Forthcoming, Proceedings of the National Academy of Sciences.

TECHNICAL REPORT
R-450
December 2015

Causal inference from big data: Theoretical foundations and the data-fusion problem

Elias Bareinboim and Judea Pearl *

*University of California, Los Angeles

Written for the Proceedings of the National Academy of Sciences

We review concepts, principles, and tools that unify current approaches to causal analysis, and attend to new challenges presented by big data. In particular, we address the problem of data-fusion – piecing together multiple datasets collected under heterogeneous conditions (i.e., different populations, regimes, and sampling methods) so as to obtain valid answers to queries of interest. The availability of multiple heterogeneous datasets presents new opportunities to big data analysts, since the knowledge that can be acquired from combined data would not be possible from any individual source alone. However, the biases that emerge in heterogeneous environments require new analytical tools. Some of these biases, including confounding, sampling selection, and cross-population biases, have been addressed in isolation, largely in restricted parametric models. We here present a general, non-parametric framework for handling these biases and, ultimately, a theoretical solution to the problem of data-fusion in causal inference tasks.

causal inference | counterfactuals | big data | confounding | external validity | meta-analysis | heterogeneity | selection bias | data integration

Introduction – Causal Inference and Big Data

The exponential growth of electronically accessible information has led some to conjecture that data alone can replace scientific knowledge in practical decision making. In this paper, we argue that the scientific paradigm that has been successful in the natural and bio-medical sciences would still be necessary for big data applications, albeit augmented with new challenges: to go beyond predictions and, utilizing information from multiple sources, provide users with reasoned recommendations for actions and policies. Clearly, the utilization of multiple sources is only realizable in a big data environment. The feasibility of meeting this challenge will be described in the context of the Structural Causal Model (SCM) framework [1, 2], which provides an easy introduction to causal inference in general.

Encoded as non-parametric structural equations, these models have led to a fruitful symbiosis between graphs and counterfactuals and has unified the potential outcome framework of Neyman, Rubin, and Robins with the econometric tradition of Haavelmo, Marschak, and Heckman. In this symbiosis, counterfactuals (or potential outcomes) emerge as natural byproducts of structural equations and serve to formally articulate research questions of interest. Graphical models, on the other hand, are used to encode scientific assumptions in a qualitative (i.e., non-parametric) and transparent language as well as to derive the logical ramifications of these assumptions, in particular, their testable implications and how they shape behavior under interventions.

One unique feature of the SCM framework, essential in big data applications, is the ability to encode mathematically the method by which data are acquired, often referred to generically as the "design." This sensibility to design, which we can label proverbially as *not all data are created equal*, is illustrated schematically through a series of scenarios depicted in Fig. 1. Each design (shown at the bottom of the figure) represents a triplet specifying the population, the regime (observational versus experimental), and the sampling method by which each dataset is generated. This formal encoding will allow us to delineate the inferences that one can draw from each design to answer the query of interest (shown at the top).

Consider the task of predicting the distribution of outcomes Y after intervening on a variable X, written $Q = P(Y = y|do(X = x))$. Assume that the information available to us comes from an observational study, in which X, Y, Z, and W are measured, and samples are selected at random. We ask for conditions under which the query Q can be inferred from the information available, which takes the form: $P(y, x, z, w)$, where Z and W are sets of observed covariates. This represents the standard task of *policy evaluation*, where controlling for confounding bias is the major issue (Task 1, Fig. 1).

Consider now Task 2 in Fig. 1 in which the goal is again to estimate the effect of the intervention $do(X = x)$ but the data available to the investigator were collected in an experimental study in which variable Z, more accessible to manipulation than X, is randomized. (*Instrumental variables* [3] are special cases of this task.) The general question in this scenario is under what conditions can randomization of variable Z be used to infer how the population would react to interventions over X. Formally, our problem is to infer $P(Y = y|do(X = x))$ from $P(y, x, w|do(Z = z))$. A non-parametric solution to these two problems will be presented in the respective *policy evaluation* section.

In each of the two previous tasks we assumed that a perfect random sample from the underlying population was drawn, which may not always be realizable. Task 3 in Fig. 1 represents a randomized clinical trial conducted on a non-representative sample of the population. Here, the information available takes the syntactic form $P(y, z, w|do(X = x), S = 1)$, and possibly $P(y, x, z, w|S = 1)$, where S is a sample selection indicator, with $S = 1$ indicating inclusion in the sample. The challenge is to estimate the effect of interest from this, far from ideal sampling condition. Formally, we ask when the target quantity $P(y|do(X = x))$ is derivable from the available information (i.e., sampling-biased distributions). The section of sample selection bias will present a solution to this problem.

Finally, the previous examples assumed that the population from which data were collected is the same as the one for which inference was intended. This is often not the case (Task 4 in Fig. 1). For example, biological experiments often use animals as substitutes for human subjects. Or, in a less obvious example, data may be available from an experimen-

Reserved for Publication Footnotes

tal study that took place several years ago, and the current population has changed in a set S of (possibly unmeasured) attributes. Our task then is to infer the causal effect at the target population, $P(y|do(X = x), S = s)$ from the information available, which now takes the form: $P(y, z, w|do(X = x))$ and $P(y, x, z, w|S = s)$. The second expression represents information obtainable from non-experimental studies on the current population, where $S = s$.

The problems represented in these archetypal examples are known as confounding bias (Tasks 1,2), sample selection bias (Task 3), and transportability bias (Task 4). The information available in each of these tasks is characterized by a different syntactic form, representing a different "design" and, naturally, each of these designs should lead to different inferences. What we shall see in subsequent sections of this paper is that the strategy of going from design to a query is the same across tasks; it follows simple rules of inference and decides, using syntactic manipulations, whether the type of data available is sufficient for the task, and, if so, how. [1]

Empowered by this strategy, the central goal of this paper will be to explicate the conditions under which causal effects can be estimated non-parametrically from multiple heterogeneous datasets. These conditions constitute the formal basis for many big data inferences since, in practice, data are never collected under idealized conditions, ready for use. The remaining of the paper is organized as follows. We start by defining structural causal models (SCMs) and stating the two fundamental laws of causal inference. We then consider respectively the problem of policy evaluation in observational and experimental settings, sampling selection bias, and data-fusion from multiple populations.

The Structural Causal Model (SCM)

At the center of the structural theory of causation lies a "structural model," M, consisting of two sets of variables, U and V, and a set F of functions that determine or simulate how values are assigned to each variable $V_i \in V$. Thus, for example, the equation

$$v_i = f_i(v, u)$$

describes a physical process by which variable V_i is *assigned* the value $v_i = f_i(v, u)$ in response to the current values, v and u, of all variables in V and U. Formally, the triplet $< U, V, F >$ defines a SCM, and the diagram that captures the relationships among the variables is called the *causal graph* G (of M). [2] The variables in U are considered "exogenous," namely, background conditions for which no explanatory mechanism is encoded in model M. Every instantiation $U = u$ of the exogenous variables uniquely determines the values of all variables in V and, hence, if we assign a probability $P(u)$ to U, it induces a probability function $P(v)$ on V. The vector $U = u$ can also be interpreted as an experimental "unit" which can stand for an individual subject, agricultural lot, or time of day. Conceptually, a unit $u = u$ should be thought of as the sum total of all relevant factors that govern the behavior of an individual or experimental circumstances.

The basic counterfactual entity in structural models is the sentence: "Y would be y had X been x in unit (or situation) $U = u$," denoted $Y_x(u) = y$. Letting M_x stand for a modified version of M, with the equation(s) of set X replaced by $X = x$, the formal definition of the counterfactual $Y_x(u)$ reads

$$Y_x(u) \triangleq Y_{M_x}(u). \qquad [1]$$

In words, the counterfactual $Y_x(u)$ in model M is defined as the solution for Y in the "modified" submodel M_x. [5] and [6] have given a complete axiomatization of structural counterfactuals, embracing both recursive and non-recursive models (see also [1, Chapter 7]).[3] Remarkably, the axioms that characterize counterfactuals in SCM coincide with those that govern potential outcomes in Rubin's causal model [8] where $Y_x(u)$ stands for the potential outcome of unit u, had u been assigned treatment $X = x$. This axiomatic agreement implies a logical equivalence of the two systems, namely any valid inference in one is also valid in the other. Their differences lie in the way assumptions are articulated and the ability of the researcher to scrutinize those assumptions and to infer their implications [2].

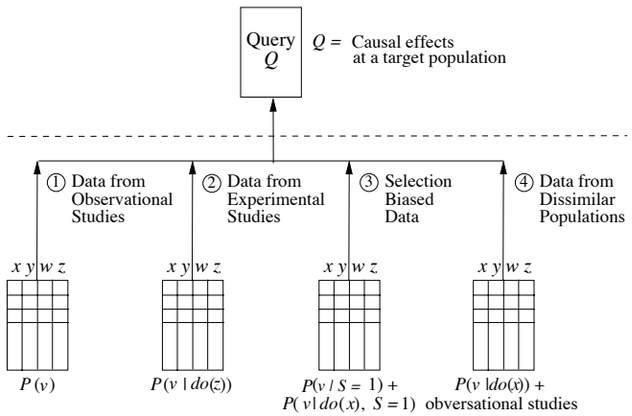

Fig. 1. Prototypical counterfactual inferences where the goal is, for example, to estimate the experimental distribution in a target population (shown at the top). Let $V = \{X, Y, Z, W\}$. There are different designs (bottom) showing that data come from non-idealized conditions, specifically: (1) from the same population under an observational regime, $P(v)$; (2) from the same population under an experimental regime when Z is randomized, $P(v|do(z))$; (3) from the same population under sampling selection bias, $P(v|S = 1)$ or $P(v|do(x), S = 1)$; (4) from a different population that is submitted to an experimental regime when X is randomized, $P(v|do(x), S = s)$, and observational studies in the target population.

Eq. (1) implies that the distribution $P(u)$ induces a well defined probability on the counterfactual event $Y_x = y$, written $P(Y_x = y)$, which is equal to the probability that a random unit u would satisfy the equation $Y_x(u) = y$. By the same reasoning, the model $< U, V, F, P(u) >$ assigns a probability to every counterfactual or combination of counterfactuals defined on the variables in V.

The two principles of causal inference. Before describing how the structural theory applies to big data inferences, it will be useful to summarize its implications in the form of two "principles," from which all other results follow.

Principle 1: "The law of structural counterfactuals."
Principle 2: "The law of structural independences."

The first principle described in Eq. (1) constitutes the semantics of counterfactuals, and instructs us how to compute counterfactuals and probabilities of counterfactuals from a

[1] An important issue that is not discussed in this paper is "measurement bias" [4].

[2] Following [1], we denote variables by capital letters and their realized values by small letters. We used family relations (e.g., children, parents, descendants) to denote the corresponding graphical relations, and we focus on directed acyclic graphs (DAG's), though many features of SCM apply to nonrecursive systems as well.

[3] The structural definition of counterfactual given in Eq. (1) was first introduced in [7].

[4] By a path we mean a consecutive edges in the graph regardless of direction. Dependencies among the U variables are represented by double-arrowed arcs, as in Fig. 3 below.

structural model. how the observed data influence the causal parameters that we aim to estimate.

Principle 2 defines how features of the model affect the data. Remarkably, regardless of the functional form of the equations in the model (F) and regardless of the distribution of the exogenous variables (U), if the model is recursive, the distribution $P(v)$ of the endogenous variables must obey certain conditional independence relations, stated roughly as follows: whenever sets X and Y are *separated* by a set Z in the graph, X is independent of Y given Z in the probability distribution. This "separation" condition, called d-separation [1, pp. 16–18], constitutes the link between the causal assumptions encoded in the graph (in the form of missing arrows) and the observed data.

Definition 1. *(d-separation)*
A set Z of nodes is said to ***block*** *a path p if either*

1. *p contains at least one arrow-emitting node that is in Z (i.e., $-Z\rightarrow$), or*
2. *p contains at least one collision node that is outside Z (i.e., $\rightarrow Z\leftarrow$) and has no descendant in Z.*

If Z blocks all *paths from set X to set Y, it is said to "d-separate X and Y," and then, variables X and Y are independent given Z, written $X \perp\!\!\!\perp Y|Z$.* [4]

D-separation implies conditional independencies for every distribution $P(v)$ that can be generated by assigning functions (F) to the variables in the graph. To illustrate, the diagram in Fig. 2(a) implies $Z_1 \perp\!\!\!\perp Y|\{X, Z_3, W_2\}$, because the conditioning set $Z = \{X, Z_3, W_2\}$ blocks all paths beween Z_1 and Y. The set $Z = \{X, Z_3, W_3\}$ however leaves the path $(Z_1 \rightarrow Z_3 \leftarrow Z_2 \rightarrow W_2 \rightarrow Y)$ unblocked (by virtue of the converging arrows (collider) at Z_3) and, so, the independence $Z_1 \perp\!\!\!\perp Y|(X, Z_3, W_3)$ is not implied by the diagram.

In the sequel, we show how these independencies help us evaluate the effect of interventions and overcome the problem of confounding bias. [5] Clearly, any attempt to predict the effects of interventions from non-experimental data must rely on causal assumptions. One of the most attractive features of the SCM framework is that those assumptions are all encoded parsimoniously in the diagram, thus, unlike "ignorability"-type assumptions [9, 10], they can be meaningfully scrutinized for scientific plausibility or be submitted to statistical tests.

Policy Evaluation and the Problem of Confounding

A central question in causal analysis is that of predicting the results of interventions, such as those resulting from medical treatments or social programs, which we denote by the symbol $do(x)$ and define using the counterfactual Y_x as[6]

$$P(y|do(x)) \triangleq P(Y_x = y) \qquad [2]$$

Figure 2(b) illustrates the submodel M_x created by the atomic intervention $do(x)$; it sets the value of X to x and thus removes the influence (arrow) of $\{W_1, Z_3\}$ on X. The set of incoming arrows towards X is sometimes called the *assignment mechanism*, and may also represent how the decision $X = x$ is made by an individual in response to natural predilections (i.e., $\{W_1, Z_3\}$), as opposed to an externally imposed assignment in a controlled experiment. [7] Furthermore, we can similarly define the result of *stratum-specific interventions* by

$$P(y|do(x), z) \triangleq P(y, z|do(x))/P(z|do(x)) = P(Y_x = y|Z_x = z) \qquad [3]$$

$P(y|do(x), z)$ captures the z-specific effect of X on Y, that is, Y's response to setting X to x among those units only for which Z responds with z. (For pre-treatment Z (e.g., sex, age, or ethnicity), those units would remain invariant to X (i.e., $Z_x = Z$).)

Recalling that any counterfactual quantity can be computed from a fully specified model $< U, V, F, P(u) >$, it follows that the interventional distributions defined in Eq. (2) and Eq.(3) can be computed directly from such a model. In practice, however, only a partially specified model is available, in the form of a graph G, and the problem arises whether the data collected can make up for our ignorance of the functions F and the probabilities $P(u)$. This is the problem of *identification*, which asks whether the interventional distribution, $P(y|do(x))$, can be estimated from the available data and the assumptions embodied in the causal graph.

In parametric settings, the question of identification amounts to asking whether some model parameter, θ, has a unique solution in terms of the parameters of P. In the nonparametric formulation, quantities such as $Q = P(y|do(x))$ should have unique solutions. The following definition captures this requirement:

Definition 2. *(Identifiability) [1, p. 77]*
A causal query Q is identifiable from distribution $P(v)$ compatible with a causal graph G, if for any two (fully specified) models M_1 and M_2 that satisfy the assumptions in G, we have

$$P_1(v) = P_2(v) \Rightarrow Q(M_1) = Q(M_2) \qquad [4]$$

In words, equality in the probabilities $P_1(v)$ and $P_2(v)$ induced by models M_1 and M_2, respectively, entails equality in the answers that these two models give to query Q. When this happens, Q depends on $P(v)$ and G only, and can therefore be expressible in terms of the parameters of $P(v)$ (i.e., regardless of the true underlying mechanisms F and randomness $P(u)$).

For queries in the form of a do-expression, for example $Q = P(y|do(x), z)$, identifiability can be decided systematically using an algebraic procedure known as the *do-calculus* [12], to be discussed next. It consists of three inference rules that permit us to manipulate interventional and observational distributions whenever certain separation conditions hold in the causal diagram G.

The rules of do-calculus. Let X, Y, Z, and W be arbitrary disjoint sets of nodes in a causal DAG G. We denote by $G_{\overline{X}}$ the graph obtained by deleting from G all arrows pointing to nodes in X (e.g., Fig. 2(b)). Likewise, we denote by $G_{\underline{X}}$ the graph obtained by deleting from G all arrows emerging from nodes in X (e.g., Fig. 2(c)). 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):

$$P(y|do(x), z, w) = P(y|do(x), w) \text{ if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}}} \qquad [5]$$

Rule 2 (Action/observation exchange):

$$P(y|do(x), do(z), w) = P(y|do(x), z, w) \text{ if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\underline{Z}}} \qquad [6]$$

Rule 3 (Insertion/deletion of actions):

$$P(y|do(x), do(z), w) = P(y|do(x), w) \text{ if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\overline{Z^*}}}, \qquad [7]$$

[5] These and other constraints implied by Principle 1 also facilitate model testing and learning [1].
[6] Alternative definitions of $do(x)$ invoking population averages only are given in [1, p. 24] and [11], which are also compatible with the results presented in this paper.
[7] This primitive operator can be used for handling stratum-specific interventions [1, Ch. 4] as well as non-compliance [1, Ch. 8] and compound interventions [1, Ch. 11.4].
[8] Such derivations are illustrated in graphical details in [1, p. 87] and in the next section.

where Z^* is the set of Z-nodes that are not ancestors of any W-node in $G_{\overline{X}}$.

To establish identifiability of a causal query Q, one needs to repeatedly apply the rules of *do*-calculus to Q, until an expression is obtained which no longer contains a *do*-operator[8]; this renders Q consistently "estimable" from nonexperimental data (henceforth, "estimable," or "unbiased," for short). The *do*-calculus was proven to be complete for queries in the form $Q = P(y|do(x), z)$ [13, 14], which means that if Q cannot be reduced to probabilities of observables by repeated application of these three rules, Q is not identifiable. We show next concrete examples of the application of the do-calculus.

Covariate selection: the back-door criterion

Consider an observational study where we wish to find the effect of treatment (X) on outcome (Y), and assume that the factors deemed relevant to the problem are structured as in Fig. 2(a); some are affecting the outcome, some are affecting the treatment, and some are affecting both treatment and response. Some of these factors may be unmeasurable, such as genetic trait or lifestyle, while others are measurable, such as gender, age, and salary level. Our problem is to select a subset of these factors for measurement and adjustment such that if we compare treated vs. untreated subjects having the same values of the selected factors, we get the correct treatment effect in that subpopulation of subjects. Such a set of factors is called a "sufficient set," "admissible set" or a set "appropriate for adjustment" (see [15, 2]). The following criterion, named "back-door" [16], provides a graphical method of selecting such a set of factors for adjustment.

Definition 3. *(admissible sets—the back-door criterion) A set Z is admissible (or "sufficient") for estimating the causal effect of X on Y if two conditions hold:*

1. *No element of Z is a descendant of X.*
2. *The elements of Z "block" all "back-door" paths from X to Y – i.e., all paths that end with an arrow pointing to X.*

Based on this criterion we see, for example that, in Fig. 2, the sets $\{Z_1, Z_2, Z_3\}$, $\{Z_1, Z_3\}$, $\{W_1, Z_3\}$, and $\{W_2, Z_3\}$ are each sufficient for adjustment, because each blocks all back-door paths between X and Y. The set $\{Z_3\}$, however, is not sufficient for adjustment because it does not block the path $X \leftarrow W_1 \leftarrow Z_1 \rightarrow Z_3 \leftarrow Z_2 \rightarrow W_2 \rightarrow Y$.

The intuition behind the back-door criterion is simple. The back-door paths in the diagram carry the "spurious associations" from X to Y, while the paths directed along the arrows from X to Y carry causative associations. If we remove the latter paths as shown in Fig. 2(c), checking whether X and Y are separated by Z amounts to verifying that Z blocks all spurious paths. This ensures that the measured association between X and Y is purely causal, namely, it correctly represents the causal effect of X on Y. Conditions for relaxing and generalizing Def. 3 are given in [1, p. 338][17, 18, 19][9].

The implication of finding a sufficient set, Z, is that stratifying on Z is guaranteed to remove all confounding bias relative to the causal effect of X on Y. In other words, it renders the effect of X on Y identifiable, via the *adjustment formula*[10]

$$P(Y = y|do(X = x)) = \sum_z P(y|x, Z = z)P(Z = z) \qquad [8]$$

Since all factors on the right-hand side of the equation are estimable (e.g., by regression) from non-experimental data, the causal effect can likewise be estimated from such data without bias. Eq. (8) differs from the conditional distribution of

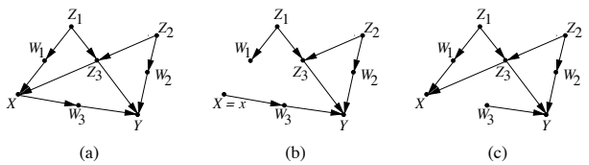

Fig. 2. (a) Graphical model illustrating d-separation and the back-door criterion. U terms are not shown explicitly. (b) Illustrating the intervention $do(X = x)$ with arrows towards X cut. (c) Illustrating the spurious paths, which pop out when we cut the outgoing edges from X, and need to be blocked if one wants to use adjustment.

Y given X, which can be written as

$$P(Y = y|X = x) = \sum_s P(y|x, Z = z)P(Z = z|x); \qquad [9]$$

the difference between these two distributions defines confounding bias.

Moreover, the back-door criterion implies an independence known as "conditional ignorability" [9], $X \perp\!\!\!\perp Y_x|Z$, and provides therefore the scientific basis for most inferences in the potential outcome framework. For example, the set of covariates that enter "propensity score" analysis [9] must constitute a back-door sufficient set, else confounding bias will arise.

The back-door criterion can be applied systematically to diagrams of any size and shape, thus freeing analysts from judging whether "X is conditionally ignorable given Z," a formidable mental task required in the potential-outcome framework. The criterion also enables the analyst to search for an optimal set of covariates—namely, a set, Z, that minimizes measurement cost or sampling variability [20, 21].

Despite its importance, adjustment for covariates (or for propensity scores) is only one tool available for estimating the effects of interventions in observational studies; more refined strategies exist which go beyond adjustment. For instance, assume that only variables $\{X, Y, W_3\}$ are observed in Fig. 2(a), so only the observational distribution $P(x, y, w_3)$ may be estimated from the samples. In this case, conditional ignorability does not hold, but an alternative strategy known as the *front-door criterion* [1, pp. 83] can be employed to yield identification. Specifically, the calculus permits rewriting the experimental distribution as:

$$P(Y = y|do(X = x)) = \sum_{w_3} P(w_3|x) \sum_{x'} P(y|x', w_3)P(x'), \qquad [10]$$

which is almost always different than Eq. (8).

Finally, in case W_3 is also not observed, only the observational distribution $P(x, y)$ can be estimated from the samples, and the calculus will discover that no reduction is feasible, which implies (by virtue of its completeness) that the target quantity is not identifiable (without further assumptions).

Identification through Auxiliary Experiments

In many applications, it is not uncommon that the quantity $Q = P(y|do(x))$ is not identifiable from the observational data alone. Imagine a researcher interested in assessing the effect (Q) of cholesterol levels (X) on heart disease (Y), assuming data about subjects diet (Z) is also collected (Fig. 3(a)). In practice, it is infeasible to control subjects cholesterol level by

[9] In particular, the criterion devised by [18] simply adds to Condition 2 of Definition 3 the requirement that X and its nondescendants (in Z) separate its descendants (in Z) from Y.

[10] Summations should be replaced by integration when applied to continuous variables.

[11] The scope of this paper is circumscribed to non-parametric analysis. Additional results can be derived whenever the researcher is willing to make parametric or functional restrictions [23, 3].

intervention, so $P(y|do(x))$ cannot be obtained from a randomized trial. Assuming, however, that an experiment can be conducted in which Z is randomized, would Q be computable given this additional piece of experimental information?

This question represents what we called Task 2 in Fig. 1, and leads to a natural extension of the identifiability problem (def. 2) in which, in addition to the standard input ($P(v)$ and G), an interventional distribution $P(v|do(z))$ is also available to help establishing $Q = P(y|do(x))$. This task can be seen as the non-parametric version of identification with instrumental variables and was named z-identification in [22]. [11]

Using the do-calculus and the assumptions embedded in Fig. 3(a), it can readily be shown that the target query Q can be transformed to read:

$$P(Y = y|do(X = x)) = P(y, x|do(z))/P(x|do(z)), \qquad [11]$$

for any level $Z = z$. Since all do-terms in Eq. (11) apply only to Z, Q is estimable from the available data. In general, it can be shown [22] that z-identifiability is feasible if and only if X intercepts all directed paths from Z to Y and $P(y|do(x))$ is identifiable in $G_{\overline{Z}}$. Note that, due to the non-parametric nature of the problem these conditions are stronger that those needed for local average treatment effect (LATE) [3] or other functionally-restricted IV applications.

Fig. 3(b) demonstrates this graphical criterion. Here Z_1 can serve as auxiliary variable because, (1) there is no directed path from Z_1 to Y in $G_{\overline{X}}$, and, (2) Z_2 is a sufficient set in $G_{\overline{Z_1}}$. The resulting expression for Q becomes:

$$P(Y = y|do(X = x)) = \sum_{z_1} P(y|x, do(z_1), z_2)P(z_2|x, z_1) \qquad [12]$$

The first factor is estimable from the experimental dataset and the second factor from the observational dataset (e.g., by regression-based methods).

Fig. 3(c) and (d) demonstrate negative examples in which Q is not estimable even when both distributions (observational and experimental) are available; each model violates the necessary conditions stated above.

Summary Result 1. (Identification in Policy Evaluation) *The analysis of policy evaluation problems has reached a fairly satisfactory state of maturity. We now possess a complete solution to the problem of identification whenever assumptions are represented in DAG form. This entails:*

- *graphical and algorithmic criteria for deciding identifiability of policy questions,*
- *automated procedures for extracting each and every identifiable estimand, and*
- *extensions to models invoking sequential dynamic decisions with unmeasured confounders.*

These results were developed in several stages over the past 20 years [16, 12, 24, 14, 22].

Sample Selection Bias

In this section, we consider the bias associated with the data-gathering process, as opposed to confounding bias that is associated with the treatment assignment mechanism. Sample selection bias (or selection bias for short) is induced by preferential selection of units for data analysis, usually governed by unknown factors including treatment, outcome, and their consequences, and represents a major obstacle to valid statistical and causal inferences. For instance, in a typical study of the effect of training program on earnings, subjects achieving higher incomes tend to report their earnings more frequently than those who earn less, resulting in biased inferences.

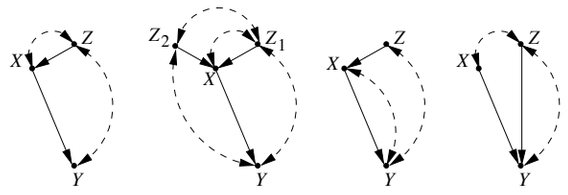

Fig. 3. Graphical models illustrating identification of $Q = P(y|do(x))$ through the use of experiments over an auxiliary variable Z. Identifiability follows from $P(x, y|do(Z = z))$ in (a), and it also requires $P(v)$ in (b). Identifiability of Q fails in (c) because Q is not identifiability in $G_{\overline{Z}}$, and is also not possible in (d) because there is a directed path not blockable by X from Z to Y.

Selection bias challenges the validity of inferences in several tasks in Artificial Intelligence [25, 26] and Statistics [27, 28] as well as in the empirical sciences (e.g., Genetics [29, 30], Economics [31, 32], and Epidemiology [33, 34]).

To illustrate the nature of preferential selection, consider the data-generating model in Fig. 4(a) in which X represents a treatment, Y represents an outcome, and S is a special (indicator) variable representing entry into the data pool – $S = 1$ means that the unit is in the sample, $S = 0$ otherwise. If our goal is, for example, to compute the population-level experimental distribution $Q = P(y|do(x))$, and the samples available are collected under preferential selection, only $P(y, x|S = 1)$ is accessible for use. Under what conditions can Q be recovered from data available under selection bias?

In the model G in Fig. 4(b) the selection process is treatment-dependent (i.e., $X \to S$), and the selection mechanism S is separated from Y by X, hence, $P(y|x) = P(y|x, S = 1)$. Moreover, given that X and Y are unconfounded, we can rewrite the l.h.s. as $P(y|x) = P(y|do(x))$, it follows that the experimental distribution is recoverable and given by $P(y|do(x)) = P(y|x, S = 1)$ [35, 36]. On the other hand, if the selection process is also outcome-dependent (Fig. 4(a)), S is not separable from Y by X in G, and Q is not recoverable by any method (without stronger assumptions) [37].

In practical settings, however, the data-gathering process may be embedded in more intricate scenarios as shown in Fig. 4(c-f), where covariates such as age, sex, socio-economic status also affect the sampling probabilities. In the model in Fig. 4(c), for example, W_1 (sex) is a driver of the treatment while also affecting the sampling process. In this case, both confounding and selection biases need to be controlled for. We can see based on Def. 3 that $\{W_1, W_2\}$, $\{W_1, W_2, Z\}$, $\{W_1, Z\}$, $\{W_2, Z\}$, and $\{Z\}$ are all back-door admissible sets, so proper for controlling confounding bias. However, only the set $\{Z\}$ is appropriate for controlling for selection bias. The reason is that when using the adjusting formula (Eq. (8)) with any set, say T, the prior distribution $P(t)$ also needs to be estimable, which is clearly not feasible for sets different than $\{Z\}$ (the only set independent of S). The proper adjustment in this case would be written as $P(y|do(x)) = \sum_z P(y|x, z, S = 1)P(z|S = 1)$, where both factors are estimable from the biased dataset.

If we apply the same rationale to Fig. 4(d) and search for a set Z that is both admissible for adjustment and also available from the biased dataset, we will fail. In a big data reality, however, additional datasets with measurements at the population level (over subsets of the variables) may be available to help computing these effects. For instance, $P(age, sex, race)$ is usually estimable from census data without selection bias.

Definition 4 (below) provides a simple extension of the back-door condition which allows us to control both selection and confounding biases by an adjustment formula.

Conditions (i-ii) assure that Z is back-door admissible, condition (iii) acts to separate the sampling mechanism S from Y, and condition (iv) guarantees that Z is measured in both population level data and biased data.

Definition 4. (Selection-backdoor criterion [37]) *Let a set Z of variables be partitioned into $Z^+ \cup Z^-$ such that Z^+ contains all non-descendants of X and Z^- the descendants of X, and let G_s stand for the graph which includes the sampling mechanism S. Z is said to satisfy the selection backdoor criterion (s-backdoor, for short) if it satisfies the following conditions:*

(i) Z^+ blocks all back door paths from X to Y in G_s,
(ii) X and Z^+ block all paths between Z^- and Y in G_s, namely, $(Z^- \perp\!\!\!\perp Y | X, Z^+)$,
(iii) X and Z block all paths between S and Y in G_s, namely, $(Y \perp\!\!\!\perp S | X, Z)$, and
(iv) Z and $Z \cup \{X, Y\}$ are measured in the unbiased and biased studies, respectively.

Theorem 1. *If Z is s-backdoor admissible, then causal effects are identified by:*

$$P(y|do(x)) = \sum_z P(y|x, z, S = 1)P(z), \qquad [13]$$

To illustrate the use of this criterion, note that any one of the sets $\{T_1, Z_3\}, \{Z_1, Z_3\}, \{Z_2, Z_3\}, \{W_2, Z_3\}$ in Fig. 4(d) satisfies conditions (i)-(ii) of Def. 4. However, the first three sets clearly do not satisfy condition (iii), but $\{W_2, Z_3\}$ does (since $Y \perp\!\!\!\perp S | \{W_2, Z_3\}$ in G). If census data are available with measurements of $\{W_2, Z_3\}$ (and biased data over $\{X, Y, W_2, Z_3\}$), condition (iv) will be satisfied, and the experimental distribution $P(y|do(x))$ is estimable through the expression $\sum_{w_2, z_3} P(y|x, w_2, z_3, S = 1)P(w_2, z_3)$.

We note that s-backdoor is a sufficient though not necessary condition for recoverability. In Fig. 4(e), for example, condition (i) is never satisfied. Nevertheless, a do-calculus derivation allows for the estimation of the experimental distribution even without an unbiased dataset [38], leading to the expression $\sum_{w_1}(P(x, y, w_1, w_2, w_3|S = 1)/P(w_2|w_1, S = 1))/\sum_{y, w_1}(P(v|S = 1)/P(w_2|w_1, S = 1))$, for any level $W_2 = w_2$.

We also should note that the odds ratio can be recovered from selection bias even in settings where the risk difference cannot [35, 36, 42].

The generalizabiity of clinical trials. The simple model of Fig. 4(f) illustrates a common pattern that assists in generalizing experimental findings from clinical trials. In such trials, confounding need not be controlled for and the major task is to generalize from non-representative samples ($S = 1$) to the population at large.

This disparity is indeed a major threat to the validity of randomized trials. Since participation cannot be mandated, we cannot guarantee that the study population would be the same as the population of interest. Specifically, the study population may consist of volunteers, who respond to financial and medical incentives offered by pharmaceutical firms or experimental teams, so, the distribution of outcomes in the study may differ substantially from the distribution of outcomes under the policy of interest.

Bearing in mind that we are in a big data context, it is not unreasonable to assume that both $P(y, z|do(x), S = 1)$ and $P(x, z, y)$ are available, and the following derivation shows how the target query in the model of Fig. 4(f) can be transformed to match these two datasets:

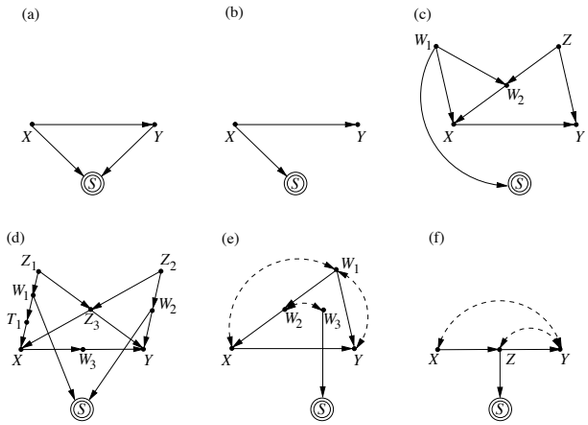

Fig. 4. Canonical models where selection is treatment-dependent in (a,b) and also outcome-dependent in (a). More complex models in which $\{W_1, W_2\}$ and $\{Z\}$ are sufficient for adjustment, but only the latter is adequate for recovering from selection bias (c). There is no sufficient set for adjustment without external data in (d,e,f). (d) Example of s-backdoor admissible set. (e,f) Structures with no s-admissible sets that require more involved recoverability strategies involving post-treatment variables.

$$\begin{aligned} P(y|do(x)) &= \sum_z P(y|do(x), z)P(z|do(x)) \\ &= \sum_z P(y|do(x), z)P(z|x) \\ &= \sum_z P(y|do(x), z, S = 1)P(z|x) \qquad [14] \end{aligned}$$

The two factors in the final expression are estimable from the available data; the first from the trial's (biased) dataset, and the second from the population level dataset.

This example demonstrates the important role that post-treatment variables (Z) play in facilitating generalizations from clinical trials. Previous analyses [10, 39, 40] have invariably relied on an assumption called "S-ignorability," i.e., $Y_x \perp\!\!\!\perp Z|S$, which states that the potential outcome Y_x is independent of the selection mechanism S in every stratum $Z = z$. When Z satisfies this assumption, generalizabiity can be accomplished by reweighing (or recalibrating) $P(z)$. Recently, however, it was shown that s-ignorability is rarely satisfied by post-treatment variables and, even when it does, reweighting will not give the correct result [41]. [12]

The derivation of Eq. (14) demonstrates that post-treatment variables can nevertheless be leveraged for the task albeit through non-conventional re-weighting formulas, which can be derived systematically by the do-calculus [38].

Summary Result 2. (Recoverability from Selection Bias)

- *The s-backdoor criterion (Def. 4) provides a sufficient condition for simultaneous recovery from both confounding and sampling selection bias.*
- *In clinical trials, causal effects can be recovered from selection bias through systematic derivations in do-calculus, leveraging both pre-treatment and post-treatment variables.*
- *More powerful recoverability methods have been developed for special classes of models [42, 37, 38].*

[12] In general, the language of ignorability is too coarse for handling post-treatment variables [41].

Transportability and the Problem of Data-fusion

In this section, we consider Task 4 (Fig. 1), the problem of extrapolating experimental findings across domains (i.e., settings, populations, environments) that differ both in their distributions and their inherent causal characteristics. This problem, called "transportability" in [43], lies at the heart of every scientific investigation since, invariably, experiments performed in one environment are intended to be used elsewhere, where conditions are likely to be different. Special cases of transportability can be found in the literature under different rubrics such as "lack of external validity" [44, 45], "heterogeneity" [46], "meta-analysis" [47, 48]. We formalize the transportability problem in non-parametric settings and show that despite glaring differences between the two populations, it might still be possible to infer causal effects at the target population by borrowing experimental knowledge from the source populations.

For instance, assume our goal is to infer the casual effect at one population from experiments conducted in a different population after noticing that the two age distributions are different. To illustrate how this task should be formally tackled, consider the data-generating model in Fig. 5(a) in which X represents a treatment, Y represents an outcome, Z represents age, and S (graphically depicted as a square) is a special variable representing the set of all unaccounted factors (e.g., proximity to the beach) that creates differences in Z (age in this case), between the source (π) and target (π^*) populations. Formally, conditioning on the event $S = s^*$ would mean that we are considering population π^*, otherwise population π is being considered. This graphical representation is called "selection diagrams". [13]

Our task is then to express the query $Q = P(y|do(x), S = s^*) = P^*(y|do(x))$ in terms of the experiments conducted in π and the observations collected in π^*, that is, $P(y, z|do(x))$ and $P^*(y, x, z)$. Conditions for accomplishing this task were derived in [43, 49, 50]. To illustrate how these conditions work in model of Fig. 5(a), note that the target quantity can be re-written as follows:

$$\begin{aligned} Q &= \sum_z P(y|do(x), S = s^*, z)P(z|S = s^*, do(x)) \\ &= \sum_z P(y|do(x), z)P(z|S = s^*, do(x)) \\ &= \sum_z P(y|do(x), z)P(z|S = s^*) \\ &= \sum_z P(y|do(x), z)P^*(z), \end{aligned} \quad [15]$$

where the first line of the derivation follows after conditioning on Z, the second line from the independence $(S \perp\!\!\!\perp Y|Z)_{G_{\overline{X}}}$ (called *s-admissibility*), the third line from the third rule of the do-calculus, and the last line from the definition of S-node. Eq. (15) is called a *transport formula* because it explicates how experimental findings in π are transported over to π^*; the first factor is estimable from π and the second from π^*.

Consider Fig. 5(b) where Z now corresponds to "language skills" (a proxy for the original variable, age, which is unmeasured). A simple derivation yields a different transport formula [43], namely

$$Q = P(y|do(x)), \quad [16]$$

In a similar fashion, one can derive a transport formula for Fig. 5(c) in which Z represents a post-treatment variable (e.g. "biomarker"), giving

$$Q = \sum_z P(y|do(x), z)P^*(z|x), \quad [17]$$

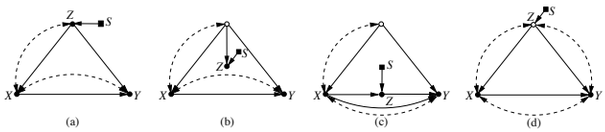

Fig. 5. Selection diagrams depicting differences between source and target populations. In (a), the two populations differ in age (Z) distributions (so S points to Z). In (b), the populations differs in how reading skills (Z) depends on age (an unmeasured variable, represented by the hollow circle) and the age distributions are the same. In (c), the populations differ in how Z depends on X. In (d), the unmeasured confounder (bidirected arrow) between Z and Y precludes transportability.

The transport formula in Eq. (17) states that to estimate the causal effect of X on Y in the target population π^*, we must estimate the z-specific effect $P(y|do(x), z)$ in π and average it over z, weighted by the conditional probability $P^*(z|x)$ estimated at π^* (instead of the traditional $P^*(z)$). Interestingly, Fig. 5(d) represents a scenario in which Q is not transportable regardless of the number of samples collected.

The models in Fig. 5 are special cases of the more general theme of deciding transportability under any causal diagram. It can be shown that transportability is feasible *if and only if* there exists a sequence of rules that transforms the query expression $Q = P(y|do(x), s^*)$ into a form where the do-operator is separated from the S-variables [49]. A complete and effective procedure was devised by [49, 50], which given any selection diagram, decides if such a sequence exists and synthesizes a transport formula whenever possible. Each transport formula determines what information need to be extracted from the experimental and observational studies and how they ought to be combined to yield an estimate of Q.

Transportability from multiple populations. A generalization of transportability theory to multi-environments when limited experiments are available in each environments led to a principled solution to the *data-fusion problem*. Data-fusion aims to combining results from many experimental and observational studies, each conducted on a different population and under a different set of conditions, so as to synthesize an aggregate measure of targeted effect size that is "better," in some sense, than any one study in isolation. This fusion problem has received enormous attention in the health and social sciences, and is typically handled by "averaging out" differences (e.g., using inverse-variance weighting), which, in general, tends to blur, rather than exploit design distinctions among the available studies.

Fortunately, using multiple "selection diagrams" to encode commonalities among studies, [51] "synthesized" an estimator that is guaranteed to provide unbiased estimate of the desired quantity, whenever such estimate exists. It is based on information that each study shares with the target environment. Remarkably, a consistent estimator can be constructed from multiple sources with limited experiment even in cases where it is not constructable from any subset of sources considered separately [52]. We summarize these results as follows:

Summary Result 3. (Transportability and Data-fusion) *We now possess complete solutions to the problem of transportability and data-fusion, which entail the following:*

- *Graphical and algorithmic criteria for deciding transportability and data-fusion in non-parametric models;*

[13] Each diagram shown in Fig. 5 constitutes indeed the overlapping of the causal diagrams of the source and target populations. More formally, each variable V_i should be supplemented with an S-node whenever the underlying function f_i or background factor U_i is different between π and π^*. If knowledge about commonalities and disparities is not available, transport across domains cannot, of course, be justified.

- *Automated procedures for extracting transport formulae specifying what needs to be collected in each of the underlying studies;*
- *An assurance that, when the algorithm fails, fusion is infeasible regardless of the sample size.*

For detailed discussions of these results, see [43, 50, 52].

Conclusion

The unification of the structural, counterfactual, and graphical approaches to causal analysis gave rise to mathematical tools that have helped to resolve a wide variety of causal inference problems, including the control of confounding, sampling bias, and cross-population bias. In this paper, we presented a general approach to these problems, based on a syntactic transformation of the query of interest into a format derivable from the available information. Tuned to nuances in design, this approach enables us to address a crucial problem in big data applications: the need to combine datasets collected under heterogeneous conditions, so as to synthesize consistent estimates of causal effects in a target population. As a by-product of this analysis, we arrived at solutions to two other long-held problems: Recovery from sampling selection bias and generalization of randomized clinical trials. We hope that the framework laid out in this paper will stimulate further research to enhance the arsenal of techniques for drawing causal inferences from big data.

1. Pearl J (2000) Causality: Models, Reasoning, and Inference. (Cambridge University Press, New York). Second ed., 2009.
2. Pearl J (2009) Causal inference in statistics: An overview. Statistics Surveys 3:96–146.
3. Angrist J, Imbens G, Rubin D (1996) Identification of causal effects using instrumental variables (with comments). Journal of the American Statistical Association 91(434):444–472.
4. Greenland S, Lash T (2008) Bias analysis in Modern Epidemiology, eds. Rothman K, Greenland S, Lash T. (Lippincott Williams & Wilkins, Philadelphia, PA), 3rd edition, pp. 345–380.
5. Galles D, Pearl J (1998) An axiomatic characterization of causal counterfactuals. Foundation of Science 3(1):151–182.
6. Halpern J (1998) Axiomatizing causal reasoning in Uncertainty in Artificial Intelligence, eds. Cooper G, Moral S. (Morgan Kaufmann, San Francisco, CA), pp. 202–210.
7. Balke A, Pearl J (1995) Counterfactuals and policy analysis in structural models in Uncertainty in Artificial Intelligence 11, eds. Besnard P, Hanks S. (Morgan Kaufmann, San Francisco), pp. 11–18.
8. Rubin D (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology 66:688–701.
9. Rosenbaum P, Rubin D (1983) The central role of propensity score in observational studies for causal effects. Biometrika 70:41–55.
10. Hotz VJ, Imbens G, Mortimer JH (2005) Predicting the efficacy of future training programs using past experiences at other locations. Journal of Econometrics 125(1-2):241–270.
11. Spirtes P, Glymour C, Scheines R (2000) Causation, Prediction, and Search. (MIT Press, Cambridge, MA), 2nd edition.
12. Pearl J (1995) Causal diagrams for empirical research. Biometrika 82(4):669–710.
13. Huang Y, Valtorta M (2006) Pearl's calculus of intervention is complete in Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence, eds. Dechter R, Richardson T. (AUAI Press, Corvallis, OR), pp. 217–224.
14. Shpitser I, Pearl J (2006) Identification of conditional interventional distributions in Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence, eds. Dechter R, Richardson T. (AUAI Press, Corvallis, OR), pp. 437–444.
15. Greenland S, Pearl J, Robins J (1999) Causal diagrams for epidemiologic research. Epidemiology 10(1):37–48.
16. Pearl J (1993) Comment: Graphical models, causality, and intervention. Statistical Science 8(3):266–269.
17. Shpitser I, VanderWeele T, Robins J (2010) On the validity of covariate adjustment for estimating causal effects in Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence. (AUAI, Corvallis, OR), pp. 527–536.
18. Pearl J, Paz A (2014) Confounding equivalence in causal equivalence. Journal of Causal Inference 2:77–93.
19. E. Perkovic, J. Textor MK, Maathuis M (2015) A complete adjustment criterion. in Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence, eds. Meila M, Heskes T. (AUAI Press, Corvallis, OR), pp. 682–691.
20. Tian J, Paz A, Pearl J (1998) Finding minimal separating sets, (Cognitive Systems Laboratory, Department of Computer Science, UCLA, CA), Technical Report R-254.
21. van der Zander B, Liskiewicz M, Textor J (2014) Constructing Separators and Adjustment Sets in Ancestral Graphs. (AUAI Press), pp. 907–916.
22. Bareinboim E, Pearl J (2012) Causal inference by surrogate experiments: z-identifiability in Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence, eds. de Freitas N, Murphy K. (AUAI Press, Corvalis, OR), pp. 113–120.
23. Chen B, Pearl J (2014) Graphical tools for linear structural equation modeling, (Department of Computer Science, University of California, Los Angeles, CA), Technical Report R-432, <http://ftp.cs.ucla.edu/pub/stat_ser/r432.pdf>. Forthcoming, Psychometrika.
24. Tian J, Pearl J (2002) A general identification condition for causal effects in Proceedings of the Eighteenth National Conference on Artificial Intelligence. (AAAI Press/The MIT Press, Menlo Park, CA), pp. 567–573.
25. Cooper G (1995) Causal discovery from data in the presence of selection bias. Proceedings of the Fifth International Workshop on Artificial Intelligence and Statistics pp. 140–150.
26. Cortes C, Mohri M, Riley M, Rostamizadeh A (2008) Sample Selection Bias Correction Theory. (Springer, Berlin, Heidelberg), pp. 38–53.
27. Whittemore A (1978) Collapsibility of multidimensional contingency tables. Journal of the Royal Statistical Society, B 40(3):328–340.
28. Kuroki M, Cai Z (2006) On recovering a population covariance matrix in the presence of selection bias. Biometrika 93(3):601–611.
29. Pirinen M, Donnelly P, Spencer C (2012) Including known covariates can reduce power to detect genetic effects in case-control studies. Nature Genetics 44:848–851.
30. Mefford J, Witte JS (2012) The covariate's dilemma. PLoS Genet 8(11):e1003096.
31. Heckman J (1979) Sample selection bias as a specification error. Econometrica 47(1):pp. 153–161.
32. Angrist JD (1997) Conditional independence in sample selection models. Economics Letters 54(2):103–112.
33. Robins J (2001) Data, design, and background knowledge in etiologic inference. Epidemiology 12(3):313–320.
34. Glymour M, Greenland S (2008) Causal diagrams in Modern Epidemiology, eds. Rothman K, Greenland S, Lash T. (Lippincott Williams & Wilkins, Philadelphia, PA), 3rd edition, pp. 183–209.
35. Greenland S, Pearl J (2011) Adjustments and their consequences – collapsibility analysis using graphical models. International Statistical Review 79(3):401–426.
36. Didelez V, Kreiner S, Keiding N (2010) Graphical models for inference under outcome-dependent sampling. Statistical Science 25(3):368–387.
37. Bareinboim E, Tian J, Pearl J (2014) Recovering from selection bias in causal and statistical inference in Proceedings of the Twenty-Eight National Conference on Artificial Intelligence, eds. Brodley C, Stone P. (AAAI Press, Menlo Park, CA), pp. 2410–2416.
38. Bareinboim E, Tian J (2015) Recovering causal effects from selection bias in Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, eds. Koenig S, Bonet B. (AAAI Press, Menlo Park, CA), pp. 3475–3481.
39. Cole SR, Stuart EA (2010) Generalizing evidence from randomized clinical trials to target populations the actg 320 trial. American journal of epidemiology 172(1):107–115.
40. Tipton E et al. (2014) Sample selection in randomized experiments: A new method using propensity score stratified sampling. Journal of Research on Educational Effectiveness 7(1):114–135.
41. Pearl J (2015) Generalizing experimental findings. Journal of Causal Inference 3(2):259–266.
42. Bareinboim E, Pearl J (2012) Controlling selection bias in causal inference in Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, eds. Lawrence N, Girolami M. (JMLR W&CP), pp. 100–108.
43. Pearl J, Bareinboim E (2014) External validity: From do-calculus to transportability across populations. Statistical Science 29(4):579–595.
44. Campbell D, Stanley J (1963) Experimental and Quasi-Experimental Designs for Research. (Wadsworth Publishing, Chicago).
45. Manski C (2007) Identification for Prediction and Decision. (Harvard University Press, Cambridge, Massachusetts).
46. Höfler M, Gloster A, Hoyer J (2010) Causal effects in psychotherapy: Counterfactuals counteract overgeneralization. Psychotherapy Research 20(6):668–679.
47. Glass GV (1976) Primary, secondary, and meta-analysis of research. Educational Researcher 5(10):pp. 3–8.
48. Hedges LV, Olkin I (1985) Statistical Methods for Meta-Analysis. (Academic Press).
49. Bareinboim E, Pearl J (2012) Transportability of causal effects: Completeness results in Proceedings of The Twenty-Sixth Conference on Artificial Intelligence, eds. Hoffmann J, Selman B. (AAAI Press, Menlo Park, CA), pp. 698–704.
50. Bareinboim E, Pearl J (2013) A general algorithm for deciding transportability of experimental results. Journal of Causal Inference 1(1):107–134.
51. Bareinboim E, Pearl J (2013) Meta-transportability of causal effects: A formal approach in Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics, eds. Carvalho C, Ravikumar P. (JMLR W&CP), pp. 135–143.
52. Bareinboim E, Pearl J (2014) Transportability from multiple environments with limited experiments: Completeness results in Advances in Neural Information Processing Systems 27 (NIPS 2014), eds. Ghahramani Z, Welling M, Cortes C, Lawrence N, Weinberger K.