Adjustment Criteria for Recovering Causal Effects from Missing Data

Mojdeh Saadati[✉][0000-0002-9092-8464] and Jin Tian[0000-0001-5313-1600]

Department of Computer Science
Iowa State University
2434 Osborn Dr, Ames, IA 50011, USA
{msaadati,jtian}@iastate.edu

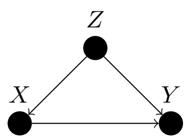
Abstract. Confounding bias, missing data, and selection bias are three common obstacles to valid causal inference in the data sciences. Covariate adjustment is the most pervasive technique for recovering casual effects from confounding bias. In this paper we introduce a covariate adjustment formulation for controlling confounding bias in the presence of missing-not-at-random data and develop a necessary and sufficient condition for recovering causal effects using the adjustment. We also introduce an adjustment formulation for controlling both confounding and selection biases in the presence of missing data and develop a necessary and sufficient condition for valid adjustment. Furthermore, we present an algorithm that lists all valid adjustment sets and an algorithm that finds a valid adjustment set containing the minimum number of variables, which are useful for researchers interested in selecting adjustment sets with desired properties.

Keywords: missing data · missing not at random · causal effect · adjustment · selection bias.

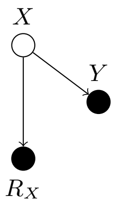
1 Introduction

Discovering causal relationships from observational data has been an important task in empirical sciences, for example, assessing the effect of a drug on curing diabetes, a fertilizer on growing agricultural products, and an advertisement on the success of a political party. One major challenge to estimating the effect of a treatment on an outcome from observational data is the existence of *confounding bias* - i.e., the lack of control on the effect of spurious variables on the outcome. This issue is formally addressed as the *identifiability problem* in [13], which concerns with computing the effect of a set of treatment variables ($\mathbf{X}$) on a set of outcome variables ($\mathbf{Y}$), denoted by $P(\mathbf{y} \mid do(\mathbf{x}))$, given observed probability distribution $P(\mathbf{V})$ and a causal graph G, where $P(\mathbf{V})$ corresponds to the observational data and G is a directed acyclic graph (DAG) representing qualitative causal relationship assumptions between variables in the domain. The effect $P(\mathbf{y} \mid do(\mathbf{x}))$ may not be equal to its probabilistic counterpart $P(\mathbf{y} \mid \mathbf{x})$ due to the existence of variables, called *covariates*, that affect both the treatments and outcomes, and the difference is known as confounding bias. For example, Fig. 1(a) shows a causal graph where variable Z is a covariate for estimating the effect of X on Y.

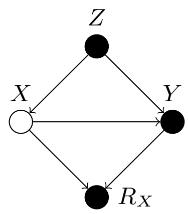
Confounding bias problem has been studied extensively in the field. In principle the identifiability problem can be solved using a set of causal inference rules called *do-calculus* [12], and complete identification algorithms have been developed [23, 5, 19]. In practice, however, the most widely used method for controlling the confounding bias is the "adjustment formula" $P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{z}} P(\mathbf{y} \mid \mathbf{x}, \mathbf{Z} = \mathbf{z})P(\mathbf{Z} = \mathbf{z})$, which dictates that the causal effect $P(\mathbf{y} \mid do(\mathbf{x}))$ can be computed by *controlling* for a set of covariates $\mathbf{Z}$. Pearl provided a back-door criterion under which a set $\mathbf{Z}$ makes the adjustment formula hold [12].



(a) Confounding bias



(b) A m-graph encoding MNAR data



(c) Confounding bias and MNAR

Fig. 1: Examples for confounding bias and MNAR

Another major challenge to valid causal inference is the missing data problem, which occurs when some variable values are missing from observed data. Missing data is a common problem in empirical sciences. Indeed there is a large literature on dealing with missing data in diverse disciplines including statistics, economics, social sciences, and machine learning. To analyze data with missing values, it is imperative to understand the mechanisms that lead to missing data. The seminal work by Rubin [15] classifies missing data mechanisms into three categories: *missing completely at random (MCAR)*, *missing at random (MAR)*, and *missing not at random (MNAR)*. Roughly speaking, the mechanism is MCAR if whether variable values are missing is completely independent of the values of variables in the data set; the mechanism is MAR when missingness is independent of the missing values given the observed values; and the mechanism is MNAR if it is neither MCAR nor MAR. For example, assume that in a study of the effect of family income (FI) and parent's education level (PE) on the quality of child's education (CE), some respondents chose not to reveal their child's education quality for various reasons. Fig. 2 shows causal graphs representing the three missing data mechanisms where R_{CE} is an indicator variable such that $R_{CE} = 0$ if the CE value is missing and $R_{CE} = 1$ otherwise. In these graphs solid circles represent always-observed variables and hollow circles represent variables that could have missing values. The model in Fig. 2(a) is MCAR, e.g., respondents decide to reveal the child's education quality based on coin-flips. The model in Fig. 2(b) is MAR, where respondents with higher family income have a higher chance of revealing the child's education quality; however whether the CE values are missing is independent of the actual values of CE given the FI value. The model in Fig. 2(c) is MNAR, where respondents with higher child's education quality have a higher chance of revealing it, i.e., whether the CE values are missing depends on the actual values of CE.

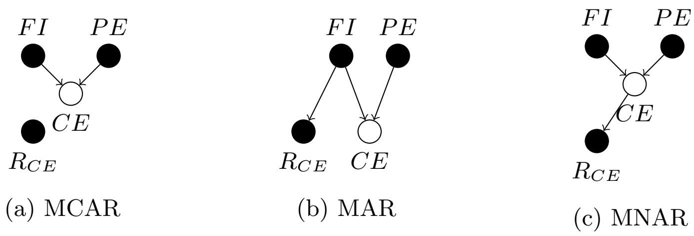


Fig. 2: Three types of missing data mechanisms

It is known that when the data is MAR, the underlying distribution is estimable from observed data with missing values. Then a causal effect is estimable if it is identifiable from the observed distribution [10]. However, if the data is MNAR, whether a probabilistic distribution or a causal effect is estimable from missing data depends closely on both the query and the exact missing data mechanisms. For example, in the MNAR model in Fig. 1(b), $P(X)$ cannot be estimated consistently even if infinite amount of data are collected, while $P(y|do(x)) = P(y|x) = P(y|x, R_X = 1)$ is estimable from missing data. On the other hand, in the MNAR model in Fig. 1(c), $P(y|do(x))$ is not estimable. In the MNAR model in Fig. 2(c), neither $P(CE)$ nor $P(CE \mid do(FI))$ can be estimated from observed data with missing values.

Various techniques have been developed to deal with missing data in statistical inference, e.g., listwise deletion [7], which requires data to be MCAR to obtain unbiased estimates, and multiple imputation [16], which requires MAR. Most of the work in machine learning makes MAR assumption and use maximum likelihood based methods (e.g. EM algorithms) [6] , with a few work explicitly incorporates missing data mechanism into the model [6, 9, 8].

The use of graphical models called *m-graphs* for inference with missing data was more recent [11]. M-graphs provide a general framework for inference with arbitrary types of missing data mechanisms including MNAR. Sufficient conditions for determining whether probabilistic queries (e.g., $P(\mathbf{y} \mid \mathbf{x})$ or $P(\mathbf{x}, \mathbf{y})$) are estimable from missing data are provided in [11, 10]. General algorithms for identifying the joint distribution have been developed in [18, 22].

The problem of identifying causal effects $P(\mathbf{y} \mid do(\mathbf{x}))$ from missing data in the causal graphical model settings has not been well studied. To the best of our knowledge the only results are the sufficient conditions given in [10]. The goal of this paper is to provide general conditions under which the causal effects can be identified from missing data using the covariate adjustment formula – the most pervasive method in practice for causal effect estimation under confounding bias.

We will also extend our results to cope with another common obstacles to valid causal inference - *selection bias*. Selection bias may happen due to preferential exclusion of part of the population from sampling. To illustrate, consider a study of the effect of diet on blood sugar. If individuals that are healthy and consume less sugar than average population are less likely to participate in the study, then the data gathered is not a faithful representation of the population and biased results will be produced. This bias cannot be removed by sampling more examples or controlling for confounding bias. Note that, in some sense, selection bias could be considered as a very special case of missing data mechanisms, where values of all of the variables are either all observed or all missing simultaneously. Missing data problem allows much richer missingness patterns such that in any particular observation, some of the variables could be observed and others could be missing. Missing data is modeled by introducing individual missingness indicators for each variable (such that $R_X = 0$ if X value is missing), while selection bias is typically modeled by introducing a single selection indicator variable (S) representing whether a unit is included in the sample or not (that is, if $S = 0$ then values of all variables are missing).

Identifying causal effects from selection bias has been studied in the literature [2, 1]. Adjustment formulas for recovering causal effects under selection bias have been introduced and complete graphical criteria have been developed [3, 4]. However these results are not applicable to the missing data problems which have much richer missingness patterns than could be modeled by selection bias. To the best of our knowledge, using adjustment for causal effect identification when the observed data suffers from missing values or both selection bias and missing values has not been studied in the causal graphical model settings. In this paper we will provide a characterization for these tasks.

Specifically, the contributions of this paper are:

1. We introduce a covariate adjustment formulation for recovering causal effects from missing data, and provide a necessary and sufficient graphical condition for when a set of covariates are valid for adjustment.
2. We introduce a covariate adjustment formulation for causal effects identification when the observed data suffer from both selection bias and missing values, and provide a necessary and sufficient graphical condition for the validity of a set of covariates for adjustment.
3. We develop an algorithm that lists *all* valid adjustment sets in polynomial delay time, and an algorithm that finds a valid adjustment set containing the minimum number of variables. The algorithms are useful for scientists to select adjustment sets with desired properties (e.g. low measurement cost).

The proofs are presented in the Appendix in [17] due to the space constraints.

2 Definitions and Related Work

Each variable will be represented with a capital letter (X) and its realized value with the small letter (x). We will use bold letters ($\mathbf{X}$) to denote sets of variables.

Structural Causal Models. The systematic analysis of confounding bias, missing data mechanisms, and selection bias requires a formal language where the characterization of the underlying data-generating model can be encoded explicitly. We use the language of Structural Causal Models (SCM) [13]. In SCMs, performing an action/intervention of setting $\mathbf{X}$=$\mathbf{x}$ is represented through the do-operator, $do(\mathbf{X}$=$\mathbf{x})$, which induces an experimental distribution $P(\mathbf{y}|do(\mathbf{x}))$, known as the causal effect of $\mathbf{X}$ on $\mathbf{Y}$. We will use do-calculus to derive causal expressions from other causal quantities. For a detailed discussion of SCMs and do-calculus, we refer readers to [13].

Each SCM M has a causal graph G associated to it, with directed arrows encoding direct causal relationships and dashed-bidirected arrows encoding the existence of an unobserved common causes. We use typical graph-theoretic terminology $Pa(\mathbf{C}), Ch(\mathbf{C}), De(\mathbf{C}), An(\mathbf{C})$ representing the union of $\mathbf{C}$ and respectively the parents, children, descendants, and ancestors of $\mathbf{C}$. We use $G_{\overline{\mathbf{C}_1}\underline{\mathbf{C}_2}}$ to denote the graph resulting from deleting all incoming edges to $\mathbf{C}_1$ and all outgoing edges from $\mathbf{C}_2$ in G. The expression $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z})_G$ denotes that $\mathbf{X}$ is d-separated from $\mathbf{Y}$ given $\mathbf{Z}$ in the corresponding causal graph G [13] (subscript may be omitted).

Missing Data and M-graphs. To deal with missing data, we use *m-graphs* introduced in [11] to represent both the data generation model and the missing data mechanisms. M-graphs enhance the causal graph G by introducing a set $\mathbf{R}$ of binary missingness indicator variables. We will also partition the set of observable variables $\mathbf{V}$ into $\mathbf{V}_o$ and $\mathbf{V}_m$ such that $\mathbf{V}_o$ is the set of variables that will be observed in all data cases and $\mathbf{V}_m$ is the set of variables that are missing in some data cases and observed in other cases. Every variable $V_i \in \mathbf{V}_m$ is associated with a variable $R_{V_i} \in \mathbf{R}$ such that, in any observed data case, $R_{V_i} = 0$ if the value of corresponding V_i is missing and $R_{V_i} = 1$ if V_i is observed. We assume that $\mathbf{R}$ variables may not be parents of variables in $\mathbf{V}$, since $\mathbf{R}$ variables are missingness indicator variables and we assume that the data generation process over $\mathbf{V}$ variables does not depend on the missingness mechanisms. For any set $\mathbf{C} \subseteq \mathbf{V}_m$, let $\mathbf{R}_\mathbf{C}$ represent the set of $\mathbf{R}$ variables corresponding to variables in $\mathbf{C}$. See Fig. 2 for examples of m-graphs, in which we use solid circles to represent always observed variables in $\mathbf{V}_o$ and $\mathbf{R}$, and hollow circles to represent partially observed variables in $\mathbf{V}_m$.

Causal Effect Identification by Adjustment. Covariate adjustment is the most widely used technique for identifying causal effects from observational data. Formally,

Definition 1 (Adjustment Formula [13]). *Given a causal graph G over a set of variables $\boldsymbol{V}$, a set $\boldsymbol{Z}$ is called* covariate adjustment *(or adjustment for short) for estimating the causal effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$, if, for any distribution $P(\boldsymbol{V})$ compatible with G, it holds that*

$$P(\boldsymbol{y} \mid do(\boldsymbol{x})) = \sum_{\boldsymbol{z}} P(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{z}) P(\boldsymbol{z}). \tag{1}$$

Pearl developed the celebrated "Backdoor Criterion" to determine whether a set is admissible for adjustment [12] given in the following:

Definition 2 (Backdoor Criterion). *A set of variables $\boldsymbol{Z}$ satisfies the back-door criterion relative to a pair of variables $(\boldsymbol{X}, \boldsymbol{Y})$ in a causal graph G if:*

a) No node in $\boldsymbol{Z}$ is a descendant of $\boldsymbol{X}$, and
b) $\boldsymbol{Z}$ blocks every path between $\boldsymbol{X}$ and $\boldsymbol{Y}$ that contains an arrow into $\boldsymbol{X}$.

Complete graphical conditions have been derived for determining whether a set is admissible for adjustment [20, 24, 14] as follows.

Definition 3 (Proper Causal Path). *A proper causal path from a node $X \in \boldsymbol{X}$ to a node $Y \in \boldsymbol{Y}$ is a causal path (i.e., a directed path) which does not intersect $\boldsymbol{X}$ except at the beginning of the path.*

Definition 4 (Adjustment Criterion [20]). *A set of variables $\boldsymbol{Z}$ satisfies the adjustment criterion relative to a pair of variables $(\boldsymbol{X}, \boldsymbol{Y})$ in a causal graph G if:*

a) No element of $\boldsymbol{Z}$ is a descendant in $G_{\overline{X}}$ of any $W \notin \boldsymbol{X}$ which lies on a proper causal path from $\boldsymbol{X}$ to $\boldsymbol{Y}$.
b) All non-causal paths between $\boldsymbol{X}$ and $\boldsymbol{Y}$ in G are blocked by $\boldsymbol{Z}$.

A set $\mathbf{Z}$ is an admissible adjustment for estimating the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ by the adjustment formula if and only if it satisfies the adjustment criterion.

3 Adjustment for Recovering Causal Effects from Missing Data

In this section we address the task of recovering a causal effect $P(\mathbf{y} \mid do(\mathbf{x}))$ from missing data given a m-graph G over observed variables $\mathbf{V} = \mathbf{V}_o \cup \mathbf{V}_m$ and missingness indicators $\mathbf{R}$. The main difference with the well studied identifiability problem [13], where we attempt to identify $P(\mathbf{y} \mid do(\mathbf{x}))$ from the joint distribution $P(\mathbf{V})$, lies in that, given data corrupted by missing values, $P(\mathbf{V})$ itself may not be recoverable. Instead, a distribution like $P(\mathbf{V}_o, \mathbf{V}_m, \mathbf{R} = 1)$ is assumed to be estimable from observed data cases in which all variables in $\mathbf{V}$ are observed (i.e., complete data cases). In general, in the context of missing data, the probability distributions in the form of $P(\mathbf{V}_o, \mathbf{W}, \mathbf{R}_{\mathbf{W}} = 1)$ for any $\mathbf{W} \subseteq \mathbf{V}_m$, called *manifest distributions*, are assumed to be estimable from observed data cases in which all variables in $\mathbf{W}$ are observed (values of variables in $\mathbf{V}_m \setminus \mathbf{W}$ are possibly missing).

The problem of recovering probabilistic queries from the manifest distributions has been studied in [11, 10, 18, 22].

We will extend the adjustment formula for identifying causal effects to the context of missing data based on the following observation which is stated in Theorem 1 in [11]:

Lemma 1 *For any* $\boldsymbol{W}_o, \boldsymbol{Z}_o \in \boldsymbol{V}_o$ *and* $\boldsymbol{W}_m, \boldsymbol{Z}_m \in \boldsymbol{V}_m$*,* $P(\boldsymbol{W}_o, \boldsymbol{W}_m \mid \boldsymbol{Z}_o, \boldsymbol{Z}_m, \boldsymbol{R}_{\boldsymbol{W}_m \cup \boldsymbol{Z}_m} = 1)$ *is recoverable.*

Formally, we introduce the adjustment formula for recovering causal effects from missing data by extending Eq. (1) as follows.

Definition 5 (M-Adjustment Formula). *Given a m-graph G over observed variables* $\boldsymbol{V} = \boldsymbol{V}_o \cup \boldsymbol{V}_m$ *and missingness indicators* $\boldsymbol{R}$*, a set* $\boldsymbol{Z} \subseteq \boldsymbol{V}$ *is called a m-adjustment (adjustment under missing data) set for estimating the causal effect of* $\boldsymbol{X}$ *on* $\boldsymbol{Y}$*, if, for every model compatible with G, it holds that*

$$P(\boldsymbol{y} \mid do(\boldsymbol{x})) = \sum_{\boldsymbol{z}} P(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{R}_{\boldsymbol{W}} = 1) P(\boldsymbol{z} \mid \boldsymbol{R}_{\boldsymbol{W}} = 1), \tag{2}$$

where $\boldsymbol{W} = \boldsymbol{V}_m \cap (\boldsymbol{X} \cup \boldsymbol{Y} \cup \boldsymbol{Z})$.

In the above formulation, we allow that the treatments $\mathbf{X}$, outcomes $\mathbf{Y}$, and covariates $\mathbf{Z}$ all could contain $\mathbf{V}_m$ variables that have missing values. Both terms on the right-hand-side of Eq. (2) are recoverable based on Lemma 1. Therefore the causal effect $P(\mathbf{y} \mid do(\mathbf{x}))$ is recoverable if it can be expressed in the form of m-adjustment.

We look for conditions under which a set $\mathbf{Z}$ is admissible as m-adjustment. Intuitively, we can start with the adjustment formula (1), consider an adjustment set as a candidate m-adjustment set, and then check for needed conditional independence relations. Based on this intuition, we obtain a straightforward sufficient condition for a set $\mathbf{Z}$ to be a m-adjustment set as follows.

Proposition 1 *A set* $\boldsymbol{Z}$ *is a m-adjustment set for estimating the causal effect of* $\boldsymbol{X}$ *on* $\boldsymbol{Y}$ *if, letting* $\boldsymbol{W} = \boldsymbol{V}_m \cap (\boldsymbol{X} \cup \boldsymbol{Y} \cup \boldsymbol{Z})$*,*

a) $\boldsymbol{Z}$ *satisfies the adjustment criterion (Def. 4),*
b) $\boldsymbol{R}_{\boldsymbol{W}}$ *is d-separated from* $\boldsymbol{Y}$ *given* $\boldsymbol{X}$*,* $\boldsymbol{Z}$*, i.e.,* $(\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{R}_{\boldsymbol{W}} \mid \boldsymbol{X}, \boldsymbol{Z})$*, and*
c) $\boldsymbol{Z}$ *is d-separated from* $\boldsymbol{R}_{\boldsymbol{W}}$*, i.e.,* $(\boldsymbol{Z} \perp\!\!\!\perp \boldsymbol{R}_{\boldsymbol{W}})$*.*

Proof. Condition (a) makes sure that the causal effect can be identified in terms of the adjustment formula Eq. (1). Then given Conditions (b) and (c), Eq. (1) is equal to Eq. (2).

However this straightforward criterion in Proposition 1 is not necessary. To witness, consider the set $\{V_{m1}, V_{m2}\}$ in Fig. 3 which satisfies the back-door criterion but not the conditions in Proposition 1 because V_{m2} is not d-separated from R_2. Still, it can be shown that $\{V_{m1}, V_{m2}\}$ is a m-adjustment set (e.g. by do-calculus derivation).

Next we introduce a complete criterion to determine whether a covariate set is admissible as m-adjustment to recover causal effects from missing data, extending the existing work on adjustment [20, 24, 3, 4, 14].

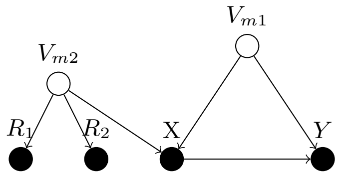


Fig. 3: In this m-graph V_{m2} is not d-separated from R_2. However, $\{V_{m2}, V_{m1}\}$ is an admissible m-adjustment set.

Definition 6 (M-Adjustment Criterion). *Given a m-graph G over observed variables $\boldsymbol{V} = \boldsymbol{V}_o \cup \boldsymbol{V}_m$ and missingness indicators $\boldsymbol{R}$, and disjoint sets of variables $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \subseteq \boldsymbol{V}$, letting $\boldsymbol{W} = \boldsymbol{V}_m \cap (\boldsymbol{X} \cup \boldsymbol{Y} \cup \boldsymbol{Z})$, $\boldsymbol{Z}$ satisfies the m-adjustment criterion relative to the pair $(\boldsymbol{X}, \boldsymbol{Y})$ if*

a) No element of $\boldsymbol{Z}$ is a descendant in $G_{\overline{X}}$ of any $W \notin \boldsymbol{X}$ which lies on a proper causal path from $\boldsymbol{X}$ to $\boldsymbol{Y}$.
b) All non-causal paths between $\boldsymbol{X}$ and $\boldsymbol{Y}$ in G are blocked by $\boldsymbol{Z}$ and $\boldsymbol{R_W}$.
c) $\boldsymbol{R_W}$ is d-separated from $\boldsymbol{Y}$ given $\boldsymbol{X}$ under the intervention of $do(\boldsymbol{x})$, i.e., $(\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{R_W} \mid \boldsymbol{X})_{G_{\overline{X}}}$.
d) Every $X \in \boldsymbol{X}$ is either a non-ancestor of $\boldsymbol{R_W}$ or it is d-separated from $\boldsymbol{Y}$ in $G_{\underline{X}}$, i.e., $\forall X \in \boldsymbol{X} \cap An(\boldsymbol{R_W}), (X \perp\!\!\!\perp \boldsymbol{Y})_{G_{\underline{X}}}$.

Theorem 1 (M-Adjustment) *A set $\boldsymbol{Z}$ is a m-adjustment set for recovering causal effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$ by the m-adjustment formula in Def. 5 if and only if it satisfies the m-adjustment criterion in Def. 6.*

Conditions (a) and (b) in Def. 6 echo the adjustment criterion in Def. 4, and it can be shown that if $\mathbf{Z}$ satisfies the m-adjustment criterion then it satisfies the adjustment criterion (using the fact that no variables in $\mathbf{R}$ can be parents of variables in $\mathbf{V}$). In other words, we only need to look for m-adjustment sets from admissible adjustment sets.

As an example consider Fig. 3. Both $\{V_{m1}\}$ and $\{V_{m1}, V_{m2}\}$ satisfy the m-adjustment criterion (and the adjustment criterion too). According to Theorem 1, $P(y \mid do(x))$ can be recovered from missing data by m-adjustment as

$$P(y \mid do(x)) = \sum_{v_{m1}} p(y \mid x, v_{m1}, R_1 = 1) P(v_{m1} \mid R_1 = 1), \tag{3}$$

$$= \sum_{v_{m1}, v_{m2}} P(y \mid x, v_{m1}, v_{m2}, R_1 = 1, R_2 = 1) P(v_{m1}, v_{m2} \mid R_1 = 1, R_2 = 1). \tag{4}$$

4 Listing M-Adjustment Sets

In the previous section we provided a criterion under which a set of variables $\mathbf{Z}$ is an admissible m-adjustment set for recovering a causal effect. It is natural to ask how to find an admissible set. In reality, it is common that more than one set of variables are admissible. In such situations it is possible that some m-adjustment sets might be preferable over others based on various aspects such as feasibility, difficulty, and cost of collecting variables. Next we first present an algorithm that systematically lists all m-adjustment sets and then present an algorithm that finds a minimum m-adjustment set. These algorithms provide flexibility for researchers to choose their preferred adjustment set based on their needs and assumptions.

4.1 Listing all admissible sets

It turns out in general there may exist exponential number of m-adjustment sets. To illustrate, we look for possible m-adjustment sets in the m-graph in Fig. 4 for recovering the causal effect $P(y \mid do(x))$ (this graph is adapted from a graph in [4]). A valid m-adjustment set $\mathbf{Z}$ needs to close all the k non-causal paths from X to Y. $\mathbf{Z}$ must contain at least one variable in $\{V_{i1}, V_{i2}, V_{i3}\}$ for each $i = 1, \ldots, k$. Therefore, to close each path, there are 7 possible $\mathbf{Z}$ sets, and for k paths, we have total 7^k $\mathbf{Z}$ sets as potential m-adjustment sets. For each of them, Conditions (c) and (d) in Def. 6 are satisfied because $(\mathbf{R} \perp\!\!\!\perp Y \mid X)_{G_{\overline{X}}}$ and X is not an ancestor of any $\mathbf{R}$ variables. We obtain that there are at least 7^k number of m-adjustment sets.

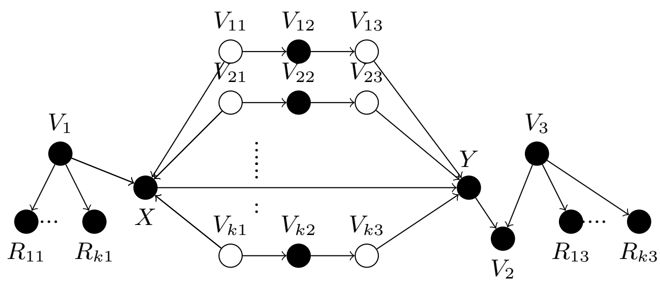


Fig. 4: An example of exponential number of m-adjustment sets

The above example demonstrates that any algorithm that lists all m-adjustment sets will be exponential time complexity. To deal with this issue, we will provide an algorithm with polynomial delay complexity [21]. Polynomial delay algorithms require polynomial time to generate the first output (or indicate failure) and the time between any two consecutive outputs is polynomial as well.

To facilitate the construction of a listing algorithm, we introduce a graph transformation called *Proper Backdoor Graph* originally introduced in [24].

Definition 7 (Proper Backdoor Graph [24]). *Let G be a causal graph, and $\boldsymbol{X}$, $\boldsymbol{Y}$ be disjoint subsets of variables. The proper backdoor graph, denoted as $G^{pbd}_{\boldsymbol{X},\boldsymbol{Y}}$, is obtained from G by removing the first edge of every proper causal path from $\boldsymbol{X}$ to $\boldsymbol{Y}$.*

Next we present an alternative equivalent formulation of the m-adjustment criterion in Def. 6 that will be useful in constructing a listing algorithm.

Definition 8 (M-Adjustment Criterion, Math. Version). *Given a m-graph G over observed variables $\boldsymbol{V} = \boldsymbol{V}_o \cup \boldsymbol{V}_m$ and missingness indicators $\boldsymbol{R}$, and disjoint sets of variables $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{Z} \subseteq \boldsymbol{V}$, letting $\boldsymbol{W} = \boldsymbol{V}_m \cap (\boldsymbol{X} \cup \boldsymbol{Y} \cup \boldsymbol{Z})$, $\boldsymbol{Z}$ satisfies the m-adjustment criterion relative to the pair ($\boldsymbol{X}$, $\boldsymbol{Y}$) if*

a) $\boldsymbol{Z} \cap Dpcp(\boldsymbol{X}, \boldsymbol{Y}) = \phi$
b) $(\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{X} \mid \boldsymbol{Z}, \boldsymbol{R}_{\boldsymbol{W}})_{G^{pbd}_{\boldsymbol{X},\boldsymbol{Y}}}$
c) $(\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{R}_{\boldsymbol{W}} \mid \boldsymbol{X})_{G_{\overline{X}}}$
d) $((\boldsymbol{X} \cap An(\boldsymbol{R}_{\boldsymbol{W}})) \perp\!\!\!\perp \boldsymbol{Y})_{G_{\underline{X}}}$

where $D_{pcp}(\boldsymbol{X}, \boldsymbol{Y}) = De((De(\boldsymbol{X})_{G_{\overline{X}}} \setminus \boldsymbol{X}) \cap An(\boldsymbol{Y})_{G_{\overline{X}}})$.

In Definition 8, $D_{pcp}(\mathbf{X}, \mathbf{Y})$, originally introduced in [24], represents the set of descendants of those variables in a proper causal path from $\mathbf{X}$ to $\mathbf{Y}$.

Proposition 2 *Definition 8 and Definition 6 are equivalent.*

Finally to help understanding the logic of the listing algorithm we introduce a definition originally introduced in [4]:

Definition 9 (Family of Separators [4]). *For disjoint sets of variables $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{E}$, and $\boldsymbol{I} \subseteq \boldsymbol{E}$, a family of separators is defined as follows:*

$$Z_{G(\boldsymbol{X},\boldsymbol{Y})}\langle \boldsymbol{I}, \boldsymbol{E}\rangle := \{\boldsymbol{Z} \mid (\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \mid \boldsymbol{Z})_G \text{ and } \boldsymbol{I} \subseteq \boldsymbol{Z} \subseteq \boldsymbol{E}\}, \tag{5}$$

which represents the set of all sets that d-separate $\boldsymbol{X}$ and $\boldsymbol{Y}$ and encompass all variables in set $\boldsymbol{I}$ but do not have any variables outside $\boldsymbol{E}$.

Algorithm 1 presents the function ListMAdj that lists all the m-adjustment sets in a given m-graph G for recovering the causal effect of $\mathbf{X}$ on $\mathbf{Y}$. We note that the algorithm uses an external function FindSep described in [24] (not presented in this paper). FindSep(G, $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{I}$, $\mathbf{E}$) will return a set in $Z_{G(\mathbf{X},\mathbf{Y})}\langle \mathbf{I}, \mathbf{E}\rangle$ if such a set exists; otherwise it returns $\perp$ representing failure.

Function ListMAdj works by first excluding all variables lying in the proper causal paths from consideration (Line 3) and then calling the function ListSepConditions (Line 4) to return all the m-adjustment sets. The function of ListSepConditions is summarized in the following proposition:

Proposition 3 (Correctness of ListSepCondition) *Given a m-graph G and sets of disjoint variables $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{E}$, and $\boldsymbol{I} \subseteq \boldsymbol{E}$, ListSepConditions lists all sets $\boldsymbol{Z}$ such that:*

$\boldsymbol{Z} \in \{\boldsymbol{Z} \mid (\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{R}_{\boldsymbol{Z}}, \boldsymbol{R}_{\boldsymbol{X}\cap \boldsymbol{V}_m}, \boldsymbol{R}_{\boldsymbol{Y}\cap \boldsymbol{V}_m})_{G^{pbd}_{\boldsymbol{X},\boldsymbol{Y}}}$ & $(\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{R}_{\boldsymbol{Z}} \mid \boldsymbol{X})_{G_{\overline{X}}}$ & $((\boldsymbol{X} \cap An(\boldsymbol{R}_{\boldsymbol{Z}})) \perp\!\!\!\perp \boldsymbol{Y})_{G_{\underline{X}}}$ & $\boldsymbol{I} \subseteq \boldsymbol{Z} \subseteq \boldsymbol{E}\}$ *where $\boldsymbol{R}_{\boldsymbol{Z}}$ is a shorthand for $\boldsymbol{R}_{\boldsymbol{Z}\cap \boldsymbol{V}_m}$.*

Algorithm 1: Listing all the m-adjustment sets

1 **Function** *ListMAdj* $(G, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{V}_o, \boldsymbol{V}_m, \boldsymbol{R})$
2 $\quad G^{pbd}_{\mathbf{X},\mathbf{Y}} \leftarrow$ compute proper back-door graph G
3 $\quad \mathbf{E} \leftarrow (\mathbf{V}_o \cup \mathbf{V}_m \cup \mathbf{R}) \setminus \{\mathbf{X} \cup \mathbf{Y} \cup D_{pcp}(\mathbf{X}, \mathbf{Y})\}$.
4 $\quad$ ListSepConditions$(G^{pbd}_{\mathbf{X},\mathbf{Y}}, \mathbf{X}, \mathbf{Y}, \mathbf{R}, \mathbf{V}_o, \mathbf{V}_m, \mathbf{I} = \{\mathbf{R}_{\mathbf{X}\cap\mathbf{V}_m} \cup \mathbf{R}_{\mathbf{Y}\cap\mathbf{V}_m}\}, \mathbf{E})$
5 **Function** *ListSepConditions* $(G, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{V}_o, \boldsymbol{V}_m, \boldsymbol{I}, \boldsymbol{E})$
6 $\quad$ **if** $(\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{R}_{\boldsymbol{I}} \mid \boldsymbol{X})_{G_{\overline{\boldsymbol{X}}}}$ *and* $((\boldsymbol{X} \cap An(\boldsymbol{R}_{\boldsymbol{I}})) \perp\!\!\!\perp \boldsymbol{Y})_{G_{\underline{\boldsymbol{X}}}}$ *and* $FindSep(G, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{I}, \boldsymbol{E}) \neq \perp$ **then**
7 $\quad\quad$ **if** $\boldsymbol{I} = \boldsymbol{E}$ **then**
8 $\quad\quad\quad$ Output$(\mathbf{I} \setminus \mathbf{R})$
9 $\quad\quad$ **else**
10 $\quad\quad\quad$ W $\leftarrow$ arbitrary variable from $\mathbf{E} \setminus (\mathbf{I} \cup \mathbf{R})$
11 $\quad\quad\quad$ **if** $W \in \boldsymbol{V}_o$ **then**
12 $\quad\quad\quad\quad$ ListSepConditions$(G, \mathbf{X}, \mathbf{Y}, \mathbf{R}, \mathbf{V}_o, \mathbf{V}_m, \mathbf{I} \cup \{W\}, \mathbf{E})$
13 $\quad\quad\quad\quad$ ListSepConditions$(G, \mathbf{X}, \mathbf{Y}, \mathbf{R}, \mathbf{V}_o, \mathbf{V}_m, \mathbf{I}, \mathbf{E} \setminus \{W\})$
14 $\quad\quad\quad$ **if** $W \in \boldsymbol{V}_m$ *and* $\boldsymbol{R}_W \in \boldsymbol{E}$ **then**
15 $\quad\quad\quad\quad$ ListSepConditions$(G, \mathbf{X}, \mathbf{Y}, \mathbf{R}, \mathbf{V}_o, \mathbf{V}_m, \mathbf{I} \cup \{W, \mathbf{R}_W\}, \mathbf{E})$
16 $\quad\quad\quad\quad$ ListSepConditions$(G, \mathbf{X}, \mathbf{Y}, \mathbf{R}, \mathbf{V}_o, \mathbf{V}_m, \mathbf{I}, \mathbf{E} \setminus \{W, \mathbf{R}_W\})$

ListSepConditions, by considering both including and not including each variable, recursively generates all subsets of $\mathbf{V}$ and for each generated set examines whether the conditions (b), (c), and (d) in Def. 8 hold or not. If those conditions were satisfied, the algorithm will return that candidate set as a m-adjustment set. ListSepConditions generates each potential set by taking advantage of back-tracking algorithm and at each recursion for a variable $W \in \mathbf{V}$ examines two cases of having W in candidate set or not. If $W \in \mathbf{V}_o$, then the algorithm examines having and not having this variable in the m-adjustment set and continues to decide about the rest of the variables in next recursion. If $W \in \mathbf{V}_m$, then the algorithm includes both W and R_W in the candidate m-adjustment set. Therefore, the algorithm considers both cases of having W, R_W and not having them in the candidate set. ListSepConditions, at the beginning of each recursion (Line 7), examines whether the candidate m-adjustment set so far satisfies the conditions (b), (c), (d) in Def. 8 or not. If any of them is not satisfied, the recursion stops for that candidate set. The function FindSep examines the existence of a set containing all variables in $\mathbf{I}$ and not having any of $\mathbf{V} \setminus \mathbf{E}$ that d-separates $\mathbf{X}$ from $\mathbf{Y}$. If this set does not exist FindSep returns $\perp$. ListSepConditions utilizes FindSep in order to check the satisfaction of condition (b) in Def. 8 for the candidate set. Since the graph G given to FindSep is a proper back-door graph, all paths between $\mathbf{X}$ and $\mathbf{Y}$ in this graph is non-causal. Therefore, if a set separates $\mathbf{X}$ and $\mathbf{Y}$ in G^{pbd}, this set blocks all non-causal paths from $\mathbf{X}$ to $\mathbf{Y}$ in G.

The following theorem states that ListMAdj lists all the m-adjustment sets in a given m-graph G for recovering the causal effect of $\mathbf{X}$ on $\mathbf{Y}$.

Theorem 2 (Correctness of ListMAdj) *Given a m-graph G and disjoint sets of variables $\boldsymbol{X}$ and $\boldsymbol{Y}$, ListMAdj returns all the sets that satisfy the m-adjustment criterion relative to ($\boldsymbol{X}$, $\boldsymbol{Y}$).*

The following results state that Algorithm 1 is polynomial delay.

Proposition 4 (Time Complexity of ListSepConditions) *ListSepConditions has a time complexity of $O(n(n+m))$ polynomial delay where n and m are the number of variables and edges in the given graph G respectively.*

Theorem 3 (Time Complexity of ListMAdj) *ListMAdj returns all the m-adjustment sets with $O(n(n+m))$ polynomial delay where n and m are the number of variables and edges in the given graph G respectively.*

4.2 Finding a Minimum M-Adjustment Set

The problem of finding a m-adjustment set with minimum number of variables is important in practice. Using a small adjustment set can reduce the computational time. The cost of measuring more variables might be another reason researchers may be interested in finding a minimum adjustment set. Next we present an algorithm that for a given graph G and disjoint sets $\mathbf{X}$ and $\mathbf{Y}$ returns a m-adjustment set with the minimum number of variables.

Algorithm 2: Find minimum size m-adjustment set

1 **Function** *FindMinAdjSet*(G, $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{V}_o$, $\boldsymbol{V}_m$, $\boldsymbol{R}$)
2   G′← compute proper back-door graph $G^{pbd}_{X,Y}$
3     $\mathbf{E} \leftarrow (\mathbf{V}_o \cup \mathbf{V}_m) \setminus \{\mathbf{X} \cup \mathbf{Y} \cup D_{pcp}(\mathbf{X}, \mathbf{Y})\}$.
4     $\mathbf{E}' \leftarrow \{E \in \mathbf{E} \mid E \in \mathbf{V}_o$ or $E \in \mathbf{V}_m$ and $(\mathbf{R}_E \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X})_{G'_{\overline{X}}}\}$
5     $\mathbf{E}'' \leftarrow \{E \in \mathbf{E}' \mid \mathbf{E} \in \mathbf{V}_o$ or $E \in \mathbf{V}_m$ and $(\mathbf{X} \cap An(\mathbf{R_E}) \perp\!\!\!\perp \mathbf{Y})_{G'_{\underline{X}}}\}$
6     $\mathbf{W} \leftarrow 1$ for all variables
7     $\mathbf{I} \leftarrow$ empty set
8     $\mathbf{N} \leftarrow$ FindMinCostSep(G', $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{I}$, $\mathbf{E}''$, $\mathbf{W}$)
9   return $\mathbf{N} \cup \mathbf{R}_N$

Function FindMinAdjSet takes a m-graph G as input and returns a m-adjustment set with minimum number of variables. The function works by first removing all variables that violate Conditions (a), (c), and (d) in the m-adjustment criterion Def. 8 in lines 2 to 5, and then calling an external function FinMinCostSep given in [24] which returns a minimum weight separator. FindMinAdjSet sets all the weights for each variable to be 1 to get a set with minimum size.

Theorem 4 (Correctness of FindMinAdjSet) *Given a m-graph G and disjoint sets of variables $\boldsymbol{X}$, and $\boldsymbol{Y}$, FindMinAdjSet returns a m-adjustment set relative to ($\boldsymbol{X}$, $\boldsymbol{Y}$) with the minimum number of variables.*

(a) Selection bias

(b) MNAR model with selection bias

ID	X	R_X	Y	S
1	1	1	0	1
2	0	1	1	1
3	NA	NA	NA	0
4	NA	0	1	1
5	NA	NA	NA	0

(c) An example data set compatible with the model in Fig. 5(b)

Fig. 5: Examples of selection bias and MNAR

Theorem 5 (Time Complexity of FindMinAdjSet) *FindMinAdjSet has a time complexity of $O(n^3)$.*

5 Adjustment from both Selection Bias and Missing Data

In Sections 3 and 4 we have addressed the task of recovering causal effects by adjustment from missing data. In practice another common issue that data scientists face in estimating causal effects is selection bias. Selection bias can be modeled by introducing a binary indicator variable S such that $S = 1$ if a unit is included in the sample, and $S = 0$ otherwise [2]. Graphically selection bias is modeled by a special hollow node S (drawn round with double border) that is pointed to by every variable in $\mathbf{V}$ that affects the process by which an unit is included in the data. In Fig. 5(a), for example, selection is affected by the treatment variable X.

In the context of selection bias, the observed distribution is $P(\mathbf{V} \mid S = 1)$, collected under seletion bias, instead of $P(\mathbf{V})$. The goal of inference is to recover the causal effect $P(\mathbf{y} \mid do(\mathbf{x}))$ from $P(\mathbf{V} \mid S = 1)$. The use of adjustment for recovering causal effects in this setting has been studied and complete adjustment conditions have been developed in [3, 4].

What if the observed data suffers from both selection bias and missing values? In the model in Fig. 5(b), for example, whether a unit is included in the sample depends on the value of the outcome Y. If a unit is included in the sample, the values of treatment X could be missing depending on the actual X values. Fig. 5(c) shows an example data set compatible with the model in Fig. 5(b) illustrating the difference between selection bias and missing data. To the best of our knowledge, causal inference under this setting has not been formally studied.

In this section, we will characterize the use of adjustment for causal effect identification when the observed data suffers from both selection bias and missing values. First we introduce an adjustment formula called *MS-adjustment* for recovering causal effect under both missing data and selection bias. Then we provide a complete condition under which a set $\mathbf{Z}$ is valid as MS-adjustment set.

Definition 10 (MS-Adjustment Formula). *Given a m-graph G over observed variables $\boldsymbol{V} = \boldsymbol{V}_o \cup \boldsymbol{V}_m$ and missingness indicators $\boldsymbol{R}$ augmented with a selection bias indicator S, a set $\boldsymbol{Z} \subseteq \boldsymbol{V}$ is called a ms-adjustment (adjustment under missing data and selection bias) set for estimating the causal effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$, if for every model compatible with G it holds that*

$$P(\boldsymbol{y} \mid do(\boldsymbol{x})) = \sum_{\boldsymbol{z}} P(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{R_W} = 1, S = 1) P(\boldsymbol{z} \mid \boldsymbol{R_W} = 1, S = 1), \tag{6}$$

where $\boldsymbol{W} = \boldsymbol{V}_m \cap (\boldsymbol{X} \cup \boldsymbol{Y} \cup \boldsymbol{Z})$.

Both terms on the right-hand-side of Eq. (6) are recoverable from selection biased data in which all variables in $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$ are observed. Therefore the causal effect $P(\mathbf{y} \mid do(\mathbf{x}))$ is recoverable if it can be expressed in the form of ms-adjustment.

Next we provide a complete criterion to determine whether a set $\mathbf{Z}$ is an admissible ms-adjustment.

Definition 11 (MS-Adjustment Criterion). *Given a m-graph G over observed variables $\boldsymbol{V} = \boldsymbol{V}_o \cup \boldsymbol{V}_m$ and missingness indicators $\boldsymbol{R}$ augmented with a selection bias indicator S, and disjoint sets of variables $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{Z}$, letting $\boldsymbol{W} = \boldsymbol{V}_m \cap (\boldsymbol{X} \cup \boldsymbol{Y} \cup \boldsymbol{Z})$, $\boldsymbol{Z}$ satisfies the ms-adjustment criterion relative to the pair ($\boldsymbol{X}$, $\boldsymbol{Y}$) if*

a) No element of $\boldsymbol{Z}$ is a descendant in $G_{\overline{\boldsymbol{X}}}$ of any $W \notin \boldsymbol{X}$ which lies on a proper causal path from $\boldsymbol{X}$ to $\boldsymbol{Y}$.
b) All non-causal paths between $\boldsymbol{X}$ and $\boldsymbol{Y}$ in G are blocked by $\boldsymbol{Z}$, $\boldsymbol{R_W}$, and S.
c) $\boldsymbol{R_W}$ and S are d-separated from $\boldsymbol{Y}$ given $\boldsymbol{X}$ under the intervention of $do(\boldsymbol{x})$. i.e., $(\boldsymbol{Y} \perp\!\!\!\perp (\boldsymbol{R_W} \cup S) \mid \boldsymbol{X})_{G_{\overline{\boldsymbol{X}}}}$
d) Every $X \in \boldsymbol{X}$ is either a non-ancestor of $\{\boldsymbol{R_W}, S\}$ or it is d-separated from $\boldsymbol{Y}$ in $G_{\underline{X}}$. i.e., $\forall X \in \boldsymbol{X} \cap An(\boldsymbol{R_W} \cup S), (X \perp\!\!\!\perp \boldsymbol{Y})_{G_{\underline{X}}}$.

Theorem 6 (MS-Adjustment) *A set $\boldsymbol{Z}$ is a ms-adjustment set for recovering causal effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$ by the ms-adjustment formula in Definition 10 if and only if it satisfies the ms-adjustment criterion in Definition 11.*

To demonstrate the application of Theorem 6, consider the causal graph in Fig. 6 where V_1 ,V_5, Y may have missing values and the selection S depends on the values of X_2. To recover the causal effect of $\{X_1, X_2\}$ on variable Y, V_1 satisfies the ms-adjustment criterion. We obtain $P(y \mid do(x_1, x_2)) = \sum_{V_1} P(y \mid x_1, x_2, V_1, S = 1, R_y = 1, R_1 = 1) P(V_1 \mid S = 1, R_y = 1, R_1 = 1)$.

We note that the two algorithms given in Section 4, for listing all m-adjustment sets and finding a minimum size m-adjustment set, can be extended to list all ms-adjustment sets and find a minimum ms-adjustment set with minor modifications.

6 Conclusion

In this paper we introduce a m-adjustment formula for recovering causal effect in the presence of MNAR data and provide a necessary and sufficient graphical condition - m-adjustment criterion for when a set of covariates are valid

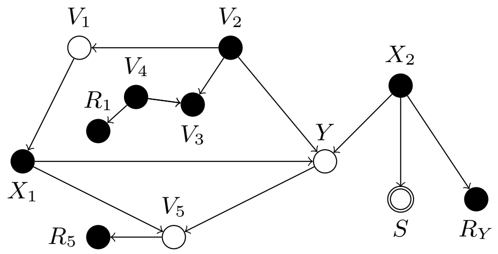


Fig. 6: An example for recovering causal effect under both selection bias and MNAR data

m-adjustment. We introduce a ms-adjustment formulation for causal effects identification in the presence of both selection bias and MNAR data and provide a necessary and sufficient graphical condition - ms-adjustment criterion for when a set of covariates are valid ms-adjustment. We develop an algorithm that lists all valid m-adjustment or ms-adjustment sets in polynomial delay time, and an algorithm that finds a valid m-adjustment or ms-adjustment set containing the minimum number of variables. The algorithms are useful for data scientists to select adjustment sets with desired properties (e.g. low measurement cost). Adjustment is the most used tool for estimating causal effect in the data sciences. The results in this paper should help to alleviate the problem of missing data and selection bias in a broad range of data-intensive applications.

Acknowledgements

This research was partially supported by NSF grant IIS-1704352 and ONR grant N000141712140.

References

1. Bareinboim, E., Tian, J.: Recovering causal effects from selection bias. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. pp. 3475–3481 (2015)
2. Bareinboim, E., Tian, J., Pearl, J.: Recovering from selection bias in causal and statistical inference. In: Proceeding of the Twenty-Eighth AAAI Conference on Artificial Intelligence. pp. 2410–2416 (2014)
3. Correa, J.D., Bareinboim, E.: Causal effect identification by adjustment under confounding and selection biases. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. pp. 3740–3746 (2017)
4. Correa, J.D., Tian, J., Bareinboim, E.: Generalized adjustment under confounding and selection biases. In: Thirty-Second AAAI Conference on Artificial Intelligence. pp. 6335–6342 (2018)
5. Huang, Y., Valtorta, M.: Identifiability in causal bayesian networks: A sound and complete algorithm. In: Proceedings of the 21st National Conference on Artificial Intelligence. vol. 2, pp. 1149–1154 (2006)

6. Koller, D., Friedman, N., Bach, F.: Probabilistic graphical models: principles and techniques. MIT press (2009)
7. Little, R.J.A., Rubin, D.B.: Statistical Analysis with Missing Data. John Wiley & Sons, Inc. (1986)
8. Marlin, B.M., Zemel, R.S., Roweis, S.T., Slaney, M.: Collaborative filtering and the missing at random assumption. In: Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence. pp. 267–275 (2007)
9. Marlin, B.M., Zemel, R.S., Roweis, S.T., Slaney, M.: Recommender systems, missing data and statistical model estimation. In: Proceedings of the 22nd International Joint Conference on Artificial Intelligence. pp. 2686–2691 (2011)
10. Mohan, K., Pearl, J.: Graphical models for recovering probabilistic and causal queries from missing data. In: Advances in Neural Information Processing Systems. pp. 1520–1528 (2014)
11. Mohan, K., Pearl, J., Tian, J.: Graphical models for inference with missing data. In: Advances in neural information processing systems. pp. 1277–1285 (2013)
12. Pearl, J.: Causal diagrams for empirical research. Biometrika **82**(4), 669–688 (1995)
13. Pearl, J.: Causality: Models, Reasoning and Inference. Cambridge University Press, 2nd edn. (2009)
14. Perkovic, E., Textor, J., Kalisch, M., Maathuis, M.H.: Complete graphical characterization and construction of adjustment sets in markov equivalence classes of ancestral graphs. The Journal of Machine Learning Research **18**(1), 8132–8193 (2017)
15. Rubin, D.: Inference and missing data. Biometrika **63**(3), 581–592 (1976)
16. Rubin, D.B.: Multiple imputations in sample surveys-a phenomenological bayesian approach to nonresponse. In: Proceedings of the survey research methods section of the American Statistical Association. vol. 1, pp. 20–34 (1978)
17. Saadati, M., Tian, J.: Adjustment criteria for recovering causal effects from missing data. Tech. rep., Department of Computer Science, Iowa State University. (2019), arXiv:1907.01654
18. Shpitser, I., Mohan, K., Pearl, J.: Missing data as a causal and probabilistic problem. In: Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence. pp. 802–811 (2015)
19. Shpitser, I., Pearl, J.: Identification of joint interventional distributions in recursive semi-markovian causal models. In: Proceedings of the National Conference on Artificial Intelligence. vol. 21, p. 1219 (2006)
20. Shpitser, I., VanderWeele, T., Robins, J.M.: On the validity of covariate adjustment for estimating causal effects. In: Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence. pp. 527–536 (2010)
21. Takata, K.: Space-optimal, backtracking algorithms to list the minimal vertex separators of a graph. Discrete Applied Mathematics **158**(15), 1660–1667 (2010)
22. Tian, J.: Recovering probability distributions from missing data. In: Proceedings of the Ninth Asian Conference on Machine Learning. vol. PMLR 77 (2017)
23. Tian, J., Pearl, J.: A general identification condition for causal effects. In: Eighteenth National Conference on Artificial Intelligence. pp. 567–573 (2002)
24. van der Zander, B., Liśkiewicz, M., Textor, J.: Constructing separators and adjustment sets in ancestral graphs. In: Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence. pp. 907–916 (2014)