m-Transportability: Transportability of a Causal Effect from Multiple Environments

Sanghack Lee and **Vasant Honavar**

Artificial Intelligence Research Laboratory
Department of Computer Science
Iowa State University
Ames, IA 50011
{shlee, honavar} at iastate.edu

Abstract

We study *m-transportability*, a generalization of *transportability*, which offers a license to use causal information elicited from experiments and observations in $m \geq 1$ *source* environments to estimate a causal effect in a given *target* environment. We provide a novel characterization of *m*-transportability that directly exploits the completeness of *do*-calculus to obtain the necessary and sufficient conditions for *m*-transportability. We provide an algorithm for deciding *m*-transportability that determines whether a causal relation is *m*-transportable; and if it is, produces a *transport formula*, that is, a recipe for estimating the desired causal effect by combining experimental information from m source environments with observational information from the target environment.

1 Introduction

Eliciting causal effects from observations and experiments is a hallmark of intelligence. Pearl (1995; 2000) introduced causal diagrams, a formal representation for combining data with causal information, and *do*-calculus (Pearl 1995; 2000; 2012), a set of three rules that constitutes a sound (Pearl 1995) and complete (Shpitser and Pearl 2006b; Huang and Valtorta 2006) inferential machinery for causal inference. The resulting framework has been used to decide the identifiability of causal effects (Tian and Pearl 2002; Shpitser and Pearl 2006a; Tian 2004), i.e., determine whether a given set of causal assumptions is sufficient for determining causal effects from observations.

The practical need to transfer causal effects elicited in one domain (setting, environment, population) e.g., a controlled laboratory setting, to a different setting presents us with the problem of transporting causal information from a source environment to a possibly different target environment. For example, one might want to know if causal relation between teaching strategies and student learning obtained by through a randomized trial in a public school in Chicago can be transported to a public school in Minneapolis that has an admittedly different population of students. Pearl and Bareinboim (2011) introduced *selection diagrams* which provide a formal representation for expressing knowledge about differences and commonalities between the source and target environments. They used the selection diagrams to provide a formal definition of *transportability* of causal effects elicited from experimental studies in a *source* environment to a *target* environment in which only an observational study is possible. Bareinboim and Pearl (2012b) described procedures to (i) decide transportability between a given source and a target environment, and (ii) when the answer is affirmative, determining the set of experimental and observational studies that must be carried out to license the desired transport. Against this background, we consider a setting where we have access to experiments and observations from $m \geq 1$ source environments and a target environment; and causal information is not transportable between each individual source environment and the target environment. In such a setting, we examine the conditions under which it is possible to transfer causal information learned collectively from the m source environments to the target environment.

The main contributions of this paper are as follows: (i) We study *m*-transportability, a generalization of transportability (Pearl and Bareinboim 2011), which offers a license to transfer causal information learned collectively from experiments and observations from m source environments to a given target environment where only observational information can be obtained. (ii) We introduce a novel technique that directly exploits the completeness of *do*-calculus (Shpitser and Pearl 2006b; Huang and Valtorta 2006) to obtain the necessary and sufficient conditions for *m*-transportability. Our results directly apply to transportability (*m*-transportability where $m = 1$) as well as identifiability which is a special case of transportability. (iii) We provide, based on a modification from Bareinboim and Pearl's algorithm for deciding transportability (Bareinboim and Pearl 2012b), a sound and complete algorithm for deciding *m*-transportability that determines whether the causal effect is *m*-transportable; and if it is, produces a transport formula, that is, a recipe for estimating the causal effect by combining the experimental information from the m source environments with observational information from the target environment.

This work was carried out independently of Bareinboim and Pearl (2013)[1] which introduces μ-transportability, which, despite some differences in the formulation, upon

---

[1]Our work was completed during 2012 and submitted to AAAI 2013 on January 22, 2013. We became aware of the work of Bareinboim and Pearl (2013) when it appeared as a Technical Re-

closer examination, turns out to be essentially identical to m-transportability. Not surprisingly, the algorithm for deciding μ-transportability is essentially identical to our algorithm for deciding m-transportability: Both algorithms are simple extensions of **sID**, the algorithm for deciding transportability (a special case of m-transportability where m=1) introduced by Bareinboim and Pearl (2012b); and both algorithms rely on the same graphical criterion to decide whether a causal relation is μ-transportable (m-transportable). However, the work described in this paper differs from that of Bareinboim and Pearl (2013) with respect to the technique used to derive the graphical criterion for lack of m-transportability (μ-transportability): Bareinboim and Pearl (2013) construct a *certificate* that serves as counter-example that establishes lack of μ-transportability. In contrast, we demonstrate the nonexistence of a transport formula (and hence lack of m-transportability) by directly exploiting the completeness of do-calculus (Shpitser and Pearl 2006b; Huang and Valtorta 2006) and standard probability manipulations.

2 Preliminaries

A *causal diagram* (Pearl 2000) G is a semi-Markovian graph or an acyclic directed mixed graph, (i.e., a graph with directed as well as bidirected edges that does not have directed cycles) which encodes a set of *causal assumptions*. We denote by $G\left[\mathbf{Y}\right]$, a subgraph of G containing nodes in $\mathbf{Y}$ and all arrows between the corresponding nodes in G. Following (Pearl 2000), we denote by $G_{\overline{\mathbf{X}}}$, the edge subgraph of G where all incoming arrows into nodes in $\mathbf{X}$ are deleted; by $G_{\underline{\mathbf{Y}}}$, the edge subgraph of G where all outgoing arrows from nodes in $\mathbf{Y}$ are deleted; and by $G_{\overline{\mathbf{X}}\underline{\mathbf{Y}}}$, the edge subgraph of G where all incoming arrows into nodes in $\mathbf{X}$ and all outgoing arrows from nodes in $\mathbf{Y}$ are deleted.

A *causal model* (Pearl 2000) is a tuple $\langle \mathbf{U}, \mathbf{V}, F \rangle$ where $\mathbf{U}$ is a set of background or hidden variables that cannot be observed or experimented on but which can influence the rest of the model; $\mathbf{V}$ is a set of observed variables $\{V_1, \ldots V_n\}$ that are determined by variables in the model, i.e., variables in $\mathbf{U} \cup \mathbf{V}$; F is a set of deterministic functions $\{f_1, \ldots, f_n\}$ where each f_i specifies the value of the observed variable V_i given the values of observable parents of V_i and the values of hidden causes of V_i. A *probabilistic causal model* (Pearl 2000) (PCM) is a tuple $M = \langle \mathbf{U}, \mathbf{V}, F, P\left(\mathbf{U}\right) \rangle$ where $P\left(\mathbf{U}\right)$ is a joint distribution over $\mathbf{U}$.

Intervention (Pearl 2000) on a set of variables $\mathbf{X} \subseteq \mathbf{V}$ of PCM $M = \langle \mathbf{U}, \mathbf{V}, F, P\left(\mathbf{U}\right) \rangle$ involves setting to $\mathbf{X} = \mathbf{x}$ and is denoted by *do*-operation $do\left(\mathbf{X} = \mathbf{x}\right)$ or simply $do\left(\mathbf{x}\right)$. A causal effect of $\mathbf{X}$ on a disjoint set of variables $\mathbf{Y} \subseteq \mathbf{V}$ is written as $P\left(\mathbf{y} | do\left(\mathbf{x}\right)\right)$ or simply $P_{\mathbf{x}}\left(\mathbf{y}\right)$. Intervention on a set of variables $\mathbf{X} \subseteq \mathbf{V}$ creates a *submodel* (Pearl 2000) $M_{\mathbf{x}}$ of M defined as follows: $M_{\mathbf{x}} = \langle \mathbf{U}, \mathbf{V}, F_{\mathbf{x}}, P\left(\mathbf{U}\right) \rangle$ where $F_{\mathbf{x}}$ is obtained by taking a set of distinct copies of functions in F and replacing the functions that determine the value of

variables in $\mathbf{X}$ by constant functions setting the variables to values $\mathbf{x}$. It is easy to see that a causal diagram G that encodes the causal assumptions of model M is modified to $G_{\overline{\mathbf{X}}}$ by intervention on $\mathbf{X}$. The causal effect of $\mathbf{X}$ on a disjoint set of variables $\mathbf{Y} \subseteq \mathbf{V}$ is said to be *identifiable* from P in G if $P_{\mathbf{x}}\left(\mathbf{y}\right)$ can be computed uniquely from the joint distribution $P(\mathbf{V})$ of the observed variables in any PCM that induces G (Pearl 2000).

Do-calculus (Pearl 1995) offers a sound and *complete* (Shpitser and Pearl 2006b; Huang and Valtorta 2006) inferential machinery for deciding *identifiability* (Tian and Pearl 2002; Shpitser and Pearl 2006a; Tian 2004) in the sense that, if a causal effect is *identifiable*, there exists a sequence of applications of the rules of *do*-calculus that transforms the causal effect formula into one that includes only observational quantities. Let G be a causal diagram and P be a distribution on G. Let $\mathbf{W}$, $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be disjoint sets of variables in G. Then, the three rules of *do*-calculus are (Pearl 1995):

- (Rule 1) Insertion/deletion of observations:
  $P_{\mathbf{x}}\left(\mathbf{y} \mid \mathbf{z}, \mathbf{w}\right) = P_{\mathbf{x}}\left(\mathbf{y} \mid \mathbf{w}\right)$ if $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{X}, \mathbf{W})_{G_{\overline{\mathbf{X}}}}$

- (Rule 2) Intervention/observation exchange:
  $P_{\mathbf{x},\mathbf{z}}\left(\mathbf{y} \mid \mathbf{w}\right) = P_{\mathbf{x}}\left(\mathbf{y} \mid \mathbf{z}, \mathbf{w}\right)$ if $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{X}, \mathbf{W})_{G_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}}$

- (Rule 3) Insertion/deletion of interventions:
  $P_{\mathbf{x},\mathbf{z}}\left(\mathbf{y} \mid \mathbf{w}\right) = P_{\mathbf{x}}\left(\mathbf{y} \mid \mathbf{w}\right)$ if $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{X}, \mathbf{W})_{G_{\overline{\mathbf{X}}, \overline{\mathbf{Z}(\mathbf{W})}}}$

where $\mathbf{Z}\left(\mathbf{W}\right)$ represents $\mathbf{Z} \setminus An\left(\mathbf{W}\right)_{G_{\overline{\mathbf{X}}}}$.[2]

3 *m*-Transportability

We proceed to formalize the notion of m-transportability, which offers a license to transfer causal information learned *collectively* from experiments and observations from m source environments to a target environment where only observational information can be obtained. The definitions of *selection diagrams* and m-transportability and the m-transportability theorem are direct adaptations of their counterparts introduced by Bareinboim and Pearl (2012b) for transportability (1-transportability). Let J denote an index set $\{1, \ldots, m\}$ for m source environments:

Definition 1 (Selection Diagram (adapted from (Bareinboim and Pearl 2012b))). Let $\mathbf{M} = \left\{M^j\right\}_{j \in J}$ and M^* be a set of causal models relative to domains $\mathbf{\Pi} = \left\{\Pi^j\right\}_{j=J}$ and Π^*, sharing a causal diagram G. The set $\mathbf{M} \cup \{M^*\}$ is said to induce a *selection diagram* D if D is constructed as follows:

1. Every edge in G is also an edge in D
2. D contains an extra edge $S_i \rightarrow V_i$ if there might exist a discrepancy $f_i^j \neq f_i^*$ or $P^j\left(U^i\right) \neq P^*\left(U^i\right)$ between M^* and some model $M^j \in \mathbf{M}$.

We denote by $\mathbf{S}$, the set of *selection variables*. We associate with each domain Π^j, a set of *domain-specific selection variables* $\mathbf{S}^j \subseteq \mathbf{S}$ where: $\mathbf{S}^j = \{S_i \mid f_i^j \neq$
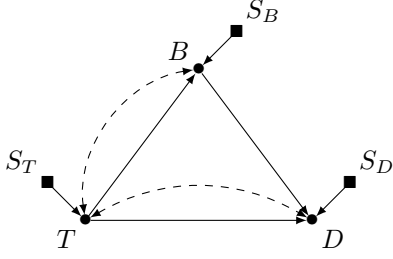
Figure 1: A selection diagram for three populations built on a causal diagram of T, B, and D. The selection variables on observable nodes T, B, and D are due to the differences on two unmeasured variables, gender and ethnicity.

f_i^* or $P^j\left(U^i\right) \neq P^*\left(U^i\right)\}$. We denote by $\mathcal{S}$, the collection of sets $\{\mathbf{S}^j\}_{j\in J}$. Given $\mathbf{W} \subseteq \mathbf{V}$ a subset of observed variables, $\mathbf{S_W}$ is given by $\{S_i \in \mathbf{S} \mid \exists_{V_i \in \mathbf{W}} S_i \rightarrow V_i\}$. A set of possible values of a selection variable S_i is $\{*\} \cup \{j \in J \mid S_i \in \mathbf{S}^j\}$. Given a selection diagram, P with selection variables setting to $\mathbf{s}^j = j$ and $\mathbf{s} \setminus \mathbf{S}^j = *$ represents P^j for $j \in J$ and P with $\mathbf{s} = *$ stands for P^*.[3]

Definition 2 (Causal Effects *m*-Transportability (adapted from (Bareinboim and Pearl 2012b))). Let D be a selection diagram and $\mathcal{S}$ a collection of sets of domain-specific selection variables relative to a set of domains $\Pi^* \cup \mathbf{\Pi}$. Let I be the information set $\left\{I\left(\Pi^j\right)\right\}_{j\in J} \cup \{I\left(\Pi^*\right)\}$ where $I\left(\Pi^j\right) = \left\{P^j\left(\mathbf{V} \setminus \mathbf{V}' \mid do\left(\mathbf{v}'\right)\right)\right\}_{\mathbf{V}' \subsetneq \mathbf{V}}$, a set of interventional distributions from Π^j for every $j \in J$, and $I\left(\Pi^*\right) = \{P^*\left(\mathbf{V}\right)\}$, an observational distribution from the target domain Π^*. The causal effect $R = P_{\mathbf{x}}\left(\mathbf{y}\right)$ is said to be *m-transportable* from $\mathbf{\Pi}$ to Π^* in D if $P_{\mathbf{x}}^*\left(\mathbf{y}\right)$ is uniquely computable from I in any model that induces D and $\mathcal{S}$.

The following example illustrates the power of *m*-transportability.

The Power of *m*-Transportability Consider the selection diagram (adapted from Figure 2(c) in (Pearl and Bareinboim 2011)) shown in Figure 1 which includes three observed variables corresponding to treatment T, disease D, and a biological marker B; two confounders, gender (between T and B) and ethnicity (between T and D). Suppose we are interested in estimating a causal effect of treatment on disease $P_t^*\left(d\right)$ in a population Π^* given observational study from Π^* and experimental studies on two different populations Π^1 and Π^2. Suppose Π^* differs from Π^1 and Π^2 in terms of their gender and ethnicity distributions respectively. These differences are denoted by a set of selection variables $\mathbf{S} = \{S_T, S_B, S_D\}$ and domain-specific selection variables $\mathbf{S}^1 = \{S_T, S_B\}$ and $\mathbf{S}^2 = \{S_T, S_D\}$. It follows from Theorems 4 and 5 of (Bareinboim and Pearl 2012b) that the causal effect $P_t^*\left(d\right)$ in the population Π^* is not transportable from Π^1 or Π^2 alone. However, it is easy to see that $P_t^*\left(d\right)$ can be computed by combining the observational and experimen-

---

[3]Unlike Bareinboim and Pearl (2013), the selection diagram D is constructed from a target domain and *all* source domains.

tal data from Π^1 and Π^2 and applying the standard rules of probability and rule 1 of *do*-calculus:

$$
\begin{aligned}
P_t^*\left(d\right) &= \sum_b P_t^*\left(d \mid b\right) P_t^*\left(b\right) \\
&= \sum_b P_t\left(d \mid b, s_D = *\right) P_t\left(b \mid s_B = *\right) \\
&= \sum_b P_t^1\left(d \mid b\right) P_t^2\left(b\right).
\end{aligned}
$$

The necessary and sufficient conditions for *m*-transportability follow immediately from the completeness (Shpitser and Pearl 2006b; Huang and Valtorta 2006; Pearl and Bareinboim 2011) of *do*-calculus.

Theorem 1 (*m*-Transportability Theorem (adapted from Theorem 1 in (Pearl and Bareinboim 2011))). *Let D be the selection diagram characterizing $\mathbf{\Pi}$ and Π^*, and $\mathcal{S}$ be a collection of sets of domain-specific selection variables in D. The effect $P_{\mathbf{x}}\left(\mathbf{y} \mid \mathbf{z}\right)$ is m-transportable from $\mathbf{\Pi}$ to Π^* if and only if the expression $P_{\mathbf{x}}\left(\mathbf{y} \mid \mathbf{z}, \mathbf{s}\right)$ is reducible, using the rules of* do-*calculus, to an expression where terms are either* do-*free or $\mathbf{S}^j$-free (selection variables that appear as conditioning variable in* do-*terms are a subset of $\mathbf{S} \setminus \mathbf{S}^j$ for some $\Pi^j \in \mathbf{\Pi}$.)*

Proof. The proof follows immediately from the completeness of *do*-calculus (Shpitser and Pearl 2006b; Huang and Valtorta 2006) and the *do*-calculus characterization lemma (Lemma 1 in (Pearl and Bareinboim 2011)) with the one minor modification with respect to when terms with *do*-operators are estimable from $\mathbf{\Pi}$: Since Π^j and Π^* differ with respect to selection variables $\mathbf{S}^j$, every *do*-term that does not contain $\mathbf{S}^j$ is estimable from Π^j. □

By applying rule 2 of *do*-calculus and marginalization (Shpitser and Pearl 2006a), we can reduce estimation of a conditional causal effect to the estimation of an unconditional causal effect. Hence, without loss of generality, we focus on *m*-transportability of unconditional causal effects.

Corollary 1. *If no finite sequence of rewritings of a causal effect using standard rules of probability and the rules of* do-*calculus yields a transport formula, then the effect is not m-transportable from source domains to the target domain.*

4 Characterizing *m*-Transportability

We proceed to characterize *m*-transportability in terms of properties of selection diagrams. We first extend the definitions of *s*-tree* and *s-hedge* (Bareinboim and Pearl 2012b) to the setting with *m* source domains instead of a single source domain. We say that a selection diagram is an *ms*-tree* if it is an s*-tree for every source domain. We define an ms-hedge by replacing s*-trees by ms*-trees in the definition of an s-hedge (Bareinboim and Pearl 2012b).

4.1 Analysis of Hedges

A hedge (Shpitser and Pearl 2006b) is an essential structure which hinders identification of a causal effect. Hedge structures were used to generate counter-examples (Shpitser and Pearl 2006b) to establish non-identifiability of causal effects in specific settings. Unlike (Shpitser and Pearl 2006b; Bareinboim and Pearl 2012b), we exploit the completeness of *do*-calculus to show the non-*m*-transportability of a causal

effect. Here, we examine hedge, s-hedge, and ms-hedge more closely to reveal some formulae that are not uniquely computable on them.

Definition 3 (Bare-Hedge)**.** Let D be a selection diagram and $\mathbf{X}$, $\mathbf{Y}$ be nonempty disjoint sets of $\mathbf{V}$. Then a triple $\langle D, \mathbf{X}, \mathbf{Y} \rangle$ is called a *bare-hedge* if D and $D \setminus \mathbf{X}$ are $\mathbf{Y}$-rooted ms*-trees forming an ms-hedge for $P_{\mathbf{x}}(\mathbf{y})$ in D.

It follows from Corollary 1 that if $P_{\mathbf{x}}(\mathbf{y})$ is not *m*-transportable, then any sequence of rewritings of $P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{s})$ must yield a formula with at least one *do*-term which is not $\mathbf{S}^j$-free. This implies that non-*m*-transportable causal effects must contain a term that cannot be eliminated to yield a transport formula. When such a term is rewritten using other terms, at least one of them must contain such a term as well. We say that the probability terms that satisfy this condition are of the definite form (defined below):

Definition 4 (Definite Form)**.** Given a bare-hedge $\langle D, \mathbf{X}, \mathbf{Y} \rangle$, a probability term is said to be of the *definite form* if and only if it is of the form $P_{\hat{\mathbf{x}}}(\mathbf{v}' \mid \mathbf{v}'', \mathbf{s}'')$ where $\mathbf{V}', \mathbf{V}'' \subset \mathbf{V}$ and (i) $\hat{\mathbf{X}} \subseteq \mathbf{X}$ and $\hat{\mathbf{X}} \neq \emptyset$, (ii) $\mathbf{V}' \cup \mathbf{V}'' \supseteq \mathbf{Y}$, (iii) $\mathbf{V}' \setminus \mathbf{X} \neq \emptyset$, and (iv) $\mathbf{S}_{\mathbf{V} \setminus \mathbf{X}} \subseteq \mathbf{S}'' \subseteq \mathbf{S}$. We will use $\Psi^D_{\mathbf{X}, \mathbf{Y}}$ to denote the set of all probability terms of the definite form and $\complement \Psi^D_{\mathbf{X}, \mathbf{Y}}$ to denote its complement.

A set of probabilities defined on a selection diagram is of the form $P_{\mathbf{v}_1}(\mathbf{v}_2 \mid \mathbf{v}_3, \mathbf{s}')$ where $\mathbf{V}_1$, $\mathbf{V}_2$, and $\mathbf{V}_3$ are disjoint subsets of observed variables, $\mathbf{V}_2$ is nonempty, and a subset $\mathbf{S}'$ of selection variables satisfies either $\mathbf{S}' = \mathbf{S}$ or $((\mathbf{S} \setminus \mathbf{S}') \perp\!\!\!\perp \mathbf{V}_2 \mid \mathbf{V}_3, \mathbf{V}_1)_{D_{\overline{\mathbf{V}_1}}}$. We choose $\mathbf{s}' = *$ or $\mathbf{s}' \setminus \mathbf{S}^j = *$ and $\mathbf{s}' \cap \mathbf{S}^j = j$ to select a population from among Π^* and $\Pi^j \in \mathbf{\Pi}$.

Hedges and Do-calculus Because hedges play an important role in the identifiability of causal effects, it is instructive to examine the interplay between a hedge and the rules of *do*-calculus. We will show a closure property of $\Psi^D_{\mathbf{X}, \mathbf{Y}}$ with respect to *do*-calculus. Towards this end, we define mutually exclusive ancestors of a DAG G with respect to a set of nodes.

Definition 5 (Mutually Exclusive Ancestors)**.** The mutually exclusive ancestors of a DAG G with respect to a subset $\mathbf{W}$ of its nodes is defined as

$$\Upsilon^G(\mathbf{W}) = \left\{ An(W)_G \setminus An(\mathbf{W} \cap an(W)_G)_G \right\}_{W \in \mathbf{w}}$$

We use $\mathscr{F}_{\mathbf{W}}$ to denote a family of sets $\Upsilon^G(\mathbf{W})$ indexed by $\mathbf{W}$. Given $\mathbf{W}' \subseteq \mathbf{W}$, we say that $\mathscr{F}_{\mathbf{W}'}$ is a subfamily of $\mathscr{F}_{\mathbf{W}}$. Given $W \in \mathbf{W}$, we use $\mathscr{F}_W$ to denote the corresponding element of the family $\mathscr{F}_{\mathbf{W}}$. Followings are its properties which are tightly related to the d-separation criterion (Pearl 1988).

Proposition 1. *Let $\mathscr{F}$ be a family $\Upsilon^{G[\mathbf{A}]}(\mathbf{B})$ given a bare-hedge $\langle G, \mathbf{X}, \mathbf{Y} \rangle$ where $\mathbf{B} \subseteq \mathbf{A}$ and $\mathbf{A} \subseteq \mathbf{V}$. Let $A \in \mathbf{A}$ and $B \in \mathbf{B}$. Following properties hold:*

(a) The union of the family is $An(\mathbf{B})_{G[\mathbf{A}]}$.

(b) The only common descendant of all elements in $\mathscr{F}_B$ is B.

(c) The family is pairwise disjoint.

(d) Each vertex in $\mathscr{F}_B$ is d-connected to B in $G[\mathscr{F}_B]$.

(e) Each pair of vertices in $\mathscr{F}_B$ is d-connected in $G[\mathscr{F}_B]$ conditioned on B.

Suppose, further that $\mathcal{C}(\mathbf{A}) = \{G[\mathbf{A}]\}$, $An(\mathbf{B})_{G[\mathbf{A}]} = \mathbf{A}$, and $\mathbf{B}$ is the union of two disjoint nonempty sets $\mathbf{C}$ and $\mathbf{D}$. Then we have:

(f) Each set in $\mathscr{F}_{\mathbf{D}}$ is d-connected to some set in $\mathscr{F}_{\mathbf{C}}$ through a bidirected path in $G[\mathbf{A}]$ conditioned on $\mathbf{D}$.

(g) Let $\mathbf{C}$ be the union of two disjoint nonempty sets $\mathbf{C}'$ and $\mathbf{C}''$. Then $\bigcup \mathscr{F}_{\mathbf{C}'}$ and $\bigcup \mathscr{F}_{\mathbf{C}''}$ are d-connected via bidirected edges in $G[\mathbf{A}]$ conditioned on $\mathbf{D}$.

Proof. Omitted. $\qquad\square$

We denote the *syntactic symmetric difference* between two probability terms $R_1 = P_{\mathbf{v}_1}(\mathbf{v}_2 \mid \mathbf{v}_3)$ and $R_2 = P_{\mathbf{w}_1}(\mathbf{w}_2 \mid \mathbf{w}_3)$ by $R_1 \ominus R_2$ which is given by $\bigcup_i \mathbf{V}_i \ominus \mathbf{W}_i$.[4] If two probability terms are equal according to *do*-calculus, then $\mathbf{v}_2 = \mathbf{w}_2$.

Lemma 1. *Let $\langle D, \mathbf{X}, \mathbf{Y} \rangle$ be a bare-hedge, ψ be a probability term in $\Psi^D_{\mathbf{X}, \mathbf{Y}}$, and ψ' be a probability term defined on D. If $\psi \ominus \psi'$ is a nonempty subset of $\mathbf{S}_{\mathbf{V} \setminus \mathbf{X}} \cup \mathbf{V} \setminus \mathbf{X}$, then conditions of three rules of* do-*calculus to impose equality between ψ and ψ' are not satisfied.*

Proof. Since ψ does not contain interventions on $\mathbf{S}_{\mathbf{V} \setminus \mathbf{X}} \cup \mathbf{V} \setminus \mathbf{X}$, we do not consider deletion or change of interventions. In the case of the selection variables, since they appear only as observations and $\mathbf{S}_{\mathbf{V} \setminus \mathbf{X}} \subseteq \mathbf{S}''$, it suffices to consider only removal of selection variables. Suppose that ψ is $P_{\hat{\mathbf{x}}}(\mathbf{v}' \mid \mathbf{v}'', \mathbf{s}'')$ and let $\mathbf{W}$ be $\psi \ominus \psi'$. In the following, we prove conditional dependencies related to rules of *do*-calculus by showing the existence of a d-connection path in a subgraph of D.

(a) Since $(\mathbf{V}' \not\perp\!\!\!\perp \mathbf{W} \mid \mathbf{V}'', \hat{\mathbf{X}}, \mathbf{S}'')_{D_{\overline{\hat{\mathbf{x}}}}}$, insertion of $\mathbf{W}$ as observations on ψ is impossible. Let $\mathscr{F}$ be $\Upsilon^{D[\mathbf{V} \setminus \mathbf{X}]}(\mathbf{V}' \cup \mathbf{V}'' \setminus \mathbf{X})$. From Proposition 1(a), it follows that either $W \in \bigcup \mathscr{F}_{\mathbf{V}' \setminus \mathbf{X}}$ or $W \in \bigcup \mathscr{F}_{\mathbf{V}'' \setminus \mathbf{X}}$. When $W \in \bigcup \mathscr{F}_{\mathbf{V}' \setminus \mathbf{X}}$, there exists a variable V' such that $W \in \mathscr{F}_{V'} \in \mathscr{F}_{\mathbf{V}' \setminus \mathbf{X}}$. From Proposition 1(d), it follows that there must be a d-connection path between W and V'. When $W \in \bigcup \mathscr{F}_{\mathbf{V}'' \setminus \mathbf{X}}$, there exists a variable V'' such that $W \in \mathscr{F}_{V''} \in \mathscr{F}_{\mathbf{V}'' \setminus \mathbf{X}}$. Because $\mathcal{C}(D[\mathbf{V} \setminus \mathbf{X}]) = \{D[\mathbf{V} \setminus \mathbf{X}]\}$ and $An(\mathbf{V}' \cup \mathbf{V}'' \setminus \mathbf{X})_{D[\mathbf{V} \setminus \mathbf{X}]} = \mathbf{V} \setminus \mathbf{X}$ from Proposition 1(f), $A \in \mathscr{F}_{V''}$ and $B \in \mathscr{F}_{V'} \in \mathscr{F}_{\mathbf{V}' \setminus \mathbf{X}}$ are d-connected for some variable A, B, and V'. Proposition 1(e) and Proposition 1(d) respectively imply the existence of d-connection paths between A and W given V'' in $\mathscr{F}_{V''}$ and between B and V' in $\mathscr{F}_{V'}$. Hence, W and V' must be d-connected in $D[\mathbf{V} \setminus \mathbf{X}]$.

(b) Since $(\mathbf{V}' \not\perp\!\!\!\perp \mathbf{W} \mid \mathbf{V}'' \setminus \mathbf{W}, \hat{\mathbf{X}}, \mathbf{S}'' \setminus \mathbf{W})_{D_{\overline{\hat{\mathbf{x}}}}}$, deletion of $\mathbf{W}$ in observation of ψ is impossible. Let $\mathbf{W}_1 = \mathbf{W} \cap \mathbf{V}''$ and $\mathbf{W}_2 = \mathbf{W} \cap \mathbf{S}''$. When $\mathbf{W}_1 \neq \emptyset$,

---

[4]The symmetric difference between two sets $\mathbf{C}$ and $\mathbf{D}$ is denoted by $\mathbf{C} \ominus \mathbf{D}$ which is given by $(\mathbf{C} \cup \mathbf{D}) \setminus (\mathbf{C} \cap \mathbf{D})$.

let $\mathscr{F}$ be $\Upsilon^{D[\mathbf{V}\setminus\mathbf{X}]}(\mathbf{W}_1 \cup \mathbf{V}' \cup \mathbf{V}'' \setminus \mathbf{X})$, then $\bigcup \mathscr{F}_{\mathbf{W}_1}$, $\bigcup \mathscr{F}_{\mathbf{V}'\setminus\mathbf{X}}$, and $\bigcup \mathscr{F}_{\mathbf{V}''\setminus\mathbf{X}}$ partitions $\mathbf{V} \setminus \mathbf{X}$. Since the three sets are confounded, $A \in \mathscr{F}_{V'} \in \mathscr{F}_{\mathbf{V}'\setminus\mathbf{X}}$ and $B \in \mathscr{F}_W \in \mathscr{F}_{\mathbf{W}_1}$ are d-connected via bidirected edges given $\mathbf{V}'' \setminus \mathbf{X}$ for some variables A, B, V' and W by Proposition 1(g). From Proposition 1(d), there are d-connection paths between V' and A and between B and W. Hence, V' and W are d-connected through A and B. When $\mathbf{W}_1 = \emptyset$ and $\mathbf{W}_2 \neq \emptyset$, for a selection variable $S_i \in \mathbf{W}_2$ let V_i be a variable S_i points to. From (a), V_i and V' are d-connected. Hence, $\mathbf{W}_2$ and $\mathbf{V}'$ must be d-connected in $D[\mathbf{W}_2 \cup \mathbf{V} \setminus \mathbf{X}]$.

(c) Since $(\mathbf{V}' \not\perp\!\!\!\perp \mathbf{W} \mid \mathbf{V}'' \setminus \mathbf{W}, \hat{\mathbf{X}}, \mathbf{S}'')_{D_{\overline{\hat{\mathbf{X}}\mathbf{W}}}}$, change of $\mathbf{W}$ in observation to interventions of ψ is impossible. Since $\Upsilon^{D[\mathbf{V}\setminus\mathbf{X}]}_{\underline{\mathbf{w}}}(\mathbf{W} \cup \mathbf{V}' \cup \mathbf{V}'' \setminus \mathbf{X})$ is identical to $\Upsilon^{D[\mathbf{V}\setminus\mathbf{X}]}(\mathbf{W}_1 \cup \mathbf{V}' \cup \mathbf{V}'' \setminus \mathbf{X})$ defined in (b) with $\mathbf{W}_2 = \emptyset$, this is a special case of (b).

(d) Since $(\mathbf{V}' \not\perp\!\!\!\perp \mathbf{W} \mid \mathbf{V}'', \hat{\mathbf{X}}, \mathbf{S}'')_{D_{\overline{\hat{\mathbf{X}}},\overline{\mathbf{W}(\mathbf{V}''\cup\mathbf{S}'')}}}$, insertion of $\mathbf{W}$ as interventions on ψ is impossible. Let $\mathbf{W}_1$ be $\mathbf{W} \setminus An(\mathbf{V}'')_D$, then $D[\mathbf{V} \setminus \mathbf{X}]_{\overline{\hat{\mathbf{X}}},\overline{\mathbf{W}(\mathbf{V}''\cup\mathbf{S}'')}} = D[\mathbf{V} \setminus \mathbf{X}]_{\overline{\mathbf{W}_1}}$. When $\mathbf{W}_1 = \emptyset$, we follow the proof of (a) since $D[\mathbf{V} \setminus \mathbf{X}]_{\overline{\mathbf{W}_1}} = D[\mathbf{V} \setminus \mathbf{X}]$. When $\mathbf{W}_1 \neq \emptyset$, $\mathbf{W}_1$ is a subset of $An(\mathbf{V}' \cap \mathbf{Y})_{D[\mathbf{V}\setminus\mathbf{X}]}$ since $\mathbf{Y}$ is a maximal root set of D and $\mathbf{V}' \cup \mathbf{V}'' \supseteq \mathbf{Y}$. A set of maximal elements[5] of $\mathbf{W}_1$ is a nonempty subset of $An(\mathbf{V}' \cap \mathbf{Y})_{D[\mathbf{V}\setminus\mathbf{X}]_{\overline{\mathbf{W}_1}}}$. Hence, $\mathbf{W}$ and $\mathbf{V}'$ are d-connected in $D[\mathbf{V} \setminus \mathbf{X}]_{\overline{\mathbf{W}_1}}$.

It follows that it is not possible to transform ψ into ψ' using *do*-calculus rules. $\square$

Lemma 2. *Let $\langle D, \mathbf{X}, \mathbf{Y}\rangle$ be a bare-hedge, ψ a probability term in $\Psi^D_{\mathbf{X},\mathbf{Y}}$, and ψ' a probability term defined on D. If ψ' is derived from ψ using* do-*calculus and $\psi \ominus \psi'$ is a subset of $\mathbf{S_X} \cup \mathbf{X}$, then ψ' is in $\Psi^D_{\mathbf{X},\mathbf{Y}}$.*

Proof. Suppose ψ is $P_{\hat{\mathbf{x}}}(\mathbf{v}' \mid \mathbf{v}'', \mathbf{s}'')$. It is easy to see that ψ' trivially satisfies the second, third, and fourth conditions of the definite form (see Definition 4). Hence, we focus our attention on the first condition and prove that the interventions $\hat{\mathbf{X}}$ in ψ can neither be (a)changed to observations nor (b) deleted.

(a) Since $(\mathbf{V}' \not\perp\!\!\!\perp \hat{\mathbf{X}} \mid \mathbf{V}'', \mathbf{S}'')_{D_{\hat{\mathbf{X}}}}$, $\hat{\mathbf{X}}$ cannot be rewritten as observations from ψ. Let $\mathscr{F} = \Upsilon^{D[\mathbf{V}]_{\hat{\mathbf{x}}}}(\mathbf{V}' \cup \mathbf{V}'' \cup \hat{\mathbf{X}})$. Then the three sets, $\bigcup \mathscr{F}_{\hat{\mathbf{X}}}$, $\bigcup \mathscr{F}_{\mathbf{V}'}$, and $\bigcup \mathscr{F}_{\mathbf{V}''}$, are confounded since $\mathcal{C}(D[\mathbf{V}]_{\hat{\mathbf{X}}}) = \{D[\mathbf{V}]_{\hat{\mathbf{X}}}\}$ which implies that $\mathscr{F}_{\hat{X}} \in \mathscr{F}_{\hat{\mathbf{X}}}$ and $\mathscr{F}_{V'} \in \mathscr{F}_{\mathbf{V}'}$ are d-connected through nodes in $\bigcup \mathscr{F}_{\mathbf{V}''}$ by Proposition 1(g). It is easy to see that extending the resulting path to the roots of $\mathscr{F}_{\hat{X}}$ and $\mathscr{F}_{V'}$, yields a d-connection path between $\hat{X}$ and V'.

---

[5]In order theory, an element X is a maximal element of a set $\mathbf{X}$ if there is no $Y \in \mathbf{X}$ such that $X < Y$. In a graph G, $X \in an(Y)_G$ implies $X < Y$.

(b) Since $(\mathbf{V}' \not\perp\!\!\!\perp \hat{\mathbf{X}} \mid \mathbf{V}'', \mathbf{S}'')_{D_{\overline{\hat{\mathbf{X}}(\mathbf{V}'',\mathbf{s})}}}$, $\hat{\mathbf{X}}$ cannot be deleted from ψ. Let $\hat{\mathbf{X}}_1 = \hat{\mathbf{X}} \setminus An(\mathbf{V}'')_D$, then $D[\mathbf{V}]_{\overline{\hat{\mathbf{X}}(\mathbf{V}'',\mathbf{S}'')}} = D[\mathbf{V}]_{\overline{\hat{\mathbf{X}}_1}}$. If $\hat{\mathbf{X}}_1 \neq \emptyset$, each maximal element of $\hat{\mathbf{X}}_1$ is d-connected to $\mathbf{V}' \cap \mathbf{Y}$ as shown in Lemma 1(d). If $\hat{\mathbf{X}}_1 = \emptyset$, then $D[\mathbf{V}]_{\overline{\hat{\mathbf{X}}_1}} = D[\mathbf{V}]$. The d-connection path between $\hat{\mathbf{X}}$ and $\mathbf{V}'$ in $D[\mathbf{V}]_{\hat{\mathbf{x}}}$ from (a) is also in $D[\mathbf{V}]$.

A proper subset of $\hat{\mathbf{X}}$ or a subset of $\mathbf{S_X}$ might be altered by *do*-calculus but $\hat{\mathbf{X}}$ cannot be changed nor deleted from ψ using *do*-calculus. Therefore, ψ' is in $\Psi^D_{\mathbf{X},\mathbf{Y}}$. $\square$

Lemma 3. *Let $\langle D, \mathbf{X}, \mathbf{Y}\rangle$ be a bare-hedge, ψ a probability term in $\Psi^D_{\mathbf{X},\mathbf{Y}}$, and ψ' a probability term defined on D. If $\psi \ominus \psi'$ intersects $\mathbf{H}_1 = \mathbf{S}_{\mathbf{V}\setminus\mathbf{X}} \cup \mathbf{V} \setminus \mathbf{X}$ and $\mathbf{H}_2 = \mathbf{S_X} \cup \mathbf{X}$, ψ cannot be rewritten as ψ' using* do-*calculus.*

Proof. Let $\mathbf{W}$ be $\psi \ominus \psi'$, $\mathbf{W}_1 = \mathbf{W} \cap \mathbf{H}_1$, and $\mathbf{W}_2 = \mathbf{W} \cap \mathbf{H}_2$. Since $\mathbf{W} \not\subseteq \hat{\mathbf{X}}$, the *do*-calculus rules for changing interventions to observations or deleting interventions are not applicable. To show that none of the other *do*-calculus rules are applicable, due to the decomposition property of the conditional independence test, it suffices to show conditional dependence between $\mathbf{V}'$ and $\mathbf{W}_1$. From Lemma 1, we know that $\mathbf{V}'$ and $\mathbf{W}_1$ are d-connected through paths that lie in a subgraph of $D[\mathbf{H}_1]$. It follows that $\mathbf{V}'$ and $\mathbf{W}_1$ are not conditionally independent regardless of the choice of $\mathbf{W}_2$. $\square$

The preceding results lead to the following lemma:

Lemma 4. *Given a bare-hedge $\langle D, \mathbf{X}, \mathbf{Y}\rangle$, a set $\Psi^D_{\mathbf{X},\mathbf{Y}}$ is closed under* do-*calculus.*

Proof. By Lemma 1–3, if a probability ψ in $\Psi^D_{\mathbf{X},\mathbf{Y}}$ is equal to a probability ψ' by *do*-calculus, then the syntactic symmetric difference between ψ and ψ' is a subset of $\mathbf{X} \cup \mathbf{S_X}$ and ψ' is also in $\Psi^D_{\mathbf{X},\mathbf{Y}}$. $\square$

This directly leads to the following corollary.

Corollary 2. *Given a bare-hedge $\langle D, \mathbf{X}, \mathbf{Y}\rangle$, the complement of $\Psi^D_{\mathbf{X},\mathbf{Y}}$ is closed under* do-*calculus.*

Hedges and Standard Probability Manipulations We examine the behavior of definite form probability associated with a bare-hedge under the standard probability manipulations.

Lemma 5. *Let $\langle D, \mathbf{X}, \mathbf{Y}\rangle$ be a bare-hedge. If a formula is derived from a probability term in $\Psi^D_{\mathbf{X},\mathbf{Y}}$ by standard probability manipulations, then the formula contains at least one probability term in $\Psi^D_{\mathbf{X},\mathbf{Y}}$.*

Proof. Let $\psi \in \Psi^D_{\mathbf{X},\mathbf{Y}}$ be expressed as $P_{\hat{\mathbf{x}}}(\mathbf{v}' \mid \mathbf{v}'', \mathbf{s}'')$. By the definition of conditional probability, $\psi = P_{\hat{\mathbf{x}}}(\mathbf{v}', \mathbf{v}'' \mid \mathbf{s}'') / P_{\hat{\mathbf{x}}}(\mathbf{v}'' \mid \mathbf{s}'')$ if $P_{\hat{\mathbf{x}}}(\mathbf{v}'' \mid \mathbf{s}'') \neq 0$ and the dividend is of the definite form. From the definition of marginal probability (or law of total probability), we have: $\psi = \sum_{\mathbf{w}} P_{\hat{\mathbf{x}}}(\mathbf{v}', \mathbf{w} \mid \mathbf{v}'', \mathbf{s}'')$.

The sum and product rules are sufficient for probabilistic inference (Jaynes and Bretthorst 2003, pg. 35). From the sum rule, we have: $\psi = 1 - \sum_{\mathbf{w} \in \complement \mathbf{v}'} P_{\hat{\mathbf{x}}}(\mathbf{w} \mid \mathbf{v}'', \mathbf{s}'')$ where $\complement \mathbf{v}'$ is the complement of event $\mathbf{v}'$. It is easy to see that this does not yield a transport formula for ψ. From the product rule, we have: $\psi = P_{\hat{\mathbf{x}}}(\mathbf{v}' \setminus \mathbf{w} \mid \mathbf{w}, \mathbf{v}'', \mathbf{s}'') P_{\hat{\mathbf{x}}}(\mathbf{w} \mid \mathbf{v}'', \mathbf{s}'')$ where $\mathbf{W} \subset \mathbf{V}'$. If $\mathbf{V}' \setminus \mathbf{W} \setminus \mathbf{X} \neq \emptyset$, then the first term of ψ is in $\Psi_{\mathbf{X}, \mathbf{Y}}^D$. Otherwise, the second term of ψ is in $\Psi_{\mathbf{X}, \mathbf{Y}}^D$. We can show similar results for other probability manipulations e.g., Bayes' theorem. $\square$

Hedges and Transport Formula

Lemma 6. *Given a bare-hedge $\langle D, \mathbf{X}, \mathbf{Y} \rangle$, none of probability terms in $\Psi_{\mathbf{X}, \mathbf{Y}}^D$ can be expressed in terms of probability terms only in $\complement \Psi_{\mathbf{X}, \mathbf{Y}}^D$.*

Proof. This follows from the closure property of $\Psi_{\mathbf{X}, \mathbf{Y}}^D$ under *do*-calculus (Lemma 4) and the behavior of probability terms in $\Psi_{\mathbf{X}, \mathbf{Y}}^D$ under standard probability manipulations (Lemma 5). $\square$

The preceding results lead to the following theorem which establishes a crucial connection between bare hedge structures and (lack of) *m*-transportability.

Theorem 2. *Let D, $\mathcal{S}$, and I be as defined in the definition of causal effects m-transportability. Let $\langle D, \mathbf{X}, \mathbf{Y} \rangle$ be a bare-hedge. Then, none of causal effect of the target domain in $\Psi_{\mathbf{X}, \mathbf{Y}}^D$ is uniquely computable from I and $\complement \Psi_{\mathbf{X}, \mathbf{Y}}^D$ in any model that induces D and $\mathcal{S}$.*

Proof. Let $^* \Psi_{\mathbf{X}, \mathbf{Y}}^D$ be a set of probability terms in $\Psi_{\mathbf{X}, \mathbf{Y}}^D$ with $\mathbf{s} = *$. Then, the union of I and $\complement \Psi_{\mathbf{X}, \mathbf{Y}}^D$ is all probability terms except $^* \Psi_{\mathbf{X}, \mathbf{Y}}^D$. By Lemma 6, any expression derived from a probability term ψ^* in $^* \Psi_{\mathbf{X}, \mathbf{Y}}^D$ contains at least one term in $^* \Psi_{\mathbf{X}, \mathbf{Y}}^D$. Thus, there is no transport formula for ψ^* and it is not uniquely computable from given information by the completeness of *do*-calculus. $\square$

5 Algorithm for *m*-Transportability

We present below, **msID**, a sound and complete algorithm for deciding m-transportability that generalizes **sID**, an algorithm for deciding transportability (or 1-transportability) (Bareinboim and Pearl 2012b), which is a modification of the causal effects identifiability algorithm (Shpitser and Pearl 2006b).[6] Instead of a single source domain as in the case of **sID**, the arguments of **msID** include m source domains; A second key difference between **sID** and **msID** has to do with the substitution of line "if $(\mathbf{S} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X})_{D_{\overline{\mathbf{X}}}}$, return $P_{\mathbf{x}}(\mathbf{y})$" (line 6 of **sID**) with "if $\exists_{j \in J} (\mathbf{S}^j \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X})_{D_{\overline{\mathbf{X}}}}$, return $P_{\mathbf{x}}^j(\mathbf{y})$" (line 6 in Figure 2) to account for the use of experimental studies from m source domains instead of a single source domain.

---

[6] As noted earlier, **msID** is, modulo some notational differences, essentially equivalent to $\boldsymbol{\mu}$**sID** (Bareinboim and Pearl 2013). We include **msID** here to make the paper self-contained.

Function msID $(\mathbf{y}, \mathbf{x}, P, D)$
1: **if** $\mathbf{x} = \emptyset$ **, return** $P(\mathbf{y})$
2: **if** $\mathbf{V} \setminus An(\mathbf{Y})_D \neq \emptyset$ **, return** msID$\big(\mathbf{y}, \mathbf{x} \cap An(\mathbf{Y})_D,$
    $P(An(\mathbf{Y})_D), D[An(\mathbf{Y})_D]\big)$
3: $\mathbf{W} \leftarrow \mathbf{V} \setminus \mathbf{X} \setminus An(\mathbf{Y})_{D_{\overline{\mathbf{x}}}}$
    if $\mathbf{W} \neq \emptyset$, **return** msID $(\mathbf{y}, \mathbf{x} \cup \mathbf{w}, P, D)$
4: **if** $\mathcal{C}(D \setminus \mathbf{X}) = \{C_1, \ldots, C_k\}$ **,**
    return $\sum_{\mathbf{v} \setminus \{\mathbf{y}, \mathbf{x}\}} \prod_i$ msID $(c_i, \mathbf{v} \setminus c_i, P, D)$
5: **if** $\mathcal{C}(D \setminus \mathbf{X}) = \{C\}$ **,**
6:     **if** $\exists_j (\mathbf{S}^j \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X})_{D_{\overline{\mathbf{x}}}}$ **, return** $P_{\mathbf{x}}^j(\mathbf{y})$
7:     **if** $\mathcal{C}(D) = \{D\}$ **,** throws **FAIL** (D, C)
8:     **if** $C \in \mathcal{C}(D)$ **, return** $\sum_{\mathbf{c} \setminus \mathbf{y}} \prod_{V_i \in C} P(v_i \mid v_D^{(i-1)})$
9:     **if** $(\exists C') C \subset C' \in \mathcal{C}(D)$ **, return** msID$(\mathbf{y}, \mathbf{x} \cap C',$
        $\prod_{V_i \in C'} P(V_i \mid V_D^{(i-1)} \cap C', v_D^{(i-1)} \setminus C'), C')$

Figure 2: An *m*-transportability algorithm **msID** to identify $P_{\mathbf{x}}^*(\mathbf{y})$ by calling **msID**$(\mathbf{y}, \mathbf{x}, P^*, D)$. A set of sC-components in a graph D is denoted by $\mathcal{C}(D)$ (Bareinboim and Pearl 2012b). We assume that I and $\mathcal{S}$ are globally defined.

Theorem 3 (Soundness)**.** *Whenever **msID**$(\mathbf{y}, \mathbf{x}, P^*, D)$ outputs an expression for $P_{\mathbf{x}}^*(\mathbf{y})$, the expression correctly estimates $P_{\mathbf{x}}^*(\mathbf{y})$.*

Proof. The correctness of steps except the line 6 of **msID** are shown in (Shpitser and Pearl 2006b; Bareinboim and Pearl 2012b). The line 6 of **msID** parallelizes the line 6 of **sID**. Hence, the correctness follows from **sID** which is due to S-admissibility of $\mathbf{X}$ along with Corollary 1 in (Pearl and Bareinboim 2011). $\square$

We proceed to show that whenever **msID** returns FAIL with an ms-hedge, the causal effect in question is indeed not *m*-transportable.

Lemma 7 (adapted from Theorem 5 in (Bareinboim and Pearl 2012b))**.** *Let D be a selection diagram induced from $\boldsymbol{\Pi}$ and $\boldsymbol{\Pi}^*$. Assume that there exists F, F' that form an ms-hedge for $P_{\mathbf{x}}(\mathbf{y})$ in $\boldsymbol{\Pi}$ and $\boldsymbol{\Pi}^*$. Then $P_{\mathbf{x}}(\mathbf{y})$ is not m-transportable from $\boldsymbol{\Pi}$ to $\boldsymbol{\Pi}^*$.*

Proof. The assumptions of the lemma, together with the definitions of bare-hedge and ms-hedge (Section 4) imply that there exists a bare-hedge $\langle D', \mathbf{X}, \mathbf{R} \rangle$ where $\mathbf{R} \subset An(\mathbf{Y})_{D_{\overline{\mathbf{x}}}}$ and D' is a subgraph of D. Then, from Lemma 6, $P_{\mathbf{x}}(\mathbf{r})$ is not *m*-transportable from $\boldsymbol{\Pi}$ to $\boldsymbol{\Pi}^*$. From Lemma 14 in (Bareinboim and Pearl 2012b) it follows that $P_{\mathbf{x}}(\mathbf{y})$ is not *m*-transportable from $\boldsymbol{\Pi}$ to $\boldsymbol{\Pi}^*$. $\square$

Lemma 8 (adapted from Theorem 7 in (Bareinboim and Pearl 2012b))**.** *Whenever **msID**$(\mathbf{y}, \mathbf{x}, P^*, D)$ returns FAIL (i.e., fails to transport $P_{\mathbf{x}}(\mathbf{y})$ from $\boldsymbol{\Pi}$ to $\boldsymbol{\Pi}^*$), there must exist $\mathbf{X}' \subseteq \mathbf{X}$, $\mathbf{Y}' \subseteq \mathbf{Y}$, such that the graph pair D, C returned by the FAIL condition of **msID** contains as edge subgraphs ms*-trees F, F' that form an ms-hedge for $P_{\mathbf{x}'}(\mathbf{y}')$.*

Proof. Since **msID** generalizes **sID**, the proof follows from arguments along the lines of those used to prove Theorem 7 in (Bareinboim and Pearl 2012b). □

Theorem 4 (Completeness). *Algorithm **msID** is complete.*

Proof. This follows from Lemma 7 and 8 and the completeness of **sID** (Bareinboim and Pearl 2012b). □

It is easy to see that the **msID** terminates; and that its runtime complexity is polynomial in the number of variables. The number of recursion is $O(|\mathbf{V}|^2)$ since (i) each recursion in line 3 (adding interventions on $\mathbf{X}$) and 4 (factorization) is called at most once; (ii) line 2 and 9 *only* reduce the size of $\mathbf{X}$; and (iii) line 1, 6, 7, and 8 are terminal. In addition, graphical tests (e.g., tests for conditional independence) can be carried out in time that is polynomial in the number of observed variables.

6   Conclusion

We introduced m-transportability which offers a license to transfer causal information learned collectively from experiments and observations from m source environments to a given target environment where only observational information can be obtained. Bareinboim and Pearl (2013) independently introduced μ-transportability, which despite some differences in the formulation, turns out to be essentially identical to m-transportability. We established the necessary and sufficient conditions for m-transportability by directly exploiting the completeness of *do*-calculus (Shpitser and Pearl 2006b; Huang and Valtorta 2006). We introduced **msID**, a sound and complete algorithm for deciding m-transportability that outputs (a) evidence of non-m-transportability if a causal effect is not m-transportable from a given set of m source environments to a specified target environment; and (b) a transport formula, that is, a recipe for estimating the desired causal effect by combining experimental information from m source environments with only observational information from the target environment, otherwise. We note that **msID** is essentially identical to **μsID**, the algorithm for deciding μ-transportability that was independently introduced by Bareinboim and Pearl (2013). This is not surprising in light of the fact that m-transportability and μ-transportability are equivalent notions; Both **msID** and **μsID** are simple extensions of **sID**, the algorithm for deciding transportability (a special case of m-transportability where m=1) introduced by Bareinboim and Pearl (2012b); Both **msID** and **μsID** rely on the same graphical criterion to decide whether a causal relation is m-transportable (μ-transportable). However, the work presented in this paper differs from that of Bareinboim and Pearl (2013) with respect to the technique used to derive the graphical criterion for lack of m-transportability (μ-transportability): Bareinboim and Pearl (2013) construct a *certificate* that serves as counter-example that establishes lack of μ-transportability. In contrast, we demonstrate the nonexistence of a transport formula (and hence lack of m-transportability) by directly exploiting the completeness of *do*-calculus (Shpitser

and Pearl 2006b; Huang and Valtorta 2006) and standard probability manipulations.

Causal effects identifiability (Galles and Pearl 1995; Tian 2004; Tian and Pearl 2002; Shpitser and Pearl 2006a; 2006b), transportability (Pearl and Bareinboim 2011; Bareinboim and Pearl 2012b), z-identifiability (Bareinboim and Pearl 2012a) (the problem of estimating in a given domain, the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ from surrogate experiments on $\mathbf{Z}$ that are more amenable to experimental manipulation in the domain than $\mathbf{X}$) are all special cases of *meta-identifiability* (Pearl 2012) which has to do with non-parametric identification of causal effects given *multiple* domains and *arbitrary* information from each domain. The results presented in this paper expand the subclass of *meta-identifiability* problems with provably correct and complete algorithmic solutions. There are several additional special cases of meta-identifiability to consider, including in particular, generalizations of z-identifiability and transportability to allow causal information from experiments, including surrogate experiments, in *multiple* source domains to be combined to facilitate the estimation of a causal effect in a target domain.

References

Bareinboim, E., and Pearl, J. 2012a. Causal inference by surrogate experiments: z-identifiability. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, 113–120. AUAI Press.

Bareinboim, E., and Pearl, J. 2012b. Transportability of causal effects: Completeness results. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, 698–704. AAAI Press.

Bareinboim, E., and Pearl, J. 2013. Meta-transportability of causal effects: A formal approach. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*. To appear.

Galles, D., and Pearl, J. 1995. Testing identifiability of causal effects. In *Proceedings of the Eleventh Annual Con-*

ference on Uncertainty in Artificial Intelligence, 185–195. Morgan Kaufmann.

Huang, Y., and Valtorta, M. 2006. Pearl's calculus of intervention is complete. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, 217–224. AUAI Press.

Jaynes, E., and Bretthorst, G. 2003. *Probability Theory: The Logic of Science*. Cambridge University Press.

Pearl, J., and Bareinboim, E. 2011. Transportability of causal and statistical relations: A formal approach. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, 247–254. AAAI Press.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

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

Pearl, J. 2000. *Causality: models, reasoning, and inference*. New York, NY, USA: Cambridge University Press.

Pearl, J. 2012. The do-calculus revisited. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, 4–11. AUAI Press.

Shpitser, I., and Pearl, J. 2006a. Identification of conditional interventional distributions. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, 437–444. AUAI Press.

Shpitser, I., and Pearl, J. 2006b. Identification of joint interventional distributions in recursive semi-markovian causal models. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence*, 1219–1226. AAAI Press.

Tian, J., and Pearl, J. 2002. A general identification condition for causal effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, 567–573. AAAI Press / The MIT Press.

Tian, J. 2004. Identifying conditional causal effects. In *Proceedings of the Twentieth Conference in Uncertainty in Artificial Intelligence*, 561–568. AUAI Press.