

## External Validity and Transportability: A Formal Approach

Pearl, Judea

*University of California, Los Angeles, Department of Computer Science*

*Los Angeles, CA (90095), United States*

*E-mail: judea@cs.ucla.edu*

Bareinboim, Elias

*University of California, Los Angeles, Department of Computer Science*

*Los Angeles, CA (90095), United States*

*E-mail: eb@cs.ucla.edu*

### Abstract

We provide a formal definition of the notion of “transportability,” or “external validity,” as a license to transfer causal information from experimental studies to a different population in which only observational studies can be conducted. We introduce a formal representation called “selection diagrams” for expressing differences and commonalities between populations of interest and, using this representation, we derive procedures for deciding whether causal effects in the target population can be inferred from experimental findings in a different population. When the answer is affirmative, the procedures identify the set of experimental and observational studies that need be conducted to license the transport.

### Introduction

Science is about generalization; conclusions that are obtained in a laboratory setting are transported and applied elsewhere, in an environment that differs in many aspects from that of the laboratory.

If the target environment is arbitrary, or drastically different from the study environment nothing can be learned from the latter. However, the fact that most experiments are conducted with the intention of applying the results elsewhere means that we usually deem the target environment sufficiently similar to the study environment to justify the transport of experimental results or their ramifications.

Remarkably, the conditions that permit such transport have not received systematic formal treatment. The standard literature on this topic, falling under rubrics such as “quasi-experiments,” “meta analysis,” and “external validity,” consists primarily of “threats,” namely, verbal narratives of what can go wrong when we try to transport results from one study to another (e.g., [SCC02, chapter 3]; [Man07]). In contrast, we seek to establish “licensing assumptions,” namely, formal conditions under which the transport of results across diverse environments is licensed from first principles.

Transportability analysis requires a formal language within which the notion of “environment” or “population” is given precise characterization, and differences among populations can be encoded and analyzed. The advent of causal diagrams [Pea95, SGS00, Pea09] provides such a language and renders the formalization of transportability possible. Using this language, this paper offers a precise definition for the notion of transportability and establishes formal conditions that, if held true, would permit us to transport results across studies, domains, environments, or populations.

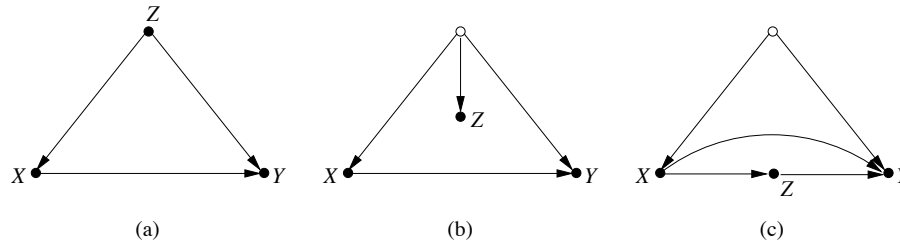


Figure 1: Causal diagrams depicting Examples 1–3. In (a)  $Z$  represents “age.” In (b)  $Z$  represents “linguistic skills” while age (hollow circle) is unmeasured. In (c)  $Z$  represents a biological marker situated between the treatment ( $X$ ) and a disease ( $Y$ ).

### Motivating Examples

To motivate our discussion and to demonstrate some of the subtle questions that transportability entails, we will consider three simple examples, graphically depicted in Fig. 1.

**Example 1** We conduct a randomized trial in Los Angeles (LA) and estimate the causal effect of treatment  $X$  on outcome  $Y$  for every age group  $Z = z$  as depicted in Fig. 1(a). We now wish to generalize the results to the population of New York City (NYC), but we find that the distribution  $P(x, y, z)$  in LA is different from the one in NYC (call the latter  $P^*(x, y, z)$ ). In particular, the average age in NYC is significantly higher than that in LA. How are we to estimate the causal effect of  $X$  on  $Y$  in NYC, denoted  $P^*(y|do(x))$ .<sup>1</sup>

If we can assume that age-specific effects  $P(y|do(x), Z = z)$  are invariant across cities, the overall causal effect in NYC should be

$$(1) \quad P^*(y|do(x)) = \sum_z P(y|do(x), z)P^*(z)$$

This *transport formula* combines experimental results obtained in LA,  $P(y|do(x), z)$ , with observational aspects of NYC population,  $P^*(z)$ , to obtain an experimental claim  $P^*(y|do(x))$  about NYC.

Our first task in this paper will be to explicate the assumptions that renders this extrapolation valid. We ask, for example, what must we assume about other confounding variables beside age, both latent and observed, for Eq. (1) to be valid, or, would the same transport formula hold if  $Z$  was not age, but some proxy for age, say, language proficiency. More intricate yet, what if  $Z$  stood for an  $X$ -dependent variable, say hyper-tension level, that stands between  $X$  and  $Y$ ? Let us examine the proxy issue first.

**Example 2** Let the variable  $Z$  in Example 1 stand for subjects language skills, which correlates with age (not measured) (see Fig. 1(b)). Given the observed disparity  $P(z) \neq P^*(z)$ , how are we to estimate the causal effect  $P^*(y|do(x))$  in NYC from the  $z$ -specific causal effect  $P(y|do(x), z)$  estimated in LA?

If the two cities enjoy identical age distributions and NYC residents acquire linguistic skills at a younger age, then, since  $Z$  has no effect whatsoever on  $X$  and  $Y$ , the inequality  $P(z) \neq P^*(z)$  can be ignored and, intuitively, the proper transport formula should be

$$(2) \quad P^*(y|do(x)) = P(y|do(x))$$

<sup>1</sup>The  $do(x)$  notation [Pea95, Pea09] interprets  $P(y|do(x))$  as the probability of outcomes  $Y = y$  in a randomized experiment where the treatment variables  $X$  take on values  $X = x$ .  $P(y|do(x), z)$  is logically equivalent to  $P(Y_x = y|Z_x = z)$  in counterfactual notation. Likewise, the diagrams used in this paper should be interpreted as parsimonious encoding of the data generating model [Pea09, p. 101], where every bi-directed arc  $X \leftrightarrow Y$  stands for a set of latent variables affecting  $X$  and  $Y$ .

If, on the other hand, the conditional probabilities  $P(z|age)$  and  $P^*(z|age)$  are the same in both cities, and the inequality  $P(z) \neq P^*(z)$  reflects genuine age differences, Eq. (2) is no longer valid, since the age difference may be a critical factor in determining how people react to  $X$ . We see, therefore, that the transport formula depends on the causal context in which distributional differences are embedded.

This example also demonstrates why the invariance of  $Z$ -specific causal effects should not be taken for granted. While justified in Example 1, with  $Z = age$ , it fails in Example 2, in which  $Z$  was equated with “language skills.” Indeed, using Fig. 1(b) for guidance, the  $Z$ -specific effect of  $X$  on  $Y$  in NYC is given by:

$$(3) \quad P^*(y|do(x), z) = \sum_{age} P(y|do(x), age)P^*(age|z)$$

Thus, if the two populations differ in the relation between age and skill, i.e.,  $P(age|z) \neq P^*(age|z)$  the skill-specific causal effect would differ as well.

**Example 3** *Examine the case where  $Z$  is a  $X$ -dependent variable, say a disease bio-marker as shown in Fig. 1(c). Assume further that the disparity  $P(z) \neq P^*(z)$  is discovered in each level of  $X$  and that, again, both the average and the  $z$ -specific causal effect  $P(y|do(x), z)$  are estimated in the LA experiment, for all levels of  $X$  and  $Z$ . Can we, based on information given, estimate the average causal effect in NYC?*

Here, Eq. (1) is wrong for two reasons. First, as in the case of age-proxy, it matters whether the disparity in  $P(z)$  represents differences in susceptibility to  $X$  or differences in propensity to receiving  $X$ . In the latter case, Eq. (2) would be valid, while in the former, more information is needed. Second, the overall causal effect is no longer a simple average of the  $z$ -specific causal effects but is given by

$$(4) \quad P^*(y|do(x)) = \sum_z P^*(y|do(x), z)P^*(z|do(x))$$

which reduces to (1) only in the special case where  $Z$  is unaffected by  $X$ , as is the case in Fig. 1(a). We shall see (Theorem 3 below) that the correct transport formula is

$$(5) \quad P^*(y|do(x)) = \sum_z P(y|do(x), z)P^*(z|x)$$

which calls for weighting the  $z$ -specific effects by  $P^*(z|x)$ , to be estimated at the target environment.

## Formalizing Transportability

### Selection Diagrams and Selection Variables

The examples above demonstrate that transportability is a causal, not statistical notion, requiring knowledge of the mechanisms, or processes, through which differences come about. To witness, every probability distribution  $P(x, y, z)$  that is compatible with the process of Fig. 1(b) is also compatible with that of Fig. 1(a) and, yet, the two processes dictate different transport formulas. Thus, to represent formally the differences between populations we must resort to a representation in which the causal mechanisms are explicitly encoded and in which population differences are represented as local modifications of those mechanisms.

To this end, we will use causal diagrams augmented with a set,  $S$ , of “selection variables,” where each member of  $S$  corresponds to a mechanism by which the two populations differ, and switching between the two populations will be represented by conditioning on different values of these  $S$  variables.

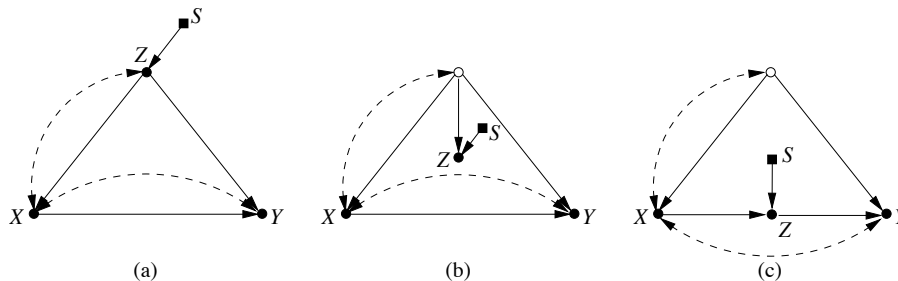


Figure 2: Selection diagrams depicting Examples 1–3. In (a) the two populations differ in age distributions. In (b) the populations differs in how  $Z$  depends on age (an unmeasured variable, represented by the hollow circle) and the age distributions are the same. In (c) the populations differ in how  $Z$  depends on  $X$ .

Formally, if  $P(v|do(x))$  stands for the distribution of a set  $V$  of variables in the experimental study (with  $X$  randomized) then we designate by  $P^*(v|do(x))$  the distribution of  $V$  if we were to conduct the study on population  $\Pi^*$  instead of  $\Pi$ . We now attribute the difference between the two to the action of a set  $S$  of selection variables, and write<sup>2</sup>

$$P^*(v|do(x)) = P(v|do(x), s^*).$$

Of equal importance is the absence of an  $S$  variable pointing to  $Y$  in Fig. 2(a), which encodes the assumption that age-specific effects are invariant across the two populations.

The variables in  $S$  represent exogenous conditions that determine the values of the variables to which they point.<sup>3</sup>

For example, the age disparity  $P(z) \neq P^*(z)$  discussed in Example 1 will be represented by the inequality  $P(z) \neq P(z|s)$  where  $S$  stands for all factors determining age differences between NYC and LA.

This graphical representation, which we will call “selection diagrams” can also represent structural differences between the two populations. For example, if the causal diagram of the study population contains an arrow between  $X$  and  $Y$ , and the one for the target population contains no such arrow, the selection diagram will be  $X \rightarrow Y \leftarrow S$  where the role of variable  $S$  is to disable the arrow  $X \rightarrow Y$  when  $S = s^*$  (i.e.,  $P(y|x, s^*) = P(y|x', s^*)$  for all  $x'$ ) and reinstate it when  $S = s$ .<sup>4</sup> Our analysis will apply therefore to all factors by which populations may differ or that may “threaten” the transport of conclusions between studies, populations, locations or environments.

For clarity, we will represent the  $S$  variables by squares, as in Fig. 2, which uses selection diagrams to encode the three examples discussed above. In particular, Fig. 2(a) and 2(b) represent, respectively, two different mechanisms responsible for the observed disparity  $P(z) \neq P^*(z)$ . The first (Fig. 2(a)) dictates transport formula (1) while the second (Fig. 2(b)) calls for direct, unadjusted transport (2).

In the extreme case, we could add selection nodes to all variables, which means that we have no reason to believe that the two populations share any mechanism in common, and this, of course would inhibit any exchange of conclusions between the two. Conversely, absence of a selection node pointing

<sup>2</sup>Alternatively, one can represent the two populations’ distributions by  $P(v|do(x), s)$ , and  $P(v|do(x), s^*)$ , respectively. The results, however, will be the same, since only the location of  $S$  enters the analysis.

<sup>3</sup>Elsewhere, we analyze  $S$  variables representing selection of units into the study pool [BP11]; there, the arrows will be pointing towards  $S$ .

<sup>4</sup>[[Pea95, Pea09, p. 71]] and [Daw02], for example, use conditioning on auxiliary variables to switch between experimental and observational studies. [Daw02] further uses such variables to represent changes in parameters of probability distributions.

to a variable, say  $Z$ , represents an assumption of invariance: the local mechanism that assigns values to  $Z$  is the same in both populations.

### Transportability: Definitions and Examples

Using selection diagrams as the basic representational language, and harnessing the concepts of intervention, *do*-calculus<sup>5</sup> and identifiability [Pea09, p. 77] we give the notion of transportability a formal definition.

#### Definition 1 (Transportability)

Given two populations, denoted  $\Pi$  and  $\Pi^*$ , characterized by probability distributions  $P$  and  $P^*$ , and causal diagrams  $G$  and  $G^*$ , respectively, a causal relation  $R$  is said to be transportable from  $\Pi$  to  $\Pi^*$  if  $R(\Pi)$  is estimable from the set  $I$  of interventions on  $\Pi$ , and  $R(\Pi^*)$  is identified from  $\{P, P^*, I, G, G^*\}$ .

Definition 1 provides a declarative characterization of transportability which, in theory, requires one to demonstrate the non-existence of two competing models, agreeing on  $\{P, P^*, I, G, G^*\}$ , and disagreeing on  $R(\Pi^*)$ . Such demonstrations are extremely cumbersome for reasonably sized models, and we seek therefore procedural criteria which, given the pair  $(G, G^*)$  will decide the transportability of any given relation directly from the structures of  $G$  and  $G^*$ . Such criteria will be developed in the sequel by breaking down a complex relation  $R$  into more elementary relations whose transportability can immediately be recognized. We will formalize the structure of this procedure in Theorem 1, followed by Definitions 2 and 3 below, which will identify two special cases where transportability is immediately recognizable.

**Theorem 1 ([PB11a, PB11b])** Let  $D$  be the selection diagram characterizing  $\Pi$  and  $\Pi^*$ , and  $S$  a set of selection variables in  $D$ . The relation  $R = P(y|do(x), z)$  is transportable from  $\Pi$  to  $\Pi^*$  if and only if the expression  $P(y|do(x), z, s)$  is reducible, using the rules of *do*-calculus (Appendix 1), to an expression in which  $S$  appears only as a conditioning variable in *do*-free terms.

#### Definition 2 (Direct Transportability)

A causal relation  $R$  is said to be directly transportable from  $\Pi$  to  $\Pi^*$ , if  $R(\Pi^*) = R(\Pi)$ .

The equality  $R(\Pi^*) = R(\Pi)$  means that  $R$  retains its validity without adjustment, as in Eq. (2). A graphical test for direct transportability of  $P(y|do(x))$  follows immediately from *do*-calculus (Appendix 1) and reads:  $(S \perp\!\!\!\perp Y|X)_{G_{\overline{X}}}$ ; i.e.,  $X$  blocks all paths from  $S$  to  $Y$  once we remove all arrows pointing to  $X$ . Indeed, such condition would allow us to eliminate  $S$  from the *do*-expression, and write:

$$R(\Pi^*) = P(y|do(x), s) = P(y|do(x)) = R(\Pi)$$

**Example 4** Figure 4(a) represents a simple example of direct transportability. Indeed, since  $S$  merely changes the mechanism by which the value  $X = x$  is selected (sometimes called “treatment assignment mechanism”), it does not change any causal effect of  $X$  [Pea09, pp. 72–73].

#### Definition 3 (Trivial Transportability)

A causal relation  $R$  is said to be trivially transportable from  $\Pi$  to  $\Pi^*$ , if  $R(\Pi^*)$  is identifiable from  $(G^*, P^*)$ .

<sup>5</sup>The three rules of *do*-calculus are illustrated in graphical details in [Pea09, p. 87] (See Appendix 1).

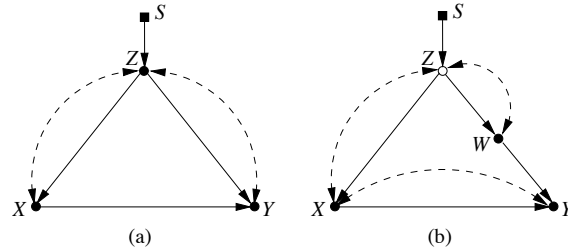


Figure 3: Selection diagrams illustrating  $S$ -admissibility. (a) has no  $S$ -admissible set while in (b),  $W$  is  $S$ -admissible.

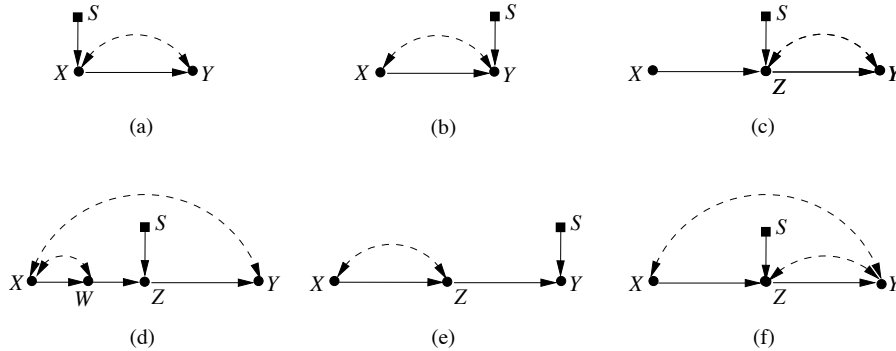


Figure 4: Selection diagrams illustrating transportability. The causal effect  $P(y|do(x))$  is (trivially) transportable in (c) but not in (b) and (f). It is transportable in (a), (d), and (e) (see Corollary 2 and Example 9).

This criterion amounts to ordinary (nonparametric) identifiability of causal relations using graphs [Pea09, p. 77]. It permits us to estimate  $R(\Pi^*)$  directly from passive observations on  $\Pi^*$ , un-aided by causal information from  $\Pi$ .

**Example 5** Let  $R$  be the causal effect  $P(y|do(x))$  and let the selection diagram be  $X \rightarrow Y \leftarrow S$ , then  $R$  is trivially transportable, since  $R(\Pi^*) = P^*(y|x)$ .

**Example 6** Let  $R$  be the causal effect  $P(y|do(x))$  and let the selection diagram of  $\Pi$  and  $\Pi^*$  be  $X \rightarrow Y \leftarrow S$ , with  $X$  and  $Y$  confounded as in Fig. 4(b), then  $R$  is not transportable, because  $P^*(y|do(x)) = P(y|do(x), s)$  cannot be decomposed into  $s$ -free or  $do$ -free expressions using  $do$ -calculus. This is the smallest graph for which the causal effect is non-transportable.

### Transportability of Causal Effects: A Graphical Criterion

We now state and prove two theorems that permit us to decide algorithmically, given a selection diagram, whether a relation is transportable between two populations, and what the transport formula should be.

**Theorem 2** ([PB11a, PB11b]) Let  $D$  be the selection diagram characterizing  $\Pi$  and  $\Pi^*$ , and  $S$  the set of selection variables in  $D$ . The  $z$ -specific causal effect  $P(y|do(x), z)$  is transportable from  $\Pi$  to  $\Pi^*$  if  $Z$   $d$ -separates  $Y$  from  $S$  in the  $X$ -manipulated version of  $D$ , that is,  $Z$  satisfies  $(Y \perp\!\!\!\perp S|Z)_{D_{\overline{X}}}$ .

**Definition 4** ( $S$ -admissibility)

A set  $T$  of variables satisfying  $(Y \perp\!\!\!\perp S|T)$  in  $D_{\overline{X}}$  will be called  $S$ -admissible.

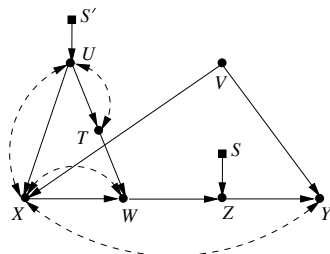


Figure 5: Selection diagram in which the causal effect is shown to be transportable in two iterations of Theorem 3.

**Corollary 1 ([PB11a, PB11b])** *The average causal effect  $P(y|do(x))$  is transportable from  $\Pi$  to  $\Pi^*$  if there exists a set  $Z$  of observed pre-treatment covariates that is  $S$ -admissible. Moreover, the transport formula is given by the weighting of Eq. (1).*

**Example 7** *The causal effect is transportable in Fig. 2(a), since  $Z$  is  $S$ -admissible, and directly transportable in Fig. 2(b) and 4(a), where the empty set is  $S$ -admissible. It is also transportable in Fig. 3(b), where  $W$  is  $S$ -admissible, but not in Fig. 3(a) where no  $S$ -admissible set exists.*

Contrasting the diagrams in Figs. 2(a) and 3(a), we witness again the crucial role of causal knowledge in facilitating transportability. These two diagrams are statistically indistinguishable in both the study and target populations, yet the former is transportable, while the latter is not.

**Corollary 2 ([PB11a, PB11b])** *Any  $S$  variable that is pointing directly into  $X$  as in Fig. 4(a), or that is  $d$ -connected to  $Y$  only through  $X$  can be ignored.*

We now generalize Theorem 2 to cases involving  $X$ -dependent  $Z$  variables, as in Fig. 2(c).

**Theorem 3 ([PB11a, PB11b])** *The causal effect  $P(y|do(x))$  is transportable from  $\Pi$  to  $\Pi^*$  if any one of the following conditions holds*

1.  $P(y|do(x))$  is trivially transportable
2. There exists a set of covariates,  $Z$  (possibly affected by  $X$ ) such that  $Z$  is  $S$ -admissible and for which  $P(z|do(x))$  is transportable
3. There exists a set of covariates,  $W$  that satisfy  $(X \perp\!\!\!\perp Y|W, S)_D$  and for which  $P(w|do(x))$  is transportable.

**Remark.**

The test entailed by Theorem 3 is recursive, since the transportability of one causal effect depends on that of another. However, given that the diagram is finite and feedback-free, the sets  $Z$  and  $W$  needed in conditions 2 and 3 would become closer and closer to  $X$ , and the iterative process will terminate after a finite number of steps. Still, Theorem 3 is not complete, as shown in [PB11a].

**Example 8** *Applying Theorem 3 to Fig. 2(c), we conclude that  $R = P(y|do(x))$  is trivially transportable, for it is identifiable in  $\Pi^*$  through the front-door criterion [Pea09].  $R$  is likewise (trivially) transportable in Fig. 4(c) (by the back-door criterion).  $R$  is not transportable however in Fig. 3(a), where no  $S$ -admissible set exists.*

**Example 9** Fig. 4(d) requires that we invoke both conditions of Theorem 3, iteratively, and yields transport formula (see [PB11a]):

$$(6) \quad P^*(y|do(x)) = \sum_z P(y|do(x), z) \sum_w P(w|do(x)) P^*(z|w)$$

The first two factors on the right are estimable in the experimental population, and the third in the observational side. Surprisingly, the joint effect  $P(y, w, z|do(x))$  need not be estimated in the experiment; a decomposition that results in improved estimation power.

A similar analysis applies to Fig. 4(e).

The model of Fig. 4(f) however does not allow for the transportability of  $P(y|do(x))$  because there is no  $S$ -admissible set in the diagram and condition 3 of Theorem 3 cannot be invoked.

**Example 10** Fig. 5 represents a more challenging selection diagram, which requires several iterations to discern transportability, and yields (see [PB11a]):

$$(7) \quad P^*(y|do(x)) = \sum_z P(y|do(x), z) \sum_w P^*(z|w) \sum_t P(w|do(x), t) P^*(t)$$

The main power of this formula is to guide the learning agent in deciding what measurements need be taken in each population. It asserts, for example, that variables  $U$  and  $V$  need not be measured, that the  $W$ -specific causal effects need not be learned in the experiment and only the conditional probabilities  $P^*(z|w)$  and  $P^*(t)$  need be learned in the target population.

## Conclusions

Given judgmental assessments of how target populations may differ from those under study, the paper offers a formal representational language for making these assessments precise and for deciding whether causal relations in the target population can be inferred from experiments conducted elsewhere. When such inference is possible, the criteria provided by Theorems 1-3 yield transport formulae, namely, principled ways of recalibrating the learned relations so as to account for differences in the populations. These formulae enable the learner to select the essential measurements in both the experimental and observational studies, and thus minimize measurement costs and sample variability.

Extending these results to observational studies, [PB11a] showed that there is also benefit in transporting statistical findings from one population to another in that it enables learners to avoid repeated measurements that are not absolutely necessary for reconstructing the relation transferred. Procedures for deciding whether such reconstruction is feasible when certain re-measurements are forbidden or too costly were shown capable of substantial savings in sample size or increase in estimation power.

A second extension of transportability analysis led to a causally principled definition of “surrogate endpoint,” namely, a variable  $Z$  such that knowing the effect of treatment on  $Z$  allows predictions of the effect of  $X$  on the more clinically relevant outcome  $Y$  [JG09]. [PB11a] have argued that a surrogate should serve not merely as a good predictor of outcomes, but also as a *robust* predictor of effects in the face of changing external conditions. Therefore, any formal definition of surrogacy must make change in conditions an integral part of the definition.<sup>6</sup> Accordingly,  $Z$  is defined as a surrogate of  $Y$  if observation of  $Z$  in  $\Pi^*$  enables the effect of  $X$  on  $Y$  to be transported from  $\Pi$  to  $\Pi^*$  without

<sup>6</sup>Traditional definitions of surrogacy [Pre89] as well as those based on “principal strata” [FR02] lack this feature and are, therefore, inadequate [Pea11].



re-measurement of  $Y$  and regardless of the mechanism responsible for variations in  $Z$ . The procedure developed from this transportability-based definition allows for the identification of valid surrogates in a complex set of causal relations.

Our analysis is based on the assumption that the investigator is in possession of sufficient knowledge to determine, at least qualitatively, where two populations may differ. In practice, such knowledge may only be partially available and, as is the case in every mathematical exercise, the benefit of the analysis lies primarily in understanding what knowledge is needed for the task to succeed and how sensitive conclusions are to knowledge that we do not possess.

## Appendix 1

The *do*-calculus [Pea95] consists of three rules that permit us to transform expressions involving *do*-operators into other expressions of this type, whenever certain conditions hold in the causal diagram  $G$ . (See footnote 1 for semantics.)

We consider a DAG  $G$  representing the data-generating model, in which each child-parent family represents a deterministic function  $x_i = f_i(pa_i, \epsilon_i)$ , where  $pa_i$  are the parents of  $X_i$  in  $G$ ; and  $\epsilon_i, i = 1, \dots, n$  are arbitrarily distributed random disturbances, representing factors that the investigator chooses not to include in the analysis.

Let  $X, Y$ , and  $Z$  be arbitrary disjoint sets of nodes in a causal DAG  $G$ . An expression of the type  $E = P(y|do(x), z)$  is said to be compatible with  $G$  if the interventional distribution described by  $E$  can be generated by parameterizing the graph with a set of functions  $f_i$  and a set of distributions of the random disturbances  $\epsilon_i, i = 1, \dots, n$ .

We denote by  $G_{\overline{X}}$  the graph obtained by deleting from  $G$  all arrows pointing to nodes in  $X$ . Likewise, we denote by  $G_{\underline{X}}$  the graph obtained by deleting from  $G$  all arrows emerging from nodes in  $X$ . To represent the deletion of both incoming and outgoing arrows, we use the notation  $G_{\overline{X}\underline{Z}}$ .

The following three rules are valid for every interventional distribution compatible with  $G$ .

**Rule 1** (Insertion/deletion of observations):

$$P(y|do(x), z, w) = P(y|do(x), w) \text{ if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}}}$$

**Rule 2** (Action/observation exchange):

$$P(y|do(x), do(z), w) = P(y|do(x), z, w) \text{ if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\underline{Z}}}$$

**Rule 3** (Insertion/deletion of actions):

$$P(y|do(x), do(z), w) = P(y|do(x), w) \text{ if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\underline{Z}(w)}},$$

where  $Z(W)$  is the set of  $Z$ -nodes that are not ancestors of any  $W$ -node in  $G_{\overline{X}}$ . The *do*-calculus was proven to be complete [SP06].

## References

- [BP11] E. Bareinboim and J. Pearl, *Controlling selection bias in causal inference*, Tech. Report R-381, <[http://ftp.cs.ucla.edu/pub/stat\\_ser/r381.pdf](http://ftp.cs.ucla.edu/pub/stat_ser/r381.pdf)>, Department of Computer Science, University of California, Los Angeles, 2011.
- [Daw02] A.P. Dawid, *Influence diagrams for causal modelling and inference*, International Statistical Review **70** (2002), 161–189.

- [FR02] C.E. Frangakis and D.B. Rubin, *Principal stratification in causal inference*, *Biometrics* **1** (2002), no. 58, 21–29.
- [JG09] M.M. Joffe and T. Green, *Related causal frameworks for surrogate outcomes*, *Biometrics* **65** (2009), 530–538.
- [Man07] C. Manski, *Identification for prediction and decision*, Harvard University Press, Cambridge, Massachusetts, 2007.
- [PB11a] J. Pearl and E. Bareinboim, *Transportability across studies: A formal approach*, Tech. Report R-372, <[http://ftp.cs.ucla.edu/pub/stat\\_ser/r372.pdf](http://ftp.cs.ucla.edu/pub/stat_ser/r372.pdf)>, Department of Computer Science, University of California, Los Angeles, 2011.
- [PB11b] ———, *Transportability of causal and statistical relations: A formal approach*, Proceedings of the Twenty-Fifth National Conference on Artificial Intelligence, AAAI Press, Menlo Park, CA, 2011, pp. –.
- [Pea95] J. Pearl, *Causal diagrams for empirical research*, *Biometrika* **82** (1995), no. 4, 669–710.
- [Pea09] ———, *Causality: Models, reasoning, and inference*, 2nd ed., Cambridge University Press, New York, 2009.
- [Pea11] ———, *Principal stratification - a goal or a tool?*, *The International Journal of Biostatistics* **7** (2011).
- [Pre89] R.L. Prentice, *Surrogate endpoints in clinical trials: definition and operational criteria*, *Statistics in Medicine* **8** (1989), 431–440.
- [SCC02] W.R. Shadish, T.D. Cook, and D.T. Campbell, *Experimental and quasi-experimental designs for generalized causal inference*, second ed., Houghton-Mifflin, Boston, 2002.
- [SGS00] P. Spirtes, C.N. Glymour, and R. Scheines, *Causation, prediction, and search*, 2nd ed., MIT Press, Cambridge, MA, 2000.
- [SP06] I. Shpitser and J Pearl, *Identification of conditional interventional distributions*, Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence (R. Dechter and T.S. Richardson, eds.), AUAI Press, Corvallis, OR, 2006, pp. 437–444.