

| Statistica Sinica Preprint No: SS-2025-0210 | |
|--|---|
| Title | Discussion of “Causal and counterfactual views of missing data models” by Razieh Nabi, Rohit Bhattacharya, Ilya Shpitser, and James M. Robins |
| Manuscript ID | SS-2025-0210 |
| URL | http://www.stat.sinica.edu.tw/statistica/ |
| DOI | 10.5705/ss.202025.0210 |
| Complete List of Authors | Alexander W. Levis and Edward H. Kennedy |
| Corresponding Authors | Edward H. Kennedy |
| E-mails | edward@stat.cmu.edu |
| Notice: Accepted author version. | |

Discussion of “Causal and counterfactual views of missing data models” by Razieh Nabi, Rohit Bhattacharya, Ilya Shpitser, and James M. Robins

Alexander W. Levis & Edward H. Kennedy

Department of Statistics & Data Science,
Carnegie Mellon University

alevis@cmu.edu; edward@stat.cmu.edu

1 Introduction

We congratulate [Nabi et al. \(2022\)](#) on their impressive and insightful paper, which illustrates the benefits of using causal/counterfactual perspectives and tools in missing data problems. This paper represents an important approach to missing-not-at-random (MNAR) problems, exploiting nonparametric independence restrictions for identification, as opposed to parametric/semiparametric models, or resorting to sensitivity analysis. Crucially, the authors represent these restrictions with missing data directed acyclic graphs (m-DAGs), which can be useful to determine identification in complex and interesting MNAR models. In this discussion we consider (i) how/whether other tools from causal inference could be useful in missing data problems, (ii) problems that combine missing data and causal inference together, and (iii) some work on estimation in one of the authors’ example MNAR models.

2 Other causal identification tools

The graphical arguments used in [Nabi et al. \(2022\)](#) stem from a powerful analogy to a widely used set of tools in the causal inference literature. In drawing this connection, their work raises the question: *which other tools and methodologies can be borrowed from the causal literature and bear fruit for missing data problems?*

The focus of [Nabi et al. \(2022\)](#) is—for the most part—on characterizing identification of the full data law, $p(L^{(1)}, R)$, under sets of (factual and counterfactual) conditional independence relations implied by a directed acyclic graph and its associated structural model. However, it is often sufficient for practical purposes to identify and estimate certain functions or functionals of this distribution, say the marginal “target” data mean functional $\mathbb{E}(L^{(1)})$, rather than the entire full law. Moreover, although nonparametric structural models and conditional independence relations represent a large and important set of possible restrictions on the full data law, other plausible structural restrictions may partially or point identify certain estimands of interest. Indeed, both *missing data instrumental variables* and *shadow variables* are ideas born out of the causal inference literature—the latter analogous to a negative control outcome

variable—which move beyond conditional independence restrictions and can yield partial or full identification of certain functionals of the target law.

Instrumental variables (IVs), while being much more widely adopted as tools for causal inference, have been increasingly studied in recent years for use in missing data problems (Tchetgen Tchetgen and Wirth, 2017; Sun et al., 2018), having been proposed for such purposes at least as early as Heckman (1979). Just as IVs can help identify causal effects in the presence of unmeasured exposure-outcome confounding, IVs for missing data can help identify aspects of the target law when the data are missing not at random (MNAR), i.e., missingness status is related to the missing variables themselves, such that adjusting for measured covariates is not sufficient. Specifically, an IV for a partially missing outcome is a completely observed cause of missingness that is unrelated to the outcome conditional on measured covariates (Tchetgen Tchetgen and Wirth, 2017). In the statistical causal inference literature, IVs have been used to partially identify causal effects (Balke and Pearl, 1997; Swanson et al., 2018; Levis et al., 2025). Point identification of the average treatment effect is possible under additional *homogeneity* assumptions, which rule out certain kinds of effect modification by the unmeasured confounders (Wang and Tchetgen Tchetgen, 2018; Hernán and Robins, 2020; Levis et al., 2024). Alternatively, under a *monotonicity* assumption, the average treatment effect among compliers—subjects who would be exposed under encouragement from the IV but not otherwise—is nonparametrically identified (Imbens and Angrist, 1994). In the missing data context, while the homogeneity approach has been employed (Tchetgen Tchetgen and Wirth, 2017), the types of nonparametrically identified local subgroup effects that result from monotonicity seem to not yet have been explored. More broadly, ideas from principal stratification (Frangakis and Rubin, 2002) may be interesting to consider for missing data problems.

As a parallel notion to IVs, the idea of negative control outcomes (Tchetgen Tchetgen, 2014; Park et al., 2024) for handling unmeasured confounding in causal inference has been recently proposed as a strategy for MNAR outcomes, under the name of “shadow variables” (Miao et al., 2015; Miao and Tchetgen Tchetgen, 2016; Li et al., 2023). A shadow variable is a completely measured predictor of the outcome that is independent of missingness status conditional on measured covariates and the potentially missing outcome. These structural requirements of course could be encoded in a DAG as in Nabi et al. (2022), but on their own are not sufficient for identification of the full law, i.e., the algorithms in Nabi et al. (2022) would (rightfully) fail. Instead, the shadow variable conditions imply certain moment restrictions on the missingness mechanism, which guarantee identification of the means of functions of the missing variable only when brought together with “bridge” or “representer” assumptions (Li et al., 2023). Interestingly, recent advances in the proximal causal inference literature make use of a pair of negative control variables to identify causal effects under unmeasured confounding (Miao et al., 2018; Cui et al., 2024). It would be interesting to consider such exploitation of multiple negative controls in the missing data context.

Lastly, while mentioned in Appendix S3 of Nabi et al. (2022), their work more broadly motivates the potential utility of single world intervention graphs (SWIGs; Richardson and Robins (2013)) in the presence of missing data. What appears particularly underexplored is the extent to which SWIGs can be used not only to elicit identification conditions for the distribution of full data variables, but also the effects of (perhaps partially missing) exposure variables. We briefly entertain this possibility in the next section.

3 Simultaneous intervention on treatment and missingness

To confine our discussion within the limits of a concrete example, consider the observational point exposure setting with partially missing exposures. Namely, suppose we observe $O = (X, R, RA, Y)$, where X are a set of baseline confounders, A is a partially missing exposure variable with $R = \mathbb{1}(A \text{ observed})$, and Y is an outcome of interest. This setting was considered in Williamson et al. (2012) and Kennedy (2020). Adopting the notational conventions of Nabi et al. (2022), we can equivalently posit the existence of $A^{(1)}$, the exposure variable under an intervention that sets $R = 1$, and write $O = (X, R, A, Y)$ where $A = RA^{(1)} + (1 - R)“?”$. Suppose interest lies in identifying and estimating $\mathbb{E}(Y^{a^{(1)}})$, i.e., the mean potential outcome under intervention that sets $A^{(1)}$ to $a^{(1)}$. Moreover, suppose that we are in the missing at random (MAR) setting where

$$R \perp\!\!\!\perp A^{(1)} \mid X, Y, \quad (1)$$

a particular example for which is illustrated in Figure 1a, adopting the graphical conventions used in Nabi et al. (2022).

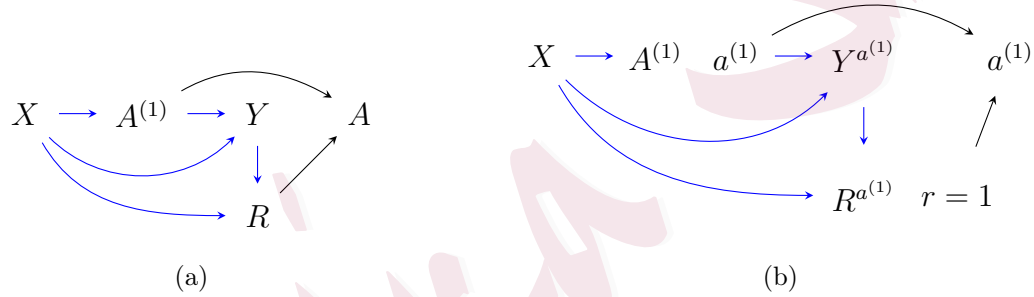


Figure 1: m-DAG and m-SWIG for missing point exposure

Under either the nonparametric structural equation model with independent errors (Pearl, 2009) or the finest fully randomized causally interpretable structured tree graph (Robins and Richardson, 2010), the m-DAG in Figure 1a implies

$$A^{(1)} \perp\!\!\!\perp Y^{a^{(1)}} \mid X. \quad (2)$$

Combining this no unmeasured confounding (NUC) condition with the MAR condition (1), Kennedy (2020) shows via Bayes' rule that

$$\mathbb{E}(Y^{a^{(1)}} \mid X) = \frac{\mathbb{E}(Y \lambda_{a^{(1)}}(X, Y) \mid X)}{\mathbb{E}(\lambda_{a^{(1)}}(X, Y) \mid X)}, \quad (3)$$

where $\lambda_a(X, Y) = \mathbb{P}[A = a \mid X, Y, R = 1]$.

Note that whereas (1) can be read off from the m-DAG immediately via d-separation, the NUC condition (2) cannot. In fact, this is a key innovation of SWIGs, in that under the above-mentioned causal models, this conditional independence on counterfactuals can be determined via d-separation after a node-splitting operation (Richardson and Robins, 2013). Consider the intervention that sets treatment to $a^{(1)}$ and the missingness indicator to 1. In

this case, to draw the corresponding SWIG, the node $A^{(1)}$ is split into random ($A^{(1)}$) and fixed ($a^{(1)}$) components, with the former retaining all incoming arrows from the original DAG, and the latter retaining all outgoing arrows; any (non-deterministic) descendant of the fixed node gets the counterfactual label $a^{(1)}$. The SWIG (or, perhaps, “m-SWIG”) corresponding to this intervention is displayed in Figure 1b, where we performed the same operation for the missingness indicator.

A couple of observations are worth noting. First, we can apparently in this example say more than (2), namely that $A^{(1)} \perp\!\!\!\perp (Y^{a^{(1)}}, R^{a^{(1)}}) \mid X$, since X is the only parent of $A^{(1)}$ in the SWIG. Interestingly, although we are in a MAR setting, the missingness indicator R is a descendant of $A^{(1)}$ (through the path $A^{(1)} \rightarrow Y \rightarrow R$) and thus the counterfactual missingness indicator $R^{a^{(1)}}$ appears on the SWIG. Second, by construction, m-DAGs and corresponding m-SWIGs contain deterministic relationships that follow from consistency, e.g., the variable A is a non-random function of $A^{(1)}$ and R . Thus, we could follow [Nabi et al. \(2022\)](#) and collapse $(A^{(1)}, A)$ as a single variable when the missingness indicator is equal to 1, though in Figure 1b we keep the deterministic $A^{A^{(1)}=a^{(1)}, r=1} \equiv a^{(1)}$ for clarity. Such deterministic relationships can subtly complicate the characterization of target law identification. For instance, the “twin networks” of [Balke and Pearl \(1994\)](#) and [Pearl \(2009\)](#) were an earlier graphical attempt to represent counterfactual independence relations, but for which d-separation failed to be complete due to the presence of deterministic relationships ([Richardson and Robins, 2013](#); [Shpitser et al., 2022](#)). It is plausible that such complications will not arise under the convention adopted by the authors whereby $L^{(1)}$ cannot be descendants of R and L . More generally, whether non-random variable dependencies have an impact on d-separation statements in m-SWIGs may be an important avenue for future inquiry.

While the above example may not be particularly interesting from an identification perspective (since (3) can be obtained by elementary means), it raises the question of whether SWIGs can provide insight into more complicated problems at the intersection of causal inference and missing data—say with partially missing confounders, treatments, and outcomes—or even problems purely in the missing data realm. The simplicity of SWIGs has made them increasingly popular in applied causal settings, and we hope to see them more fully investigated for missing data.

4 Statistical estimation considerations

Moving away from the usual missing at random assumptions can yield non-standard identification formulas for target parameters. As a result, efficient nonparametric estimation in these alternative paradigms often remains unexplored. We briefly illustrate some interesting challenges by considering estimation of mean outcomes in the permutation model first proposed by [Robins \(1997\)](#) and used by [Nabi et al. \(2022\)](#) as a primary example.

We relabel $Y \equiv L_1$ and $X \equiv L_2$, so that in the authors’ motivating example, $Y^{(1)}$ is the (possibly unmeasured) HIV status and $X^{(1)}$ denotes HIV risks and fears. The missing data DAG for this model is reproduced in Figure 2 below.

Under this model we have $R_1 \perp\!\!\!\perp Y^{(1)} \mid X^{(1)}$ and $R_2 \perp\!\!\!\perp (Y^{(1)}, X^{(1)}) \mid Y, R_1$; [Nabi et al.](#)

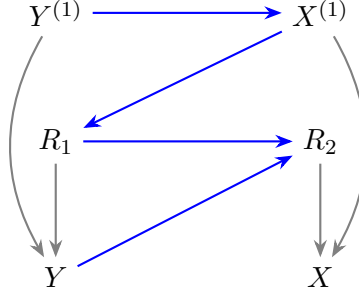


Figure 2: Directed acyclic graph for the permutation missingness model.

(2022) show how this implies that the joint distribution of $X^{(1)}$ and $Y^{(1)}$ is identified as

$$\begin{aligned} p_{X^{(1)}, Y^{(1)}}(x, y) &= \frac{p(x, y, R_1 = 1, R_2 = 1) \left\{ p(x \mid R_1 = 0, R_2 = 1) \mathbb{P}[R_1 = 0] + \int_{\mathcal{Y}} p(x \mid R_1 = 1, Y = y', R_2 = 1) p(R_1 = 1, y') dy' \right\}}{\mathbb{P}[R_2 = 1 \mid R_1 = 1, Y = y] \int_{\mathcal{Y}} p(x \mid R_1 = 1, Y = y', R_2 = 1) p(R_1 = 1, y') dy'} \\ &= p(R_1 = 1, y) p(x \mid R_1 = 1, Y = y, R_2 = 1) \left\{ 1 + \frac{p(x \mid R_1 = 0, R_2 = 1) \mathbb{P}[R_1 = 0]}{\int_{\mathcal{Y}} p(x \mid R_1 = 1, Y = y', R_2 = 1) p(R_1 = 1, y') dy'} \right\}. \end{aligned}$$

Suppose we do not wish to estimate the entire joint distribution, but instead $\psi = \mathbb{E}(Y^{(1)})$, e.g., the proportion of HIV-positive subjects in the previous example. The next proposition gives the identifying expression for this parameter under the permutation model.

Proposition 1. Assume the permutation missing model in Figure 2. Then $\psi = \mathbb{E}(Y^{(1)})$ is given by

$$\psi = \mathbb{P}(R_1 = 1) \mathbb{E}(Y \mid R_1 = 1) + \mathbb{P}(R_1 = 0) \theta,$$

where

$$\theta = \mathbb{E}(Y^{(1)} \mid R_1 = 0) = \mathbb{E} \left(\frac{\mathbb{E}\{Y/\zeta(Y) \mid R_1 = 1, R_2 = 1, X\}}{\mathbb{E}\{1/\zeta(Y) \mid R_1 = 1, R_2 = 1, X\}} \mid R_1 = 0, R_2 = 1 \right)$$

for $\zeta(Y) = \mathbb{P}[R_2 = 1 \mid R_1 = 1, Y]$.

Proof. By identification of the full law above, we have

$$\begin{aligned} \psi &= \int_{\mathcal{Y}} y \int_{\mathcal{X}} p_{X^{(1)}, Y^{(1)}}(x, y) dx dy \\ &= \int_{\mathcal{Y}} y p(R_1 = 1, y) dy \\ &\quad + \mathbb{P}[R_1 = 0] \int_{\mathcal{X}} p(x \mid R_1 = 0, R_2 = 1) \frac{\int_{\mathcal{Y}} y p(y \mid R_1 = 1) p(x \mid R_1 = 1, Y = y, R_2 = 1) dy}{\int_{\mathcal{Y}} p(y \mid R_1 = 1) p(x \mid R_1 = 1, Y = y, R_2 = 1) dy} dx \\ &= \mathbb{P}[R_1 = 1] \cdot \mathbb{E}(Y \mid R_1 = 1) \\ &\quad + \mathbb{P}[R_1 = 0] \cdot \mathbb{E} \left(\frac{\int_{\mathcal{Y}} y p(y \mid R_1 = 1) p(X \mid R_1 = 1, Y = y, R_2 = 1) dy}{\int_{\mathcal{Y}} p(y \mid R_1 = 1) p(X \mid R_1 = 1, Y = y, R_2 = 1) dy} \mid R_1 = 0, R_2 = 1 \right) \end{aligned}$$

Now by Bayes' rule $p(X \mid R_1 = 1, Y = y, R_2 = 1) = \frac{p(y \mid R_1=1, R_2=1, X) p(X \mid R_1=1, R_2=1)}{p(y \mid R_1=1, R_2=1)}$ and $\frac{p(y \mid R_1=1)}{p(y \mid R_1=1, R_2=1)} = \frac{\mathbb{P}[R_2=1 \mid R_1=1]}{\mathbb{P}[R_2=1 \mid R_1=1, Y]}$, which gives the result. \square

Remark 1. We use a particular parametrization that ensures all nuisances can be represented by conditional expectations (at the expense of preserving variation independence); alternative parametrizations are of course also possible (and later we consider a different parametrization when Y is binary).

Next we establish the nonparametric influence function of the functional θ , the variance of which represents a local asymptotic minimax lower bound for estimation of this quantity.

Lemma 1. Let $\zeta(Y) = \mathbb{P}[R_2 = 1 \mid R_1 = 1, Y]$ as before and define

$$\begin{aligned}\alpha(X) &= \mathbb{E}(1/\zeta(Y) \mid R_1 = 1, R_2 = 1, X) \\ \beta(X) &= \mathbb{E}(Y/\zeta(Y) \mid R_1 = 1, R_2 = 1, X) \\ \gamma(X) &= \mathbb{P}[R_1 = 1 \mid R_2 = 1, X] \\ \delta(Y) &= \mathbb{E}\left(\frac{1}{\alpha(X)} \frac{1 - \gamma(X)}{\gamma(X)} \left\{Y - \frac{\beta(X)}{\alpha(X)}\right\} \mid R_1 = 1, R_2 = 1, Y\right).\end{aligned}$$

Then the functional $\theta = \mathbb{E}(\beta(X)/\alpha(X) \mid R_1 = 0, R_2 = 1)$ satisfies the von Mises expansion

$$\theta(\bar{\mathbb{P}}) - \theta(\mathbb{P}) = \int \varphi(o; \bar{\mathbb{P}}) d(\bar{\mathbb{P}} - \mathbb{P})(o) + R_\theta(\bar{\mathbb{P}}; \mathbb{P}),$$

with influence function

$$\begin{aligned}\varphi(O; \mathbb{P}) &= \frac{1}{\mathbb{P}[R_1 = 0, R_2 = 1]} \left[(1 - R_1)R_2 \left\{ \frac{\beta(X)}{\alpha(X)} - \theta \right\} + \frac{R_1}{\alpha(X)} \frac{R_2}{\zeta(Y)} \frac{1 - \gamma(X)}{\gamma(X)} \left\{ Y - \frac{\beta(X)}{\alpha(X)} \right\} \right. \\ &\quad \left. - R_1 \left(\frac{R_2}{\zeta(Y)} - 1 \right) \delta(Y) \right],\end{aligned}$$

and remainder

$$\begin{aligned}R_\theta(\bar{\mathbb{P}}; \mathbb{P}) &= \left\{ 1 - \frac{\mathbb{P}[R_1 = 0, R_2 = 1]}{\bar{\mathbb{P}}[R_1 = 0, R_2 = 1]} \right\} \{ \theta(\bar{\mathbb{P}}) - \theta(\mathbb{P}) \} \\ &\quad + \frac{1}{\bar{\mathbb{P}}[R_1 = 0, R_2 = 1]} \mathbb{E}_{\bar{\mathbb{P}}} \left(R_2 \left\{ \frac{\bar{\beta}(X)}{\bar{\alpha}(X)} - \frac{\beta(X)}{\alpha(X)} \right\} \left\{ [1 - \gamma(X)] - \frac{\alpha(X)}{\bar{\alpha}(X)} \frac{\gamma(X)}{\bar{\gamma}(X)} [1 - \bar{\gamma}(X)] \right\} \right. \\ &\quad - \frac{R_1}{\bar{\alpha}(X)} R_2 \frac{1 - \bar{\gamma}(X)}{\bar{\gamma}(X)} \left\{ \frac{\bar{\beta}(X)}{\bar{\alpha}(X)} - \frac{\beta(X)}{\alpha(X)} \right\} \left\{ \frac{1}{\bar{\zeta}(Y)} - \frac{1}{\zeta(Y)} \right\} \\ &\quad - R_1 R_2 \left\{ \frac{1}{\bar{\zeta}(Y)} - \frac{1}{\zeta(Y)} \right\} \{ \bar{\delta}(Y) - \delta(Y) \} \\ &\quad \left. + R_1 R_2 \left(Y - \frac{\beta(X)}{\alpha(X)} \right) \left\{ \frac{1}{\bar{\zeta}(Y)} - \frac{1}{\zeta(Y)} \right\} \left\{ \frac{1}{\bar{\alpha}(X)} \frac{1 - \bar{\gamma}(X)}{\bar{\gamma}(X)} - \frac{1}{\alpha(X)} \frac{1 - \gamma(X)}{\gamma(X)} \right\} \right).\end{aligned}$$

Proof. We omit details, but the result follows from calculations similar to those discussed for example in Section 4 of [Kennedy \(2022\)](#). \square

The results above are given for arbitrary real-valued Y . When $Y \in \{0, 1\}$ is binary, e.g., HIV status in the motivating example, then many of the expressions simplify substantially. The following results illustrate this.

Corollary 1. *Define*

$$\rho = \frac{\text{odds}(Y = 1 \mid R_1 = R_2 = 1)}{\text{odds}(Y = 1 \mid R_1 = 1)}$$

$$\xi(x) = \text{odds}(Y = 1 \mid X = x, R_1 = R_2 = 1).$$

When Y is binary then the quantity θ from Proposition 1 is given by

$$\theta = \mathbb{E} \left\{ \frac{\xi(X)}{\rho + \xi(X)} \mid R_1 = 0, R_2 = 1 \right\}.$$

The above result aids interpretation of the functional θ in the binary case. Specifically, first note that the conditional odds in the expression can also be written as a product of a density ratio and marginal odds, i.e.,

$$\begin{aligned} \xi(X) &= \text{odds}(Y = 1 \mid X, R_1 = R_2 = 1) \\ &= \frac{d\mathbb{P}(X \mid R_1 = R_2 = 1, Y = 1)}{d\mathbb{P}(X \mid R_1 = R_2 = 1, Y = 0)} \times \text{odds}(Y = 1 \mid R_1 = R_2 = 1). \end{aligned}$$

Therefore denoting the density ratio by

$$\lambda(x) = \frac{d\mathbb{P}(x \mid R_1 = R_2 = 1, Y = 1)}{d\mathbb{P}(x \mid R_1 = R_2 = 1, Y = 0)}$$

we can also write

$$\theta = \mathbb{E} \left\{ \frac{\text{odds}(Y = 1 \mid R_1 = 1)\lambda(X)}{1 + \text{odds}(Y = 1 \mid R_1 = 1)\lambda(X)} \mid R_1 = 0, R_2 = 1 \right\}.$$

Now note that the term

$$\text{odds}(Y = 1 \mid R_1 = 1)\lambda(X)$$

can be interpreted via Bayes' theorem as a posterior odds, where the prior odds, $\text{odds}(Y = 1 \mid R_1 = 1)$, comes from the $R_1 = 1$ group (those initially tested in the HIV example), while the likelihood ratio $\lambda(x)$ comes from the $R_1 = R_2 = 1$ group (those tested and for whom data on risks/fears were measured). Then the quantity

$$\frac{\text{odds}(Y = 1 \mid R_1 = 1)\lambda(X)}{1 + \text{odds}(Y = 1 \mid R_1 = 1)\lambda(X)}$$

is just this posterior odds converted to the probability scale. Therefore this quantity reflects the posterior probability of a positive HIV test, combining the prior information from the initial testing, with the likelihood ratio from the second round of data collection on risks/fears. Finally the outer expectation standardizes to the $R_1 = 0, R_2 = 1$ group (those not tested but with measured risks/fears).

The next result gives the influence function for θ in the binary case, under a different parametrization from the influence function result above for the general Y case. For simplicity, and to focus ideas, we also consider the setting where the odds ratio ρ is known. We discuss how the influence function and corresponding estimators are still useful when the odds ratio is unknown after the result.

Corollary 2. Define the odds ratio ρ and conditional odds $\xi(x)$ as in Corollary 1, and let

$$\varpi(x) = \frac{d\mathbb{P}(x \mid R_1 = 0, R_2 = 1)}{d\mathbb{P}(x \mid R_1 = R_2 = 1)}.$$

When Y is binary and ρ is known, the functional θ from Proposition 1 satisfies the von Mises expansion

$$\theta(\bar{\mathbb{P}}) - \theta(\mathbb{P}) = \int \varphi(o; \bar{\mathbb{P}}) d(\bar{\mathbb{P}} - \mathbb{P})(o) + R_\theta(\bar{\mathbb{P}}; \mathbb{P}),$$

with influence function $\varphi(O; \mathbb{P})$ equal to

$$\rho \left\{ \frac{1 + \xi(X)}{\rho + \xi(X)} \right\}^2 \frac{R_1 R_2 \varpi(X)}{\mathbb{P}[R_1 = R_2 = 1]} \left\{ Y - \frac{\xi(X)}{1 + \xi(X)} \right\} + \frac{(1 - R_1) R_2}{\mathbb{P}[R_1 = 0, R_2 = 1]} \left\{ \frac{\xi(X)}{\rho + \xi(X)} - \theta \right\}$$

and remainder $R_\theta(\bar{\mathbb{P}}; \mathbb{P})$ given by

$$\begin{aligned} & \frac{\mathbb{E}(R_1 R_2)}{\mathbb{E}(R_1 R_2)} \int \left(\frac{\xi}{\rho + \xi} - \frac{\bar{\xi}}{\rho + \bar{\xi}} \right) (\bar{\varpi} - \varpi) dP_{11} + \frac{\mathbb{E}(R_1 R_2)}{\mathbb{E}(R_1 R_2)} \int \left(S_{2g} + \frac{S_{2f}\rho}{(\rho + \bar{\xi})^2} \right) \bar{\varpi} dP_{11} \\ & + \left\{ \frac{\mathbb{E}((1 - R_1) R_2)}{\mathbb{E}((1 - R_1) R_2)} - \frac{\mathbb{E}(R_1 R_2)}{\mathbb{E}(R_1 R_2)} \right\} \int \left(\frac{\bar{\xi}}{\rho + \bar{\xi}} - \frac{\xi}{\rho + \xi} \right) dP_{01} \\ & + \left\{ 1 - \frac{\mathbb{E}((1 - R_1) R_2)}{\mathbb{E}((1 - R_1) R_2)} \right\} (\bar{\theta} - \theta), \end{aligned}$$

for (S_{2f}, S_{2g}) second-order terms defined in the appendix, and $dP_{st} = d\mathbb{P}(x \mid R_1 = s, R_2 = t)$.

Proof. The influence function can be obtained using the facts that (i) for $g(\xi) = \frac{\xi}{\rho + \xi}$ we have $g'(\xi) = \frac{\rho}{(\rho + \xi)^2}$, and (ii) the influence function for $\xi(x)$ when X is discrete is given by

$$\frac{\{1 + \xi(X)\}^2 \mathbb{1}(X = x) R_1 R_2}{d\mathbb{P}(x \mid R_1 = R_2 = 1) \mathbb{P}[R_1 = R_2 = 1]} \left\{ Y - \mathbb{P}[Y = 1 \mid X, R_1 = R_2 = 1] \right\},$$

together with calculations similar to those discussed in Section 4 of Kennedy (2022). A derivation of the remainder expression is given in the appendix. \square

The above motivates the one-step/double-machine-learning estimator

$$\hat{\theta} = \mathbb{P}_n \left\{ \left(\frac{1 + \hat{\xi}}{\rho + \hat{\xi}} \right)^2 \frac{(R_1 R_2) \rho \hat{\varpi}}{\mathbb{P}_n(R_1 = R_2 = 1)} \left(Y - \frac{\hat{\xi}}{1 + \hat{\xi}} \right) + \frac{(1 - R_1) R_2}{\mathbb{P}_n(R_1 = 0, R_2 = 1)} \left(\frac{\hat{\xi}}{\rho + \hat{\xi}} \right) \right\}.$$

When the odds ratio ρ is unknown, one can simply use the above estimator with the plug-in estimator $\hat{\rho}$ replacing ρ . Since ρ can be estimated with simple sample averages, the plug-in is \sqrt{n} -consistent and asymptotically normal, and the resulting estimator of θ will just have an extra asymptotically linear term.

References

- Balke, A. and Pearl, J. (1994), “Probabilistic evaluation of counterfactual queries,” in *Proceedings of the 12th Conference on Artificial Intelligence-Volume 1*, Menlo Park, CA: MIT Press, pp. 230–237.
- (1997), “Bounds on treatment effects from studies with imperfect compliance,” *Journal of the American Statistical Association*, 92, 1171–1176.
- Cui, Y., Pu, H., Shi, X., Miao, W., and Tchetgen Tchetgen, E. (2024), “Semiparametric proximal causal inference,” *Journal of the American Statistical Association*, 119, 1348–1359.
- Frangakis, C. E. and Rubin, D. B. (2002), “Principal stratification in causal inference,” *Biometrics*, 58, 21–29.
- Heckman, J. J. (1979), “Sample selection bias as a specification error,” *Econometrica: Journal of the econometric society*, 153–161.
- Hernán, M. A. and Robins, J. M. (2020), *Causal inference: what if*, Boca Raton: Chapman & Hill/CRC.
- Imbens, G. W. and Angrist, J. D. (1994), “Identification and Estimation of Local Average Treatment Effects,” *Econometrica*, 62, 467–475.
- Kennedy, E. H. (2020), “Efficient nonparametric causal inference with missing exposure information,” *The International Journal of Biostatistics*, 16.
- (2022), “Semiparametric doubly robust targeted double machine learning: a review,” *arXiv preprint arXiv:2203.06469*.
- Levis, A. W., Bonvini, M., Zeng, Z., Keele, L., and Kennedy, E. H. (2025), “Covariate-assisted bounds on causal effects with instrumental variables,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, qkaf028.
- Levis, A. W., Kennedy, E. H., and Keele, L. (2024), “Nonparametric identification and efficient estimation of causal effects with instrumental variables,” *arXiv preprint arXiv:2402.09332*.
- Li, W., Miao, W., and Tchetgen Tchetgen, E. (2023), “Non-parametric inference about mean functionals of non-ignorable non-response data without identifying the joint distribution,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85, 913–935.
- Miao, W., Geng, Z., and Tchetgen Tchetgen, E. J. (2018), “Identifying causal effects with proxy variables of an unmeasured confounder,” *Biometrika*, 105, 987–993.
- Miao, W., Liu, L., Tchetgen, E. T., and Geng, Z. (2015), “Identification, doubly robust estimation, and semiparametric efficiency theory of nonignorable missing data with a shadow variable,” *arXiv preprint arXiv:1509.02556*.
- Miao, W. and Tchetgen Tchetgen, E. J. (2016), “On varieties of doubly robust estimators under missingness not at random with a shadow variable,” *Biometrika*, 103, 475–482.
- Nabi, R., Bhattacharya, R., Shpitser, I., and Robins, J. (2022), “Causal and counterfactual views of missing data models,” *arXiv preprint arXiv:2210.05558*.

- Park, C., Richardson, D. B., and Tchetgen Tchetgen, E. J. (2024), “Single proxy control,” *Biometrics*, 80, ujae027.
- Pearl, J. (2009), *Causality*, Cambridge University Press.
- Richardson, T. S. and Robins, J. M. (2013), “Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality,” *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, 128, 2013.
- Robins, J. M. (1997), “Non-response models for the analysis of non-monotone non-ignorable missing data,” *Statistics in medicine*, 16, 21–37.
- Robins, J. M. and Richardson, T. S. (2010), “Alternative graphical causal models and the identification of direct effects,” Tech. rep., Center for Statistics and the Social Sciences, University of Washington.
- Shpitser, I., Richardson, T. S., and Robins, J. M. (2022), “Multivariate counterfactual systems and causal graphical models,” in *Probabilistic and causal inference: The works of Judea Pearl*, New York, NY: Association for Computing Machinery, pp. 813–852.
- Sun, B., Liu, L., Miao, W., Wirth, K., Robins, J., and Tchetgen, E. J. T. (2018), “Semi-parametric estimation with data missing not at random using an instrumental variable,” *Statistica Sinica*, 28, 1965.
- Swanson, S. A., Hernán, M. A., Miller, M., Robins, J. M., and Richardson, T. S. (2018), “Partial identification of the average treatment effect using instrumental variables: review of methods for binary instruments, treatments, and outcomes,” *Journal of the American Statistical Association*, 113, 933–947.
- Tchetgen Tchetgen, E. (2014), “The control outcome calibration approach for causal inference with unobserved confounding,” *American journal of epidemiology*, 179, 633–640.
- Tchetgen Tchetgen, E. J. and Wirth, K. E. (2017), “A general instrumental variable framework for regression analysis with outcome missing not at random,” *Biometrics*, 73, 1123–1131.
- Wang, L. and Tchetgen Tchetgen, E. (2018), “Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 80, 531–550.
- Williamson, E., Forbes, A., and Wolfe, R. (2012), “Doubly robust estimators of causal exposure effects with missing data in the outcome, exposure or a confounder,” *Statistics in Medicine*, 31, 4382–4400.