}.$$

The identification of $\mathbb{P}(Y = y \mid U = c, D = 0)$ follows the same logic as in Assumption $2ZU$, as detailed in Section B.2.6. ■

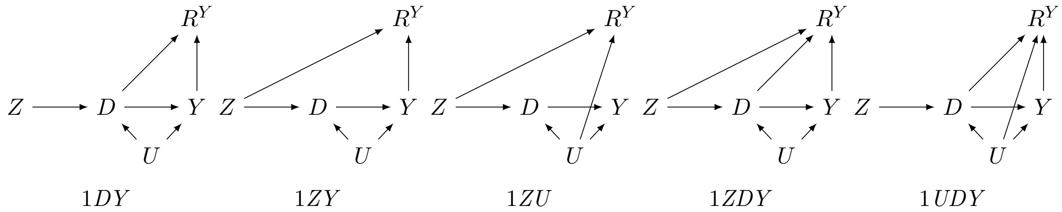


Figure 8: The DAGs for the mechanisms for missingness in the outcome that do not identify the CACE. Section C.1 gives a counterexample for each.

Appendix C. Counterexamples

This appendix collects the counterexamples referenced in Sections 4, 5, and 6. In each case, we exhibit two parameter configurations that induce the same observable data probabilities but different values of the CACE. The CACE is therefore not identifiable under the stated mechanism.

C.1 Counterexamples for the Missing Outcome Models

Figure 8 shows the DAGs. Assumptions $1DY$ and $1ZY$ identify the CACE under two-sided noncompliance, so we give counterexamples under one-sided noncompliance. Assumption $1ZDY$ contains both $1DY$ and $1ZY$, which already fail under one-sided noncompliance, so we give a counterexample under two-sided noncompliance. Assumptions $1ZU$ and $1UDY$ fail under both one-sided and two-sided noncompliance. For these, we give counterexamples under one-sided noncompliance, which is the special case of two-sided noncompliance with no always-takers. Define

$$\begin{aligned}\mathbb{P}_{dy1|z} &= \mathbb{P}(D=d, Y=y, R^Y=1 \mid Z=z),\\ \mathbb{P}_{d+0|z} &= \mathbb{P}(D=d, R^Y=0 \mid Z=z).\end{aligned}$$

C.1.1 Assumption $1DY$

For a binary Y with one-sided noncompliance, we consider the following observable data probabilities:

$$(\mathbb{P}_{011|0}, \mathbb{P}_{001|0}, \mathbb{P}_{011|1}, \mathbb{P}_{001|1}, \mathbb{P}_{111|1}, \mathbb{P}_{101|1}, \mathbb{P}_{0+0|0}, \mathbb{P}_{0+0|1}, \mathbb{P}_{1+0|1}) = \left(\frac{1}{4}, \frac{1}{6}, \frac{1}{4}, \frac{1}{12}, \frac{1}{48}, \frac{1}{32}, \frac{7}{12}, \frac{5}{12}, \frac{19}{96}\right).$$

Define the parameters:

$$\begin{aligned}\Theta_{dy|z} &= \mathbb{P}(D=d, Y=y \mid Z=z) \text{ for } (z,d,y) = (0,0,1), (1,0,1), (1,0,0), (1,1,0),\\ \Theta_{R^Y|dy} &= \mathbb{P}(R^Y=1 \mid D=d, Y=y) \text{ for } d=0,1 \text{ and } y=0,1.\end{aligned}$$

The observable data probabilities relate to the parameters as follows:

$$\mathbb{P}_{011|0} = \Theta_{01|0}\Theta_{R^Y|01},$$

$$
\begin{aligned}
\mathbb{P}_{001|0} &= (1-\Theta_{01|0})\Theta_{R^Y|00},\\
\mathbb{P}_{011|1} &= \Theta_{01|1}\Theta_{R^Y|01},\\
\mathbb{P}_{001|1} &= \Theta_{00|1}\Theta_{R^Y|00},\\
\mathbb{P}_{111|1} &= (1-\Theta_{01|1}-\Theta_{00|1}-\Theta_{10|1})\Theta_{R^Y|11},\\
\mathbb{P}_{101|1} &= \Theta_{10|1}\Theta_{R^Y|10},\\
\mathbb{P}_{0+0|0} &= \Theta_{01|0}(1-\Theta_{R^Y|01}) + (1-\Theta_{01|0})(1-\Theta_{R^Y|00}),\\
\mathbb{P}_{0+0|1} &= \Theta_{01|1}(1-\Theta_{R^Y|01}) + \Theta_{00|1}(1-\Theta_{R^Y|00}),\\
\mathbb{P}_{1+0|1} &= \Theta_{10|1}(1-\Theta_{R^Y|10}) + (1-\Theta_{01|1}-\Theta_{00|1}-\Theta_{10|1})(1-\Theta_{R^Y|11}).
\end{aligned}
$$

These observable data probabilities are consistent with $\Theta_{01|0}=\frac{1}{2}$, $\Theta_{00|1}=\frac{1}{4}$, $\Theta_{01|1}=\frac{1}{2}$, $\Theta_{10|1}=\frac{1}{8}$, $\Theta_{R^Y|00}=\frac{1}{3}$, $\Theta_{R^Y|01}=\frac{1}{2}$, $\Theta_{R^Y|10}=\frac{1}{4}$, $\Theta_{R^Y|11}=\frac{1}{6}$, and the CACE $=\frac{1}{2}$. They are also consistent with $\Theta_{01|0}=\frac{1}{2}$, $\Theta_{00|1}=\frac{1}{4}$, $\Theta_{01|1}=\frac{1}{2}$, $\Theta_{10|1}=\frac{1}{12}$, $\Theta_{R^Y|00}=\frac{1}{3}$, $\Theta_{R^Y|01}=\frac{1}{2}$, $\Theta_{R^Y|10}=\frac{3}{8}$, $\Theta_{R^Y|11}=\frac{1}{8}$, and the CACE $=\frac{2}{3}$. Therefore, the CACE cannot be uniquely identified.

C.1.2 Assumption $1ZY$

For a binary Y with one-sided noncompliance, we consider the following observable data probabilities:

$$
(\mathbb{P}_{011|0},\mathbb{P}_{001|0},\mathbb{P}_{011|1},\mathbb{P}_{001|1},\mathbb{P}_{111|1},\mathbb{P}_{101|1},\mathbb{P}_{0+0|0},\mathbb{P}_{0+0|1},\mathbb{P}_{1+0|1}) = \left(\frac{1}{4},\frac{1}{6},\frac{1}{12},\frac{1}{16},\frac{1}{48},\frac{1}{32},\frac{7}{12},\frac{29}{48},\frac{19}{96}\right).
$$

Define the parameters:

$$
\begin{aligned}
\Theta_{dy|z} &= \mathbb{P}(D=d, Y=y \mid Z=z) \text{ for } (z,d,y) = (0,0,1),(1,0,1),(1,0,0),(1,1,0),\\
\Theta_{R^Y|zy} &= \mathbb{P}(R^Y=1 \mid Z=z, Y=y) \text{ for } z=0,1 \text{ and } y=0,1.
\end{aligned}
$$

The observable data probabilities relate to the parameters as follows:

$$
\begin{aligned}
\mathbb{P}_{011|0} &= \Theta_{01|0}\Theta_{R^Y|01},\\
\mathbb{P}_{001|0} &= (1-\Theta_{01|0})\Theta_{R^Y|00},\\
\mathbb{P}_{011|1} &= \Theta_{01|1}\Theta_{R^Y|11},\\
\mathbb{P}_{001|1} &= \Theta_{00|1}\Theta_{R^Y|10},\\
\mathbb{P}_{111|1} &= (1-\Theta_{01|1}-\Theta_{00|1}-\Theta_{10|1})\Theta_{R^Y|11},\\
\mathbb{P}_{101|1} &= \Theta_{10|1}\Theta_{R^Y|10},\\
\mathbb{P}_{0+0|0} &= \Theta_{01|0}(1-\Theta_{R^Y|01}) + (1-\Theta_{01|0})(1-\Theta_{R^Y|00}),\\
\mathbb{P}_{0+0|1} &= \Theta_{01|1}(1-\Theta_{R^Y|11}) + \Theta_{00|1}(1-\Theta_{R^Y|10}),\\
\mathbb{P}_{1+0|1} &= \Theta_{10|1}(1-\Theta_{R^Y|10}) + (1-\Theta_{01|1}-\Theta_{00|1}-\Theta_{10|1})(1-\Theta_{R^Y|11}).
\end{aligned}
$$

These observable data probabilities are consistent with $\Theta_{01|0}=\frac{1}{2}$, $\Theta_{00|1}=\frac{1}{4}$, $\Theta_{01|1}=\frac{1}{2}$, $\Theta_{10|1}=\frac{1}{8}$, $\Theta_{R^Y|00}=\frac{1}{3}$, $\Theta_{R^Y|01}=\frac{1}{2}$, $\Theta_{R^Y|10}=\frac{1}{4}$, $\Theta_{R^Y|11}=\frac{1}{6}$, and the CACE $=\frac{1}{2}$. They are also consistent with $\Theta_{01|0} = \frac{2}{3}$, $\Theta_{00|1} = \frac{1}{4}$, $\Theta_{01|1} = \frac{1}{2}$, $\Theta_{10|1} = \frac{1}{8}$, $\Theta_{R^Y|00} = \frac{1}{2}$, $\Theta_{R^Y|01} = \frac{3}{8}$, $\Theta_{R^Y|10} = \frac{1}{4}$, $\Theta_{R^Y|11} = \frac{1}{6}$, and the CACE $= -\frac{1}{6}$. Therefore, the CACE cannot be uniquely identified.

C.1.3 Assumption 1ZU

For a binary Y with one-sided noncompliance, we consider the following observable data probabilities:

$$(\mathbb{P}_{011|0}, \mathbb{P}_{001|0}, \mathbb{P}_{011|1}, \mathbb{P}_{001|1}, \mathbb{P}_{111|1}, \mathbb{P}_{101|1}, \mathbb{P}_{0+0|0}, \mathbb{P}_{0+0|1}, \mathbb{P}_{1+0|1}) = \left(\frac{1}{8}, \frac{3}{16}, \frac{1}{16}, \frac{1}{16}, \frac{3}{64}, \frac{9}{64}, \frac{11}{16}, \frac{1}{8}, \frac{9}{16}\right).$$

Define the parameters:

$$\begin{aligned}
\Theta_n &= \mathbb{P}(U = n), \\
\Theta_{Y|ud} &= \mathbb{P}(Y = 1 \mid U = u, D = d) \text{ for } (u, d) = (n, 0), (c, 1), (c, 0), \\
\Theta_{R^Y|zu} &= \mathbb{P}(R^Y = 1 \mid Z = z, U = u) \text{ for } z = 0, 1 \text{ and } u = n, c.
\end{aligned}$$

The observable data probabilities relate to the parameters as follows:

$$\begin{aligned}
\mathbb{P}_{011|0} &= \Theta_{Y|n0}\Theta_{R^Y|0n}\Theta_n + \Theta_{Y|c0}\Theta_{R^Y|0c}(1 - \Theta_n), \\
\mathbb{P}_{001|0} &= (1 - \Theta_{Y|n0})\Theta_{R^Y|0n}\Theta_n + (1 - \Theta_{Y|c0})\Theta_{R^Y|0c}(1 - \Theta_n), \\
\mathbb{P}_{011|1} &= \Theta_{Y|n0}\Theta_{R^Y|1n}\Theta_n, \\
\mathbb{P}_{001|1} &= (1 - \Theta_{Y|n0})\Theta_{R^Y|1n}\Theta_n, \\
\mathbb{P}_{111|1} &= \Theta_{Y|c1}\Theta_{R^Y|1c}(1 - \Theta_n), \\
\mathbb{P}_{101|1} &= (1 - \Theta_{Y|c1})\Theta_{R^Y|1c}(1 - \Theta_n), \\
\mathbb{P}_{0+0|0} &= (1 - \Theta_{R^Y|0n})\Theta_n + (1 - \Theta_{R^Y|0c})(1 - \Theta_n), \\
\mathbb{P}_{0+0|1} &= (1 - \Theta_{R^Y|1n})\Theta_n, \\
\mathbb{P}_{1+0|1} &= (1 - \Theta_{R^Y|1c})(1 - \Theta_n).
\end{aligned}$$

These observable data probabilities are consistent with $\Theta_n = \frac{1}{4}$, $\Theta_{Y|n0} = \frac{1}{2}$, $\Theta_{Y|c0} = \frac{1}{3}$, $\Theta_{Y|c1} = \frac{1}{4}$, $\Theta_{R^Y|0n} = \frac{1}{2}$, $\Theta_{R^Y|1n} = \frac{1}{2}$, $\Theta_{R^Y|0c} = \frac{1}{4}$, $\Theta_{R^Y|1c} = \frac{1}{4}$, and the CACE $= -\frac{1}{12}$. They are also consistent with $\Theta_n = \frac{1}{4}$, $\Theta_{Y|n0} = \frac{1}{2}$, $\Theta_{Y|c0} = \frac{3}{8}$, $\Theta_{Y|c1} = \frac{1}{4}$, $\Theta_{R^Y|0n} = \frac{1}{4}$, $\Theta_{R^Y|1n} = \frac{1}{2}$, $\Theta_{R^Y|0c} = \frac{1}{3}$, $\Theta_{R^Y|1c} = \frac{1}{4}$, and the CACE $= -\frac{1}{8}$. Therefore, the CACE cannot be uniquely identified.

C.1.4 Assumption 1ZDY

For a binary Y with two-sided noncompliance, we consider the following observable data probabilities:

$$\begin{aligned}
&(\mathbb{P}_{011|0}, \mathbb{P}_{001|0}, \mathbb{P}_{111|0}, \mathbb{P}_{101|0}, \mathbb{P}_{011|1}, \mathbb{P}_{001|1}, \mathbb{P}_{111|1}, \mathbb{P}_{101|1}, \mathbb{P}_{0+0|0}, \mathbb{P}_{1+0|0}, \mathbb{P}_{0+0|1}, \mathbb{P}_{1+0|1}) \\
&= \left(\frac{1}{8}, \frac{1}{12}, \frac{1}{16}, \frac{1}{32}, \frac{1}{4}, \frac{1}{48}, \frac{1}{16}, \frac{1}{16}, \frac{13}{24}, \frac{5}{32}, \frac{17}{48}, \frac{1}{4}\right).
\end{aligned}$$

Define the parameters:

$$\Theta_{dy|z} = \mathbb{P}(D = d, Y = y \mid Z = z) \text{ for } (z, d, y) = (0,0,0), (0,1,1), (0,1,0), (1,0,1), (1,0,0), (1,1,0),$$
$$\Theta_{R^Y|zdy} = \mathbb{P}(R^Y = 1 \mid Z = z, D = d, Y = y) \text{ for } z = 0,1,\ d = 0,1, \text{ and } y = 0,1.$$

The observable data probabilities relate to the parameters as follows:

$$\begin{aligned}
\mathbb{P}_{011|0} &= (1 - \Theta_{00|0} - \Theta_{11|0} - \Theta_{10|0})\Theta_{R^Y|001},\\
\mathbb{P}_{001|0} &= \Theta_{00|0}\Theta_{R^Y|000},\\
\mathbb{P}_{111|0} &= \Theta_{11|0}\Theta_{R^Y|011},\\
\mathbb{P}_{101|0} &= \Theta_{10|0}\Theta_{R^Y|010},\\
\mathbb{P}_{011|1} &= \Theta_{01|1}\Theta_{R^Y|101},\\
\mathbb{P}_{001|1} &= \Theta_{00|1}\Theta_{R^Y|100},\\
\mathbb{P}_{111|1} &= (1 - \Theta_{01|1} - \Theta_{00|1} - \Theta_{10|1})\Theta_{R^Y|111},\\
\mathbb{P}_{101|1} &= \Theta_{10|1}\Theta_{R^Y|110},\\
\mathbb{P}_{0+0|0} &= (1 - \Theta_{00|0} - \Theta_{11|0} - \Theta_{10|0})(1 - \Theta_{R^Y|001}) + \Theta_{00|0}(1 - \Theta_{R^Y|000}),\\
\mathbb{P}_{1+0|0} &= \Theta_{11|0}(1 - \Theta_{R^Y|011}) + \Theta_{10|0}(1 - \Theta_{R^Y|010}),\\
\mathbb{P}_{0+0|1} &= \Theta_{01|1}(1 - \Theta_{R^Y|101}) + \Theta_{00|1}(1 - \Theta_{R^Y|100}),\\
\mathbb{P}_{1+0|1} &= (1 - \Theta_{01|1} - \Theta_{00|1} - \Theta_{10|1})(1 - \Theta_{R^Y|111}) + \Theta_{10|1}(1 - \Theta_{R^Y|110}).
\end{aligned}$$

These observable data probabilities are consistent with $\Theta_{10|0} = \frac{1}{8}$, $\Theta_{11|0} = \frac{1}{8}$, $\Theta_{00|0} = \frac{1}{4}$, $\Theta_{00|1} = \frac{1}{8}$, $\Theta_{01|1} = \frac{1}{2}$, $\Theta_{10|1} = \frac{1}{8}$, $\Theta_{R^Y|000} = \frac{1}{3}$, $\Theta_{R^Y|001} = \frac{1}{4}$, $\Theta_{R^Y|010} = \frac{1}{4}$, $\Theta_{R^Y|011} = \frac{1}{2}$, $\Theta_{R^Y|100} = \frac{1}{6}$, $\Theta_{R^Y|101} = \frac{1}{2}$, $\Theta_{R^Y|110} = \frac{1}{2}$, $\Theta_{R^Y|111} = \frac{1}{4}$, and the CACE $= 1$. They are also consistent with $\Theta_{10|0} = \frac{1}{8}$, $\Theta_{11|0} = \frac{1}{8}$, $\Theta_{00|0} = \frac{1}{3}$, $\Theta_{00|1} = \frac{1}{4}$, $\Theta_{01|1} = \frac{3}{8}$, $\Theta_{10|1} = \frac{1}{6}$, $\Theta_{R^Y|000} = \frac{1}{4}$, $\Theta_{R^Y|001} = \frac{3}{10}$, $\Theta_{R^Y|010} = \frac{1}{4}$, $\Theta_{R^Y|011} = \frac{1}{2}$, $\Theta_{R^Y|100} = \frac{1}{12}$, $\Theta_{R^Y|101} = \frac{2}{3}$, $\Theta_{R^Y|110} = \frac{3}{8}$, $\Theta_{R^Y|111} = \frac{3}{10}$, and the CACE $= \frac{1}{3}$. Therefore, the CACE cannot be uniquely identified.

C.1.5 Assumption $1UDY$

For a binary Y with one-sided noncompliance, we consider the following observable data probabilities:

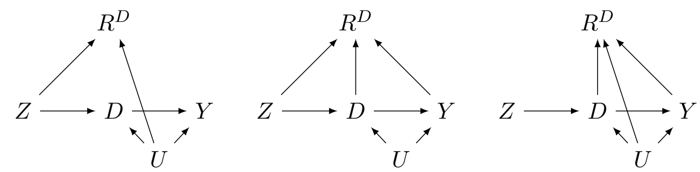
$$(\mathbb{P}_{011|0}, \mathbb{P}_{001|0}, \mathbb{P}_{011|1}, \mathbb{P}_{001|1}, \mathbb{P}_{111|1}, \mathbb{P}_{101|1}, \mathbb{P}_{0+0|0}, \mathbb{P}_{0+0|1}, \mathbb{P}_{1+0|1}) = \left(\frac{3}{16}, \frac{3}{16}, \frac{1}{8}, \frac{1}{16}, \frac{1}{18}, \frac{1}{12}, \frac{5}{8}, \frac{5}{16}, \frac{13}{36}\right).$$

Define the parameters:

$$\begin{aligned}
\Theta_n &= \mathbb{P}(U = n),\\
\Theta_{Y|ud} &= \mathbb{P}(Y = 1 \mid U = u, D = d) \text{ for } (u, d) = (n, 0), (c, 1), (c, 0),\\
\Theta_{R^Y|udy} &= \mathbb{P}(R^Y = 1 \mid U = u, D = d, Y = y) \text{ for } (u, d) = (n, 0), (c, 1), (c, 0) \text{ and } y = 0, 1.
\end{aligned}$$

The observable data probabilities relate to the parameters as follows:

$$\mathbb{P}_{011|0} = \Theta_{Y|n0}\Theta_{R^Y|n01}\Theta_n + \Theta_{Y|c0}\Theta_{R^Y|c01}(1 - \Theta_n),$$



$2ZU$ $\quad$ $2ZDY$ $\quad$ $2UDY$

Figure 9: The DAGs for the mechanisms for missingness in the treatment that do not identify the CACE. Section C.2 gives a counterexample for each.

$$
\begin{aligned}
\mathbb{P}_{001|0} &= (1-\Theta_{Y|n0})\Theta_{R^Y|n00}\Theta_n + (1-\Theta_{Y|c0})\Theta_{R^Y|c00}(1-\Theta_n),\\
\mathbb{P}_{011|1} &= \Theta_{Y|n0}\Theta_{R^Y|n01}\Theta_n,\\
\mathbb{P}_{001|1} &= (1-\Theta_{Y|n0})\Theta_{R^Y|n00}\Theta_n,\\
\mathbb{P}_{111|1} &= \Theta_{Y|c1}\Theta_{R^Y|c11}(1-\Theta_n),\\
\mathbb{P}_{101|1} &= (1-\Theta_{Y|c1})\Theta_{R^Y|c10}(1-\Theta_n),\\
\mathbb{P}_{0+0|0} &= \Theta_{Y|n0}(1-\Theta_{R^Y|n01})\Theta_n + \Theta_{Y|c0}(1-\Theta_{R^Y|c01})(1-\Theta_n)\\
&\quad + (1-\Theta_{Y|n0})(1-\Theta_{R^Y|n00})\Theta_n + (1-\Theta_{Y|c0})(1-\Theta_{R^Y|c00})(1-\Theta_n),\\
\mathbb{P}_{0+0|1} &= \Theta_{Y|n0}(1-\Theta_{R^Y|n01})\Theta_n + (1-\Theta_{Y|n0})(1-\Theta_{R^Y|n00})\Theta_n,\\
\mathbb{P}_{1+0|1} &= \Theta_{Y|c1}(1-\Theta_{R^Y|c11})(1-\Theta_n) + (1-\Theta_{Y|c1})(1-\Theta_{R^Y|c10})(1-\Theta_n).
\end{aligned}
$$

These observable data probabilities are consistent with $\Theta_n = \frac{1}{2}$, $\Theta_{Y|n0} = \frac{1}{2}$, $\Theta_{Y|c0} = \frac{1}{4}$, $\Theta_{Y|c1} = \frac{1}{3}$, $\Theta_{R^Y|n00} = \frac{1}{4}$, $\Theta_{R^Y|n01} = \frac{1}{2}$, $\Theta_{R^Y|c00} = \frac{1}{3}$, $\Theta_{R^Y|c01} = \frac{1}{2}$, $\Theta_{R^Y|c10} = \frac{1}{4}$, $\Theta_{R^Y|c11} = \frac{1}{3}$, and the CACE $= \frac{1}{12}$. They are also consistent with $\Theta_n = \frac{1}{2}$, $\Theta_{Y|n0} = \frac{3}{4}$, $\Theta_{Y|c0} = \frac{1}{2}$, $\Theta_{Y|c1} = \frac{4}{9}$, $\Theta_{R^Y|n00} = \frac{1}{2}$, $\Theta_{R^Y|n01} = \frac{1}{3}$, $\Theta_{R^Y|c00} = \frac{1}{2}$, $\Theta_{R^Y|c01} = \frac{1}{4}$, $\Theta_{R^Y|c10} = \frac{3}{10}$, $\Theta_{R^Y|c11} = \frac{1}{4}$, and the CACE $= -\frac{1}{18}$. Therefore, the CACE cannot be uniquely identified.

C.2 Counterexamples for the Missing Treatment Models

Figure 9 shows the DAGs. Assumption $2ZU$ identifies the CACE under one-sided noncompliance, so we give a counterexample under two-sided noncompliance. Assumptions $2ZDY$ and $2UDY$ fail under both one-sided and two-sided noncompliance. For these, we give counterexamples under one-sided noncompliance, which is the special case of two-sided noncompliance with no always-takers. Define

$$
\begin{aligned}
\mathbb{P}_{dy1|z} &= \mathbb{P}(D=d, Y=y, R^D=1 \mid Z=z),\\
\mathbb{P}_{y+0|z} &= \mathbb{P}(Y=y, R^D=0 \mid Z=z).
\end{aligned}
$$

C.2.1 Assumption $2ZU$

For a binary Y with two-sided noncompliance, we consider the following observable data probabilities:

$$(\mathbb{P}_{011|0}, \mathbb{P}_{001|0}, \mathbb{P}_{011|1}, \mathbb{P}_{001|1}, \mathbb{P}_{111|0}, \mathbb{P}_{101|0}, \mathbb{P}_{111|1}, \mathbb{P}_{101|1}, \mathbb{P}_{1+0|0}, \mathbb{P}_{0+0|0}, \mathbb{P}_{1+0|1}, \mathbb{P}_{0+0|1})$$

$$= \left(\frac{5}{48}, \frac{5}{48}, \frac{1}{12}, \frac{1}{12}, \frac{1}{16}, \frac{1}{16}, \frac{7}{96}, \frac{11}{96}, \frac{1}{3}, \frac{1}{3}, \frac{29}{96}, \frac{11}{32}\right).$$

Define the parameters:

$$\begin{aligned}
\Theta_u &= \mathbb{P}(U = u) \text{ for } u = a, n, \\
\Theta_{Y|ud} &= \mathbb{P}(Y = 1 \mid U = u, D = d) \text{ for } (u, d) = (a, 1), (n, 0), (c, 1), (c, 0), \\
\Theta_{R^D|zu} &= \mathbb{P}(R^D = 1 \mid Z = z, U = u) \text{ for } z = 0, 1 \text{ and } u = a, n, c.
\end{aligned}$$

The observable data probabilities relate to the parameters as follows:

$$\begin{aligned}
\mathbb{P}_{011|0} &= \Theta_{Y|n0}\Theta_n\Theta_{R^D|0n} + \Theta_{Y|c0}(1 - \Theta_n - \Theta_a)\Theta_{R^D|0c}, \\
\mathbb{P}_{001|0} &= (1 - \Theta_{Y|n0})\Theta_n\Theta_{R^D|0n} + (1 - \Theta_{Y|c0})(1 - \Theta_n - \Theta_a)\Theta_{R^D|0c}, \\
\mathbb{P}_{011|1} &= \Theta_{Y|n0}\Theta_n\Theta_{R^D|1n}, \\
\mathbb{P}_{001|1} &= (1 - \Theta_{Y|n0})\Theta_n\Theta_{R^D|1n}, \\
\mathbb{P}_{111|0} &= \Theta_{Y|a1}\Theta_a\Theta_{R^D|0a}, \\
\mathbb{P}_{101|0} &= (1 - \Theta_{Y|a1})\Theta_a\Theta_{R^D|0a}, \\
\mathbb{P}_{111|1} &= \Theta_{Y|a1}\Theta_a\Theta_{R^D|1a} + \Theta_{Y|c1}(1 - \Theta_n - \Theta_a)\Theta_{R^D|1c}, \\
\mathbb{P}_{101|1} &= (1 - \Theta_{Y|a1})\Theta_a\Theta_{R^D|1a} + (1 - \Theta_{Y|c1})(1 - \Theta_n - \Theta_a)\Theta_{R^D|1c}, \\
\mathbb{P}_{1+0|0} &= \Theta_{Y|n0}\Theta_n(1 - \Theta_{R^D|0n}) + \Theta_{Y|c0}(1 - \Theta_n - \Theta_a)(1 - \Theta_{R^D|0c}) + \Theta_{Y|a1}\Theta_a(1 - \Theta_{R^D|0a}), \\
\mathbb{P}_{0+0|0} &= (1 - \Theta_{Y|n0})\Theta_n(1 - \Theta_{R^D|0n}) + (1 - \Theta_{Y|c0})(1 - \Theta_n - \Theta_a)(1 - \Theta_{R^D|0c}) \\
&\quad + (1 - \Theta_{Y|a1})\Theta_a(1 - \Theta_{R^D|0a}), \\
\mathbb{P}_{1+0|1} &= \Theta_{Y|n0}\Theta_n(1 - \Theta_{R^D|1n}) + \Theta_{Y|a1}\Theta_a(1 - \Theta_{R^D|1a}) + \Theta_{Y|c1}(1 - \Theta_n - \Theta_a)(1 - \Theta_{R^D|1c}), \\
\mathbb{P}_{0+0|1} &= (1 - \Theta_{Y|n0})\Theta_n(1 - \Theta_{R^D|1n}) + (1 - \Theta_{Y|a1})\Theta_a(1 - \Theta_{R^D|1a}) \\
&\quad + (1 - \Theta_{Y|c1})(1 - \Theta_n - \Theta_a)(1 - \Theta_{R^D|1c}).
\end{aligned}$$

These observable data probabilities are consistent with $\Theta_n = \frac{1}{2}$, $\Theta_a = \frac{1}{4}$, $\Theta_{Y|n0} = \frac{1}{2}$, $\Theta_{Y|c0} = \frac{1}{2}$, $\Theta_{Y|a1} = \frac{1}{2}$, $\Theta_{Y|c1} = \frac{1}{3}$, $\Theta_{R^D|0n} = \frac{1}{4}$, $\Theta_{R^D|1n} = \frac{1}{3}$, $\Theta_{R^D|0a} = \frac{1}{2}$, $\Theta_{R^D|1a} = \frac{1}{4}$, $\Theta_{R^D|0c} = \frac{1}{3}$, $\Theta_{R^D|1c} = \frac{1}{2}$, and the CACE $= -\frac{1}{6}$. They are also consistent with $\Theta_n = \frac{1}{3}$, $\Theta_a = \frac{1}{3}$, $\Theta_{Y|n0} = \frac{1}{2}$, $\Theta_{Y|c0} = \frac{1}{2}$, $\Theta_{Y|a1} = \frac{1}{2}$, $\Theta_{Y|c1} = \frac{3}{8}$, $\Theta_{R^D|0n} = \frac{1}{4}$, $\Theta_{R^D|1n} = \frac{1}{2}$, $\Theta_{R^D|0a} = \frac{3}{8}$, $\Theta_{R^D|1a} = \frac{1}{16}$, $\Theta_{R^D|0c} = \frac{3}{8}$, $\Theta_{R^D|1c} = \frac{1}{2}$, and the CACE $= -\frac{1}{8}$. Therefore, the CACE cannot be uniquely identified.

C.2.2 Assumption $2ZDY$

For a binary Y with one-sided noncompliance, we consider the following observable data probabilities:

$$\begin{aligned}
&(\mathbb{P}_{011|0}, \mathbb{P}_{001|0}, \mathbb{P}_{011|1}, \mathbb{P}_{001|1}, \mathbb{P}_{111|1}, \mathbb{P}_{101|1}, \mathbb{P}_{1+0|0}, \mathbb{P}_{0+0|0}, \mathbb{P}_{1+0|1}, \mathbb{P}_{0+0|1}) \\
&= \left(\frac{1}{8}, \frac{1}{6}, \frac{1}{4}, \frac{1}{24}, \frac{1}{64}, \frac{1}{16}, \frac{3}{8}, \frac{1}{3}, \frac{23}{64}, \frac{13}{48}\right).
\end{aligned}$$

Define the parameters:

$$\Theta_{dy|z} = \mathbb{P}(D = d, Y = y \mid Z = z) \text{ for } (z, d, y) = (0,0,1), (1,0,1), (1,0,0), (1,1,0),$$
$$\Theta_{R^D|zdy} = \mathbb{P}(R^D = 1 \mid Z = z, D = d, Y = y) \text{ for } (z, d) = (0,0), (1,0), (1,1) \text{ and } y = 0, 1.$$

The observable data probabilities relate to the parameters as follows:

$$\begin{aligned}
\mathbb{P}_{011|0} &= \Theta_{01|0}\Theta_{R^D|001},\\
\mathbb{P}_{001|0} &= (1-\Theta_{01|0})\Theta_{R^D|000},\\
\mathbb{P}_{011|1} &= \Theta_{01|1}\Theta_{R^D|101},\\
\mathbb{P}_{001|1} &= \Theta_{00|1}\Theta_{R^D|100},\\
\mathbb{P}_{111|1} &= (1-\Theta_{01|1}-\Theta_{00|1}-\Theta_{10|1})\Theta_{R^D|111},\\
\mathbb{P}_{101|1} &= \Theta_{10|1}\Theta_{R^D|110},\\
\mathbb{P}_{1+0|0} &= \Theta_{01|0}(1-\Theta_{R^D|001}),\\
\mathbb{P}_{0+0|0} &= (1-\Theta_{01|0})(1-\Theta_{R^D|000}),\\
\mathbb{P}_{1+0|1} &= \Theta_{01|1}(1-\Theta_{R^D|101}) + (1-\Theta_{01|1}-\Theta_{00|1}-\Theta_{10|1})(1-\Theta_{R^D|111}),\\
\mathbb{P}_{0+0|1} &= \Theta_{00|1}(1-\Theta_{R^D|100}) + \Theta_{10|1}(1-\Theta_{R^D|110}).
\end{aligned}$$

These observable data probabilities are consistent with $\Theta_{01|0} = \frac{1}{2}$, $\Theta_{00|1} = \frac{1}{4}$, $\Theta_{01|1} = \frac{1}{2}$, $\Theta_{10|1} = \frac{1}{8}$, $\Theta_{R^D|000} = \frac{1}{3}$, $\Theta_{R^D|001} = \frac{1}{4}$, $\Theta_{R^D|100} = \frac{1}{6}$, $\Theta_{R^D|101} = \frac{1}{2}$, $\Theta_{R^D|110} = \frac{1}{2}$, $\Theta_{R^D|111} = \frac{1}{8}$, and the CACE $= \frac{1}{2}$. They are also consistent with $\Theta_{01|0} = \frac{1}{2}$, $\Theta_{00|1} = \frac{1}{8}$, $\Theta_{01|1} = \frac{3}{8}$, $\Theta_{10|1} = \frac{1}{4}$, $\Theta_{R^D|000} = \frac{1}{3}$, $\Theta_{R^D|001} = \frac{1}{4}$, $\Theta_{R^D|100} = \frac{1}{3}$, $\Theta_{R^D|101} = \frac{2}{3}$, $\Theta_{R^D|110} = \frac{1}{4}$, $\Theta_{R^D|111} = \frac{1}{16}$, and the CACE $= \frac{1}{4}$. Therefore, the CACE cannot be uniquely identified.

C.2.3 Assumption $2UDY$

For a binary Y with one-sided noncompliance, we consider the following observable data probabilities:

$$\begin{aligned}
&(\mathbb{P}_{011|0}, \mathbb{P}_{001|0}, \mathbb{P}_{011|1}, \mathbb{P}_{001|1}, \mathbb{P}_{111|1}, \mathbb{P}_{101|1}, \mathbb{P}_{1+0|0}, \mathbb{P}_{0+0|0}, \mathbb{P}_{1+0|1}, \mathbb{P}_{0+0|1})\\
&= \left(\frac{3}{16}, \frac{3}{16}, \frac{1}{8}, \frac{1}{16}, \frac{1}{18}, \frac{1}{12}, \frac{3}{16}, \frac{7}{16}, \frac{17}{72}, \frac{7}{16}\right).
\end{aligned}$$

Define the parameters:

$$\Theta_n = \mathbb{P}(U = n),$$
$$\Theta_{Y|ud} = \mathbb{P}(Y = 1 \mid U = u, D = d) \text{ for } (u, d) = (n, 0), (c, 1), (c, 0),$$
$$\Theta_{R^D|udy} = \mathbb{P}(R^D = 1 \mid U = u, D = d, Y = y) \text{ for } (u, d) = (n, 0), (c, 1), (c, 0) \text{ and } y = 0, 1.$$

The observable data probabilities relate to the parameters as follows:

$$\begin{aligned}
\mathbb{P}_{011|0} &= \Theta_{Y|n0}\Theta_n\Theta_{R^D|n01} + \Theta_{Y|c0}(1-\Theta_n)\Theta_{R^D|c01},\\
\mathbb{P}_{001|0} &= (1-\Theta_{Y|n0})\Theta_n\Theta_{R^D|n00} + (1-\Theta_{Y|c0})(1-\Theta_n)\Theta_{R^D|c00},
\end{aligned}$$

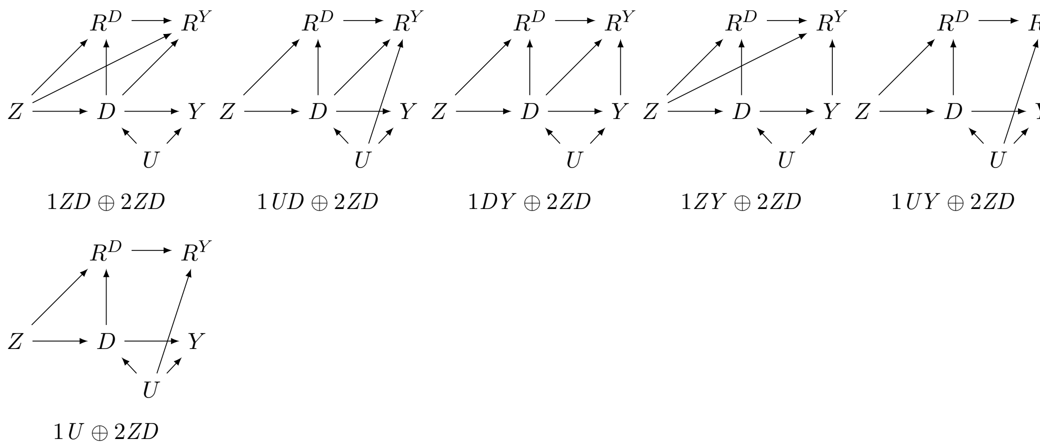


Figure 10: The DAGs for the combined mechanisms that do not identify the CACE. Section C.3 gives a counterexample for each.

$$
\begin{aligned}
\mathbb{P}_{011|1} &= \Theta_{Y|n0}\Theta_n\Theta_{R^D|n01},\\
\mathbb{P}_{001|1} &= (1-\Theta_{Y|n0})\Theta_n\Theta_{R^D|n00},\\
\mathbb{P}_{111|1} &= \Theta_{Y|c1}(1-\Theta_n)\Theta_{R^D|c11},\\
\mathbb{P}_{101|1} &= (1-\Theta_{Y|c1})(1-\Theta_n)\Theta_{R^D|c10},\\
\mathbb{P}_{1+0|0} &= \Theta_{Y|n0}\Theta_n(1-\Theta_{R^D|n01}) + \Theta_{Y|c0}(1-\Theta_n)(1-\Theta_{R^D|c01}),\\
\mathbb{P}_{0+0|0} &= (1-\Theta_{Y|n0})\Theta_n(1-\Theta_{R^D|n00}) + (1-\Theta_{Y|c0})(1-\Theta_n)(1-\Theta_{R^D|c00}),\\
\mathbb{P}_{1+0|1} &= \Theta_{Y|n0}\Theta_n(1-\Theta_{R^D|n01}) + \Theta_{Y|c1}(1-\Theta_n)(1-\Theta_{R^D|c11}),\\
\mathbb{P}_{0+0|1} &= (1-\Theta_{Y|n0})\Theta_n(1-\Theta_{R^D|n00}) + (1-\Theta_{Y|c1})(1-\Theta_n)(1-\Theta_{R^D|c10}).
\end{aligned}
$$

These observable data probabilities are consistent with $\Theta_n = \frac{1}{2}$, $\Theta_{Y|n0} = \frac{1}{2}$, $\Theta_{Y|c0} = \frac{1}{4}$, $\Theta_{Y|c1} = \frac{1}{3}$, $\Theta_{R^D|n00} = \frac{1}{4}$, $\Theta_{R^D|n01} = \frac{1}{2}$, $\Theta_{R^D|c00} = \frac{1}{3}$, $\Theta_{R^D|c01} = \frac{1}{2}$, $\Theta_{R^D|c10} = \frac{1}{4}$, $\Theta_{R^D|c11} = \frac{1}{3}$, and the CACE $= \frac{1}{12}$. They are also consistent with $\Theta_n = \frac{1}{3}$, $\Theta_{Y|n0} = \frac{1}{2}$, $\Theta_{Y|c0} = \frac{5}{16}$, $\Theta_{Y|c1} = \frac{3}{8}$, $\Theta_{R^D|n00} = \frac{3}{8}$, $\Theta_{R^D|n01} = \frac{3}{4}$, $\Theta_{R^D|c00} = \frac{3}{11}$, $\Theta_{R^D|c01} = \frac{3}{10}$, $\Theta_{R^D|c10} = \frac{1}{5}$, $\Theta_{R^D|c11} = \frac{2}{9}$, and the CACE $= \frac{1}{16}$. Therefore, the CACE cannot be uniquely identified.

C.3 Counterexamples for the Missing Treatment and Outcome Models

Figure 10 shows the DAGs. Assumptions $1ZD\oplus 2ZD$, $1UD\oplus 2ZD$, and $1UY\oplus 2ZD$ fail under both one-sided and two-sided noncompliance. For these, we give counterexamples under one-sided noncompliance, which is the special case of two-sided noncompliance with no always-takers. Assumptions $1DY \oplus 2ZD$ and $1ZY \oplus 2ZD$ contain Assumptions $1DY$ and $1ZY$, respectively, which already fail under one-sided noncompliance, so we give counterexamples under two-sided noncompliance. Assumption $1U \oplus 2ZD$ identifies the CACE under one-sided noncompliance, so we give a counterexample under two-sided noncompliance. Define

$$
\mathbb{P}_{dy11|z} = \mathbb{P}(D = d, Y = y, R^D = 1, R^Y = 1 \mid Z = z),
$$

$$\begin{aligned}
\mathbb{P}_{y+01|z} &= \mathbb{P}(Y = y, R^D = 0, R^Y = 1 \mid Z = z),\\
\mathbb{P}_{d1+0|z} &= \mathbb{P}(D = d, R^D = 1, R^Y = 0 \mid Z = z),\\
\mathbb{P}_{+0+0|z} &= \mathbb{P}(R^D = 0, R^Y = 0 \mid Z = z).
\end{aligned}$$

C.3.1 Assumption $1ZD \oplus 2ZD$

For a binary Y with one-sided noncompliance, we consider the following observable data probabilities:

$$\begin{aligned}
&(\mathbb{P}_{0111|0}, \mathbb{P}_{0011|0}, \mathbb{P}_{0111|1}, \mathbb{P}_{0011|1}, \mathbb{P}_{1111|1}, \mathbb{P}_{1011|1},\\
&\ \mathbb{P}_{1+01|0}, \mathbb{P}_{0+01|0}, \mathbb{P}_{1+01|1}, \mathbb{P}_{0+01|1}, \mathbb{P}_{01+0|0}, \mathbb{P}_{01+0|1}, \mathbb{P}_{11+0|1}, \mathbb{P}_{+0+0|0}, \mathbb{P}_{+0+0|1})\\
&= \left(\frac{1}{16}, \frac{1}{16}, \frac{1}{32}, \frac{1}{16}, \frac{1}{96}, \frac{1}{96}, \frac{1}{8}, \frac{1}{8}, \frac{13}{192}, \frac{17}{192}, \frac{1}{8}, \frac{9}{32}, \frac{1}{24}, \frac{1}{2}, \frac{13}{32}\right).
\end{aligned}$$

Define the parameters:

$$\begin{aligned}
\Theta_{dy|z} &= \mathbb{P}(D = d, Y = y \mid Z = z) \text{ for } (z, d, y) = (0, 0, 1), (1, 0, 1), (1, 0, 0), (1, 1, 0),\\
\Theta_{R^D|zd} &= \mathbb{P}(R^D = 1 \mid Z = z, D = d) \text{ for } (z, d) = (0, 0), (1, 0), (1, 1),\\
\Theta_{R^Y|zdr^D} &= \mathbb{P}(R^Y = 1 \mid Z = z, D = d, R^D = r^D) \text{ for } (z, d) = (0, 0), (1, 0), (1, 1) \text{ and } r^D = 0, 1.
\end{aligned}$$

The observable data probabilities relate to the parameters as follows:

$$\begin{aligned}
\mathbb{P}_{0111|0} &= \Theta_{01|0}\Theta_{R^D|00}\Theta_{R^Y|001},\\
\mathbb{P}_{0011|0} &= (1 - \Theta_{01|0})\Theta_{R^D|00}\Theta_{R^Y|001},\\
\mathbb{P}_{0111|1} &= \Theta_{01|1}\Theta_{R^D|10}\Theta_{R^Y|101},\\
\mathbb{P}_{0011|1} &= \Theta_{00|1}\Theta_{R^D|10}\Theta_{R^Y|101},\\
\mathbb{P}_{1111|1} &= (1 - \Theta_{01|1} - \Theta_{00|1} - \Theta_{10|1})\Theta_{R^D|11}\Theta_{R^Y|111},\\
\mathbb{P}_{1011|1} &= \Theta_{10|1}\Theta_{R^D|11}\Theta_{R^Y|111},\\
\mathbb{P}_{1+01|0} &= \Theta_{01|0}(1 - \Theta_{R^D|00})\Theta_{R^Y|000},\\
\mathbb{P}_{0+01|0} &= (1 - \Theta_{01|0})(1 - \Theta_{R^D|00})\Theta_{R^Y|000},\\
\mathbb{P}_{1+01|1} &= \Theta_{01|1}(1 - \Theta_{R^D|10})\Theta_{R^Y|100} + (1 - \Theta_{01|1} - \Theta_{00|1} - \Theta_{10|1})(1 - \Theta_{R^D|11})\Theta_{R^Y|110},\\
\mathbb{P}_{0+01|1} &= \Theta_{00|1}(1 - \Theta_{R^D|10})\Theta_{R^Y|100} + \Theta_{10|1}(1 - \Theta_{R^D|11})\Theta_{R^Y|110},\\
\mathbb{P}_{01+0|0} &= (1 - \Theta_{01|0})\Theta_{R^D|00}(1 - \Theta_{R^Y|001}) + \Theta_{01|0}\Theta_{R^D|00}(1 - \Theta_{R^Y|001}),\\
\mathbb{P}_{01+0|1} &= \Theta_{00|1}\Theta_{R^D|10}(1 - \Theta_{R^Y|101}) + \Theta_{01|1}\Theta_{R^D|10}(1 - \Theta_{R^Y|101}),\\
\mathbb{P}_{11+0|1} &= \Theta_{10|1}\Theta_{R^D|11}(1 - \Theta_{R^Y|111}) + (1 - \Theta_{01|1} - \Theta_{00|1} - \Theta_{10|1})\Theta_{R^D|11}(1 - \Theta_{R^Y|111}),\\
\mathbb{P}_{+0+0|0} &= (1 - \Theta_{01|0})(1 - \Theta_{R^D|00})(1 - \Theta_{R^Y|000}) + \Theta_{01|0}(1 - \Theta_{R^D|00})(1 - \Theta_{R^Y|000}),\\
\mathbb{P}_{+0+0|1} &= \Theta_{00|1}(1 - \Theta_{R^D|10})(1 - \Theta_{R^Y|100}) + \Theta_{01|1}(1 - \Theta_{R^D|10})(1 - \Theta_{R^Y|100})\\
&\quad + \Theta_{10|1}(1 - \Theta_{R^D|11})(1 - \Theta_{R^Y|110}) + (1 - \Theta_{01|1} - \Theta_{00|1} - \Theta_{10|1})(1 - \Theta_{R^D|11})(1 - \Theta_{R^Y|110}).
\end{aligned}$$

These observable data probabilities are consistent with $\Theta_{01|0} = \frac{1}{2}$, $\Theta_{00|1} = \frac{1}{2}$, $\Theta_{01|1} = \frac{1}{4}$, $\Theta_{10|1} = \frac{1}{8}$, $\Theta_{R^D|00} = \frac{1}{4}$, $\Theta_{R^D|10} = \frac{1}{2}$, $\Theta_{R^D|11} = \frac{1}{4}$, $\Theta_{R^Y|001} = \frac{1}{2}$, $\Theta_{R^Y|101} = \frac{1}{4}$, $\Theta_{R^Y|111} = \frac{1}{3}$,

$\Theta_{R^Y|000} = \frac{1}{3}$, $\Theta_{R^Y|100} = \frac{1}{6}$, $\Theta_{R^Y|110} = \frac{1}{2}$, and the CACE $= -\frac{1}{2}$. They are also consistent with $\Theta_{01|0} = \frac{1}{2}$, $\Theta_{00|1} = \frac{1}{3}$, $\Theta_{01|1} = \frac{1}{6}$, $\Theta_{10|1} = \frac{1}{4}$, $\Theta_{R^D|00} = \frac{1}{4}$, $\Theta_{R^D|10} = \frac{3}{4}$, $\Theta_{R^D|11} = \frac{1}{8}$, $\Theta_{R^Y|001} = \frac{1}{2}$, $\Theta_{R^Y|101} = \frac{1}{4}$, $\Theta_{R^Y|111} = \frac{1}{3}$, $\Theta_{R^Y|000} = \frac{1}{3}$, $\Theta_{R^Y|100} = \frac{1}{2}$, $\Theta_{R^Y|110} = \frac{3}{14}$, and the CACE $= -\frac{1}{6}$. Therefore, the CACE cannot be uniquely identified.

C.3.2 Assumption $1UD \oplus 2ZD$

For a binary Y with one-sided noncompliance, we consider the following observable data probabilities:

$$\begin{aligned}
&(\mathbb{P}_{0111|0}, \mathbb{P}_{0011|0}, \mathbb{P}_{0111|1}, \mathbb{P}_{0011|1}, \mathbb{P}_{1111|1}, \mathbb{P}_{1011|1},\\
&\ \mathbb{P}_{1+01|0}, \mathbb{P}_{0+01|0}, \mathbb{P}_{1+01|1}, \mathbb{P}_{0+01|1}, \mathbb{P}_{01+0|0}, \mathbb{P}_{01+0|1}, \mathbb{P}_{11+0|1}, \mathbb{P}_{+0+0|0}, \mathbb{P}_{+0+0|1})\\
&= \left(\frac{11}{192}, \frac{19}{192}, \frac{1}{128}, \frac{1}{128}, \frac{1}{16}, \frac{1}{16}, \frac{5}{192}, \frac{1}{24}, \frac{5}{64}, \frac{5}{64}, \frac{11}{32}, \frac{3}{64}, \frac{1}{8}, \frac{83}{192}, \frac{17}{32}\right).
\end{aligned}$$

Define the parameters:

$$\begin{aligned}
\Theta_n &= \mathbb{P}(U = n),\\
\Theta_{Y|ud} &= \mathbb{P}(Y = 1 \mid U = u, D = d) \text{ for } (u, d) = (n, 0), (c, 1), (c, 0),\\
\Theta_{R^D|zd} &= \mathbb{P}(R^D = 1 \mid Z = z, D = d) \text{ for } (z, d) = (0, 0), (1, 0), (1, 1),\\
\Theta_{R^Y|udr^D} &= \mathbb{P}(R^Y = 1 \mid U = u, D = d, R^D = r^D) \text{ for } (u, d) = (n, 0), (c, 1), (c, 0) \text{ and } r^D = 0, 1.
\end{aligned}$$

The observable data probabilities relate to the parameters as follows:

$$\begin{aligned}
\mathbb{P}_{0111|0} &= \Theta_{Y|n0}\Theta_n\Theta_{R^D|00}\Theta_{R^Y|n01} + \Theta_{Y|c0}(1-\Theta_n)\Theta_{R^D|00}\Theta_{R^Y|c01},\\
\mathbb{P}_{0011|0} &= (1-\Theta_{Y|n0})\Theta_n\Theta_{R^D|00}\Theta_{R^Y|n01} + (1-\Theta_{Y|c0})(1-\Theta_n)\Theta_{R^D|00}\Theta_{R^Y|c01},\\
\mathbb{P}_{0111|1} &= \Theta_{Y|n0}\Theta_n\Theta_{R^D|10}\Theta_{R^Y|n01},\\
\mathbb{P}_{0011|1} &= (1-\Theta_{Y|n0})\Theta_n\Theta_{R^D|10}\Theta_{R^Y|n01},\\
\mathbb{P}_{1111|1} &= \Theta_{Y|c1}(1-\Theta_n)\Theta_{R^D|11}\Theta_{R^Y|c11},\\
\mathbb{P}_{1011|1} &= (1-\Theta_{Y|c1})(1-\Theta_n)\Theta_{R^D|11}\Theta_{R^Y|c11},\\
\mathbb{P}_{1+01|0} &= \Theta_{Y|n0}\Theta_n(1-\Theta_{R^D|00})\Theta_{R^Y|n00} + \Theta_{Y|c0}(1-\Theta_n)(1-\Theta_{R^D|00})\Theta_{R^Y|c00},\\
\mathbb{P}_{0+01|0} &= (1-\Theta_{Y|n0})\Theta_n(1-\Theta_{R^D|00})\Theta_{R^Y|n00} + (1-\Theta_{Y|c0})(1-\Theta_n)(1-\Theta_{R^D|00})\Theta_{R^Y|c00},\\
\mathbb{P}_{1+01|1} &= \Theta_{Y|n0}\Theta_n(1-\Theta_{R^D|10})\Theta_{R^Y|n00} + \Theta_{Y|c1}(1-\Theta_n)(1-\Theta_{R^D|11})\Theta_{R^Y|c10},\\
\mathbb{P}_{0+01|1} &= (1-\Theta_{Y|n0})\Theta_n(1-\Theta_{R^D|10})\Theta_{R^Y|n00} + (1-\Theta_{Y|c1})(1-\Theta_n)(1-\Theta_{R^D|11})\Theta_{R^Y|c10},\\
\mathbb{P}_{01+0|0} &= \Theta_n\Theta_{R^D|00}(1-\Theta_{R^Y|n01}) + (1-\Theta_n)\Theta_{R^D|00}(1-\Theta_{R^Y|c01}),\\
\mathbb{P}_{01+0|1} &= \Theta_n\Theta_{R^D|10}(1-\Theta_{R^Y|n01}),\\
\mathbb{P}_{11+0|1} &= (1-\Theta_n)\Theta_{R^D|11}(1-\Theta_{R^Y|c11}),\\
\mathbb{P}_{+0+0|0} &= \Theta_n(1-\Theta_{R^D|00})(1-\Theta_{R^Y|n00}) + (1-\Theta_n)(1-\Theta_{R^D|00})(1-\Theta_{R^Y|c00}),\\
\mathbb{P}_{+0+0|1} &= \Theta_n(1-\Theta_{R^D|10})(1-\Theta_{R^Y|n00}) + (1-\Theta_n)(1-\Theta_{R^D|11})(1-\Theta_{R^Y|c10}).
\end{aligned}$$

These observable data probabilities are consistent with $\Theta_n = \frac{1}{4}$, $\Theta_{Y|n0} = \frac{1}{2}$, $\Theta_{Y|c0} = \frac{1}{3}$, $\Theta_{Y|c1} = \frac{1}{2}$, $\Theta_{R^D|00} = \frac{1}{2}$, $\Theta_{R^D|10} = \frac{1}{4}$, $\Theta_{R^D|11} = \frac{1}{3}$, $\Theta_{R^Y|n00} = \frac{1}{6}$, $\Theta_{R^Y|c00} = \frac{1}{8}$, $\Theta_{R^Y|c10} = \frac{1}{4}$,

$\Theta_{R^Y|n01} = \frac{1}{4}$, $\Theta_{R^Y|c01} = \frac{1}{3}$, $\Theta_{R^Y|c11} = \frac{1}{2}$, and the CACE $= \frac{1}{6}$. They are also consistent with $\Theta_n = \frac{23}{60}$, $\Theta_{Y|n0} = \frac{1}{2}$, $\Theta_{Y|c0} = \frac{4}{13}$, $\Theta_{Y|c1} = \frac{1}{2}$, $\Theta_{R^D|00} = \frac{1}{2}$, $\Theta_{R^D|10} = \frac{15}{92}$, $\Theta_{R^D|11} = \frac{15}{37}$, $\Theta_{R^Y|n00} = \frac{13}{92}$, $\Theta_{R^Y|c00} = \frac{39}{296}$, $\Theta_{R^Y|c10} = \frac{2449}{8096}$, $\Theta_{R^Y|n01} = \frac{1}{4}$, $\Theta_{R^Y|c01} = \frac{13}{37}$, $\Theta_{R^Y|c11} = \frac{1}{2}$, and the CACE $= \frac{5}{26}$. Therefore, the CACE cannot be uniquely identified.

C.3.3 Assumption $1UY \oplus 2ZD$

For a binary Y with one-sided noncompliance, we consider the following observable data probabilities:

$$\begin{aligned}
&(\mathbb{P}_{0111|0}, \mathbb{P}_{0011|0}, \mathbb{P}_{0111|1}, \mathbb{P}_{0011|1}, \mathbb{P}_{1111|1}, \mathbb{P}_{1011|1},\\
&\ \mathbb{P}_{1+01|0}, \mathbb{P}_{0+01|0}, \mathbb{P}_{1+01|1}, \mathbb{P}_{0+01|1}, \mathbb{P}_{01+0|0}, \mathbb{P}_{01+0|1}, \mathbb{P}_{11+0|1}, \mathbb{P}_{+0+0|0}, \mathbb{P}_{+0+0|1})\\
&= \left(\frac{1}{24}, \frac{3}{64}, \frac{1}{96}, \frac{1}{128}, \frac{1}{48}, \frac{1}{64}, \frac{11}{192}, \frac{3}{32}, \frac{41}{384}, \frac{19}{192}, \frac{79}{192}, \frac{17}{384}, \frac{41}{192}, \frac{67}{192}, \frac{185}{384}\right).
\end{aligned}$$

Define the parameters:

$$\begin{aligned}
\Theta_n &= \mathbb{P}(U = n),\\
\Theta_{Y|ud} &= \mathbb{P}(Y = 1 \mid U = u, D = d) \text{ for } (u, d) = (n, 0), (c, 1), (c, 0),\\
\Theta_{R^D|zd} &= \mathbb{P}(R^D = 1 \mid Z = z, D = d) \text{ for } (z, d) = (0, 0), (1, 0), (1, 1),\\
\Theta_{R^Y|uyr^D} &= \mathbb{P}(R^Y = 1 \mid U = u, Y = y, R^D = r^D) \text{ for } u = n, c,\ y = 0, 1, \text{ and } r^D = 0, 1.
\end{aligned}$$

The observable data probabilities relate to the parameters as follows:

$$\begin{aligned}
\mathbb{P}_{0111|0} &= \Theta_{Y|n0}\Theta_n\Theta_{R^D|00}\Theta_{R^Y|n11} + \Theta_{Y|c0}(1-\Theta_n)\Theta_{R^D|00}\Theta_{R^Y|c11},\\
\mathbb{P}_{0011|0} &= (1-\Theta_{Y|n0})\Theta_n\Theta_{R^D|00}\Theta_{R^Y|n01} + (1-\Theta_{Y|c0})(1-\Theta_n)\Theta_{R^D|00}\Theta_{R^Y|c01},\\
\mathbb{P}_{0111|1} &= \Theta_{Y|n0}\Theta_n\Theta_{R^D|10}\Theta_{R^Y|n11},\\
\mathbb{P}_{0011|1} &= (1-\Theta_{Y|n0})\Theta_n\Theta_{R^D|10}\Theta_{R^Y|n01},\\
\mathbb{P}_{1111|1} &= \Theta_{Y|c1}(1-\Theta_n)\Theta_{R^D|11}\Theta_{R^Y|c11},\\
\mathbb{P}_{1011|1} &= (1-\Theta_{Y|c1})(1-\Theta_n)\Theta_{R^D|11}\Theta_{R^Y|c01},\\
\mathbb{P}_{1+01|0} &= \Theta_{Y|n0}\Theta_n(1-\Theta_{R^D|00})\Theta_{R^Y|n10} + \Theta_{Y|c0}(1-\Theta_n)(1-\Theta_{R^D|00})\Theta_{R^Y|c10},\\
\mathbb{P}_{0+01|0} &= (1-\Theta_{Y|n0})\Theta_n(1-\Theta_{R^D|00})\Theta_{R^Y|n00} + (1-\Theta_{Y|c0})(1-\Theta_n)(1-\Theta_{R^D|00})\Theta_{R^Y|c00},\\
\mathbb{P}_{1+01|1} &= \Theta_{Y|n0}\Theta_n(1-\Theta_{R^D|10})\Theta_{R^Y|n10} + \Theta_{Y|c1}(1-\Theta_n)(1-\Theta_{R^D|11})\Theta_{R^Y|c10},\\
\mathbb{P}_{0+01|1} &= (1-\Theta_{Y|n0})\Theta_n(1-\Theta_{R^D|10})\Theta_{R^Y|n00} + (1-\Theta_{Y|c1})(1-\Theta_n)(1-\Theta_{R^D|11})\Theta_{R^Y|c00},\\
\mathbb{P}_{01+0|0} &= \Theta_{Y|n0}\Theta_n\Theta_{R^D|00}(1-\Theta_{R^Y|n11}) + \Theta_{Y|c0}(1-\Theta_n)\Theta_{R^D|00}(1-\Theta_{R^Y|c11})\\
&\quad + (1-\Theta_{Y|n0})\Theta_n\Theta_{R^D|00}(1-\Theta_{R^Y|n01}) + (1-\Theta_{Y|c0})(1-\Theta_n)\Theta_{R^D|00}(1-\Theta_{R^Y|c01}),\\
\mathbb{P}_{01+0|1} &= \Theta_{Y|n0}\Theta_n\Theta_{R^D|10}(1-\Theta_{R^Y|n11}) + (1-\Theta_{Y|n0})\Theta_n\Theta_{R^D|10}(1-\Theta_{R^Y|n01}),\\
\mathbb{P}_{11+0|1} &= \Theta_{Y|c1}(1-\Theta_n)\Theta_{R^D|11}(1-\Theta_{R^Y|c11}) + (1-\Theta_{Y|c1})(1-\Theta_n)\Theta_{R^D|11}(1-\Theta_{R^Y|c01}),\\
\mathbb{P}_{+0+0|0} &= \Theta_{Y|n0}\Theta_n(1-\Theta_{R^D|00})(1-\Theta_{R^Y|n10}) + \Theta_{Y|c0}(1-\Theta_n)(1-\Theta_{R^D|00})(1-\Theta_{R^Y|c10})\\
&\quad + (1-\Theta_{Y|n0})\Theta_n(1-\Theta_{R^D|00})(1-\Theta_{R^Y|n00}) + (1-\Theta_{Y|c0})(1-\Theta_n)(1-\Theta_{R^D|00})(1-\Theta_{R^Y|c00}),
\end{aligned}$$

$$\begin{aligned}\mathbb{P}_{+0+0|1} = &\Theta_{Y|n0}\Theta_n(1-\Theta_{R^D|10})(1-\Theta_{R^Y|n10}) + (1-\Theta_{Y|n0})\Theta_n(1-\Theta_{R^D|10})(1-\Theta_{R^Y|n00}) \\ &+ \Theta_{Y|c1}(1-\Theta_n)(1-\Theta_{R^D|11})(1-\Theta_{R^Y|c10}) + (1-\Theta_{Y|c1})(1-\Theta_n)(1-\Theta_{R^D|11})(1-\Theta_{R^Y|c00}).\end{aligned}$$

These observable data probabilities are consistent with $\Theta_n = \frac{1}{4}$, $\Theta_{Y|n0} = \frac{1}{2}$, $\Theta_{Y|c0} = \frac{1}{3}$, $\Theta_{Y|c1} = \frac{1}{2}$, $\Theta_{R^D|00} = \frac{1}{2}$, $\Theta_{R^D|10} = \frac{1}{4}$, $\Theta_{R^D|11} = \frac{1}{3}$, $\Theta_{R^Y|n00} = \frac{1}{6}$, $\Theta_{R^Y|n10} = \frac{1}{4}$, $\Theta_{R^Y|c00} = \frac{1}{3}$, $\Theta_{R^Y|c10} = \frac{1}{3}$, $\Theta_{R^Y|n01} = \frac{1}{4}$, $\Theta_{R^Y|n11} = \frac{1}{3}$, $\Theta_{R^Y|c01} = \frac{1}{8}$, $\Theta_{R^Y|c11} = \frac{1}{6}$, and the CACE $= \frac{1}{6}$. They are also consistent with $\Theta_n = \frac{3}{8}$, $\Theta_{Y|n0} = \frac{1}{2}$, $\Theta_{Y|c0} = \frac{1}{10}$, $\Theta_{Y|c1} = \frac{1}{4}$, $\Theta_{R^D|00} = \frac{1}{2}$, $\Theta_{R^D|10} = \frac{1}{6}$, $\Theta_{R^D|11} = \frac{2}{5}$, $\Theta_{R^Y|n00} = \frac{1}{12}$, $\Theta_{R^Y|n10} = \frac{25}{48}$, $\Theta_{R^Y|c00} = \frac{11}{36}$, $\Theta_{R^Y|c10} = \frac{13}{48}$, $\Theta_{R^Y|n01} = \frac{1}{4}$, $\Theta_{R^Y|n11} = \frac{1}{3}$, $\Theta_{R^Y|c01} = \frac{1}{12}$, $\Theta_{R^Y|c11} = \frac{1}{3}$, and the CACE $= \frac{3}{20}$. Therefore, the CACE cannot be uniquely identified.

C.3.4 Assumption $1DY \oplus 2ZD$

For a binary Y with two-sided noncompliance, we consider the following observable data probabilities:

$$\begin{aligned}&(\mathbb{P}_{0111|0}, \mathbb{P}_{0011|0}, \mathbb{P}_{1111|0}, \mathbb{P}_{1011|0}, \mathbb{P}_{0111|1}, \mathbb{P}_{0011|1}, \mathbb{P}_{1111|1}, \mathbb{P}_{1011|1},\\ &\mathbb{P}_{1+01|0}, \mathbb{P}_{0+01|0}, \mathbb{P}_{1+01|1}, \mathbb{P}_{0+01|1}, \mathbb{P}_{01+0|0}, \mathbb{P}_{11+0|0}, \mathbb{P}_{01+0|1}, \mathbb{P}_{11+0|1}, \mathbb{P}_{+0+0|0}, \mathbb{P}_{+0+0|1})\\ &= \left(\frac{1}{32}, \frac{5}{64}, \frac{1}{64}, \frac{1}{8}, \frac{1}{48}, \frac{1}{32}, \frac{1}{32}, \frac{1}{6}, \frac{17}{192}, \frac{19}{192}, \frac{11}{96}, \frac{1}{12}, \frac{9}{64}, \frac{7}{64}, \frac{7}{96}, \frac{17}{96}, \frac{5}{16}, \frac{29}{96}\right).\end{aligned}$$

Define the parameters:

$$\begin{aligned}\Theta_{dy|z} &= \mathbb{P}(D=d, Y=y \mid Z=z) \text{ for } (z,d,y) = (0,0,0), (0,1,1), (0,1,0), (1,0,1), (1,0,0), (1,1,0),\\ \Theta_{R^D|zd} &= \mathbb{P}(R^D = 1 \mid Z=z, D=d) \text{ for } z=0,1 \text{ and } d=0,1,\\ \Theta_{R^Y|dyr^D} &= \mathbb{P}(R^Y=1 \mid D=d, Y=y, R^D=r^D) \text{ for } d=0,1,\ y=0,1, \text{ and } r^D=0,1.\end{aligned}$$

The observable data probabilities relate to the parameters as follows:

$$\begin{aligned}\mathbb{P}_{0111|0} &= (1-\Theta_{00|0}-\Theta_{11|0}-\Theta_{10|0})\Theta_{R^D|00}\Theta_{R^Y|011},\\ \mathbb{P}_{0011|0} &= \Theta_{00|0}\Theta_{R^D|00}\Theta_{R^Y|001},\\ \mathbb{P}_{1111|0} &= \Theta_{11|0}\Theta_{R^D|01}\Theta_{R^Y|111},\\ \mathbb{P}_{1011|0} &= \Theta_{10|0}\Theta_{R^D|01}\Theta_{R^Y|101},\\ \mathbb{P}_{0111|1} &= \Theta_{01|1}\Theta_{R^D|10}\Theta_{R^Y|011},\\ \mathbb{P}_{0011|1} &= \Theta_{00|1}\Theta_{R^D|10}\Theta_{R^Y|001},\\ \mathbb{P}_{1111|1} &= (1-\Theta_{01|1}-\Theta_{00|1}-\Theta_{10|1})\Theta_{R^D|11}\Theta_{R^Y|111},\\ \mathbb{P}_{1011|1} &= \Theta_{10|1}\Theta_{R^D|11}\Theta_{R^Y|101},\\ \mathbb{P}_{1+01|0} &= (1-\Theta_{00|0}-\Theta_{11|0}-\Theta_{10|0})(1-\Theta_{R^D|00})\Theta_{R^Y|010} + \Theta_{11|0}(1-\Theta_{R^D|01})\Theta_{R^Y|110},\\ \mathbb{P}_{0+01|0} &= \Theta_{00|0}(1-\Theta_{R^D|00})\Theta_{R^Y|000} + \Theta_{10|0}(1-\Theta_{R^D|01})\Theta_{R^Y|100},\\ \mathbb{P}_{1+01|1} &= \Theta_{01|1}(1-\Theta_{R^D|10})\Theta_{R^Y|010} + (1-\Theta_{01|1}-\Theta_{00|1}-\Theta_{10|1})(1-\Theta_{R^D|11})\Theta_{R^Y|110},\\ \mathbb{P}_{0+01|1} &= \Theta_{00|1}(1-\Theta_{R^D|10})\Theta_{R^Y|000} + \Theta_{10|1}(1-\Theta_{R^D|11})\Theta_{R^Y|100},\end{aligned}$$

$$
\begin{aligned}
\mathbb{P}_{01+0|0} &= \Theta_{00|0}\Theta_{R^D|00}(1-\Theta_{R^Y|001}) + (1-\Theta_{00|0}-\Theta_{11|0}-\Theta_{10|0})\Theta_{R^D|00}(1-\Theta_{R^Y|011}),\\
\mathbb{P}_{11+0|0} &= \Theta_{10|0}\Theta_{R^D|01}(1-\Theta_{R^Y|101}) + \Theta_{11|0}\Theta_{R^D|01}(1-\Theta_{R^Y|111}),\\
\mathbb{P}_{01+0|1} &= \Theta_{00|1}\Theta_{R^D|10}(1-\Theta_{R^Y|001}) + \Theta_{01|1}\Theta_{R^D|10}(1-\Theta_{R^Y|011}),\\
\mathbb{P}_{11+0|1} &= \Theta_{10|1}\Theta_{R^D|11}(1-\Theta_{R^Y|101}) + (1-\Theta_{01|1}-\Theta_{00|1}-\Theta_{10|1})\Theta_{R^D|11}(1-\Theta_{R^Y|111}),\\
\mathbb{P}_{+0+0|0} &= \Theta_{00|0}(1-\Theta_{R^D|00})(1-\Theta_{R^Y|000}) + (1-\Theta_{00|0}-\Theta_{11|0}-\Theta_{10|0})(1-\Theta_{R^D|00})(1-\Theta_{R^Y|010})\\
&\quad + \Theta_{10|0}(1-\Theta_{R^D|01})(1-\Theta_{R^Y|100}) + \Theta_{11|0}(1-\Theta_{R^D|01})(1-\Theta_{R^Y|110}),\\
\mathbb{P}_{+0+0|1} &= \Theta_{00|1}(1-\Theta_{R^D|10})(1-\Theta_{R^Y|000}) + \Theta_{01|1}(1-\Theta_{R^D|10})(1-\Theta_{R^Y|010})\\
&\quad + \Theta_{10|1}(1-\Theta_{R^D|11})(1-\Theta_{R^Y|100}) + (1-\Theta_{01|1}-\Theta_{00|1}-\Theta_{10|1})(1-\Theta_{R^D|11})(1-\Theta_{R^Y|110}).
\end{aligned}
$$

These observable data probabilities are consistent with $\Theta_{10|0}=\frac{3}{8}$, $\Theta_{11|0}=\frac{1}{8}$, $\Theta_{00|0}=\frac{5}{16}$, $\Theta_{00|1}=\frac{1}{8}$, $\Theta_{01|1}=\frac{1}{8}$, $\Theta_{10|1}=\frac{1}{2}$, $\Theta_{R^D|00}=\frac{1}{2}$, $\Theta_{R^D|01}=\frac{1}{2}$, $\Theta_{R^D|10}=\frac{1}{2}$, $\Theta_{R^D|11}=\frac{1}{2}$, $\Theta_{R^Y|001}=\frac{1}{2}$, $\Theta_{R^Y|011}=\frac{1}{3}$, $\Theta_{R^Y|101}=\frac{2}{3}$, $\Theta_{R^Y|111}=\frac{1}{4}$, $\Theta_{R^Y|000}=\frac{1}{3}$, $\Theta_{R^Y|010}=\frac{1}{2}$, $\Theta_{R^Y|100}=\frac{1}{4}$, $\Theta_{R^Y|110}=\frac{2}{3}$, and the CACE $=\frac{1}{4}$. They are also consistent with $\Theta_{10|0}=\frac{3}{8}$, $\Theta_{11|0}=\frac{1}{8}$, $\Theta_{00|0}=\frac{5}{16}$, $\Theta_{00|1}=\frac{1}{6}$, $\Theta_{01|1}=\frac{1}{6}$, $\Theta_{10|1}=\frac{4}{9}$, $\Theta_{R^D|00}=\frac{1}{2}$, $\Theta_{R^D|01}=\frac{1}{2}$, $\Theta_{R^D|10}=\frac{3}{8}$, $\Theta_{R^D|11}=\frac{9}{16}$, $\Theta_{R^Y|001}=\frac{1}{2}$, $\Theta_{R^Y|011}=\frac{1}{3}$, $\Theta_{R^Y|101}=\frac{2}{3}$, $\Theta_{R^Y|111}=\frac{1}{4}$, $\Theta_{R^Y|000}=\frac{1}{3}$, $\Theta_{R^Y|010}=\frac{5}{9}$, $\Theta_{R^Y|100}=\frac{1}{4}$, $\Theta_{R^Y|110}=\frac{7}{12}$, and the CACE $=\frac{11}{24}$. Therefore, the CACE cannot be uniquely identified.

C.3.5 Assumption $1ZY \oplus 2ZD$

For a binary Y with two-sided noncompliance, we consider the following observable data probabilities:

$$
\begin{aligned}
&(\mathbb{P}_{0111|0}, \mathbb{P}_{0011|0}, \mathbb{P}_{1111|0}, \mathbb{P}_{1011|0}, \mathbb{P}_{0111|1}, \mathbb{P}_{0011|1}, \mathbb{P}_{1111|1}, \mathbb{P}_{1011|1},\\
&\ \mathbb{P}_{1+01|0}, \mathbb{P}_{0+01|0}, \mathbb{P}_{1+01|1}, \mathbb{P}_{0+01|1}, \mathbb{P}_{01+0|0}, \mathbb{P}_{11+0|0}, \mathbb{P}_{01+0|1}, \mathbb{P}_{11+0|1}, \mathbb{P}_{+0+0|0}, \mathbb{P}_{+0+0|1})\\
&= \left(\frac{5}{96}, \frac{3}{64}, \frac{1}{96}, \frac{3}{64}, \frac{1}{64}, \frac{1}{24}, \frac{1}{64}, \frac{1}{12}, \frac{1}{8}, \frac{1}{8}, \frac{1}{6}, \frac{7}{64}, \frac{29}{192}, \frac{13}{192}, \frac{13}{192}, \frac{17}{192}, \frac{3}{8}, \frac{79}{192}\right).
\end{aligned}
$$

Define the parameters:

$$
\begin{aligned}
\Theta_{dy|z} &= \mathbb{P}(D=d, Y=y \mid Z=z) \text{ for } (z,d,y) = (0,0,0),(0,1,1),(0,1,0),(1,0,1),(1,0,0),(1,1,0),\\
\Theta_{R^D|zd} &= \mathbb{P}(R^D=1 \mid Z=z, D=d) \text{ for } z=0,1 \text{ and } d=0,1,\\
\Theta_{R^Y|zyr^D} &= \mathbb{P}(R^Y=1 \mid Z=z, Y=y, R^D=r^D) \text{ for } z=0,1,\ y=0,1,\ \text{and } r^D=0,1.
\end{aligned}
$$

The observable data probabilities relate to the parameters as follows:

$$
\begin{aligned}
\mathbb{P}_{0111|0} &= (1-\Theta_{00|0}-\Theta_{11|0}-\Theta_{10|0})\Theta_{R^D|00}\Theta_{R^Y|011},\\
\mathbb{P}_{0011|0} &= \Theta_{00|0}\Theta_{R^D|00}\Theta_{R^Y|001},\\
\mathbb{P}_{1111|0} &= \Theta_{11|0}\Theta_{R^D|01}\Theta_{R^Y|011},\\
\mathbb{P}_{1011|0} &= \Theta_{10|0}\Theta_{R^D|01}\Theta_{R^Y|001},\\
\mathbb{P}_{0111|1} &= \Theta_{01|1}\Theta_{R^D|10}\Theta_{R^Y|111},\\
\mathbb{P}_{0011|1} &= \Theta_{00|1}\Theta_{R^D|10}\Theta_{R^Y|101},
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{P}_{1111|1} &= (1-\Theta_{01|1}-\Theta_{00|1}-\Theta_{10|1})\Theta_{R^D|11}\Theta_{R^Y|111},\\
\mathbb{P}_{1011|1} &= \Theta_{10|1}\Theta_{R^D|11}\Theta_{R^Y|101},\\
\mathbb{P}_{1+01|0} &= (1-\Theta_{00|0}-\Theta_{11|0}-\Theta_{10|0})(1-\Theta_{R^D|00})\Theta_{R^Y|010}+\Theta_{11|0}(1-\Theta_{R^D|01})\Theta_{R^Y|010},\\
\mathbb{P}_{0+01|0} &= \Theta_{00|0}(1-\Theta_{R^D|00})\Theta_{R^Y|000}+\Theta_{10|0}(1-\Theta_{R^D|01})\Theta_{R^Y|000},\\
\mathbb{P}_{1+01|1} &= \Theta_{01|1}(1-\Theta_{R^D|10})\Theta_{R^Y|110}+(1-\Theta_{01|1}-\Theta_{00|1}-\Theta_{10|1})(1-\Theta_{R^D|11})\Theta_{R^Y|110},\\
\mathbb{P}_{0+01|1} &= \Theta_{00|1}(1-\Theta_{R^D|10})\Theta_{R^Y|100}+\Theta_{10|1}(1-\Theta_{R^D|11})\Theta_{R^Y|100},\\
\mathbb{P}_{01+0|0} &= \Theta_{00|0}\Theta_{R^D|00}(1-\Theta_{R^Y|001})+(1-\Theta_{00|0}-\Theta_{11|0}-\Theta_{10|0})\Theta_{R^D|00}(1-\Theta_{R^Y|011}),\\
\mathbb{P}_{11+0|0} &= \Theta_{10|0}\Theta_{R^D|01}(1-\Theta_{R^Y|001})+\Theta_{11|0}\Theta_{R^D|01}(1-\Theta_{R^Y|011}),\\
\mathbb{P}_{01+0|1} &= \Theta_{00|1}\Theta_{R^D|10}(1-\Theta_{R^Y|101})+\Theta_{01|1}\Theta_{R^D|10}(1-\Theta_{R^Y|111}),\\
\mathbb{P}_{11+0|1} &= \Theta_{10|1}\Theta_{R^D|11}(1-\Theta_{R^Y|101})+(1-\Theta_{01|1}-\Theta_{00|1}-\Theta_{10|1})\Theta_{R^D|11}(1-\Theta_{R^Y|111}),\\
\mathbb{P}_{+0+0|0} &= \Theta_{00|0}(1-\Theta_{R^D|00})(1-\Theta_{R^Y|000})+(1-\Theta_{00|0}-\Theta_{11|0}-\Theta_{10|0})(1-\Theta_{R^D|00})(1-\Theta_{R^Y|010})\\
&\quad+\Theta_{10|0}(1-\Theta_{R^D|01})(1-\Theta_{R^Y|000})+\Theta_{11|0}(1-\Theta_{R^D|01})(1-\Theta_{R^Y|010}),\\
\mathbb{P}_{+0+0|1} &= \Theta_{00|1}(1-\Theta_{R^D|10})(1-\Theta_{R^Y|100})+\Theta_{01|1}(1-\Theta_{R^D|10})(1-\Theta_{R^Y|110})\\
&\quad+\Theta_{10|1}(1-\Theta_{R^D|11})(1-\Theta_{R^Y|100})+(1-\Theta_{01|1}-\Theta_{00|1}-\Theta_{10|1})(1-\Theta_{R^D|11})(1-\Theta_{R^Y|110}).
\end{aligned}
$$

These observable data probabilities are consistent with $\Theta_{10|0}=\frac{3}{8}$, $\Theta_{11|0}=\frac{1}{8}$, $\Theta_{00|0}=\frac{3}{16}$, $\Theta_{00|1}=\frac{1}{8}$, $\Theta_{01|1}=\frac{1}{8}$, $\Theta_{10|1}=\frac{1}{2}$, $\Theta_{R^D|00}=\frac{1}{2}$, $\Theta_{R^D|01}=\frac{1}{4}$, $\Theta_{R^D|10}=\frac{1}{2}$, $\Theta_{R^D|11}=\frac{1}{4}$, $\Theta_{R^Y|001}=\frac{1}{2}$, $\Theta_{R^Y|011}=\frac{1}{3}$, $\Theta_{R^Y|101}=\frac{2}{3}$, $\Theta_{R^Y|111}=\frac{1}{4}$, $\Theta_{R^Y|000}=\frac{1}{3}$, $\Theta_{R^Y|010}=\frac{1}{2}$, $\Theta_{R^Y|100}=\frac{1}{4}$, $\Theta_{R^Y|110}=\frac{2}{3}$, and the CACE $=-\frac{1}{4}$. They are also consistent with $\Theta_{10|0}=\frac{1}{4}$, $\Theta_{11|0}=\frac{1}{12}$, $\Theta_{00|0}=\frac{1}{4}$, $\Theta_{00|1}=\frac{1}{8}$, $\Theta_{01|1}=\frac{1}{8}$, $\Theta_{10|1}=\frac{1}{2}$, $\Theta_{R^D|00}=\frac{3}{8}$, $\Theta_{R^D|01}=\frac{3}{8}$, $\Theta_{R^D|10}=\frac{1}{2}$, $\Theta_{R^D|11}=\frac{1}{4}$, $\Theta_{R^Y|001}=\frac{1}{2}$, $\Theta_{R^Y|011}=\frac{1}{3}$, $\Theta_{R^Y|101}=\frac{2}{3}$, $\Theta_{R^Y|111}=\frac{1}{4}$, $\Theta_{R^Y|000}=\frac{2}{5}$, $\Theta_{R^Y|010}=\frac{2}{5}$, $\Theta_{R^Y|100}=\frac{1}{4}$, $\Theta_{R^Y|110}=\frac{2}{3}$, and the CACE $=-\frac{3}{10}$. Therefore, the CACE cannot be uniquely identified.

C.3.6 Assumption $1U\oplus 2ZD$

For a binary Y with two-sided noncompliance, we consider the following observable data probabilities:

$$
\begin{aligned}
&(\mathbb{P}_{0111|0},\mathbb{P}_{0011|0},\mathbb{P}_{1111|0},\mathbb{P}_{1011|0},\mathbb{P}_{0111|1},\mathbb{P}_{0011|1},\mathbb{P}_{1111|1},\mathbb{P}_{1011|1},\\
&\ \mathbb{P}_{1+01|0},\mathbb{P}_{0+01|0},\mathbb{P}_{1+01|1},\mathbb{P}_{0+01|1},\mathbb{P}_{01+0|0},\mathbb{P}_{11+0|0},\mathbb{P}_{01+0|1},\mathbb{P}_{11+0|1},\mathbb{P}_{+0+0|0},\mathbb{P}_{+0+0|1})\\
&=\left(\frac{1}{32},\frac{5}{96},\frac{1}{96},\frac{1}{96},\frac{1}{32},\frac{1}{32},\frac{5}{288},\frac{5}{288},\frac{47}{576},\frac{7}{64},\frac{59}{576},\frac{59}{576},\frac{1}{6},\frac{1}{24},\frac{1}{16},\frac{13}{144},\frac{143}{288},\frac{157}{288}\right).
\end{aligned}
$$

Define the parameters:

$$
\begin{aligned}
\Theta_u &= \mathbb{P}(U=u) \text{ for } u=a,n,\\
\Theta_{Y|ud} &= \mathbb{P}(Y=1\mid U=u,D=d) \text{ for } (u,d)=(a,1),(n,0),(c,1),(c,0),\\
\Theta_{R^D|zd} &= \mathbb{P}(R^D=1\mid Z=z,D=d) \text{ for } z=0,1 \text{ and } d=0,1,\\
\Theta_{R^Y|ur^D} &= \mathbb{P}(R^Y=1\mid U=u,R^D=r^D) \text{ for } u=a,n,c \text{ and } r^D=0,1.
\end{aligned}
$$

The observable data probabilities relate to the parameters as follows:

$$
\begin{aligned}
\mathbb{P}_{0111|0} &= \Theta_{Y|n0}\Theta_n\Theta_{R^D|00}\Theta_{R^Y|n1} + \Theta_{Y|c0}(1-\Theta_n-\Theta_a)\Theta_{R^D|00}\Theta_{R^Y|c1},\\
\mathbb{P}_{0011|0} &= (1-\Theta_{Y|n0})\Theta_n\Theta_{R^D|00}\Theta_{R^Y|n1} + (1-\Theta_{Y|c0})(1-\Theta_n-\Theta_a)\Theta_{R^D|00}\Theta_{R^Y|c1},\\
\mathbb{P}_{1111|0} &= \Theta_{Y|a1}\Theta_a\Theta_{R^D|01}\Theta_{R^Y|a1},\\
\mathbb{P}_{1011|0} &= (1-\Theta_{Y|a1})\Theta_a\Theta_{R^D|01}\Theta_{R^Y|a1},\\
\mathbb{P}_{0111|1} &= \Theta_{Y|n0}\Theta_n\Theta_{R^D|10}\Theta_{R^Y|n1},\\
\mathbb{P}_{0011|1} &= (1-\Theta_{Y|n0})\Theta_n\Theta_{R^D|10}\Theta_{R^Y|n1},\\
\mathbb{P}_{1111|1} &= \Theta_{Y|a1}\Theta_a\Theta_{R^D|11}\Theta_{R^Y|a1} + \Theta_{Y|c1}(1-\Theta_n-\Theta_a)\Theta_{R^D|11}\Theta_{R^Y|c1},\\
\mathbb{P}_{1011|1} &= (1-\Theta_{Y|a1})\Theta_a\Theta_{R^D|11}\Theta_{R^Y|a1} + (1-\Theta_{Y|c1})(1-\Theta_n-\Theta_a)\Theta_{R^D|11}\Theta_{R^Y|c1},\\
\mathbb{P}_{1+01|0} &= \Theta_{Y|n0}\Theta_n(1-\Theta_{R^D|00})\Theta_{R^Y|n0} + \Theta_{Y|a1}\Theta_a(1-\Theta_{R^D|01})\Theta_{R^Y|a0}\\
&\quad + \Theta_{Y|c0}(1-\Theta_n-\Theta_a)(1-\Theta_{R^D|00})\Theta_{R^Y|c0},\\
\mathbb{P}_{0+01|0} &= (1-\Theta_{Y|n0})\Theta_n(1-\Theta_{R^D|00})\Theta_{R^Y|n0} + (1-\Theta_{Y|a1})\Theta_a(1-\Theta_{R^D|01})\Theta_{R^Y|a0}\\
&\quad + (1-\Theta_{Y|c0})(1-\Theta_n-\Theta_a)(1-\Theta_{R^D|00})\Theta_{R^Y|c0},\\
\mathbb{P}_{1+01|1} &= \Theta_{Y|n0}\Theta_n(1-\Theta_{R^D|10})\Theta_{R^Y|n0} + \Theta_{Y|a1}\Theta_a(1-\Theta_{R^D|11})\Theta_{R^Y|a0}\\
&\quad + \Theta_{Y|c1}(1-\Theta_n-\Theta_a)(1-\Theta_{R^D|11})\Theta_{R^Y|c0},\\
\mathbb{P}_{0+01|1} &= (1-\Theta_{Y|n0})\Theta_n(1-\Theta_{R^D|10})\Theta_{R^Y|n0} + (1-\Theta_{Y|a1})\Theta_a(1-\Theta_{R^D|11})\Theta_{R^Y|a0}\\
&\quad + (1-\Theta_{Y|c1})(1-\Theta_n-\Theta_a)(1-\Theta_{R^D|11})\Theta_{R^Y|c0},\\
\mathbb{P}_{01+0|0} &= \Theta_n\Theta_{R^D|00}(1-\Theta_{R^Y|n1}) + (1-\Theta_n-\Theta_a)\Theta_{R^D|00}(1-\Theta_{R^Y|c1}),\\
\mathbb{P}_{11+0|0} &= \Theta_a\Theta_{R^D|01}(1-\Theta_{R^Y|a1}),\\
\mathbb{P}_{01+0|1} &= \Theta_n\Theta_{R^D|10}(1-\Theta_{R^Y|n1}),\\
\mathbb{P}_{11+0|1} &= \Theta_a\Theta_{R^D|11}(1-\Theta_{R^Y|a1}) + (1-\Theta_n-\Theta_a)\Theta_{R^D|11}(1-\Theta_{R^Y|c1}),\\
\mathbb{P}_{+0+0|0} &= \Theta_n(1-\Theta_{R^D|00})(1-\Theta_{R^Y|n0}) + \Theta_a(1-\Theta_{R^D|01})(1-\Theta_{R^Y|a0})\\
&\quad + (1-\Theta_n-\Theta_a)(1-\Theta_{R^D|00})(1-\Theta_{R^Y|c0}),\\
\mathbb{P}_{+0+0|1} &= \Theta_n(1-\Theta_{R^D|10})(1-\Theta_{R^Y|n0}) + \Theta_a(1-\Theta_{R^D|11})(1-\Theta_{R^Y|a0})\\
&\quad + (1-\Theta_n-\Theta_a)(1-\Theta_{R^D|11})(1-\Theta_{R^Y|c0}).
\end{aligned}
$$

These observable data probabilities are consistent with $\Theta_n = \frac{1}{4}$, $\Theta_a = \frac{1}{4}$, $\Theta_{Y|n0} = \frac{1}{2}$, $\Theta_{Y|c0} = \frac{1}{4}$, $\Theta_{Y|a1} = \frac{1}{2}$, $\Theta_{Y|c1} = \frac{1}{2}$, $\Theta_{R^D|00} = \frac{1}{3}$, $\Theta_{R^D|01} = \frac{1}{4}$, $\Theta_{R^D|10} = \frac{1}{2}$, $\Theta_{R^D|11} = \frac{1}{6}$, $\Theta_{R^Y|n0} = \frac{1}{4}$, $\Theta_{R^Y|a0} = \frac{1}{2}$, $\Theta_{R^Y|c0} = \frac{1}{6}$, $\Theta_{R^Y|n1} = \frac{1}{2}$, $\Theta_{R^Y|a1} = \frac{1}{3}$, $\Theta_{R^Y|c1} = \frac{1}{4}$, and the CACE $= \frac{1}{4}$. They are also consistent with $\Theta_n = \frac{1}{4}$, $\Theta_a = \frac{7}{16}$, $\Theta_{Y|n0} = \frac{1}{2}$, $\Theta_{Y|c0} = \frac{1}{8}$, $\Theta_{Y|a1} = \frac{1}{2}$, $\Theta_{Y|c1} = \frac{1}{2}$, $\Theta_{R^D|00} = \frac{4}{9}$, $\Theta_{R^D|01} = \frac{1}{7}$, $\Theta_{R^D|10} = \frac{1}{2}$, $\Theta_{R^D|11} = \frac{1}{6}$, $\Theta_{R^Y|n0} = \frac{11}{312}$, $\Theta_{R^Y|a0} = \frac{31}{78}$, $\Theta_{R^Y|c0} = \frac{16}{75}$, $\Theta_{R^Y|n1} = \frac{1}{2}$, $\Theta_{R^Y|a1} = \frac{1}{3}$, $\Theta_{R^Y|c1} = \frac{1}{5}$, and the CACE $= \frac{3}{8}$. Therefore, the CACE cannot be uniquely identified.

References

J. D. Angrist, G. W. Imbens, and D. B. Rubin. Identification of causal effects using instrumental variables (with discussion). *Journal of the American Statistical Association*, 91 (434):444–455, 1996.

A. Balke and J. Pearl. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439):1171–1176, 1997.

R. Calvi, A. Lewbel, and D. Tommasi. LATE with missing or mismeasured treatment. *Journal of Business & Economic Statistics*, 40(4):1701–1717, 2022.

D. Card. Using geographic variation in college proximity to estimate the return to schooling. Working Paper 4483, National Bureau of Economic Research, Cambridge, MA, 1993.

H. Chen, Z. Geng, and X. H. Zhou. Identifiability and estimation of causal effects in randomized trials with noncompliance and completely nonignorable missing data. *Biometrics*, 65(3):675–682, 2009a.

H. Chen, Z. Geng, and X. H. Zhou. Rejoinder. *Biometrics*, 65(3):689–691, 2009b.

H. Chen, P. Ding, Z. Geng, and X.-H. Zhou. Semiparametric inference of the complier average causal effect with nonignorable missing outcomes. *ACM Transactions on Intelligent Systems and Technology*, 7(2):19:1–19:15, 2015.

P. Ding. *A First Course in Causal Inference*. Chapman and Hall/CRC, New York, 2024.

O. Eren and S. Ozbeklik. The effect of noncognitive ability on the earnings of young men: A distributional analysis with measurement error correction. *Labour Economics*, 24:293–304, 2013.

R. E. Fay. Causal models for patterns of nonresponse. *Journal of the American Statistical Association*, 81(394):354–365, 1986.

C. E. Frangakis and D. B. Rubin. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika*, 86(2):365–379, 1999.

M. Frölich and M. Huber. Treatment evaluation with multiple outcome periods under endogeneity and attrition. *Journal of the American Statistical Association*, 109(508): 1697–1711, 2014.

M. Huber. On the plausibility of the latent ignorability assumption. *Econometrics*, 9(4): 47, 2021.

K. Imai. Statistical analysis of randomized experiments with non-ignorable missing binary outcomes: An application to a voting experiment. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 58(1):83–104, 2009.

G. W. Imbens and J. D. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475, 1994.

M. A. Jay, L. McGrath-Lone, and R. Gilbert. Data resource: The National Pupil Database (NPD). *International Journal of Population Data Science*, 4(1):1101, 2019.

B. Jo, E. M. Ginexi, and N. S. Ialongo. Handling missing data in randomized experiments with noncompliance. *Prevention science : the official journal of the Society for Prevention Research*, 11(4):384–396, 2010.

Y. Li, W. Miao, I. Shpitser, and E. J. Tchetgen Tchetgen. A self-censoring model for multivariate nonignorable nonmonotone missing data. *Biometrics*, 79(4):3203–3214, 2023.

K. J. Lui and K. C. Chang. Notes on odds ratio estimation for a randomized clinical trial with noncompliance and missing outcomes. *Journal of Applied Statistics*, 37(12): 2057–2071, 2010.

W. Q. Ma, Z. Geng, and Y. H. Hu. Identification of graphical models for nonignorable nonresponse of binary outcomes in longitudinal studies. *Journal of Multivariate Analysis*, 87(1):24–45, 2003.

A. Mattei and F. Mealli. Application of the principal stratification approach to the Faenza randomized experiment on breast self-examination. *Biometrics*, 63(2):437–446, 2007.

F. Mealli, G. W. Imbens, S. Ferro, and A. Biggeri. Analyzing a randomized trial on breast self-examination with noncompliance and missing outcomes. *Biostatistics*, 5(2):207–222, 2004.

K. Mohan and J. Pearl. Graphical models for processing missing data. *Journal of the American Statistical Association*, 116(534):1023–1037, 2021.

W. K. Newey and J. L. Powell. Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578, 2003.

T. Q. Nguyen, M. C. Carlson, and E. A. Stuart. Identification of complier and noncomplier average causal effects in the presence of latent missing-at-random (LMAR) outcomes: a unifying view and choices of assumptions. *Biostatistics*, 25(4):978–996, 2024.

A. J. O'Malley and S. L. Normand. Likelihood methods for treatment noncompliance and subsequent nonresponse in randomized trials. *Biometrics*, 61(2):325–334, 2005.

Y. Peng, R. J. Little, and T. E. Raghunathan. An extended general location model for causal inferences from data subject to noncompliance and missing values. *Biometrics*, 60 (3):598–607, 2004.

B. Rai. Efficient estimation with missing data and endogeneity. *Econometric Reviews*, 42 (2):220–239, 2023.

P. Schochet, J. Burghardt, and S. Glazerman. National Job Corps Study: The impacts of Job Corps on participants' employment and related outcomes. Technical report, Mathematica Policy Research, Princeton, NJ, 2001.

P. Z. Schochet, J. Burghardt, and S. McConnell. National Job Corps Study and longer-term follow-up study: Impact and benefit-cost findings using survey and summary earnings records data. Technical report, Employment and Training Administration, U.S. Department of Labor, Washington, DC, 2006.

D. S. Small and J. Cheng. Discussion of "identifiability and estimation of causal effects in randomized trials with noncompliance and completely nonignorable missing data". *Biometrics*, 65(3):682–691, 2009.

L. Taylor and X. H. Zhou. Multiple imputation methods for treatment noncompliance and nonresponse in randomized clinical trials. *Biometrics*, 65(1):88–95, 2009.

S. von Hinke, G. Davey Smith, D. A. Lawlor, C. Propper, and F. Windmeijer. Genetic markers as instrumental variables. *Journal of Health Economics*, 45:131–148, 2016.

S. von Hinke Kessler Scholder, G. L. Wehby, S. Lewis, and L. Zuccolo. Alcohol exposure in utero and child academic achievement. *Economic Journal*, 124(576):634–667, 2014.

X. H. Zhou and S. M. Li. ITT analysis of randomized encouragement design studies with missing data. *Statistics in Medicine*, 25(16):2737–2761, 2006.

L. Zuccolo, N. Fitz-Simon, R. Gray, S. M. Ring, K. Sayal, G. D. Smith, and S. J. Lewis. A non-synonymous variant in ADH1B is strongly associated with prenatal alcohol use in a European sample of pregnant women. *Human Molecular Genetics*, 18(22):4457–4466, 2009.

S. Zuo, D. Ghosh, P. Ding, and F. Yang. Mediation analysis with the mediator and outcome missing not at random. *Journal of the American Statistical Association*, 120(550):794–804, 2025.