Causal Inference With Outcome-Dependent Missingness And Self-Censoring

Abstract

We consider missingness in the context of causal inference when the outcome of interest may be missing. If the outcome directly affects its own missingness status, i.e., it is "self-censoring", this may lead to severely biased causal effect estimates. Miao et al. [2015] proposed the shadow variable method to correct for bias due to self-censoring, however, verifying the required model assumptions can be difficult. Here, we propose a test based on a randomized incentive variable offered to encourage reporting of the outcome that can be used to verify identification assumptions that are sufficient to correct for both self-censoring and confounding bias. Concretely, the test confirms whether a given set of pre-treatment covariates are sufficient to block all backdoor paths between the treatment and outcome as well as all paths between the treatment and missinginess indicator after conditioning on the outcome. We show that under these conditions, the causal effect is identified by using the treatment as a shadow variable, and it leads to an intuitive inverse probability weighting estimator that uses a product of the treatment and response weights. We evaluate the efficacy of our test and downstream estimator via simulations.

1 INTRODUCTION

"Self-censoring" is a type of missingness-not-at-random (MNAR) phenomenon that poses a particularly difficult obstacle to valid inference in settings with missing data. Researchers have begun to address MNAR missing data problems, i.e., problems where the probability of missingness depends on variables that exhibit missingness themselves. However, the task of mitigating bias due to self-censoring remains relatively unexplored [Mohan and Pearl, 2021]. Here, we consider outcome-dependent self-censoring—situations where the outcome determines its own missingness—in the context of computing causal effects. This kind of self-censoring is quite common in practice, for example, in studies where the outcome is an attribute associated with social stigma (such as drug use or risky sexual behaviors) or in general in settings where the outcome is ascertained by voluntary survey response.

Recent work in missing data uses directed acyclic graphs (DAGs) to represent substantive assumptions about causal relations among variables, including indicators of missingness [Daniel et al., 2012, Mohan et al., 2013, Mohan and Pearl, 2021]. Within this framework, Bhattacharya et al. [2019] and Nabi et al. [2020] derive a sound and complete criterion for when it is possible to recover the full data law of a missing data DAG model in the presence of MNAR missingness and unmeasured confounding. However, this criterion specifically excludes self-censoring, i.e., self-censoring prevents non-parametric identification of the full data law unless additional (non-structural) assumptions are made about the data generating process.

When the full structure of the DAG is unknown, the possibility of unmeasured confounding, or latent variables, poses further challenges to observational causal inference as it complicates the process of finding a valid set of covariates to adjust for [Shpitser et al., 2012]. Methods for covariate adjustment under MNAR data have been proposed by Saadati and Tian [2019] and Yang et al. [2019]; however, these methods require full knowledge of the structure of the missing data DAG and do not allow for self-censoring on the outcome. To recover the full data law given a self-censoring outcome and under a certain completeness condition, Miao et al. [2015] propose the use of a *shadow variable*, a variable that satisfies a relevance assumption with respect to the outcome and an exclusion restriction with respect to the outcome's missingness indicator. However, empirical tests to verify the shadow variable assumptions as well as the validity of downstream covariate adjustment remains an open problem that we seek to address in this work.

Submitted to the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023). **To be used for reviewing only**.

Here, we build on the work by Miao et al. [2015] on identification under self-censoring and Entner et al. [2013] on covariate selection to propose a method for simultaneously recovering the target law – the joint distribution of variables as if there were no missingness – and a valid covariate adjustment set in the presence of a self-censoring outcome.

We propose a test based on a randomized incentive variable offered to encourage reporting of the outcome, which can be used to verify identification assumptions that are sufficient to correct for both self-censoring and confounding bias. The test proceeds in two stages. The first stage confirms dependence between the incentive for reporting and the missingness indicator of the outcome. The second stage executes a search for a pre-treatment covariate and a covariate adjustment set that satisfies three conditions – (i) independence between the treatment and the incentive conditional on the outcome, missingness indicator of the outcome, and the covariate adjustment set, (ii) dependence between the pre-treatment covariate and the missingness indicator of the outcome while conditioning on the covariate adjustment set, and (iii) independence between the same variables and adjustment set in test (ii) while additionally conditioning on the treatment. We prove that, when these conditions hold, we may use the treatment as a valid shadow variable and have a valid covariate adjustment set. We derive the corresponding identifying functional for the causal effect and propose an intuitive inverse probability weighting estimator that uses a product of the treatment and response weights. We evaluate the efficacy of our method via simulations.

Related work on self-censoring: Sportisse et al. [2020] propose an imputation method for self-censored data that assumes factorization according to a certain latent variable DAG and parametric models for the missingness process. Mohan et al. [2018] propose methods for recovery of the full data law when all variables are discrete and when certain matrices corresponding to conditional probability tables are invertible. Duarte et al. [2021] propose an algorithm for computing bounds on the causal effect in the discrete setting, which may converge to point identification in certain cases. d'Haultfoeuille [2010] and Tchetgen Tchetgen and Wirth [2017] propose instrumental variable methods that place homogeneity restrictions on the missingness process in addition to requiring the presence of a valid instrument. Though the randomized incentive variable we consider could be used as an instrumental variable, the homogeneity restriction is untestable and can be restrictive in many real-world settings. To our knowledge, the testability of identifying assumptions of the shadow variable method and covariate adjustment under self-censoring has not been explored before.

2 MOTIVATING EXAMPLE

For our motivating example, we describe a hypothetical study inspired by Turner et al. [2009] on evaluating the effect of public health programs on encouraging safe sex practices. Consider an observational study where researchers offer a sexual education program on encouraging condom use. That is, enrollment into this program is not randomized. In a follow-up survey, participants may also choose not to disclose their post-program condom use habits, precisely because of the opinions they hold on the practice. Hence, to estimate the causal effect of the program on improving condom use, the researchers must overcomes challenges related to both confounding and self-censoring.

To incentivize response to answer sensitive questions regarding sexual behavior, the researchers randomly assigned study participants to be surveyed via a phone interview conducted by a real human being or an automated program called the Telephone Audio Computer-Assisted Self-Interviewing (T-ACASI) program. Turner et al. [2009] applied this randomized incentive strategy to successfully increase response rates on questions pertaining to drug use or risky sexual behaviors in a study conducted in Baltimore, USA. One can also imagine other randomized incentives offered to increase response[1], such as random lotteries for gift cards. Such incentives are often offered as part of studies that have a survey component. However, in most cases, the incentives can only encourage response but not guarantee it, and so the issue of self-censoring still persists.

In this work, we aim to simultaneously address the challenges of covariate selection and overcoming self-censoring, or non-response bias, inherent in observational studies that ask respondents sensitive questions.

3 MODEL AND PROBLEM SETUP

We assume the causal structure of the system is represented via a directed acyclic graph $\mathcal{G}$ defined over a set of vertices $\mathbf{V} = \{A, Y^{(1)}, R_Y, Y, I\} \cup \mathbf{W} \cup \mathbf{U}$, where A represents the treatment variable, $Y^{(1)}$ represents the outcome had we – possibly contrary to fact – been able to observe it, R_Y represents the corresponding binary missingness indicator for the outcome, Y represents the factual observed outcome – which may be either a numeric value or "?" if the observation is missing, I represents the incentive variable, $\mathbf{W}$ represents an observed set of pre-treatment covariates, and $\mathbf{U}$ denotes a set of unobserved (latent) covariates. The entire set $\mathbf{V}$ is assumed to be causally sufficient, i.e., there are no additional unmeasured common causes of any two variables in $\mathbf{V}$.

The above setup describes a causal Bayesian network over the variables $\mathbf{V}$ [Spirtes et al., 2000, Pearl, 2009], with an added restriction on the relationship between the missing variable $Y^{(1)}$, its missingness indicator, and the observed

[1]The effect of the incentive on response need not be monotonic. Our method simply relies on some correlation between the incentive and response, which we can test using the observed data.

outcome: $Y = Y^{(1)}$ when $R_Y = 1$ and $Y = ?$ when $R_Y = 0$. This is the missing data version of the *consistency* assumption in causal inference [Nabi et al., 2022]. Due to the deterministic nature of this relation, we omit drawing the observed outcome Y in our figures without loss of generality.

Independences implied in the full data law $p(\mathbf{V})$ can be read off from $\mathcal{G}$ via the d-separation criterion [Pearl, 1988, Mohan and Pearl, 2021]. Here, we make a *faithfulness* assumption [Spirtes et al., 2000], which states that any independencies in the distribution $p(\mathbf{V})$ must correspond to d-separation statements in $\mathcal{G}$. Formally, $A \perp\!\!\!\perp B \mid \mathbf{C}$ in $p(\mathbf{V})$ if and only if $A \perp\!\!\!\perp_{\text{d-sep}} B \mid \mathbf{C}$ in $\mathcal{G}$. For tests that involve conditioning on the missing outcome $Y^{(1)}$, we also assume an extension of faithfulness used in causal discovery procedures for missing data settings [Tu et al., 2019]. This assumption states that any independences that exist conditional on R_Y in $p(\mathbf{V})$ must also hold in the observed data, i.e., conditional on $R_Y = 1$. Formally, we assume $A \perp\!\!\!\perp B \mid \mathbf{C}, R_Y$ in $p(\mathbf{V})$ if and only if $A \perp\!\!\!\perp B \mid \mathbf{C}, R_Y = 1$ in $p(\mathbf{V})$, when $Y^{(1)}$ is one of A, B, or an element of the conditioning set $\mathbf{C}$.

The above assumptions are commonly used across most graphical model selection procedures [Spirtes et al., 2000]. We now list and provide brief justification for additional structural assumptions important for our method. Let $\text{pa}_{\mathcal{G}}(V)$ denote the parents of V in the graph $\mathcal{G}$. We assume that the data are generated from a distribution $p(\mathbf{V})$ that is Markov and faithful with respect to a DAG $\mathcal{G}$ that satisfies the structural assumptions M1-M4.

(M1) The only parents of Y are $Y^{(1)}$ and R_Y, i.e., $\text{pa}_{\mathcal{G}}(Y) = \{Y^{(1)}, R_Y\}$.

(M2) There is an edge from A to $Y^{(1)}$ and an edge from $Y^{(1)}$ to R_Y (i.e. the causal path $A \rightarrow Y^{(1)} \rightarrow R_Y$ exists in the graph).

(M3) The incentive I is randomly assigned (i.e. $\text{pa}_{\mathcal{G}}(I) = \emptyset$) and may only be a parent of the missingness indicator R_Y.

(M4) R_Y is not a parent of any variables in $\mathbf{V}$, A is not a parent of any variables in $\mathbf{W} \cup \mathbf{U}$, and $Y^{(1)}$ cannot be a parent of A nor any variables in $\mathbf{W} \cup \mathbf{U}$. That is, we have an ordering, $\{I\} \cup \mathbf{W} \cup \mathbf{U} < A < Y^{(1)} < R_Y$.

Assumption M1 and disallowing R_Y from having any children in assumption M4 are standard restrictions in missing data DAG models [Mohan et al., 2013]. We require assumption M2 as it simplifies our empirical tests. However, M2 is a relatively mild assumption as the existence of these edges is the primary motivation for applying our method. Assumption M3 makes sure that I is a valid proxy variable for designing indirect tests about the validity of the treatment as a shadow variable. Finally, assumption M4 states that $\mathbf{W} \cup \mathbf{U}$ are all pre-treatment variables.

Figure 1 graphically displays assumptions M1-M4. The red edges are assumed to exist whereas the blue edges may or not exist. We draw Y here to illustrate M1, but we will omit it and the red dashed edges in all figures going forward (due to the deterministic nature of its relation with $Y^{(1)}$ and R_Y.)

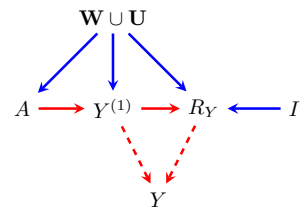


Figure 1: Graph depicting assumptions M1-M4.

Target of Inference

Under missingness, a causal effect of the treatment on the outcome corresponds to a contrast between potential outcomes $Y^{(A=a,R_Y=1)}$ and $Y^{(A=a',R_Y=1)}$, where $Y^{(A=a,R_Y=1)}$ denotes the value of the outcome had the treatment been set to some value a via intervention and had the outcome been observed. Moving forward, we use $Y^{(a,1)}$ and $Y^{(a',1)}$ for brevity. Our target of inference is the average causal effect (ACE), i.e., the mean difference $\mathbb{E}[Y^{(a,1)} - Y^{(a',1)}]$.[2] In the next section we derive an identification formula for the ACE in terms of the observed data distribution $p(A, Y, R_Y, \mathbf{W})$ based on the backdoor adjustment formula from Pearl [1995] and the shadow variable method proposed by Miao et al. [2015].

4 IDENTIFICATION

In this section we demonstrate of the ACE assuming we are given a valid backdoor adjustment set and shadow variable adjustment set $\mathbf{Z} \subset \mathbf{W}$. Note that we use a strict subset relation as we will use at least one of the remaining pre-treatment covariates for verification of the identifying assumptions. We can also ignore the incentive variable I here because it does not play a role in identification, only in the verification of the assumptions surrounding $\mathbf{Z}$ later on.

In the following, we focus on identification of $\mathbb{E}[Y^{(a,1)}]$; identification of $\mathbb{E}[Y^{(a',1)}]$ follows similarly. We perform identification in two steps. In the first step, we assume we have access to the underlying counterfactual $Y^{(1)}$. A set $\mathbf{Z}$ is said to be a valid backdoor adjustment set relative to the treatment A and outcome $Y^{(1)}$ if

(B1) $\mathbf{Z}$ does not contain any variables that are causal descendants of A, and

[2] Similar notation for representing potential outcomes under missingness has been used in Nabi et al. [2022]. This can also be expressed using do-notation as in Saadati and Tian [2019].

(B2) A and $Y^{(1)}$ are d-separated when conditioning on $\mathbf{Z}$ in a modified graph where all outgoing edges from A are deleted.

If $\mathbf{Z}$ is a valid backdoor adjustment set, then we have the following backdoor adjustment formula in terms of the full data law involving $Y^{(1)}$ [Pearl, 1995].

$$\mathbb{E}[Y^{(a,1)}] = \sum_{\mathbf{Z}} \mathbb{E}[Y^{(1)} \mid A = a, \mathbf{Z}] \times p(\mathbf{Z}) \tag{1}$$

However, due to missingness, we have access to only the margin $p(A, Y, R_Y, \mathbf{Z})$, which does not include $Y^{(1)}$. Under self-censoring, the counterfactual expression $\mathbb{E}[Y^{(1)} \mid A = a, \mathbf{Z}]$ is not equal to the observed regression $\mathbb{E}[Y \mid A = a, \mathbf{Z}, R_Y = 1]$ since $Y^{(1)} \not\perp\!\!\!\perp R_Y \mid A, \mathbf{Z}$ from assumption M2, i.e., due to self-censoring.

Next, we define shadow variables as described by Miao et al. [2015]. A variable S is a valid *shadow variable* if it is a fully observed variable that satisfies the following independence relations.

(S1) $S \not\perp\!\!\!\perp Y^{(1)} \mid R_Y = 1, \mathbf{Z}$, and

(S2) $S \perp\!\!\!\perp R_Y \mid Y^{(1)}, \mathbf{Z}$.

A shadow variable S helps us recover the target law as if there were no missingness and the propensity score of the missingness indicator $p(R_Y = 1 \mid A, Y^{(1)}, \mathbf{Z})$ from the margin $p(A, Y, R_Y, \mathbf{Z})$. We refer to the set $\mathbf{Z} \subset \mathbf{W}$ that satisfies conditions S1 and S2 as the *shadow variable adjustment set*. In our method, we verify whether the treatment A is a valid shadow variable and directly use it as such.

To identify (1), it is sufficient to identify the joint distribution of $p(A, Y^{(1)}, \mathbf{Z})$. By the chain rule of probability we have,

$$p(A, Y^{(1)}, \mathbf{Z}) = \frac{p(A, Y^{(1)}, \mathbf{Z}, R_Y = 1)}{p(R_Y = 1 \mid A, Y^{(1)}, \mathbf{Z})}. \tag{2}$$

The numerator is a function of observed data due to consistency. Identification of this joint then reduces to identification of the propensity score $p(R_Y = 1 \mid A, Y^{(1)}, \mathbf{Z})$. If we are able to verify that A is a valid shadow variable, then the independence $A \perp\!\!\!\perp R_Y \mid Y^{(1)}, \mathbf{Z}$ will hold from S2. Hence, $p(R_Y = 1 \mid A, Y^{(1)}, \mathbf{Z}) = p(R_Y = 1 \mid Y^{(1)}, \mathbf{Z})$.

Following Miao et al. [2015], we use an odds ratio factorization of the propensity score to perform the identification of $p(R_Y = 1 \mid Y^{(1)}, \mathbf{Z})$. Without loss of generality we pick any arbitrary value y_0 to be the reference value for the outcome and $R_Y = 0$ to be the reference value of the missingness indicator. The odds ratio between the outcome $Y^{(1)}$ and missingness indicator R_Y is 1 whenever either variable is at its reference value. Let $\pi_0(\mathbf{Z})$ denote the propensity score for R_Y at the reference value y_0, i.e., $\pi_0 := p(R_Y = 1 \mid Y^{(1)} = y_0, \mathbf{Z})$. Let $\eta(Y^{(1)}, \mathbf{Z})$ denote the conditional odds ratio function relating $Y^{(1)}$ and R_Y at values where $R_Y = 1$, i.e., $\eta(Y^{(1)}, \mathbf{Z}) := \mathrm{OR}(R_Y = 1, Y^{(1)} \mid \mathbf{Z})$. Then the odds ratio factorization of the propensity score can be written as,

$$p(R_Y = 1 \mid Y^{(1)}, \mathbf{Z}) = \frac{\pi_0(\mathbf{Z})}{\pi_0(\mathbf{Z}) + \eta(Y^{(1)}, \mathbf{Z})(1 - \pi_0(\mathbf{Z}))} \tag{3}$$

Whenever $Y^{(1)} = y_0$, $\eta(Y^{(1)}, \mathbf{Z}) = 1$, and we have that $p(R_Y = 1 \mid Y^{(1)} = y_0, \mathbf{Z}) = \pi_0(\mathbf{Z})$. At any other value of $Y^{(1)}$, $\eta(Y^{(1)}, \mathbf{Z}) \neq 1$, and the odds ratio factorization of the propensity score will return a different value.

When A is a valid shadow variable and $p(Y^{(1)} \mid R_Y = 1, A, \mathbf{Z})$ satisfies a widely used completeness condition[3], Miao et al. [2015] show that π_0 and η (and hence the propensity score) are identified from the observed data.

Consider Figure 2, which demonstrates how certain sets of pre-treatment covariates might be sufficient to adjust for confounding but not missingness or vice versa. In this DAG, the set $\{W_1\}$ satisfies B1 and B2. However, it does not satisfy S1 and S2 when considering the treatment A as the shadow variable because of the open collider at $Y^{(1)}$ on the path $A \rightarrow Y^{(1)} \leftarrow W_2 \rightarrow R_Y$. On the other hand, the set $\{W_2\}$ satisfies S1 and S2 when considering the treatment A as the shadow variable, but the backdoor path through W_1 remains open. Only the set $\mathbf{Z} = \{W_1, W_2\}$ satisfies all of B1, B2, S1, and S2 simultaneously. We now present our main identification results that relies on a set of covariates that satisfy all these conditions simultaneously.

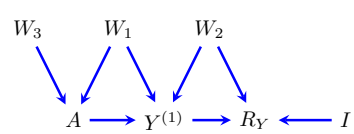


Figure 2: Graph demonstrating how different subsets of $\mathbf{W}$ satisfy the identifying assumptions B1, B2, S1, and S2.

Theorem 1. *Under structural assumptions M1-M4 and completeness of $p(Y^{(1)} \mid R_Y = 1, A, \mathbf{Z})$, if $\mathbf{Z}$ satisfies B1 and B2 and A is a valid shadow variable satisfying S1 and S2 conditional on $\mathbf{Z}$, then the expected value of the counterfactual outcome $\mathbb{E}[Y^{(a,1)}]$ is identified from the observed data distribution $p(A, Y, \mathbf{Z}, R_Y)$ as follows:*

$$\mathbb{E}[Y^{(a,1)}] = \mathbb{E}\left[\frac{R_Y \times \mathbb{I}(A = a) \times Y}{p(R_Y = 1 | Y^{(1)}, \mathbf{Z}) \times p(A = a | \mathbf{Z})}\right] \tag{4}$$

A full proof of Theorem 1 is provided in the Appendix, but the intuition is simple. The use of R_Y in the numerator of

[3]For all square-integrable functions $h(A, Y^{(1)})$, $\mathbb{E}[h(A, Y^{(1)}) \mid R_Y = 1, A, \mathbf{Z}] = 0$ almost surely if and only if $h(A, Y^{(1)}) = 0$ almost surely (see Newey and Powell [2003] or Section 3 of Miao et al. [2015] for more details).

(4) ensures that we use only observed rows of data. Further, under our assumptions, both propensity scores in the denominator are identified as functions of the observed data; $p(A = a \mid \mathbf{Z})$ depends on only observed data quantities and $p(R_Y = 1 \mid Y^{(1)}, \mathbf{Z})$ is identified via the odds ratio factorization in (3) with the treatment as a shadow variable. We can then apply standard laws of probability to show that (4) is equivalent to the full data law adjustment functional in (1). Since $\mathbf{Z}$ is assumed to be a valid backdoor adjustment set, this further implies equivalence of (4) to the counterfactual mean $\mathbb{E}[Y^{(a,1)}]$.

The above identification argument relies on the absence of certain edges from the graph $\mathcal{G}$ in order to satisfy assumptions S1, S2, B1, and B2. Figure 3 shows the edges that, if present, preclude identification. In the figure we use bidirected edges as a shorthand for a path involving unmeasured variables. For example, $A \leftrightarrow Y^{(1)}$ is shorthand for the presence of a d-connecting path $A \leftarrow U_1 \cdots U_k \rightarrow Y^{(1)}$, where $U_1, \ldots, U_k$ are unmeasured variables in $\mathbf{U}$. In the figure, it is possible for the treatment A and a set $\mathbf{Z} \subset \mathbf{W}$ to satisfy the shadow variable conditions S1 and S2 whenever the red dashed edges are absent. Similarly, it is possible for $\mathbf{Z} \subset \mathbf{W}$ to satisfy the backdoor conditions B1 and B2 whenever the green dashed edge is absent. In essence, our method in the next section tests for the absence of these edges.

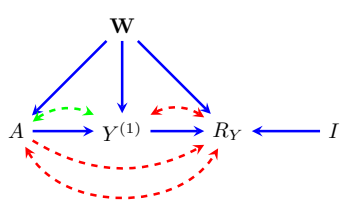


Figure 3: Red and green dashed edges impede identification.

5 TESTS FOR IDENTIFICATION CONDITIONS

Entner et al. [2013] propose a two-stage test to verify whether a set of pre-treatment covariates $\mathbf{Z}$ satisfy B1 and B2 for backdoor adjustment when there is no missing data or selection bias. However, from the previous section we have seen that, to overcome self-censoring, $\mathbf{Z}$ must also satisfy the shadow variable conditions S1 and S2 for some candidate shadow variable. Testing S2 using observed data is impossible in general, though, as such a test involves conditioning on the counterfactual variable $Y^{(1)}$. Here, we augment the test from Entner et al. [2013] so that it is capable of testing for the backdoor adjustment conditions in addition to the shadow variable assumptions when considering the treatment A as a shadow variable and only using observed data quantities. This allows us to validate the crucial identifying assumptions laid out in the previous section.

5.1 METHOD

The two-stage method described in Entner et al. [2013] searches for a $W \in \mathbf{W}$ and $\mathbf{Z} \subseteq \mathbf{W} \setminus \{W\}$ such that (i) $W \not\perp\!\!\!\perp Y \mid \mathbf{Z}$ and (ii) $W \perp\!\!\!\perp Y \mid \mathbf{Z}, A$. They assume the same partial order of variables as in assumption M4 except that the variables I and R_Y are not present in the graph. In addition, Y is a fully observed variable. Entner et al. [2013] proved that, when conditions (i) and (ii) hold, $\mathbf{Z}$ is a valid backdoor adjustment set for the causal effect of A on Y.

For our method, we propose to test if

(C1) $I \not\perp\!\!\!\perp R_Y$

and then search for some $W \in \mathbf{W}$ and $\mathbf{Z} \subseteq \mathbf{W} \setminus \{W\}$ such that

(C2) $A \perp\!\!\!\perp I \mid Y^{(1)}, R_Y = 1, \mathbf{Z}$

(C3) $W \not\perp\!\!\!\perp R_Y \mid \mathbf{Z}$

(C4) $W \perp\!\!\!\perp R_Y \mid A, \mathbf{Z}$

Conditions C1 and C2 are used to confirm the absence of the red dashed edges in Figure 3 that would prevent us from using A as a valid shadow variable; conditions C3 and C4 are a modification of the tests from Entner et al. [2013] that indirectly confirm the validity of $\mathbf{Z}$ as a backdoor adjustment set for the counterfactual outcome by using the missingness indicator to formulate the tests instead. This is formalized in the theorem below.

Theorem 2. *When C1-C4 hold in a distribution $p(\mathbf{V})$ that is Markov and faithful w.r.t a causal DAG $\mathcal{G}$ satisfying assumptions M1-M4, $\mathbf{Z}$ is a valid backdoor adjustment set for the causal effect of A on $Y^{(1)}$, and A is a valid shadow variable with $\mathbf{Z}$ as a shadow variable adjustment set.*

Proof. From Entner et al. [2013], if C3 and C4 hold, $\mathbf{Z}$ is a valid backdoor adjustment set for computing the effect of A on R_Y. From assumptions M2 and M4, all backdoor paths between A and $Y^{(1)}$, i.e., $A \leftarrow \cdots \rightarrow Y^{(1)}$, also extend into a backdoor path $A \leftarrow \cdots \rightarrow Y^{(1)} \rightarrow R_Y$ to R_Y. Thus, if $\mathbf{Z}$ blocks all backdoor paths between A and R_Y, it also blocks all backdoor paths between A and $Y^{(1)}$, making $\mathbf{Z}$ a valid adjustment set for the effect of A on $Y^{(1)}$.

Next, we prove that $\mathbf{Z}$ can be used for shadow variable adjustment using A as the shadow variable. Condition S1 holds trivially as we have $A \rightarrow Y^{(1)}$ according to assumption M2. We now prove S2 also holds: that $A \perp\!\!\!\perp R_Y \mid Y^{(1)}, \mathbf{Z}$. First, under assumption M2, C1 ensures that the randomized incentive I has a directed edge to R_Y and no other outgoing edges. In order for A to be a valid shadow variable, all paths between A and R_Y conditional on $Y^{(1)}$ and $\mathbf{Z}$ must be blocked. The following 4 cases cover all possible paths between A and R_Y given our assumptions.

1. Causal paths from A to R_Y. One possible causal path $A \rightarrow Y^{(1)} \rightarrow R_Y$ exists by assumption M4, but it is blocked by conditioning on $Y^{(1)}$. The second possibility of $A \rightarrow R_Y$ cannot exist because it would imply an open path between I and A conditional on $Y^{(1)}, R_Y = 1$, and $\mathbf{Z}$, contradicting C2.
2. Paths of the form $A \rightarrow Y^{(1)} \leftarrow \ldots \rightarrow R_Y$. Any open paths of this form contradict C2, as it implies the dependence $A \not\perp\!\!\!\perp I \mid Y^{(1)}, R_Y, \mathbf{Z}$.
3. Backdoor paths of the form $A \leftarrow \ldots \rightarrow R_Y$ that do not contain $Y^{(1)}$ as a collider. As previously noted, all such backdoor paths are blocked by $\mathbf{Z}$ based on conditions C3 and C4.
4. Backdoor paths containing $Y^{(1)}$ as a collider, i.e., $A \leftarrow \ldots \rightarrow Y^{(1)} \leftarrow \ldots \rightarrow R_Y$. All such paths are still blocked despite conditioning on $Y^{(1)}$ as $\mathbf{Z}$ blocks all backdoor paths between A and $Y^{(1)}$.

Therefore, $\mathbf{Z}$ is a valid backdoor adjustment set for the causal effect of A on $Y^{(1)}$, fulfilling B1 and B2, and it is also a valid shadow variable adjustment set for A to be a valid shadow variable, fulfilling S1 and S2 for A. □

We use Figure 4 to illustrate an example of how our tests proceed. Let us begin by only considering the solid blue edges. First, we confirm that C1 holds. Next, none of the conditions C2-C4 can be satisfied by using $\mathbf{Z} = \emptyset$ and $W = W_i$ for $i = 1, 2, 3$. When considering singleton adjustment sets, we get that, for $W = W_3$ and $\mathbf{Z} = \{W_1\}$, condition C2 is satisfied. However, this set does not satisfy C3 and C4 because of the open collider at W_1 that introduces an open backdoor path between A and $Y^{(1)}$. Finally for adjustment sets of size 2, we get that, when $W = W_3$ and $\mathbf{Z} = \{W_1, W_2\}$, all conditions are satisfied, providing the correct conclusion that the effect is identified via (4) using $\mathbf{Z} = \{W_1, W_2\}$.

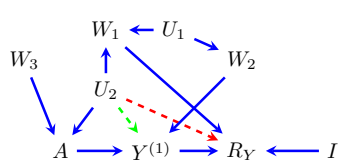


Figure 4: $W = W_3$, $\mathbf{Z} = \{W_1, W_2\}$

Next, consider the same DAG in Figure 4 with the green dashed edge being present. Because there is unmeasured confounding between the treatment and outcome variables, there is no possible $\mathbf{Z}$ that can be a valid backdoor adjustment set. Therefore, C3 and C4 can never be satisfied. Next, let us consider a DAG where the red dashed edge is present. The unmeasured confounding between A and R_Y violates S2 when using A as a shadow variable and also creates a collider path between A and I that ensures that C2 will never be satisfied. In both cases, our method correctly concludes that no adjustment set is possible.

5.2 LIMITATIONS

There exist DAGs where our identification strategy works but where our method is not able to detect the existence of a valid backdoor and shadow variable adjustment set. Consider Figure 5 where W_3 is now a confounder between the treatment and outcome. Despite $\mathbf{Z} = \{W_1, W_2, W_3\}$ fulfilling all the critical assumptions set up in Theorem 1, our method will incorrectly conclude that there is no valid adjustment set because there exists no $W \in \mathbf{W}$ that can be used as an auxiliary variable to test C3 and C4. This same limitation exists in the method proposed by Entner et al. [2013] in settings without missing data.

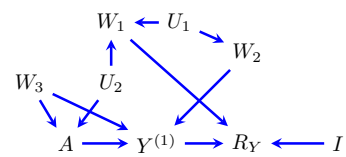


Figure 5: Setting where identification is possible using $\mathbf{Z} = \{W_1, W_2, W_3\}$, but there are no observed independence constraints that our method can use to verify this.

Furthermore, the incentive variable I is assumed to be a randomized variable in our method. In existing observational studies, it may not always be possible to identify such a variable. However, it may be possible to relax this condition such that I is conditionally randomized based on a set of pretreatment covariates. This could potentially allow for more flexibility in the identification of a valid incentive variable, but we do not pursue this line of inquiry here.

6 ESTIMATION PROCEDURE

The first step in a practical estimation procedure is to verify C1. If this test fails, then we are unable to empirically verify the validity of candidate adjustment sets using the given incentive I. If it succeeds, Algorithm 1 then searches over all possible assignments for W and $\mathbf{Z}$ to see if there is a combination of assignments that fulfills C2-C4. The search in Algorithm 1 proceeds in a style similar to the PC algorithm [Spirtes et al., 2000] for causal discovery (and the order of tests described for Figure 4) where conditioning sets of size 0 are tested first, followed by sets of size 1, and so on. Such an exponential search is unavoidable in situations where no valid adjustment set exists.[4] We use likelihood ratio tests for all tests in Algorithm 1; however, suitable non-parametric tests such as kernel conditional independence tests can also be applied [Zhang et al., 2011]. If Algorithm 1 returns a set $\mathbf{Z}$, then we have found a set satisfying B1, B2, S1, and S2, and we may use A as a valid shadow variable and $\mathbf{Z}$

[4] If desired, one may also perform a search for candidate adjustment sets using only a subset of all possible subsets of $\mathbf{W}$.

Algorithm 1 for finding a valid adjustment set $\mathbf{Z}$.

```
for each W ∈ W do
    Z_f ← W \ {W}
    for i from 0 to |Z_f| do
        Z_s ← all possible subsets of Z_f with size i
        for each Z ∈ Z_s do
            if C2-C4 are true using W and Z then
                return Z
            end if
        end for
    end for
end for
return "no adjustment set found"
```

as a valid adjustment set. Otherwise, we conclude that no adjustment set could be validated using our tests.

If we find a valid set $\mathbf{Z}$, we then estimate the distribution $p(A \mid \mathbf{Z})$ using the full dataset as it involves only fully observed variables. As the treatment is typically binary, we use a logistic regression model in our procedure. More flexible models such as generalized additive models or random forests are also possible depending on the sample size.

Next, we estimate the propensity score for R_Y, $p(R_Y = 1 \mid Y^{(1)}, \mathbf{Z})$. Let the cardinality of $\mathbf{Z}$ be $|\mathbf{Z}| = k$. In our estimation procedure, we use the odds ratio factorization of the propensity score in (3) with the parameterizations $\pi_0(\mathbf{Z}) = \text{expit}(\beta_1 Z_1 + \beta_2 Z_2 \ldots \beta_k Z_k)$ and $\eta(Y^{(1)}, \mathbf{Z}) = \exp(\gamma Y^{(1)})$. In total, we estimate $k+1$ parameters to recover the propensity score of R_Y; therefore, we need $k+1$ mean 0 estimating equations of the form $\mathbb{E}[(\frac{R_Y}{p(R_Y=1|Y^{(1)},\mathbf{Z};\boldsymbol{\beta},\gamma)} - 1) \times h(A, \mathbf{Z})]$. A simple choice for the first k equations is to use $h(A, \mathbf{Z}) = Z_i$ for each $Z_i \in \mathbf{Z}$. For the final equation, we use $h(A, \mathbf{Z}) = \overline{A}$, where $\overline{A}$ is the mean of the variable A. This gives us the following system of equations:

$$\mathbb{E}\left[\left(\frac{R_Y}{p(R_Y=1|Y^{(1)},\mathbf{Z};\boldsymbol{\beta},\gamma)} - 1\right)\begin{bmatrix} Z_1 \\ Z_2 \\ \ldots \\ Z_k \\ \overline{A} \end{bmatrix}\right] = 0 \qquad (5)$$

Because the expression $\frac{R_Y}{p(R_Y=1|Y^{(1)},\mathbf{Z};\boldsymbol{\beta},\gamma)}$ is 0 for each individual in the dataset where $R_Y = 0$, the system of estimating equations only uses observed rows of $Y^{(1)}$.

Using the estimated parameters for $p(A \mid \mathbf{Z})$ and $p(R_Y = 1 \mid Y^{(1)}, \mathbf{Z})$, we get estimates of both propensity scores for each row of data. Since inverse probability weighting estimators can be unstable due to large weights, we clip these propensity scores between values $p_{low} = 0.01$ and $p_{high} = 0.99$ [Hernán and Robins, 2010, Crump et al., 2009]. Finally, we estimate the causal effect by taking the empirical average for the expectation shown in identifying functional (4) for treatment values a and a'.

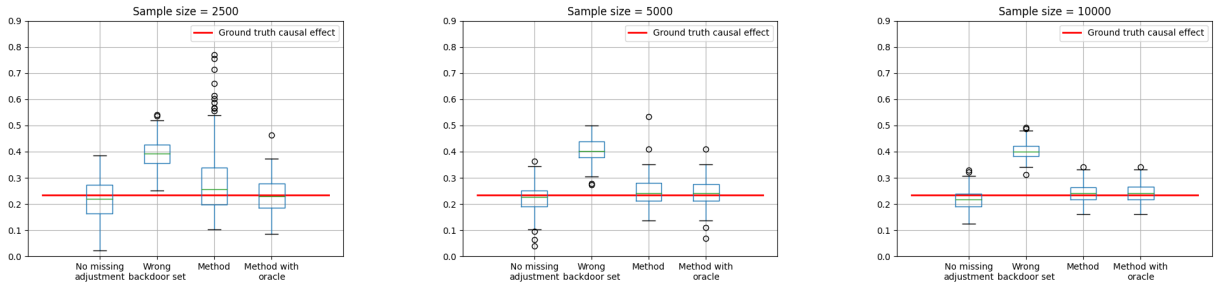
A summary of our procedure is as follows: (i) Test C1, if it holds, proceed to (ii), else, terminate the algorithm. (ii) Test C2-C4 using Algorithm 1, if it returns a set $\mathbf{Z}$, proceed to (iii), else terminate the algorithm. (iii) Estimate the propensity scores for R_Y and A and plug them into an inverse probability weighted estimator for the ACE based on the identifying functional in (4).

7 SIMULATION STUDY

For our simulations, we generate data according to the graph shown in Figure 7 and modifications of it that violate the shadow variable or backdoor conditions. We generate the pre-treatment covariates $\mathbf{W}$ from a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$ such that the off-diagonal entries, corresponding to the covariance of error terms, are non-zero for the pairs (W_2, W_3), (W_2, W_4), and (W_3, W_4). This is equivalent to a structural equation model with correlated errors where we have unmeasured confounding of the form $W_2 \leftrightarrow W_3$, $W_2 \leftrightarrow W_4$, and $W_3 \leftrightarrow W_4$. We generate $A, Y^{(1)}$, and R_Y as binary variables as functions of their parents in Figure 7. To generate R_Y, we use the odds ratio factorization and parameterization specified in section 6. The incentive variable I is normally distributed with a mean of 0 and variance 2. Precise details of the data generating process are provided in the Appendix.

Our first set of experiments focuses on evaluating the effectiveness of Algorithm 1 for finding a valid adjustment set when such a set exists and to correctly identify that no valid adjustment set exists when no such set exists. Before describing the experiments, we define key terms for evaluating accuracy. A *true positive* (TP) occurs when an adjustment set exists and the algorithm identifies the correct set of covariates. If the algorithm does not find an adjustment set or returns an incorrect one, then this is a *false negative* (FN). A *true negative* (TN) occurs when no possible adjustment set exists and the algorithm correctly finds no adjustment set. If the algorithm detects an adjustment set when no such set exists, this is considered a *false positive* (FP). *Sensitivity* is defined as $\frac{\#\text{TP}}{\#\text{TP}+\#\text{FN}}$, and *specificity* is defined as $\frac{\#\text{TN}}{\#\text{TN}+\#\text{FP}}$.

The experiment proceeds as follows. We first run 200 trials with data generated according to Figure 7 where Algorithm 1 should return $\mathbf{Z} = \{W_2, W_3, W_4\}$. We then run 200 trials where the algorithm should return no adjustment set as no valid set exists using data generated according to a DAG where we add edge $A \rightarrow R_Y$ to Figure 7 with probability 0.5 or where we treat W_4 as a latent variable with probability 0.5. The size of the dataset for each of these trials ranges from 500 to 10,000, and we use a significance level of $\alpha = 0.05$ for our tests. However, note that the effective sample size for some tests is roughly 60% of the full sample size due to the missingness of the outcome. Table 1 summarizes the results. In the Appendix, we also report results of this experiment using $\alpha = 0.01$ and $\alpha = 0.1$.



(a) (b) (c)

Figure 6: Simulation results for estimating the average causal effect using different methods and sample sizes.

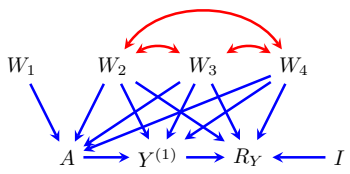


Figure 7: Graph used in our simulations.

Sample Size	Sensitivity	Specificity
500	0.011	0.350
2500	0.577	0.556
5000	0.812	0.818
10000	0.930	0.925

Table 1: Results of covariate search experiment.

The second set of experiments uses data from Figure 7 to evaluate the effectiveness of our method for downstream estimation. We compare these estimates to the bias introduced by either failing to adjust for missing data but using the correct backdoor adjustment set, or by failing to use a valid backdoor adjustment set but correctly adjusting for missing data. To estimate the ACE without adjusting for missing data, we subset the dataset to only rows of data where the outcome is observed and use a standard inverse probability weighting estimator. To estimate the ACE while using an invalid backdoor adjustment set, we use a subset of the correct adjustment set – $\{W_2, W_3\}$ – after correctly adjusting for missing data. We compare these two estimates with one obtained by running our full procedure described in Section 6 with $\alpha = 0.05$. We also generate estimates obtained using an independence oracle in Algorithm 1 to emphasize the importance of reliable conditional independence tests. We run 200 trials each for the sample sizes 500, 2,500, 5,000, and 10,000.

Results for 500 samples exhibit high bias due to inaccuracy of the tests and are reported in the Appendix in the interest of space. The red line in Figure 6 shows the ground truth causal effect. At sample size 10,000, the search algorithm correctly identifies the adjustment set for 93% of the trials, so the estimates from the method are nearly identical to the estimates from using the independence oracle. As the sample size increases, estimation from failing to adjust for missing data and confounding bias converge to biased values for the ACE. The method that ignores missingness produces reasonable estimates at low sample sizes, but it shows asymptotic convergence to a biased estimate. As expected, estimates obtained by using our full pipeline converge to the ground truth causal effect as sample size increases. Python code implementing our method and to reproduce our numerical studies are provided in the supplementary materials and will be publicly available upon publication.

8 CONCLUSION

In this paper, we discussed methods for causal effect estimation with a self-censoring outcome and when the underlying causal structure of the graph is unknown. We proved that when a set $\mathbf{Z} \subset \mathbf{W}$ satisfies both the backdoor adjustment set and the shadow variable adjustment set criteria, then identification of the average causal effect is possible through an inverse probability weighting functional. We further described a series of tests that can be used to empirically identify such a valid set $\mathbf{Z}$ using the observed data. We presented a simple search algorithm that uses our tests as a subroutine and estimates the average causal effect using an inverse probability weighting estimator if a valid set $\mathbf{Z}$ is found. Finally, we concluded with experiments based on synthetic data that demonstrate the accuracy of our search algorithm and estimation procedure.

We make progress on identifying situations where it is possible to overcome self-censoring and confounding to compute an unbiased estimate for the average causal effect. However, some open questions remain: To what extent can we relax the assumption that the incentive I must be randomized, and can we design semiparametric estimation strategies for the tests and final estimation piece that exhibit desirable statistical properties, such as robustness to model mispecification and lower asymptotic variance?

References

Rohit Bhattacharya, Razieh Nabi, Ilya Shpitser, and James M. Robins. Identification in missing data models represented by directed acyclic graphs. In *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence*. AUAI Press, 2019.

Richard K. Crump, V. Joseph Hotz, Guido W. Imbens, and Oscar A. Mitnik. Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, 96(1): 187–199, 2009.

Rhian M. Daniel, Michael G. Kenward, Simon N. Cousens, and Bianca L. De Stavola. Using causal diagrams to guide analysis in missing data problems. *Statistical Methods in Medical Research*, 21(3):243–256, 2012.

Guilherme Duarte, Noam Finkelstein, Dean Knox, Jonathan Mummolo, and Ilya Shpitser. An automated approach to causal inference in discrete settings. *arXiv preprint arXiv:2109.13471*, 2021.

Xavier d'Haultfoeuille. A new instrumental method for dealing with endogenous selection. *Journal of Econometrics*, 154(1):1–15, 2010.

Doris Entner, Patrik Hoyer, and Peter Spirtes. Data-driven covariate selection for nonparametric estimation of causal effects. In *Artificial Intelligence and Statistics*, pages 256–264. PMLR, 2013.

Miguel A. Hernán and James M. Robins. *Causal Inference: What If*. CRC Boca Raton, FL, 2010.

Wang Miao, Lan Liu, Eric Tchetgen Tchetgen, and Zhi Geng. Identification, doubly robust estimation, and semiparametric efficiency theory of nonignorable missing data with a shadow variable. *arXiv preprint arXiv:1509.02556*, 2015.

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

Karthika Mohan, Judea Pearl, and Jin Tian. Graphical models for inference with missing data. *Advances in neural information processing systems*, 26, 2013.

Karthika Mohan, Felix Thoemmes, and Judea Pearl. Estimation with incomplete data: The linear case. In *Proceedings of the International Joint Conferences on Artificial Intelligence Organization*, 2018.

Razieh Nabi, Rohit Bhattacharya, and Ilya Shpitser. Full law identification in graphical models of missing data: Completeness results. In *Proceedings of the 37th International Conference on Machine Learning*, pages 7153–7163. PMLR, 2020.

Razieh Nabi, Rohit Bhattacharya, Ilya Shpitser, and James Robins. Causal and counterfactual views of missing data models. *arXiv preprint arXiv:2210.05558*, 2022.

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

Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.

Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.

Judea Pearl. *Causality*. Cambridge University Press, 2009.

Mojdeh Saadati and Jin Tian. Adjustment criteria for recovering causal effects from missing data. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 561–577. Springer, 2019.

Ilya Shpitser, Tyler VanderWeele, and James M Robins. On the validity of covariate adjustment for estimating causal effects. *arXiv preprint arXiv:1203.3515*, 2012.

Peter L. Spirtes, Clark N. Glymour, and Richard Scheines. *Causation, prediction, and search*. MIT Press, 2000.

Aude Sportisse, Claire Boyer, and Julie Josse. Estimation and imputation in probabilistic principal component analysis with missing not at random data. *Advances in Neural Information Processing Systems*, 33:7067–7077, 2020.

Eric J. Tchetgen Tchetgen and Kathleen E. Wirth. A general instrumental variable framework for regression analysis with outcome missing not at random. *Biometrics*, 73(4): 1123–1131, 2017.

Ruibo Tu, Cheng Zhang, Paul Ackermann, Karthika Mohan, Hedvig Kjellström, and Kun Zhang. Causal discovery in the presence of missing data. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1762–1770. PMLR, 2019.

Charles F. Turner, Alia Al-Tayyib, Susan M Rogers, Elizabeth Eggleston, Maria A. Villarroel, Anthony M. Roman, James R. Chromy, and Phillip C. Cooley. Improving epidemiological surveys of sexual behaviour conducted by telephone. *International Journal of Epidemiology*, 38(4): 1118–1127, 2009.

Shu Yang, Linbo Wang, and Peng Ding. Causal inference with confounders missing not at random. *Biometrika*, 106 (4):875–888, 2019.

Kun Zhang, Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. Kernel-based conditional independence test and application in causal discovery. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*, pages 804–813, 2011.