Missing Data as a Causal and Probabilistic Problem

Ilya Shpitser
Mathematical Sciences
University of Southampton
Southampton, UK SO14 6WD
i.shpitser@soton.ac.uk

Karthika Mohan
Dept. of Computer Science
Univ. of California, Los Angeles
Los Angeles, CA 90095
karthika@cs.ucla.edu

Judea Pearl
Dept. of Computer Science
Univ. of California, Los Angeles
Los Angeles, CA 90095
judea@cs.ucla.edu

Abstract

Causal inference is often phrased as a missing data problem – for every unit, only the response to observed treatment assignment is known, the response to other treatment assignments is not. In this paper, we extend the converse approach of [7] of representing missing data problems to causal models where only interventions on missingness indicators are allowed. We further use this representation to leverage techniques developed for the problem of identification of causal effects to give a general criterion for cases where a joint distribution containing missing variables can be recovered from data actually observed, given assumptions on missingness mechanisms. This criterion is significantly more general than the commonly used “missing at random” (MAR) criterion, and generalizes past work which also exploits a graphical representation of missingness. In fact, the relationship of our criterion to MAR is not unlike the relationship between the ID algorithm for identification of causal effects [22,18], and conditional ignorability [13].

1 INTRODUCTION

Missing data is a ubiquitous problem in data analysis, and can arise due to imperfect data collection, or various types of censoring, for instance via loss to followup, or death. In addition, causal inference can be viewed as a missing data problem, since the fundamental problem of causal inference [4] is that for every unit only the response to observed treatment assignment is known, the responses to other, hypothetical treatment assignments are not known.

Handling missing data entails either dealing with a latent variable model or finding plausible assumptions under which recoverability, that is unbiased inferences about all cases from the observed cases, is possible. Well-known approaches of the former type include fitting a latent variable model via gradient descent [17], the EM algorithm [1], or performing Monte Carlo averaging via multiple imputation [16]. Well-known approaches of the latter type include the Kaplan-Meier estimator in survival analysis [5], and adjustments based on Missing Completely At Random (MCAR), and Missing At Random (MAR) assumptions [15].

While methods based on inference in a latent variable model are more generally applicable, they are also methodologically and computationally challenging. At the same time, recoverability methods based on MCAR and MAR rely on strong assumptions on how missingness comes about. When neither MCAR nor MAR holds, data is said to be Missing Not At Random (MNAR), and in this case a characterization of recoverability is an open problem, although many sufficient conditions for recoverability are known [7,6].

In this paper, we take the converse view to “causality as missing data,” and view missing data as a particular type of partly causal, and partly probabilistic inference problem [2,7]. We then represent this problem using partly causal, and partly probabilistic graphical models, and exploit techniques developed for similar models in the context of identification of causal effects to develop a general algorithm for recoverability under MNAR. In fact, the relationship between our algorithm and MAR is not unlike the relationship between the ID algorithm for identification of causal effects [22,18,19], and the conditional ignorability assumption in causal inference [13].

The paper is organized as follows. We introduce the notation and concepts we will need in section 2. In section 3 we use missingness graphs and missingness models to formally define missing data as a type of causal inference problem where only interventions on certain variables are allowed. We introduce recoverability and give examples of where recoverability is possible in MNAR settings in section 4. We introduce a general algorithm for recoverability we call MID in section 5 and show it is sound. Section 6 illustrates a complex case where the entire recursive structure of MID is necessary. Section 7 discusses non-recoverability, and section 8 contains our conclusions.
2 PRELIMINARIES

Variables are capital letters, values are small letters. Variable sets are bold capital letters, value sets are bold small letters. A state space for a variable \( A \) is \( \mathcal{X}_A \). A state space for a set of variables \( A \) is the Cartesian product of the individual state spaces: \( \mathcal{X}_A \equiv \times_{A \in A} \mathcal{X}_A \). For a set of values \( a \), and \( B \subseteq A \), denote by \( \mathbf{a}_B \) a projection of \( a \) to \( B \). Denote \( \mathbf{a}_B \) as a shorthand for \( a_{\{B\}} \). We will denote a vector of 0s as \( \mathbf{0} \). \( \mathbf{0}_B \) means "a set of 0 values to \( B \)."

2.1 GRAPH THEORY AND NOTATION

A directed graph consists of a set of nodes and directed arrows (\( \rightarrow \)) connecting pairs of nodes. A mixed graph consists of a set of nodes and directed and/or bidirected arrows (\( \leftrightarrow \)) connecting pairs of nodes. A path is a sequence of distinct edges where any edge in a sequence that ends in a node \( A \) implies the subsequent edge must start with \( A \), and each such node \( A \) may only occur at most once in this way in the sequence. A directed path from a node \( A \) to a node \( B \) is a path consisting of directed edges where all edges on the path point away from \( A \) and towards \( B \).

If the edge \( X \rightarrow Y \) exists in a graph \( \mathcal{G} \), we say \( X \) is a parent of \( Y \) and \( Y \) is a child of \( X \). If a directed path from \( X \) to \( Y \) exists in \( \mathcal{G} \), we say \( X \) is an ancestor of \( Y \), and \( Y \) is a descendant of \( X \). We denote by \( \text{pa}_\mathcal{G}(A), \text{ch}_\mathcal{G}(A), \text{de}_\mathcal{G}(A), \text{nd}_\mathcal{G}(A) \) the sets of parents, children, descendants, and ancestors, respectively. These are defined disjunctively for sets, e.g., \( \text{pa}_\mathcal{G}(A) = \cup_{A \in A} \text{pa}_\mathcal{G}(A) \). Let \( \text{fa}_\mathcal{G}(A) = \text{pa}_\mathcal{G}(A) \cap \{A\} \), \( \text{pa}_\mathcal{G}(A) = \text{pa}_\mathcal{G}(A) \setminus \{A\} \), \( \text{nd}_\mathcal{G}(A) = \text{de}_\mathcal{G}(A) \cap \text{pa}_\mathcal{G}(A) \). Given a graph \( \mathcal{G} \), we say a vertex set \( A \) is ancestral if \( \text{an}_\mathcal{G}(A) = A \). By convention, in any directed graph, \( A \in \text{ch}(A) \) and \( \text{de}(A) \). A directed graph is said to have a directed cycle if there is \( X, Y \) such that \( X \in \text{an}(Y) \cap \text{ch}(Y) \). A directed graph without such cycles is called a directed acyclic graph (DAG).

A conditional DAG (CDAG) \( \mathcal{G}(V \mid W) \) is a DAG with vertices \( V \cup W \) with the property that \( \text{pa}_\mathcal{G}(W) = \emptyset \). We will denote vertices in \( V \) as circles, and vertices in \( W \) as squares. Note that we do not require that all \( V \in V \) must have parents. We simply distinguish certain parentless nodes in \( \mathcal{G} \) as \( W \). We will interpret vertices in \( V \) as associated with random variables and vertices in \( W \) as associated with variables that have been "set to a constant" in some way. One example of a CDAG is a mutilated graph that arises in the analysis of interventional distributions. When considering d-separation on vertices in \( V \) in a CDAG \( \mathcal{G} \), we will treat it as ordinary d-separation in a DAG, except all nodes in \( W \) are implicitly conditioned on.

If vertices not in \( W \) in a CDAG correspond to a variable partition into observed and missing variables, we will explicitly denote the set of vertices corresponding to missing variables as \( M \), and the other vertices as \( O \), like so: \( \mathcal{G}(O, M \mid W) \). A CDAG where \( W \) is empty is written as \( \mathcal{G}(V) \) or \( \mathcal{G}(O, M) \) as a shorthand.

A conditional acyclic directed mixed graph (CADMG) \( \mathcal{G}(V \mid W) \) is a mixed graph with two types of edges \( \rightarrow \) and \( \leftrightarrow \) with no directed cycles, where no arrowhead may point to an element of \( W \). We will sometimes omit variables from CDAGs and CADMGs if they are obvious to avoid notation clutter, e.g. we will write \( \mathcal{G}(V \mid W) \) simply as \( \mathcal{G} \). Given a CDAG \( \mathcal{G}(O, M \mid W) \), define \( \mathcal{G}_B(\mathcal{G}) \) to be an edge subgraph obtained from \( \mathcal{G} \) by removing all arrows pointing away from \( B \).

Define a latent projection of \( \mathcal{G}(O, M \mid W) \) onto \( O \cup W \) to be a CADMG \( \mathcal{G}(O, M \mid W) \equiv \mathcal{G}^\dagger(O \mid W) \) such that for any \( V_1, V_2 \in O \cup W \):

- There is an edge \( V_1 \rightarrow V_2 \) if and only if there is a directed path \( V_1 \rightarrow \ldots \rightarrow V_2 \) in \( \mathcal{G}(O, M \mid W) \) with all intermediate nodes in \( M \).
- There is an edge \( V_1 \leftrightarrow V_2 \) if and only if there is a marginally d-connected path \( V_1 \leftarrow \ldots \leftarrow V_2 \) in \( \mathcal{G}(O, M \mid W) \) with all intermediate nodes in \( M \).

Latent projections are a simplified representation of an infinitely large class of hidden variable CDAGs with structural features in common. In this paper, we use them only to simplify the statements and proofs of our results. The results themselves will always be about models represented by DAGs (and CDAGs).

Given a CDAG \( \mathcal{G}(V \mid W) \), and \( A \subseteq V \cup W \), define \( \mathcal{G}_A(V \mid W) \equiv \mathcal{G}(V \cap A \mid W \cap A) \) to be a subgraph of \( \mathcal{G} \) containing the vertex set \( A \) and any edge in \( \mathcal{G} \) between elements in \( A \).

Given a CADMG \( \mathcal{G}(V \mid W) \), and \( V \in V \), define the district (or c-component [22, 28]) of \( V \) in \( \mathcal{G}(V \mid W) \) to be \( \text{dis}_{\mathcal{G}}(V) = \{A \in V \mid V \leftarrow \ldots \leftarrow A\} \). The set of districts of \( \mathcal{G}(V \mid W) \) is denoted by \( \mathcal{D}(\mathcal{G}(V \mid W)) \), and it partitions \( V \).

For any \( V \in O \) in a CDAG \( \mathcal{G}(O, M \mid W) \) where for every \( M \in M \), \( \text{de}_\mathcal{G}(M) \cap O \neq \emptyset \), define the clan of \( V \) as \( \text{clan}_\mathcal{G}(V) = \text{an}_{\mathcal{G}\cap_{V=M}}(\mathcal{D}_V) \), where \( \mathcal{D}_V = \text{dis}_{\mathcal{G}\cap_{O}}(V) \). For example, in \( \mathcal{G} \) shown in Fig. [1](c), where \( \{X, W\} \) are missing, \( \text{clan}_\mathcal{G}(R_X) = \text{clan}_\mathcal{G}(S_W) = \{W, R_X, S_W\} \), and \( \text{clan}_\mathcal{G}(R_W) = \text{clan}_\mathcal{G}(S_X) = \{X, R_W, S_X\} \).

For any \( D \in \mathcal{D}(\mathcal{G}(O, M \mid W)) \), and \( D_1, D_2 \in D \), \( \text{clan}_\mathcal{G}(D_1) = \text{clan}_\mathcal{G}(D_2) \). Thus we will write \( \text{clan}_\mathcal{G}(D) = \text{clan}_\mathcal{G}(D) \), for any \( D \in D \). In fact, the set of clans partitions \( O \cup M \) in \( \mathcal{G} \) with the property above.

Given a CDAG \( \mathcal{G} \), a total ordering < on vertices in \( \mathcal{G} \) is topological given \( \mathcal{G} \) if \( A < B \) implies \( A \notin \text{de}_\mathcal{G}(B) \). Given an ordering < topological given \( \mathcal{G} \), define for any vertex \( V \),
in $G$, $\text{pre}_G(V) = \{W \neq V | W \prec V\}$. Given $\prec$ topological for $G$ with a vertex set $V$, if there is a subgraph $G'$ of $G$ with a vertex set $V' \subseteq V$, we will view $\prec$ with respect to $G'$ as the natural subordering restricted to $V'$. Note that this subordering will also be topological with respect to $G'$.

A counterfactual (potential outcome) $Y(a)$ is a response $Y$ to a hypothetical assignment of a set of treatments $A$ to values $a$. Given a set of potential outcomes $Y_1(a), \ldots, Y_k(a)$, where $Y = \{Y_1, \ldots, Y_k\}$, we may consider a joint distribution

$$p(Y_1, \ldots, Y_k)(a) \equiv p(Y(a)) \equiv p(Y | do(a)).$$

The do(.) notation is discussed extensively in \cite{10}.

3 MISSING GRAPHS AND MISSINGNESS MODELS

Given a CDAG $G(V | W)$, we say $p_W(V)$ (a mapping from $X_W$ to $p(V)$) is Markov relative to $G$ if

$$p_w(V) = \prod_{V \in V} p_w(V | pa_G(V) \setminus W), \quad (1)$$

and each term $p_w(V | pa_G(V) \setminus W)$ only depends on $W \cap pa_V(V)$.

**Definition 1 (missingness graph)** Given a DAG $G(O, M)$, a DAG $G^m$ is called a missingness graph for $G$ if $G^m$ has the vertex set $O \cup M \cup R_M \cup S_M$, where $R_M = \{R_M | M \in M\}$, $S_M = \{S_M | M \in M\}$, $G = G^m_{O,M}$, and for all $M$ in $M$, $pa_G(S_M) = \{M, R_M\}$, $ch_G(M) = \emptyset$, and $ch_G(R_M) \cap (O \cup M) = \emptyset$.

By convention, if $M = \emptyset$, then $S_M = R_M = \emptyset$. We will refer to $O \cup R_M \cup S_M$ as $V$, and to $V \cup M$ as $A$. We call elements of $R_M$ indicators, and elements of $S_M$ proxies.

Define $\mathcal{M}(G^m(A))$ to be the missingness model for a missingness graph $G^m(A)$ as a set of distributions $\{p(A)\}$ over the following set of counterfactuals $A \equiv \{A(r) | R \subseteq R_M, r \in X_R\}$, such that $(\forall M \in M) R_{R_M} = \{0, 1\}, X_{S_M} = X_M \cup \{\text{missing}\}$, and the missingness mechanism that determines the value of $S_M$ is as follows: $S_M(0_{R_M}) = M$ and $S_M(1_{R_M}) = \text{missing}$. In addition: $(\forall R \subseteq R_M, r \in X_R, V \in A), \quad V(r) \perp \{\text{nd}_G(R_M)(V)\}(r) | \{pa_G(R_M)(V)\}(r). \quad (2)$

To obtain the set $\mathcal{A}$, we first define

$$\{A(r) | r \in X_{R_M} \} \equiv \{S_M(r), O, M | r \in X_{R_M}\}$$

and obtain the others via modified recursive substitution as in definition 43 in \cite{10}, pp. 100-101.

A missingness model is thus really a particular type of a graphical causal model where we only define interventions on a subset of variables $\mathcal{I}$. In particular, we allow $G(O, M)$ to represent an ordinary hidden variable statistical model. (2) is just the DAG local Markov property linking $p(A(r))$ and $G^m_{R_M}$, for every $r$. If we had chosen to split variables in $R$ into random and intervened versions, and display both explicitly in the graph rather than only displaying the random version of variables, and keeping intervened versions implicit, as we do in $G^m_{R_M}$ we would end up with Single World Intervention Graphs (SWIGs), and the appropriate local Markov property for those graphs, as discussed in \cite{11}.

Standard results on DAG models imply (2) is equivalent to (1) for $p(A(r))$ and $G^m_{R_M}$ (if we let $W = \emptyset$, and keep fixed versions of $R$ implicit in the graph). We may also let $W = R$, and treat $R$ as a split node as in a SWIG.

4 RECOVERABILITY

We call $p(V)$ the manifest distribution. A functional of $p(A)$, $f(p(A))$ is said to be recoverable given $p(V)$ in $G^m$ if there is a functional $g$ of $p(V)$, such that $f(p(A)) = g(p(V))$ for every element of $M(G^m)$. In this paper, we will concentrate on recoverability of $p(O \cup M)$, although many other kinds of recoverability problems are also interesting, for instance recovering the causal effect in a causal model with missingness.

We explicitly represent missingness as a causal inference problem because this allows us to rephrase recoverability as identifiability of causal effects. If we were allowed to assign $R_M$ without affecting other variables, we could use proxies $S_M$ to recover the behavior of the underlying missing variables $M$, due to the following result.

**Lemma 1** In a DAG $G$ where $M \neq \emptyset$, for any $p(A) \in \mathcal{M}(G^m(V, M))$, and $R_M \in R_M$, $p(Y) \in \mathcal{M}(G^m(V \cup M, M \cup M) | V \cup M, S_M \cup M, R_M)$, where $Y$ is

$$\{ \{V \cup M \setminus \{R_M\}\}(r, 0_{R_M}) | R \subseteq R_M, r \in X_R\}.$$

**Proof:** $\{V \cup M\}(r, 0_{R_M})$ obeys (2) for $G^m_{R_M}(R_M)$. Since $A \setminus \{R_M\}$ is ancestral in $G^m_{R_M} | \{R_M\}$, $(V \cup M \setminus \{R_M\})(r, 0_{R_M})$ obeys (2) for $G^m_{R_M} | \{R_M\} \setminus \{R_M\}$. Our conclusion follows since $M = S_M(0_{R_M}).$ 

In other words, fixing $R_M$ to 0 gives a new model where $M$ is effectively observed since $M = S_M(0_{R_M})$. This implies that if we were able to fix all of $R_M$, we could recover $p(O \cup M)$.

**Corollary 1** $p(\{O, S_M\}(0_{R_M})) = p(O \cup M)$ for any $G^m$, and any $p(A) \in \mathcal{M}(G^m)$.

This corollary implies that our recoverability problem is solved by expressing a particular interventional distribution
as a function of the manifest in a restricted causal model. We will attack this problem via two standard results for causal models that hold in restricted causal models as well, as shown in [11], propositions 45 and 46.

**Theorem 1** For any \( p(A) \in \mathcal{M}(G^m(V, M)) \), and (\( \forall R \subseteq R_M, r \in X_R \)).

\[
p(A(r)) = \prod_{V \in A} p(V \mid pa_{G^m}(V) \setminus R, r \in X_R). \tag{3}
\]

**Theorem 2** For any \( p(A) \in \mathcal{M}(G^m(V, M)) \), and (\( \forall R \subseteq R_M, r \in X_R \)).

\[
p(\{(V \cup M) \setminus R\}(r) \mid r) = p((V \cup M) \setminus R \mid r). \tag{4}
\]

\( [3] \) is known as the truncated factorization [10], manipulated distribution [21], or the g-formula [12, 4] is known as the consistency property.

We now illustrate how constraints of the missingness model encoded by \( G^m \), as well as \( [3] \) and \( [4] \) lead to recoverability.

### 4.1 EXAMPLES OF RECOVERABILITY

Consider Fig. [1] where \( X, C, W \) may possibly be high-dimensional. In Fig. [1](a), \( X \) is missing according to a mechanism governed by an independent proxy \( R_X \), so

\[
p(X) = p(S_X(0_{R_X})) = p(S_X \mid R_X = 0).
\]

The assumption present in this model which allows us to recover the underlying missing variable, namely \( (S_X(0_{R_X}) \perp \!\!\!\perp R_X) \), is known as **missing completely at random** (MCAR) assumption[4] This assumption is the missingness analogue of ignorability (lack of confounding between the missingness indicator \( R_X \) and the proxy \( S_X(r) \) under assignment \( r \) to \( R_X \)).

In Fig. [1](b), \( X \) is missing according to a mechanism governed by a proxy \( R_X \) which has a (statistical) dependence on \( X \) through \( C \), which is a fully observed variable. In this case,

\[
p(X, C) = p(S_X(0_{R_X}) \mid C)p(C) = p(S_X \mid R_X = 0, C)p(C).
\]

The assumption present in this model which allows us to recover the underlying missing variable, namely \( (S_X(0_{R_X}) \perp \!\!\!\perp R_X \mid C) \) is known as the **missing at random** (MAR) assumption. This assumption is the missingness analogue of **conditional ignorability** (lack of confounding between the indicator \( R_X \) and the proxy \( S_X(r) \) under assignment \( r \) to \( R_X \) given that we conditioned on a set of variables \( C \)).

In Fig. [1](c), it is not the case that

\[
\{S_W(0_{R_W}), S_X(0_{R_X})\} \perp \!\!\!\perp \{R_X, R_W\}.
\]

**Figure 1**: (a) A missingness model satisfying the **missing completely at random** (MCAR) assumption. (b) A missingness model satisfying the **missing at random** (MAR) assumption. (c) A missingness model where missingness is **not at random** (MNAR), but where recoverability is nevertheless possible.

That is, data on \( X, W \) is not missing completely at random (nor at random, since there is no fully observed variable to screen off the dependence of proxies under indicator assignment from indicators.) Nevertheless, despite the fact that data on \( p(X, W) \) is missing not at random (MNAR), we now show that \( p(X, W) \) is recoverable. We will exploit the fact that the missingness model implies

\[
\{S_W(0_{R_W}), R_X(0_{R_X})\} \perp \!\!\!\!\!\perp \{S_X(0_{R_X}), R_W\}. \tag{5}
\]

It is not difficult to show that \( p(R_W, R_X, S_W, S_X) \) is equal to

\[
p(\{S_W, R_X\}(R_W)) \cdot p(S_X(R_X), R_W) = \frac{p(S_W \mid R_X, R_W)p(R_X \mid R_W) \cdot p(S_X \mid R_X, R_W)p(R_W)}{p(S_X \mid R_X, R_W)p(R_W)}
\]

This implies \( p(X, W) = p(X)p(W) \) is equal to

\[
\left( \sum_{R_X} p(\{S_W, R_X\}(0_{R_W})) \right) \cdot \left( \sum_{R_W} p(S_X(0_{R_X}), R_W) \right) = \left( \sum_{R_W} p(S_W \mid 0_{R_W}) \right) \cdot \left( \sum_{R_W} p(S_X \mid 0_{R_X}, R_W)p(R_W) \right)
\]

The key to this example is the joint independence \( [5] \); independences of this type arise in hidden variable DAG models. We give an example later where recoverability is based not on an ordinary independence, but on a generalized independence, or Verma constraint [23][20]. In the following sections, we give a general recursive scheme for solving recoverability problems under MNAR using these types of constraints.

### 4.2 KNOWN RESULTS FOR MISSINGNESS GRAPHS

Recently [7] and [6] have used missingness graphs to derive conditions for recoverability when data is MNAR. In particular, the following characterization appears in [7] (as theorem 2).
Theorem 3 For any \( p(h) \in \mathcal{M}(G^m(V, M)) \), if no elements of \( R_M \) are adjacent in \( G^m(V, M) \), then \( p(O \cup M) \) is recoverable from \( p(O, S_M, 0_{R_M}) \) if and only if \( M \notin pa_g(R_M) \) for any \( M \in M \). Moreover, \( p(O \cup M) \) is equal to

\[
p(O, S_M, 0_{R_M}) = \prod_{R_M \in R_M} p\left(0_{R_M} \mid pa_g(R_M) \setminus M, S_{pa_g(R_M) \cap M}, 0_{R_{pa_g(R_M) \cap M}}\right).
\]

This result can be generalized in three directions. We may consider recoverability of other queries than \( p(O \cup M) \), for instance causal effects or marginal distributions. Finally, we may consider cases where elements of \( R_M \) are adjacent. This case is important because it represents important classes of missingness such as monotonic missingness due to loss to followup. A unit that drops out of a longitudinal study at time \( t \) often remains dropped out at times \( t+1, \ldots \). In our framework, we would code this by requiring that for all \( t' > t \), \( R_{M_t} = 1 \) if \( R_{M_{t-1}} = 1 \), where \( M_t \) is unit’s status at time \( t \). But this coding is only possible if indicators are allowed to be adjacent in the graph. In addition, allowing indicators to be adjacent allows us to model non-monotone missing data, where a unit may be missing at a particular time \( t \), but then becomes observed at a later time \( t + k \).

In this paper, we consider the problem of recovering \( p(O \cup M) \) given that every missing variable has an indicator and a proxy (e.g. no completely hidden variables), and that indicators \( R_M \) are allowed to be adjacent. We give a recoverability algorithm that generalizes earlier work in this setting.

5 A GENERAL RECOVERABILITY ALGORITHM

The algorithm, which we call MID, works as follows. It tries, for every \( R_M \in R_M \), to recover

\[
p(0_{R_M} \mid pa_gm(R_M) \setminus R_M, 0_{pa_gm(R_M) \cap R_M})
\]

via a subroutine DIR. If every such conditional distribution is recovered, MID recovers \( p(O \cup M) \) via [5]. Otherwise MID fails.

The subroutine DIR (so named for its resemblance to the way the ID algorithm operates when identifying controlled direct effects) has three cases. The first case, which is sufficient for obtaining the soundness part of Theorem 3, attempts to check if indicators for missing parents of \( R_M \) are non-parental non-descendants of \( R_M \), in which case recoverability of the conditional distribution for \( R_M \) is immediate.

Otherwise, DIR uses the other two cases to isolate \( R_M \) and its parents into smaller subproblems based on a particular type of ancestral set \( A^1 \), or the clan \( D^1 \) of \( R_M \). DIR is recursive, which means the input must also keep track of a set \( W \) representing variables the clan subproblem ends up depending on.

The situation is somewhat analogous to the way in which the ID algorithm attempts to identify controlled direct effects \( p(Y \mid do(Y_{pa_g(Y)})) = p(Y(Y_{pa_g(Y)})) \), with three major differences. First, we are attempting identification in a setting where some variables start off being treated as hidden, but in the course of the recursion of DIR become observed due to fixing indicators to 0. In ID variables are always either hidden or observed and do not change status. Second, since we are only allowed to intervene on indicators, we are attempting to identify

\[
p(R_M(0_{R_M} \cap pa_g(R_M)) \mid \{pa_g(R_M) \setminus R_M\}(0_{R_M} \cap pa_g(R_M)))
\]

Finally, there is not necessarily a fully interventional interpretation for the intermediate objects \( p_W(\cdot) \) that arise during the execution of DIR, since \( W \) may contain elements outside \( R_M \). This is a necessary consequence of our insistence on not imposing a causal model on \( p(M \cup O) \).

5.1 SOUNDNESS

MID and its subroutine DIR appear below as algorithm 1. In this section, we prove that MID is sound.

Corollary [1] implies that if we were able to express \( p(\{O, S_M\}(0_{R_M})) \) as a function of the manifest distribution, we would solve the recoverability problem for \( p(O \cup M) \). If we happen to know

\[
p(0_{R_M} \mid pa_gm(R_M) \setminus R_M, 0_{R_M} \cap pa_gm(R_M))
\]

for every \( R_M \in R_M \) as a function of the manifest, this would suffice due to the following result.

Lemma 2 Under \( \mathcal{M}(G^m) \), if for every \( R_M \in R_M \),

\[
p(0_{R_M} \mid pa_gm(R_M) \setminus R_M, 0_{pa_gm(R_M) \cap R_M})
\]

is a functional \( f_{R_M}(\cdot) \) of \( p(O, S_M, 0_{R_M}) \), then

\[
p(\{O, S_M\}(0_{R_M})) = \prod_{R_M \in R_M} f_{R_M}(p(O, S_M, 0_{R_M})).
\]

Proof: \( p(\{O, S_M\}(0_{R_M})) = \sum_{A} p(\{A \mid R_M\}(0_{R_M})) \).

\[
p(\{A \mid R_M\}(0_{R_M})) = \prod_{R_M \in R_M} f_{R_M}(p(O, S_M, 0_{R_M}))
\]

is implied by [3]. But no denominator is a function of \( M \), so we can apply the sum to the numerator first.  2 Finding functionals \( f_{R_M}(\cdot) \) for every \( R_M \) in order to apply Lemma 2 is the job of the subroutine DIR.
Algorithm 1 $G^m(V, M)$ a missingness graph, $p(V)$ a manifest distribution from $p(\mathcal{A}) \in \mathcal{M}(G^m(V, M))$, $p\omega(V)$ a family of manifest distributions from elements of $p(\mathcal{A}) \in \mathcal{M}(G^m(V, M))$, $< $ a topological order on $G^m$.

procedure MID($G^m(V, M), p(V)$)
for each $R_M \in R_M$,

$p(0_{R_M} | \\{pa_G^m(R_M) \setminus R_M, 0_{pa_G^m(R_M) \cap R_M}\})$
$= \text{DIR}(G^m, p, R_M)$

if $(\exists R_M \in R_M)$, s.t. $\text{DIR}(G^m, p, R_M) = \emptyset$,

return “cannot recover.”

else return

$p(0, S_M, 0_{R_M})$
$\prod_{R_M \in R_M} p(0_{R_M} | \\{pa_G^m(R_M) \setminus R_M, 0_{pa_G^m(R_M) \cap R_M}\})$.

end procedure

procedure DIR($G^m(V, M | W), p_W(V), R_M$)
if $R_M \cap (\{pa_G^m(R_M) \cap W\}) \subseteq \text{ndp}_G^m(R_M)$, return

$p_W(0_{R_M} | \\{pa_G^m(R_M) \setminus (M \cup W \cup R_M), 0_{R_M \cap (pa_G^m(R_M) \cap W)}\}$.

else $A^I \leftarrow \{R_M\}$,

while $(\text{ang}_{G^m}(A^I) \cup S_{\text{ang}_{G^m}(A^I)} \cap M \not\subseteq A^I)$ do

$A^I \leftarrow \text{ang}_{G^m}(A^I) \cup S_{\text{ang}_{G^m}(A^I)} \cap M$.

if $A^I \subseteq A$,

return $\text{DIR}(G^m_{A^I}, p_W \cap A^I(V \cap A^I), R_M)$.

if $D \subseteq \emptyset$,

$Z^I \leftarrow \text{pa}_{G^m_{\emptyset}}(V)(D) \cap R_M$

$Y^I \leftarrow \text{pa}_{G^m_{\emptyset}}(V)(D) \setminus R_M$

$M^I_D \leftarrow \{M \in (M \cap D^I) | R_M \in Z^I\}$

$M^I_{D^I} \leftarrow (M \cap D^I) \setminus M^I_D$

$V^I \leftarrow D \cup M^I_{D^I}$

$G^m \leftarrow G^m_{D^I}(V^I, M^I_{D^I})$

$p_W(V^I(D)) \leftarrow \prod_{V \in D} p_W(V | \\{\text{prep}_{G^m}(V^I \cap Z^I), 0_{\text{prep}_{G^m}(V^I \cap Z^I)}\}$.

return $\text{DIR}(G^m, p_{V^I}(D), R_M)$

end if

return $\emptyset$.

end procedure

Soundness of DIR

The subroutine DIR invoked by MID aims to recover $f_{R_M}(p) = p(0_{R_M} | \\{pa_G^m(R_M) \setminus R_M, 0_{pa_G^m(R_M) \cap R_M}\})$ by recursively attempting to restrict $R_M$ and $pa_G^m(R_M)$ to either an appropriate ancestral subset containing these vertices, or an appropriate clan of $G^m$, and, in the base case, exploiting the independence structure, and properties of the subproblem that is left.

To prove the soundness of DIR, we must establish, by induction on algorithm structure, certain results about the subproblems it considers. We will represent subproblems as a pair consisting of a CDAG $G^m$ that is a subgraph of the original graph $G^m$, and a conditional fragment of the missingness model which can be viewed as a set of all interventional distributions relevant to the subproblem, which also possibly depend on variables $W$ from larger subproblems.

Given an element $p(\mathcal{A}) \in \mathcal{M}(G^m(A), B \subseteq A$, and $W \subseteq A \setminus B$, a conditional fragment of $p(\mathcal{A})$ with respect to $B$ and $W$, denoted by $\mathcal{F}_{W, B}$, is a mapping from elements $w$ in $X_W$ to

$\mathcal{F}_{W, B} \equiv \{p_w(B(r)w) | R \subseteq R_M \cap B, r \in X_R\}$.

Note that we cannot view $p_w(B(r)w)$ as a joint response of $B$ to an intervention setting $R \cup W$ to $r \cup w$, because $W$ may contain elements outside $R$ that we are not allowed to intervene on.

For each call to DIR, we want to show that all interventional distributions in the input fragment are Markov with respect to the appropriately modified input graph, that we have enough information in the subproblem to possibly obtain $f_{R_M}(p)$, and that the manifest distribution of the fragment for the current (inner) call can be obtained from the manifest distribution of the fragment for the previous (outer) call.

Definition 2 $\mathcal{F}_{W, B}$ is causal Markov relative to a CDAG $G^m(B | W)$ if $(\forall w \in X_W, p_w(B(r)w) \in \mathcal{F}_{W, B})$, $p_w(B(r)w)$ is Markov relative to $G^m(B | W)_R$.

This definition is how we will relate fragments and corresponding subgraphs, and the following two results establish this relationship for the two recursive cases relevant for DIR.

Lemma 3 For $\mathcal{F}_{W, A}$ causal Markov relative to $G^m(A | W)$, let $D \subseteq D(G^m(A, W))$, $D^I \equiv \text{claq}_{G^m}(D)$, $W^I \equiv \text{pa}_{G^m}(D) \cup W$, $\mathcal{F}_{W^I, D^I} \equiv \{p\omega_w(D^I(r)w) | r \in X_R, R \subseteq R_M \cap D\}$ is causal Markov relative to $G^m_{pa_G^m(D^I)}(D^I | W^I)$, where for any $w \in X_W^I$.

$\mathcal{F}_{W^I, D^I} \equiv \{p\omega_w(D^I(r)w) | r \in X_R, R \subseteq R_M \cap D\}$ is causal Markov relative to $G^m_{pa_G^m(D^I)}(D^I | W^I)$, where for any $w \in X_W$, $\mathcal{F}_{W^I, D^I} \equiv \{p\omega_w(D^I(r)w) | r \in X_R, R \subseteq R_M \cap D\}$ is causal Markov relative to $G^m_{pa_G^m(D^I)}(D^I | W^I)$, where for any $w \in X_W$. $\mathcal{F}_{W^I, D^I} \equiv \{p\omega_w(D^I(r)w) | r \in X_R, R \subseteq R_M \cap D\}$ is causal Markov relative to $G^m_{pa_G^m(D^I)}(D^I | W^I)$, where for any $w \in X_W$. $\mathcal{F}_{W^I, D^I} \equiv \{p\omega_w(D^I(r)w) | r \in X_R, R \subseteq R_M \cap D\}$ is causal Markov relative to $G^m_{pa_G^m(D^I)}(D^I | W^I)$, where for any $w \in X_W$.
**Proof:** For any CDAG $\mathcal{G}(O, M | W)$, $\text{fac}_\mathcal{G}([\text{cl}(D)])$ is equal to $\text{cl}(D) \cup \text{pa}_\mathcal{G}(\text{cl}(D))$, for any $D \in \mathcal{D}(O)$. The proof is now immediate. Elements $\tilde{p}_w(D_1(r_{w1})$ of each $\mathcal{F}_{w_1, D_1}$ are Markov relative to $(G_{\text{factor}}(D_1))_{\mathbb{R}}$ by construction. The definition of $\tilde{p}_w(D_1(r_{w1})$ implies it is the same object for any $w$ consistent with $w_1$. \hfill $\Box$

**Lemma 4** For $\mathcal{F}_{W, A}$ causal Markov relative to $G_{\mathcal{M}}(A | W)$, let $V_1^\dagger \subseteq A \cup W$ be ancestral, $W_1^\dagger = W \cap V_1^\dagger$, $A_1^\dagger = A \cap V_1^\dagger$. Then for any $w_1^\dagger \in \mathcal{X}_{W_1^\dagger}$, $\mathcal{F}_{w_1^\dagger, D_1^\dagger} \equiv \{ \tilde{p}_w(A_1^\dagger(r_{w1}) | r \in \mathcal{R}_R, R \subseteq R_M \cap A_1^\dagger \}$ is causal Markov relative to $G_{A_1^\dagger}(A_1^\dagger | W_1^\dagger)$, for every $w$ consistent with $w_1^\dagger$.

**Proof:** Immediate. Elements $\tilde{p}_w(A_1^\dagger(r_{w1})$ of each $\mathcal{F}_{w_1^\dagger, A_1^\dagger}$ are Markov relative to $(G_{A_1^\dagger})_{\mathbb{R}}$ by construction. The definition of $\tilde{p}_w(A_1^\dagger(r_{w1})$ implies it is the same object for any $w$ consistent with $w_1^\dagger$. \hfill $\Box$

The next two results re-express $p(R_M | \text{pa}_g(R_M))$ from a function of the larger fragment of the outer recursive call to a function of the smaller fragment of the inner call.

**Lemma 5** Assume $\mathcal{F}_{W, A}$ is causal Markov relative to $G_{\mathcal{M}}(A | W)$, and $\mathcal{F}_{W_1, D_1}$ is defined as in Lemma 3. Then for any $R_M \in D_1$, $p_{W_1}(R_M | \text{pa}_g(R_M))$ is equal to $p_{W_1}(R_M | \text{pa}_g_{(D_1)}(R_M))$.

**Proof:** Since $R_M \in D_1$, this follows by Lemma 3. That is, $p_{W_1}(R_M | \text{pa}_g_{(D_1)}(R_M))$ is equal to $p_{W_1}(R_M | \text{pa}_g(R_M) \setminus W_1^\dagger)$, which is equal to $p_{W_1}(R_M | \text{pa}_g(R_M) \setminus W_1^\dagger)$. \hfill $\Box$

**Lemma 6** Assume $\mathcal{F}_{W, A}$ is causal Markov relative to $G_{\mathcal{M}}(A | W)$, and $\mathcal{F}_{W_1, A_1}$ is defined as in Lemma 3. Then for any $R_M \in A_1$, $p_{W_1}(R_M | \text{pa}_g(R_M))$ is equal to $p_{W_1}(R_M | \text{pa}_g(A_1))$.

**Proof:** Since $R_M \in A_1$, this follows by Lemma 3. That is, $p_{W_1}(R_M | \text{pa}_g(A_1))$ is equal to $p_{W_1}(R_M | \text{pa}_g(R_M))$. \hfill $\Box$

The next two results express the analogue of the manifest distribution of the smaller fragment as a function of the manifest distribution of the larger fragment. We assume $M_{D_1^\dagger}, M_{D_1^\dagger}, V_1^\dagger, Z_1^\dagger, Y_1^\dagger$, and $\mathcal{G}_{\mathcal{M}}$ are defined as in the distinct case of DIR. Let $W_1^\dagger = Y_1^\dagger \cup Z_1^\dagger$, and $O_1^\dagger = D \cap O$.

**Lemma 7** Assume $\mathcal{F}_{W, A}$ is causal Markov relative to $G_{\mathcal{M}}(A | W)$, and $\mathcal{F}_{W_1, D_1}$ is defined as in Lemma 3.

Then the marginal $p_{Y_1^\dagger, O_1^\dagger}(O_1^\dagger, M_{D_1^\dagger}, S_{M_{D_1^\dagger}}, R_{M_{D_1^\dagger}})$ of $p_{W_1}(D_1^\dagger) \in \mathcal{F}_{W_1, D_1}$ is equal to

$$\prod_{V \in V_1^\dagger} p_W(V | \text{pa}_g(V) \setminus (V \cup Z_1^\dagger)) \cdot 0_{\text{pa}_g(V) \cup (V \cap Z_1^\dagger)}.$$

**Proof:** Fix $w$ and $w_1^\dagger$ consistent with $w$, such that $w_1^\dagger = 0$. We get the following set of equalities, where the first is by assumption on missingness models, the second by $\mathcal{F}$, the third by definition, the fourth by Lemma 3 and the last by standard results on district factorization of hidden variable DAG models found in [22]. We can range over all possible $w_1^\dagger$ in $V_1^\dagger$, the last expression reduces to $\prod_{V \in V_1^\dagger} p_W(V | \text{pa}_g(V) \setminus (V \cup Z_1^\dagger)) \cdot 0_{\text{pa}_g(V) \cup (V \cap Z_1^\dagger)}$.

Then $p_{W_1}(O_1^\dagger, M_{D_1^\dagger}, S_{M_{D_1^\dagger}}, R_{M_{D_1^\dagger}}) = p_{W_1}(O_1^\dagger, S_{M_{D_1^\dagger}}, R_{M_{D_1^\dagger}})$.

$$p_{W_1}(O_1^\dagger, S_{M_{D_1^\dagger}}, R_{M_{D_1^\dagger}}) = \sum_{M_{D_1^\dagger}} p_{W_1}(O_1^\dagger, S_{M_{D_1^\dagger}}, R_{M_{D_1^\dagger}}) \cdot p_{W_1}(V | w_1^\dagger = 0, \text{pa}_g(V) \setminus (W_1^\dagger \setminus W_1^\dagger) \cup \text{pa}_g(V) \setminus W_1^\dagger) \cdot 0_{\text{pa}_g(V) \cup (V \cap W_1^\dagger)}.$$
Proof: We get the following set of equalities, where the first follows by assumption, and the fact that \( p_w(A) \) is Markov relative to \( G^m \), the second is by the properties of the missingness model, and the third is by (4):

\[
p_w\left(0_{RM}\mid p_{\bar{G}^m}(R_M) \setminus (W \cup R_M)\right) = p_{\bar{G}^m}(R_M) \setminus (M \cup W \cup R_M)
\]

\[
p_w\left(0_{RM}\mid M \setminus p_{\bar{G}^m}(R_M) \setminus (W \cup R_M)\right) = p_{\bar{G}^m}(R_M) \setminus (M \cup W \cup R_M),
\]

\[
p_w\left(0_{RM}\mid M \setminus p_{\bar{G}^m}(R_M) \setminus (W \cup R_M)\right) = p_{\bar{G}^m}(R_M) \setminus (M \cup W \cup R_M)
\]

We now show the main result of this paper.

\[
\rho_{\bar{G}^m}(R_M) \setminus (M \cup W \cup R_M) = p_{\bar{G}^m}(R_M) \setminus (M \cup W \cup R_M)
\]

Define an automorphism from vertex sets in \( G^m \), \( \rho_{\bar{G}^m}(B) \), as \( an_{G^m}(\{R_M\} \cup B) \). Let \( A^I \) be the fixed point of \( \rho_{\bar{G}^m} \) with the starting input of the empty set.

Lemma 10 \( A^I \) is an ancestral set in \( G^m \).

Proof: A simple proof by contradiction follows by definition of \( \rho_{\bar{G}^m} \).

We now show the main result of this paper.

Theorem 4 \( \text{MID} \) is sound.

Proof: Assuming \( \text{DIR} \) returns the answer for every \( R_M \in \text{RM} \), Corollary [1] and Lemma [2] ensure that \( \text{MID} \) recovers \( p(M \cup O) \) from \( p(V) \).

The soundness of \( \text{DIR} \) follows by induction on the recursive call structure. The inductive hypothesis is that the input conditional fragment \( F_{\bar{W},A} \) is causal Markov relative to the appropriate graph derived from the input graph \( G^m \), that the input manifest \( \tilde{p}_W(V) \) is the function of the original manifest \( p(V) \), and that \( \tilde{p}_W(R_M \mid p_{\bar{G}^m}(R_M) \setminus W) = p(R_M \mid p_{\bar{G}^m}(R_M)) \).

The base case trivially holds for the original inputs to \( \text{DIR} \). If the inductive hypothesis is true, and \( \text{DIR} \) returns after the first conditional, soundness follows by Lemma [9].

If \( \text{DIR} \) returns after the second conditional, then Lemma [10] ensures the constructed set \( A^I \) is ancestral, and the induction for the following recursive call is maintained via Lemmas [4][8] and [6].

If \( \text{DIR} \) returns after the third conditional, Lemma [5] ensures \( F_{\bar{W},D}^I \) is causal Markov relative to \( G_{fa_{G^m}}(D^I) \) for all values of \( \bar{W} \), including those that set \( Z^I \) to 0. Lemma [3] and the inductive hypothesis ensures \( p_{\bar{W}^I}(R_M \mid p_{fa_{G^m}(D^I)}) \) for all values of \( \bar{W^I} \). Finally, Lemma [7] ensures the manifest for the recursive call is a function of the input manifest. In fact, because we set \( Z^I \) to 0, properties of missingness models ensure we can treat \( M^{G^m} \) as observed in subsequent recursive calls, which means we no longer need to consider \( S_{M^{G^m}} \).

Since induction follows for all cases, so does our conclusion. }

\section{6 A COMPLEX RECOVERABLE EXAMPLE}

We now work through an example where all cases of \( \text{MID} \) and \( \text{DIR} \) are necessary. Consider the graph shown in Fig. 2(a). Here \( C \) and \( D \) are shown in green to indicate that they are fully observed. This is a more complex version of the example in Fig. 1(c). Unlike that case, here, there are no conditional independences that hold between proxies and indicators. However, if we were to divide by \( p(D \mid C) \) and sum out \( C \), in the resulting distribution \( p_{D}(S_A,S_B,R_A,R_B,A,B) \), for any fixed value \( d \) of \( D \), we would have

\[
\{(S_A(0_{RA}), R_B) \perp \perp \{S_B(0_{RB}), R_A(0_{RA})\}\}
\]

This is a type of Verma constraint [23] or generalized independence constraint [20].

Our goal is to recover \( p(A, B, C, D) \) given the missingness model corresponding to this graph, and in particular the above constraint. We must recover \( p(0_{RB} \mid A, D) \) and \( p(0_{RA} \mid 0_{RB}, B, D) \) from \( p(R_A, R_B, S_A, S_B, C, D) \).

In either case, we note that \( D \) is not an element of \( \text{cl}_{G^m}^c(R_A) = \text{cl}_{G^m}(R_B) \), which implies we can use the clan case of \( \text{DIR} \) and consider a subproblem shown in Fig. 2(b), with the corresponding manifest \( \tilde{p}_D(R_A,R_B,S_A,S_B,C) = p(S_A,S_B,R_A,R_B \mid D, C)p(C) \). In the new subproblem (for either \( R_A \) or \( R_B \)), \( C \) is not a part of the ancestral set \( A^I \) constructed by \( \text{DIR} \) in the ancestral case, so we consider a new subproblem shown in Fig. 2(c), with the corresponding manifest \( \tilde{p}_D(R_A,R_B,S_A,S_B) = \sum_c \tilde{p}_D(S_A,S_B,R_A,R_B \mid D, C)p(C) \). This new subproblem now resembles the example in Fig. 1(c), and is solved similarly. In particular, we recover \( p(0_{RB} \mid D, A) \) as

\[
\frac{\sum_{R_B} \tilde{p}_D(S_A \mid 0_{RA}, 0_{RB})\tilde{p}_D(0_{RB})}{\sum_{R_B} \tilde{p}_D(S_A \mid 0_{RA}, 0_{RB})\tilde{p}_D(0_{RB})}
\]

and \( p(0_{RA} \mid 0_{RB}, B, D) \) as \( \tilde{p}_D(0_{RA} \mid 0_{RB}, S_B) \). We then obtain \( p(A, B, C, D) \) by dividing the manifest distribution for observed cases \( p(0_{RA}, 0_{RB}, S_A, S_B, C, D) \) by the above two probabilities.
Figure 2: (a) An example where recoverability is possible via MID. (b),(c) Graphs corresponding to subproblems considered by MID in recovering $p(A, B, C, D)$.

7 NONRECOVERABILITY

The generality of MID naturally raises the question of whether it is complete, that is whether whenever it outputs “cannot recover” then it is possible to construct two elements of the missingness model that agree on the manifest but disagree on the underlying joint distribution. We leave this difficult question aside in this paper in the interests of space, but note that an approach similar to one used to show completeness for causal effects identification [18] seems promising. That is, use MID as a guide for constructing a “zoo” of structures where recoverability does not seem to be possible, and then construct a general method for showing non-recoverability for this “zoo.”

Some results on non-recoverability do exist. For example, it can be shown that $p(A)$ is not recoverable in the missingness model with the graph in Fig. 3(a) [2], and similarly that $p(A, B)$ is not recoverable in the missingness model with the graph in Fig. 3(b). Characterization of non-recoverability is an open problem.

8 DISCUSSION AND CONCLUSIONS

We have represented missing data as a type of a restricted causal inference problem. Using the machinery of graphical causal models, we have given a general algorithm for recoverability of a joint distribution in MNAR settings. Though we do not require this, our formalism allows the joint distribution we recover to come from a statistical, rather than a causal model – all causal assumptions may be restricted to the missingness model governing the behavior of proxies of missing variables under interventions on indicators. We show that the MCAR, MAR, MNAR taxonomy is not sufficiently granular to classify cases where recoverability is possible. In particular, there are MNAR examples where constraints akin to Verma constraints permit recoverability.

Aside from the algorithm, our formalism allows us to seamlessly integrate issues of identification of causal effects, and recoverability. For instance, it is known that in the graph shown in Fig. 3(c) (where we treat $\leftrightarrow$ edges as indicating the presence of an unobserved parent), $p(Y)$ is not recoverable. However, if the graph on $C, A, Y$ represents a causal model, we can show that $p(Y(a))$ is recoverable. In particular

$$p(Y(a)) = p(S_Y(a, 0_{R_Y} | a, c)p(c)$$

A similar observation appears in [6], example 3.

By explicitly representing missingness via an intervenable indicator, and a proxy as a response to this intervention, our formalism allows us to reason explicitly about the interpretation of censoring by death using the existing language of interventions. That is if $S_X$ is observed patient history, and $1_{RX}$ implies it is missing due to the patient dying, then we may either disallow considering $S_X(0_{RX})$ (e.g. “resurrecting the patient”) for that patient, allow $S_X(0_{RX})$, but treat it as making statements about exchangeable but different patients who happened to be alive that transfer over to the dead patient in a hypothetical alternative history where the patient never died, and so on.

Note that if we assume a known relationship $p(S_M(r_{RM}) | M)$ between $M$ and $S_M(r_{RM})$ other than direct equality, we can use the approach in this paper to address certain coarsening [3] and measurement error settings. We do not consider these extensions explicitly here for space reasons, but they are straightforward.

Acknowledgements

This research was supported in parts by grants from NIH R01 AI104459-01A1, NSF #IIS-1302448 and ONR #N00014-10-1-0933 and #N00014-13-1-0153.

References


