Received: 5 September 2022 | Revised: 28 October 2022 | Accepted: 31 October 2022

DOI: 10.1002/psp4.12894

TUTORIAL

An introduction to causal inference for pharmacometricians

James A. Rogers[1] | **Hugo Maas**[2] | **Alejandro Pérez Pitarch**[2]

[1]Metrum Research Group, Tariffville, Connecticut, USA

[2]Boehringer Ingelheim Pharma GmbH & Co. KG, Ingelheim am Rhein, Germany

Correspondence
James A. Rogers, Metrum Research Group, 2 Tunxis Road, Suite 112, Tariffville, CT 06081, USA.
Email: jimr@metrumrg.com

Abstract

As formal causal inference begins to play a greater role in disciplines that intersect with pharmacometrics, such as biostatistics, epidemiology, and artificial intelligence/machine learning, pharmacometricians may increasingly benefit from a basic fluency in foundational causal inference concepts. This tutorial seeks to orient pharmacometricians to three such fundamental concepts: potential outcomes, *g*-formula, and directed acyclic graphs (DAGs).

BACKGROUND AND OBJECTIVES

Pharmacometrics, as a science that is concerned with the effects of both real and hypothetical interventions, necessarily involves causal reasoning. Causal reasoning is implicit whenever a scientist uses the word "because," "effect," or "confounding" to name just a few of the telltale signs. In recognizing the value of "mechanistic" models[1–3] and in recognizing the importance of theory (as distinct from direct empiricism),[4] pharmacometrics inherently prioritizes models with causal interpretability. Moreover, pharmacometric analyses typically involve nonrandomized comparisons (e.g., comparing predicted outcomes at a high exposure to predicted effects at a low exposure when exposure is not randomized or predicted outcomes for patients with normal renal function to predicted outcomes for patients with impaired renal function when renal function status is—necessarily—not randomized), which is precisely the setting that motivates most causal inference research.

In contrast to colloquial causal reasoning (e.g., simply using words such as "because," "effect," and "confounding"), the term *causal inference* will be used here to specifically connote *formal* (i.e., mathematized) causal reasoning. Several influential publications in pharmacometrics made extensive use of causal formalisms, including instrumental variables and potential outcomes notation.[5–7] However, aside from those seminal efforts, the pharmacometric literature has been mostly devoid of explicit causal inference. To clarify, pharmacometric models often rely on mathematical representations of pharmacological and biological processes, but the explicit mathematization of causal questions is far less common. Although pharmacometrics has undoubtedly flourished even without the benefit of an explicit causal lens, there are several signs that a resurgent awareness of causal inference in pharmacometrics would be timely. These signs include:

- The application of analytic techniques arising from causal inference research, such as propensity-based matching, in exposure–response analyses for regulatory decision making.[8]
- The use of directed acyclic graphs (DAGs) to articulate the nature of causal confounding of exposure–response in immuno-oncology.[9]
- The use of DAGs for covariate selection in epidemiological studies of relevance to pharmacometrics, for

example, for characterizing longitudinal progression toward end-stage renal disease[10] or for characterizing overall survival in oncology in response to immune checkpoint inhibitors (CPIs).[11,12]

- The application of interpretable artificial intelligence/machine-learning (AI/ML) algorithms (e.g., with interpretation assisted by Shapley values) to population pharmacokinetic modeling[13] and prediction of relapse and related disease activity in multiple sclerosis,[14] contemporaneous with an increased recognition of the interpretive value of formal causal frameworks in AI/ML research.[15–18]
- The advent of real-world evidence (RWE) usage in pharmacometric analyses,[19] contemporaneous with a growing body of guidance for the use of RWE that advocates for the use of causal DAGs.[20]
- The increasingly favorable environment for employing external or synthetic control arms in clinical trials, with the intent of generating estimates of (causal) treatment effects using methods that approximate the effects of randomization.[21]
- A renewed focus in pharmacometrics on what to estimate[22] (as distinct from how to do the estimation), contemporaneous with a broader recognition that most estimands are best expressed as causal quantities.[23]

Our objective in this tutorial is therefore to provide an introductory exposition of three fundamental causal concepts that we deem to be particularly relevant for pharmacometric research: potential outcomes notation, *g*-formula, and causal DAGs. It is beyond the scope of this effort to provide a complete guide to the application of these concepts. Our present intent is rather to lay the groundwork for future application-oriented instruction by establishing the fundamental terminological and notation distinctions that are used in causal inference.

POTENTIAL OUTCOMES NOTATION

Quantities of interest (i.e., estimands) in pharmacometric research are very often *causal* estimands, that is, they are quantities summarizing what would happen in a population if a particular intervention were enacted. Potential outcomes notation greatly facilitates the mathematical expression of such estimands. The pioneering use of potential outcomes notation in pharmacometrics went hand in hand with efforts to prioritize *what* questions as distinct from *how* questions.[6,7,22]

The representational value of potential outcomes notation may be understood by contrasting it with standard notational conventions for conditional and unconditional probability statements. In what follows, we make this distinction linguistically; a more precise understanding can be gained by exploring the same distinction using the R code in Supplementary Material S1.

In relation to an outcome Y_i and an intervention A_i for subjects indexed by i:

- Y_i (with a capital "Y") refers to the as-yet-uncertain outcome that will occur for subject i. Strictly speaking, Y_i conveys only notional uncertainty and not necessarily future tense, but using the future tense is perhaps the easiest linguistic approach to expressing uncertainty in Y_i; one can then read the expression $P\left(Y_i = y\right)$ as, "the probability that the outcome for subject i will be y."
- $Y_i \mid A_i = a$ refers to the as-yet-uncertain (or notionally uncertain) outcome that will occur for subject i given that, in the system under observation, subject i receives treatment a.
- Y_i^a (the "potential outcome" under intervention a, sometimes equivalently notated as $Y_i(a)$) refers to the outcome that would occur (modal verb rather than future tense) if one intervened in the observational system to assign treatment a to subject i.[24]

It is self-evident that Y_i^a is not necessarily equal to Y_i. What may be less obvious is that Y_i^a is also not necessarily equal to $Y_i \mid A_i = a$. The intended distinction is best understood by thinking about the as-yet-uncertain quantities epistemically: if one learns that a subject, in the natural course of affairs, has taken a medication as a matter of voluntary initiative, one's knowledge about that subject's outcome changes, but not in the same way that one's knowledge changes upon learning that the subject took the medication after being exogenously enjoined to do so as a matter protocol, prescription, or policy.

For illustration, consider a hypothetical oncology scenario where $a = 0$ represents standard of care, $a = 1$ represents treatment with a novel agent under development, and Y represents objective tumor response, taking value $Y = 1$ for a complete recovery and $Y = 0$ otherwise. For simplicity, suppose the novel agent is being developed as a first-line therapy. In reality, each patient can only receive one first-line therapy, but causal inference frameworks encourage us to think about "counterfactuals"[25–27] and not merely what is practically observable. (The term *counterfactual outcome* is often used interchangeably with "potential outcome.") One might naturally assume (even though one can never verify this empirically) that:

- Some patients would completely recover on either treatment, corresponding to $Y_i^{a=0} = 1$ and $Y_i^{a=1} = 1$.
- Some patients would not completely recover on either treatment, corresponding to $Y_i^{a=0} = 0$ and $Y_i^{a=1} = 0$.

- Some patients would completely recover on standard of care but not on the novel agent, corresponding to $Y_i^{a=0} = 1$ and $Y_i^{a=1} = 0$.
- Some patients would completely recover on the novel agent but not on standard of care, corresponding to $Y_i^{a=0} = 0$ and $Y_i^{a=1} = 1$.

Supplementary Material S1 includes the R code to simulate an observational study with these features. In that code, Y^0 and Y^1 are represented as the complete ("unfiltered") columns Y0 and Y1, whereas the conditional variables $Y \mid A = 0$ and $Y \mid A = 1$ are obtained by taking row subsets of (a.k.a. "filtering") the column of observations YOBS (which corresponds to Y in our mathematical notation). In this simulated environment, one can explore the differing characteristics of the distribution of Y^0 and the distribution of $Y \mid A = 0$ (or similarly for Y^1 and $Y \mid A = 1$).

The distinction between Y_i^a and $Y_i \mid A_i = a$ lies at the heart of causal inference, which generally seeks to make inferences about the distributions of Y^0 and Y^1 in an entire population, even though Y_i^1 and Y_i^0 are never observed in the same subject. (In the words that Plato attributes to Heraclitus: "... all things move and nothing remains still ... you cannot step twice into the same stream.")[28] The challenge arises because, to estimate the marginal (whole population) distribution of Y^1, one would need to estimate the conditional distributions of both $Y^1 \mid A = 1$ and $Y^1 \mid A = 0$ (and only estimation of the former is straightforward) and to estimate the marginal distribution of Y^0, one would need estimates of the conditional distributions of both $Y^0 \mid A = 1$ and $Y^0 \mid A = 0$ (and only estimation of the latter is straightforward). The missing conditional distributions can only be estimated under certain conditional exchangeability assumptions, as discussed in the next section. Supposing that one does make the requisite assumptions to allow estimation of the joint distribution of (Y^0, Y^1), that joint distribution may then be summarized in any number of ways. In the causal inference literature, there is often a special focus on $E[Y^1 - Y^0]$, but this is just one particular quantity that can be derived from the joint distribution of Y^0 and Y^1; an estimand of the form $P(Y^1 < q)$ may be of greater interest in many pharmacometric applications and is explored in the next section.

G-FORMULA

The causal logic of *g*-formula and adjustment sets

The gap between what we want to know (e.g., the distribution of $Y^{a=1}$) and what we can actually observe (e.g., the distribution of $Y \mid A = 1$) presents a challenge. This challenge can be addressed by finding conditions that allow for *conditional exchangeability* (also referred to as "conditional ignorability" in this context). Focusing specifically for the moment on the distribution of $Y^1 \mid A = 0$, the key is to find covariates or conditions L such that $(Y^1 \mid A = 0, L)$ (which we do not observe) would be expected to have the same distribution as $(Y^1 \mid A = 1, L)$ (which we do "observe" when a consistency assumption holds)[29] and similarly, conditions such that $(Y^0 \mid A = 0, L) \sim (Y^0 \mid A = 1, L)$. In words, the challenge is to find covariates such that, once those covariate values and the assigned treatment is known, there is no additional value in knowing the treatment status toward which the subject would have been naturally inclined. These requirements are typically summarized as $(Y^{a=0}, Y^{a=1}) \perp A \mid L$, where $\perp$ signifies independence and the set of covariates L is referred to as a sufficient *adjustment set* if this criterion is satisfied. (As we will see later, DAGs provide a mechanism to find and evaluate potential adjustment sets.)

Under the conditions described previously, one can in fact estimate the entire marginal (whole population) distribution of Y^1 and Y^0 using what is known as the *g*-formula. In the causal inference literature, *g*-formula is most often derived with reference to an expected value such as $E[Y^a]$, but we offer a derivation here in relation to $P(Y^a < q)$ (for arbitrary q), as population quantiles and tail probabilities are often of particular interest in pharmacometric applications. For the simple case of non–time-varying treatment variables (a.k.a. "point exposures"), the relationship and its derivation are as follows:

$$P(Y^a < q) = E_L[P(Y^a < q \mid L)] \quad \text{Iterated expectation} \quad (1)$$

$$= E_L[P(Y^a < q \mid A = a, L)] \quad \text{Conditional exchangeability given L} \quad (2)$$

$$= E_L[P(Y < q \mid A = a, L)] \quad \text{Consistency assumption} \quad (3)$$

g-formula in pharmacometrics

The last expression in the previous derivation consists of terms that pharmacometric modelers typically estimate, although the connection may not yet be obvious. To begin with, we will suppose the usual scenario where $P(L)$ (the multivariate covariate distribution in the target population) is estimated with an empirical distribution with observed (multivariate) covariate values at $l_1, \ldots, l_N$ and where we have a model that allows estimation of $P(Y < q \mid A = a, L = l_i)$ for each value of l_i, where i indexes subjects. In that case, the estimate version of the preceding expectation can be expressed as:

$$\frac{1}{N}\sum_{i=1}^{N} \hat{P}(Y < q \mid A = a, L = l_i)$$

(The "model" for $P(L)$ in this case is simply a point mass of $1/N$ at each of the observed l_i values). This intuitive operation—generating predictions conditional on treatment and covariates and then averaging those predictions over a covariate distribution—will be familiar to pharmacometricians, who often refer to the procedure simply as "population simulation."[30] The same procedure is also referred to in epidemiology as "standardization."[31] Although the *g*-formula operation is straightforward in the context of point exposures, it is important to note that this approach *generalizes* (whence the "*g*-" in *g*-formula) to more complex settings with time-varying exposures and treatment-confounder feedback. These more complex settings were in fact the motivation for Robins' seminal 1986 article,[32] which inaugurated research on *g*-methods.

In pharmacometrics, the probability in the previous *g*-formula summand will typically be replaced by a simulation-based estimate, with simulations typically generated from a parametric nonlinear mixed-effects model (such a model would be referred to in the causal inference literature as an "outcome model" or a "Q model"[33]). Specifically, for each subject i, let $y^*_{i1}, \ldots y^*_{iM}$ be values simulated from the model for $\left(Y \mid A = a, L = l_i\right)$, with M suitably large to accurately characterize the simulation distribution. Then:

$$\widehat{P}\left(Y < q \mid A = a, L = l_i\right) = \frac{1}{M}\sum_{j=1}^{M} \mathbf{1}_{\left[y^*_{ij} < q\right]}$$

In summary, the simulation-based estimate of $P\left(Y^1 < q\right)$ is:

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{j=1}^{M} \mathbf{1}_{\left[y^*_{ij} < q\right]}$$

More commonly, a simplification is employed to avoid computing this as a nested sum. In the simplified version, one samples with replacement a large number of times M from the empirical distribution for L to obtain $l^*_1, \ldots, l^*_M$ and then simulates each y^*_i from the model for $\left(Y \mid A, L = l^*_i\right)$, finally computing the estimate as:

$$\frac{1}{M}\sum_{i=1}^{M} \mathbf{1}_{\left[y^*_i < q\right]}$$

Statistical biases and causal biases

When the estimate $\frac{1}{N}\sum_{i=1}^{N}\widehat{P}\left(Y < q \mid A = a, L = l_i\right)$ is biased relative to the true value (or "estimand") $P(Y^a < q)$, the reason(s) for the bias can be categorized according to the following scheme:

- "Statistical biases" related to the outcome model, contributing to the difference between the expectation of the estimator $\widehat{P}(Y < q \mid A = a, L)$ and its "statistical estimand" $P(Y < q \mid A = a, L)$. This is typically the meaning of the term *model misspecification* in pharmacometrics, and model diagnostics in pharmacometrics typically only aspire to investigate deficiencies of this nature.
- "Statistical biases" in the model for $P(L)$, that is, in the model for the multivariate covariate distribution in the target population. As noted already, a common practice is to simply use the empirical covariate distribution in the available sample, although it is often contestable whether this adequately reflects the target population. More targeted nonparametric approaches may leverage an epidemiological database such as the National Health and Nutrition Examination Survey database,[34] and parametric approaches have been proposed as well.[35]
- The "causal bias" attributable to the difference between the statistical estimand $E_L\left[P(Y < q \mid A = a, L)\right]$ and the causal estimand $P(Y^a < q)$. Specifically, as is evident in Step 2 of our derivation of *g*-formula, these two quantities will fail to be equal if the conditional exchangeability condition is not met, that is, if L is not a sufficient adjustment set. In terms of pharmacometric population simulation, this difference arises if simulations from the outcome model do not reflect the true distribution of $\left(Y^1 \mid A = 0, L = l_i\right)$ or $\left(Y^0 \mid A = 1, L = l_i\right)$, that is, when counterfactual outcomes cannot be correctly simulated. This failure mode is often conceived of as arising from "unmeasured covariates," but a failure to achieve conditional exchangeability can also arise from *including* certain types of covariates in the adjustment set (resulting in selection biases). In general, suspected biases arising from an inadequate adjustment set can be articulated using causal DAGs, as discussed in the next section. Depending on what has been measured, some biases of this type may be remediated by simply modifying the adjustment set. When unmeasured (and/or unmeasurable) confounders are hypothesized to exist, the likely magnitude of the bias may be evaluated through sensitivity analyses,[7] and Bayesian frameworks may be particularly appealing for this purpose.[36] Notwithstanding the value of such sensitivity analyses, we emphasize that the most essential features of the causal bias problem—defining what one wants to estimate and identifying the most likely sources of causal bias in that estimation—can be articulated without incurring the overhead of a simulation framework and/or a formal Bayesian framework. To this end, it is the opinion of the authors that causal DAGs are particularly helpful, a position that we elaborate in the next section.

Other adjustment strategies

The implicit use of *g*-formula is perhaps the most common strategy in pharmacometrics for obtaining covariate-adjusted estimates of causal effects, but it is not the only such strategy. Adjustment methodologies based on propensity scores (especially inverse propensity weighting) are an attractive alternative for removing causal biases. Propensity-based methods are out of scope for the current tutorial, except to note that such methods also involve covariate adjustment (because propensity scores are themselves functions of covariates) and that they rely on the same assumption of conditional exchangeability that *g*-formula relies on.[37]

CAUSAL DAGS

Schematics and mathematical structures

At a cursory level, the meaning of a causal DAG is likely to be intuitive: variables (represented graphically as nodes) have a dependence structure relative to each other, and directed edges (arrows) in some way signify those dependencies. For example, letting *AUC* denote the area under a pharmacokinetic concentration versus time curve, the DAG "*Dose* → *AUC* → *Outcome*" intuitively conveys the assumption that the effect of *Dose* on *Outcome* is in some sense mediated via *AUC*.

Although the intuitive schematic value of DAGs is very important, we emphasize that DAGs are also mathematical structures that have specific logical and statistical implications. The DAG "*Dose* → *AUC* → *Outcome*" would specifically imply that intervening to change *Dose* while holding *AUC* constant would not result in any change in *Outcome*. (Depending on the specific definition of *Dose* and *AUC*, this could be a very strong and contestable assumption: for example, if different formulations and/or routes of administration were in play, these alternate dosing strategies could result in the same *AUC* but with other dispositional differences—reflected by different maximum concentrations (C_{max}), for example—that might entail a different *Outcome* distribution.) This specific causal implication of the DAG would further entail the statistical implication that *Dose* and *Outcome* are conditionally independent given *AUC*.

Although the preceding example is fairly trivial, it serves to illustrate that a formal causal DAG will typically entail specific probabilistic conditional independencies.[26] This suggests the essential connection between DAGs and the conditional exchangeability assumption presented in the previous section: conditional exchangeability of counterfactual outcomes is a specific type of conditional independence, and a hypothesized DAG can be analyzed to determine whether this type of conditional independence is likely to obtain in a given situation.

In summary, a DAG serves two purposes:

- It provides an explicit representation of one's primary causal assumptions (so that those assumptions can be publicly critiqued and debated), and
- When analyzed as a mathematical structure, it can be used to deduce the consequences of those assumptions. In particular, it allows one to assess whether the requirement of conditional exchangeability—essential to causal effect estimation—is logically consistent with one's primary causal assumptions.

The latter use of DAGs, wherein they are treated as mathematical structures that can be subjected to formal analysis, distinguishes them from many superficially similar diagrams that are used in pharmacometrics for merely schematic purposes.[30] In the context of causal effect estimation, the relevant mathematical deductions involve the identification of particular paths through the DAG known as "backdoor paths." Before examining backdoor paths in detail, we first consider some principles of DAG construction that are essential if a DAG is to support such logical deductions.

DAG completeness

A DAG is said to represent the complete causal structure between a treatment and an outcome if all sources of dependence between the treatment and outcome are explained by causal links.[38] In practice this means that the following conditions must hold:

- Treatment and outcome themselves must be represented. We state this requirement explicitly to emphasize that a DAG is complete or incomplete only *in relation to a given question* (i.e., no DAG needs to be complete *as such*, in the sense of encapsulating all causal knowledge on a topic). "Treatment" in this case should be understood in the most general sense, that is, as the variable whose causal effect is of primary interest. For example, "renal impairment status" could be the "treatment" variable in a context where the causal effects of renal impairment were of interest, notwithstanding a lack of interest in, for example, the effects of kidney transplants. We acknowledge that this terminology is potentially confusing in pharmacometric applications; to make matters worse, the "treatment" variable is variously referred to as the "exposure" variable (this is the convention used by DAGitty, for example[39]); this

potentially induces even greater confusion because exposure (in the sense of, e.g., the plasma concentration of a drug) is in fact the "outcome" in a pharmacokinetic model!

- For any two variables already on the graph, all common causes of those two variables are also represented.[38]
- All selection variables are represented. A selection variable is a variable that causes distributional differences in the available data compared to the target population. For example, if the available data are composed of several studies, some of which only enrolled male subjects, the distribution of sex in the available data is not likely to represent the target population. In terms of patient-level variables, study enrollment is the "cause" of a patient's inclusion or exclusion in the available data, so this status should be included as a selection node. The inclusion of study enrollment as a node would then further entail that sex, if causally related to the outcome, be included as well (because sex would then be a common cause of two variables already on the DAG: study enrollment status and the outcome). The inclusion of selection variables is particularly important because these are nodes on which one has inevitably conditioned (the very act of obtaining the data involves conditioning on the selection node). When the selection variable is also a "collider" node, that is, a variable that is a causal descendant of two other nodes on the graph, conditioning on the collider node will induce a noncausal statistical dependency between the two other variables.[40] For example, the first exposition of this phenomenon considered the bias induced when analyzing only hospitalized patients to study the association between two variables that were themselves determinants of hospitalization.[41] Representing all selection variables ensures that collider-induced dependencies are represented.

The aforementioned conditions are minimally sufficient and may be used to simplify DAG representations.[42] On the other hand, the inclusion of additional variables that are not strictly entailed by these conditions is not harmful. Indeed, more verbose representations may be helpful initially to elicit an intelligible causal narrative from subject matter experts.

Not all variables represented in a DAG need to be observable. In fact, it is very important to represent any variable entailed by the conditions listed previously, whether observed or not.[43] For example, as represented in Figure 1b, medication status may be a function of access to health care, which may in turn be a function of socioeconomic status, and socioeconomic status may affect diet, which in turn affects hemoglobin A1c. To the degree that such relationships are plausible, one should represent them on the DAG—at least initially—whether they are observable or not. In the representation of Figure 1, rectangles are used to represent observable variables, and ovals are used to represent unobservable variables. Having elicited this causal narrative in terms of partially unobserved variables, one could then revisit the minimal requirements listed previously and consider removing the *Access* and *Diet* variables because neither of these is a common cause of both treatment and outcome, while necessarily retaining socioeconomic status because it *is* a common cause of both treatment and outcome (Figure 1c).

Determinism and stochasticity

As one might expect, an arrow emanating from node X and pointing at node Y indicates a belief that X is a cause of Y. More precisely, this means X is an argument in an unseen function (or "law of nature") that determines Y.

Technically, this functional interpretation of arrows only allows for the representation of deterministic relationships between variables, as described by Pearl: "This quasi-deterministic functional model mirrors Laplace's conception of nature (Laplace 1814), according to which of [sic] nature's laws are deterministic, and randomness surfaces merely due to our ignorance of the underlying boundary conditions."[26] Nonetheless, a DAG can represent unexplained "random" variation in any variable by representing its dependence on unobserved and parent-less "background" or "exogenous" variables, often denoted with a "U" (Figure 1a). Each exogenous "U variable" represents an amalgam of unknown causal factors that influences the endogenous variables. Our ignorance of the exogenous factors may induce an apparent randomness in the endogenous variables, but the relationships represented by the arrows are presumed to be deterministic. This conceptual framework allows us to assign a consistent meaning to the arrows (they *always* represent deterministic relationships), even while allowing the nodes to represent variables with unexplained ("random") variability. In many contexts (and in most of our exposition), DAGs are simplified by leaving "U variables" unrepresented and implicit. Such omission is, however, only valid when the "U variable" in question has only a single immediate descendant on the graph; per the completeness criterion, "U variables" that are direct causes of more than one of the depicted nodes must themselves be depicted.

An essential point in relation to the arrows (a.k.a. "edges") is that the presence of an arrow reflects *possible* causal influence. As stated previously, an arrow *always* represents a functional relationship, but sometimes it may be the "constant" function, $Y = f(X) = c$ that does not depend on its arguments. As such, the inclusion of an arrow is essentially noncommittal, whereas the *absence* of an arrow

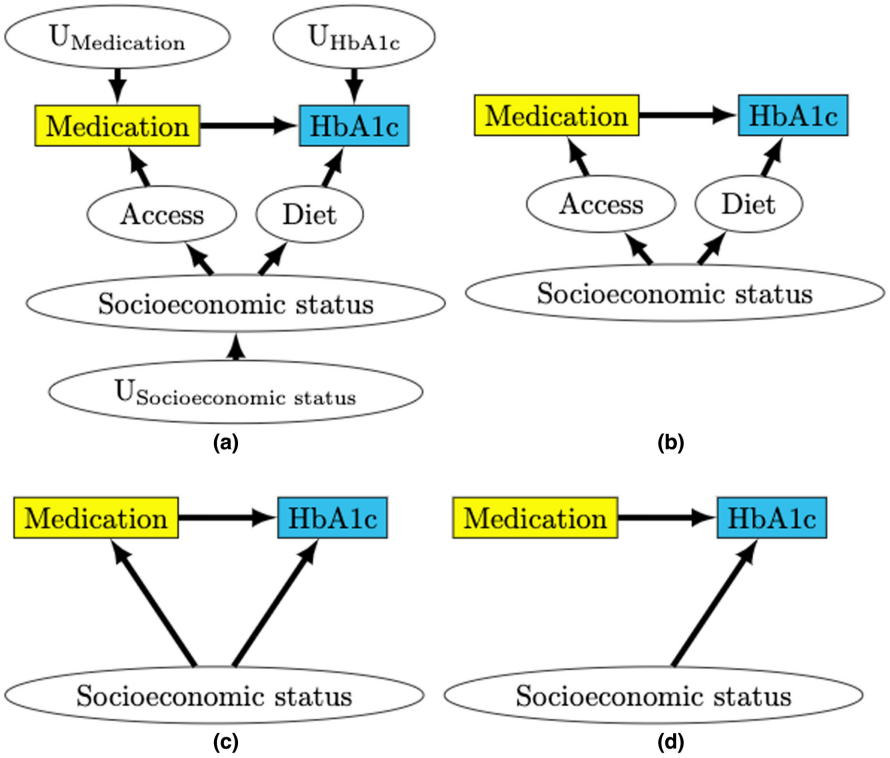

FIGURE 1 Directed acyclic graphs (DAGs) representing plausible sources of statistical dependence between a treatment variable ("Medication," represented in yellow) and an outcome ("HbA1c", represented in blue). Unspecified exogenous background variables are sometimes represented in a DAG with "U variables", as in DAG (a). Representation of such exogenous causes may be helpful during the initial development of a plausible causal structure, but for analytic purposes the DAG may be simplified by removing these variables as long as none of them is a common cause of two more variables remaining on the graph, resulting in DAG (b). Similarly, causal intermediaries such as (in this case) Diet and Access may be removed when they are not common causes of two or more variables remaining on the graph, resulting in the further simplified DAG (c). In a randomized treatment context, the further simplification reflected in DAG (d) would be justified, and in fact socioeconomic status could be removed from the DAG altogether (not shown). HbA1c, hemoglobin A1c.

encodes a commitment to the potentially very strong assumption of "no causal influence." Of course, it is generally not practically possible to rule out causal influence with certainty (except in a randomized setting, where arrows into the randomized treatment variable can be removed, as in Figure 1d). However, if (as we advocate) one's goal in creating DAGs is simply to articulate reasonable beliefs and deduce the consequences of those beliefs, thereby identifying the most likely biases, then the use of arrows to represent highly speculative causal connections becomes unnecessary.

Total causal effects and direct causal effects

Causal inference research has developed standard terminology that distinguishes between two types of causal effects: total and direct. Absent the ability to clearly articulate this distinction, two or more quantitative scientists may unwittingly use the same word ("effect") to talk about two different concepts, resulting in confused debates about whether a given effect estimate is biased or not. We introduce this distinction using the definitions provided by Pearl[44]:

Definition of total causal effect: "[The total causal effect] measures the probability that response variable Y would take on the value y when X is set to x by external intervention. This probability function is what we normally assess in a controlled experiment in which X is randomized and in which the distribution of Y is estimated for each level x of X."[44]

Definition of direct causal effect: "The term 'direct effect' is meant to quantify an influence that is not mediated by other variables in the model or, more accurately, the sensitivity of Y to changes in X while all other factors in the analysis are held fixed. Naturally, holding those factors fixed would sever all causal paths from X to Y with the exception of the direct link $X \rightarrow Y$, which is not intercepted by any intermediaries."[44]

We revisit this distinction and depict it with a DAG in the pharmacokinetic example at the end of this article.

What DAGs do not represent

In their most typical implementation (the one we wish to promote), DAGs do not attempt to represent any of the following:

- The direction of effects (the direction of causality is of course represented, but whether the average causal effect corresponds to a positive or negative association is not represented).
- The magnitude of effects.
- The statistical distributions of the variables.
- Interactions. When a variable has more than one cause, the DAG representational scheme simply represents arbitrary multiple-argument functional dependence without distinguishing whether the multiple effects are additive, super-additive, multiplicative, and so on.

Given the aforementioned limitations, DAGs may seem to provide an overly laconic representational system. Scientists are of course free to embellish DAGs in whatever manner they like, but we emphasize that even the minimal assumptions about causal structure that are encoded in a typical DAG are sufficient to generate powerful insights regarding potential biases and adjustment strategies. Significantly, assumptions about causal structure are often available from subject matter experts at the earliest planning stages, when it may be neither necessary nor desirable to speculate about the directions and magnitudes of effects or about the nature of interactions.

Practical guidance for DAG development

In the context of use that we are proposing, DAG creation begins by eliciting assumptions that subject matter experts already have (implicitly) and encoding those assumptions graphically to make them explicit. Such elicitation will generally require directed questioning by a quantitative scientist with an understanding of the eventual use of the DAG. To elicit these assumptions as effectively as possible from subject matter experts, we recommend applying the following principles, which summarize a number of points made in the preceding subsections, and which reflect a combination of personal experience and published recommendations[43]:

- Relax and have fun drawing. For a DAG development process to be successful, all participants in the process should be aware that the goal is simply to articulate shared assumptions and deduce the consequences of those assumptions, thereby identifying the most likely biases that attend (or will attend) an analysis. Because the goal is not to "prove" that a particular causal interpretation will be valid, it is not essential that any particular DAG be identified as "the correct one." Iterative and collaborative DAG development may be fostered with an interactive tool such as DAGitty.[39] With this iterative approach in mind, DAG development may begin at the earliest stages of analysis planning and/or study design, when the scientific questions are still being formulated and refined, with the understanding that early DAG iterations may represent only tentative and personal opinions, whereas more mature iterations of the DAG should ideally represent a mature and well-researched consensus.
- Know where to begin. Begin by representing the intervention variable (a.k.a. the "treatment" or "exposure" variable) and the outcome.
- Respect the rules. As the term itself implies, a DAG must be both directed and acyclic. For a graph to be *directed*, each edge must be an arrow pointing in only one direction. As a practical matter, bidirectional arrows are sometimes used, but this is merely a shorthand implying an unspecified common parent, that is, "$X \leftrightarrow Y$" does not imply that "X is a cause of Y and Y is a cause of X" but, rather, is used as a convenient shorthand for "$X \leftarrow U \rightarrow Y$ (for some unspecified U)". An *acyclic* graph must of course be devoid of directed cycles, corresponding to the tautological premise that no variable may be a cause of itself (not even indirectly). Any feedback loops must therefore be represented using distinct nodes to represent the same variables at different timepoints. For example, an adaptive dosing regimen in response to adverse events (AE) cannot be represented as $Dose \leftrightarrow AE$, but must instead be represented as $Dose_1 \rightarrow AE_1 \rightarrow Dose_2 \rightarrow AE_2 \ldots$ (with subscripts used to distinguish the same variables at different points in time). Anecdotally, the essential causal structure of such a feedback loop can often be revealed by representing only two distinct timepoints.
- Think carefully about the data that you do not see. Any selection process that may affect which records we see in the data and which we do not should be reflected by a selection node. For example, if the available data only represent study completers, completion status should be a node.
- When in doubt, draw the arrow. The absence of an arrow represents the strong assumption that the originating node does not exert causal influence on the receiving node. The presence of an arrow merely indicates *possible* causal influence.
- Focus first on what you believe to be true. Do not limit depicted nodes to measurable quantities only. Of course, one can only adjust for covariates that are measured, but depicting latent (unmeasured, and perhaps unmeasurable) variables is often helpful to articulate a

plausible causal narrative. The explicit representation of one's implicit assumptions should be prioritized at the outset, deferring concerns with the feasibility of particular covariate adjustments until a later stage.

- When the time is right, be practical. Notwithstanding the previous point, there is generally no harm in representing latent "true causes" with measurable proxies. In a pharmacokinetic context, for example, estimated glomerular filtration rate (EGFR) will generally suffice as a proxy for the true glomerular filtration rate (GFR); EGFR can therefore be represented as a "cause" of other variables, even though the estimate per se does not participate in any physiological process.
- Know when to stop. Progress toward representation of all likely common causes (including common indirect causes) and all common effects (including common indirect effects) of the intervention variable and the outcome variable. In practice, the exercise may be considered complete when the most likely common causes and effects are represented, stopping before the addition of nodes relationships becomes highly speculative.

Backdoor paths

Once a treatment A and an outcome Y are identified and a complete DAG is developed that includes potential covariates L, the conditional exchangeability condition $\left(Y^{a=0}, Y^{a=1}\right) \perp A \mid L$ can be evaluated to determine if L is a sufficient adjustment set. We note that this assessment will only be correct if the DAG is complete in the technical sense described previously. Significantly, we have now pivoted slightly with our terminology: whereas we introduced conditional exchangeability as an *assumption*, we now refer to it as a *condition* to emphasize that its truth or falsity can be derived from the more primary causal assumptions represented in a DAG. Specifically, the condition can be evaluated by identifying "backdoor paths" and determining whether those paths are "open" or "closed" conditional on L.

For the practitioner in pharmacometrics who simply wishes to apply these concepts, it may suffice to recognize that reliable tools exist to automatically analyze backdoor paths and thereby identify and evaluate adjustment sets.[39] Nonetheless, technical definitions of the most central concepts are uncomplicated (if somewhat abstract and not particularly supportive of intuition). For convenience, we therefore provide several verbatim excerpts from Greenland and Pearl[38]:

Definition of back-door path: "A back-door path from X to Y is a path that begins with a parent of X (i.e., leaves X from a 'backdoor') and ends at Y."[38] [Such a path need not be directed, i.e., any sequence of adjacent edges can be used to compose a backdoor path, regardless of the direction of the arrows.]

Definition of collider: "A variable is a collider on the path if the path enters and leaves the variable via arrowheads (a term suggested by the collision of causal forces at the variable)."[38]

Definition of open/blocked: in the absence of conditioning, "a path is *open* or *unblocked* at noncolliders and *closed* or *blocked* at colliders."[38] Conditioning on a variable reverses its blocking status: "Conditioning on a variable C closes open paths that pass through C. Conversely, conditioning on C opens paths that were blocked only at C or at an ancestral collider A."[38]

Definition of adjustment set: "A set S [L, in our notation] then satisfies the back-door criterion [and is therefore a sufficient adjustment set, in our terminology] with respect to X and Y if (a) S contains no descendant of X and (b) there are no open back-door paths from X to Y after conditioning on S."[38]

Fuller exposition of these concepts is provided by Greenland and Pearl.[38] Considering again the perspective of the applied practitioner, the essential point is to recognize that algorithmic analyses of backdoor paths can be used to ensure a logical consistency between the qualitative beliefs of subject matter experts (as reflected in a DAG elicited from those subject matter experts) and the set of covariates used by a quantitative modeler (as reflected by the covariate effects in a formal statistical model).

TABLE 1 Baseline patient disposition and (fictionalized) overall response rates (ORRs) for a phase I oncology trial

Treatment history		Phase I treatment	ORR
PD1i (first line)	Chemo (second line)	Novel Tx (third line)	30%
	PD1i + chemo (first line)	Novel Tx (second line)	20%

Note: ORR varied as a function of Tx history (30% for sequential therapy vs. 20% for combination therapy), perhaps suggesting that the patients' ability to respond to the novel Tx was modified by Tx history.

Abbreviations: PD1i, programmed death 1 inhibitor; Tx, treatment.

APPLICATIONS

Detecting collider bias

An exploratory analysis from a single-arm, phase I, immuno-oncology trial revealed an intriguing association. By design, all enrolled patients had received the investigational therapy; about half of them had a treatment history

of combination therapy (programmed cell death protein 1 [PD1] inhibitor *plus* chemotherapy) and half had a treatment history of sequential monotherapy (PD1 inhibitor *then* chemotherapy). The (fictionalized and anonymized) overall response rates (ORRs) for these two groups are represented in Table 1, revealing a higher success rate (30%) for those with a history of sequential therapy compared with the rate for those with a history of combination therapy (20%).

This observed difference gave rise to a hypothesis that seemed to have some plausibility: prior combination therapy might have resulted in a greater immunosuppressive effect compared with prior sequential therapy, making patients less likely to respond to the novel immunotherapy (a "detrimental effect modification" hypothesis). Such a hypothesis is consistent with the observation that some (although not all) types of chemotherapy are associated with lymphocytopenia.[45] On the other hand, it was also recognized that some chemotherapeutic effects are mediated through the immune system,[46] and in that sense the opposite trend might have been expected a priori, that is, one would at least expect prior chemotherapy to be synergistic with prior PD1 inhibition and to be perhaps synergistic with the novel immunotherapy as well (in the latter case, this would be a "beneficial effect modification" hypothesis).

Naïvely, one might suppose that the observed phase I results provide some support for the detrimental effect modification hypothesis, but the nonrandomized nature of the comparison requires careful consideration. To characterize the potential biases in this nonrandomized comparison, a DAG was developed, represented in Figure 2. In addition to the variables already described, this DAG includes CPI resistance as a cause of both prior tumor response and tumor response in the phase I trial. (For patients with a history of sequential therapy, prior tumor response may be understood as either failure on *both* of their prior therapies or else success on at *at least one* of those therapies.) The DAG includes an arrow from the CPI resistance node to the prior tumor response node because patients with higher levels of CPI resistance are less likely to respond to PD1 inhibitors (PD1 inhibition is a type of checkpoint inhibition). The novel therapeutic regimen also included a PD1 inhibitor as one of its components, so an arrow is drawn from CPI resistance to phase I tumor response as well.

Analysis of backdoor paths in the DAG revealed that (unavoidable) conditioning on prior tumor response had opened up a "backdoor path,"[25] a fact that can be verified manually using the definition or programmatically using the DAGitty R package or via interactive use of the DAGitty web application.[39] Programmatic verification is illustrated in Supplementary Material S1 for this article.

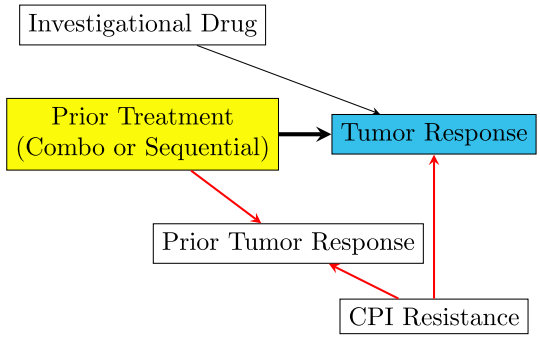

FIGURE 2 A directed acyclic graph developed to characterize potential biases in an analysis of a phase I oncology study. Elicitation from subject matter experts indicated that checkpoint inhibitor (CPI) resistance would have potentially affected both prior tumor response and tumor response in the phase I study. In the absence of any conditioning, the backdoor path through prior response would be closed and so would not introduce any bias. However, the phase I study (necessarily) only enrolled patients who had failed prior therapies, which amounts to conditioning on prior tumor response(s). Conditioning on this collider node opens the backdoor path (shown in red), signaling that the apparent affect of treatment history is biased.

The existence of this backdoor path implies that the observed association between prior treatment and tumor response is biased with respect to the true causal relationship. This particular backdoor path would be eliminated if we also conditioned on CPI resistance (which could be done via covariate adjustment), but for present purposes we will suppose that this adjustment was not made.

The essential intuition related to this backdoor path is that prior failure on either of the historical regimens may be a marker for CPI resistance, but perhaps not to the same degree. As already mentioned, chemotherapy is expected to be synergistic with PD1 inhibition. If true, this would have created a selection effect: those in the current phase I study who failed to respond to prior combination therapy may have been harder to treat (i.e., may have had greater CPI resistance) from the very beginning—prior to any treatment whatsoever—than those in the current study who failed to respond to prior sequential therapy. To explain the observed association, therefore, it suffices to hypothesize that prior combination therapy was more efficacious than prior sequential therapy in CPI-resistant patients, without invoking any effect modification hypotheses. Indeed, retrospective feedback from key opinion leaders was that there was no plausible explanation for a better response in the cohort that previously received sequential treatment and that there may have been a selection bias for better patients in that cohort (without commenting on the mechanism of the selection bias).

We note that the DAG-based analysis could also have been applied before any data were available, that is, it could have been recognized at the outset that the phase I study would not be capable of generating evidence in support of (or against) an effect modification hypothesis. Or, more subtly, it could have been recognized at the outset that the phase I study would only be capable of providing evidence for or against effect modification to the extent that one could successfully adjust for CPI resistance. Among other things, this would highlight—at the planning stages—the importance of having a good measure of CPI resistance.

Interpreting effects for correlated covariates

In any regression model, the effect estimate for a given covariate may change substantially when a correlated covariate is added to the model. This phenomenon is very familiar to the pharmacometrics community and has led to various correlation-based "rules of thumb" for flagging correlations that could compromise the interpretation of effect estimates.[47] Such rules of thumb, proposed initially as guides to prompt careful reflection, have unfortunately been interpreted at times as definitive proscriptions against ever including highly correlated covariates in a model. From a causal perspective, this proscriptive stance seems to labor under a false premise that estimates of covariate effects should always strive to be unbiased with respect to total causal effects, an unnecessary requirement in many cases.

For example, in modeling the pharmacokinetics of a drug that is at least partially renally cleared, it would be common to consider both age and EGFR (assumed here to be expressed per unit of body surface area) as potential covariates. The correlation between these two variables can easily exceed (in absolute value) rule-of-thumb thresholds such as 0.6, potentially prompting less-experienced pharmacometric modelers to agonize over the simultaneous inclusion of both covariates. More illuminating guidance can be offered in such cases by drawing a DAG based on prior knowledge, analyzing it for backdoor paths, and adjusting analyses and interpretations accordingly. Such a DAG in this case is likely to reflect the prior knowledge that the aging processes typically causes a decline in renal function (reflected by the arrow from age to EGFR in Figure 3) and might also (depending on the drug) reflect a prior belief that other age-related processes also affect exposure (reflected by the arrow directly from age to exposure; note that for simplicity we have not specified other mediating variables such as would be involved in, e.g., a hepatic clearance mechanism; recall that the DAG needs to encapsulate our prior causal knowledge only to the extent required by the completeness criteria discussed previously).

In such a context, there are at least two distinct causal questions related to renal impairment that are potentially of interest:

Q1. What is the causal effect *of EGFR (expressed per unit of body surface area)* on exposure? For example, how would exposures differ if—an entirely hypothetical scenario—we could somehow intervene to improve EGFR while holding everything else constant? (Strictly speaking, an *estimate* of GFR is not a cause of physiological or pharmacokinetic processes, but to simplify matters we will gloss over the difference between GFR and EGFR.)

Q2. What is the causal effect *of dose level* on exposure in a clinically identifiable subpopulation defined by an EGFR threshold, for example, what would exposures be in the subpopulation if they received a lower dose?

We begin by noting that Q1, despite its hypothetical nature, is a question of definite import, as implied by US Food and Drug Administration guidance to evaluate the potential for confounding due to baseline factors such as age when estimating the pharmacokinetic effects of renal function[48] (a concern with confounding would be incoherent if the causal effect of renal impairment per se were not of interest). Having established this, we now examine question Q1 with the benefit of a causal lens. Assuming one accepts the basic causal assumptions embedded in the DAG in Figure 3a on the basis of prior knowledge, it follows that there is a "backdoor path" through age (shown in red in Figure 3) that connects EGFR to exposure. This backdoor path can only be closed by conditioning on age, and we can only condition on age if we include this as a covariate in our model (the existence of this backdoor path can be discovered and/or confirmed using DAGitty in the event that one does not want to work manually from the definitions). The need to adjust for age would only be obviated if we (a priori) eliminated the arrow connecting age to exposure, corresponding to a strong prior assumption that the effect of age on exposure is entirely mediated by its effect on renal function (this relatively strong assumption could be warranted in a context where there was strong prior evidence of an entirely renal clearance mechanism but would probably be unjustified otherwise). Thus, far from proscribing the simultaneous inclusion of these two correlated covariates, the DAG-based analysis shows that (in the absence of a strong assumption) the inclusion of both covariates is necessary if one wishes to estimate the causal effect of EGFR on exposure.

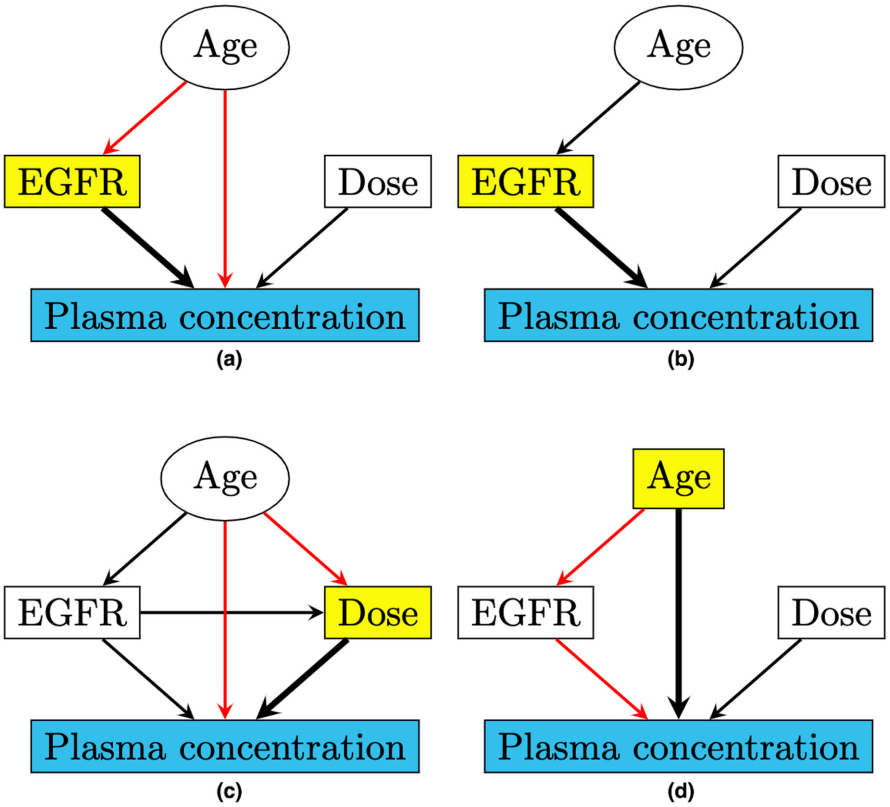

FIGURE 3 Directed acyclic graphs (DAGs) for determining whether to include the correlated covariates age and estimated glomerular filtration rate (EGFR) in a population pharmacokinetic model. If prior knowledge suggests that the effects of age will be partially but not entirely mediated by EGFR, then failure to adjust for age leaves open a backdoor path (depicted in red), signaling a potentially biased estimate for the effect of EGFR on plasma concentration (a). This backdoor path would be closed by adjusting for age, that is, including age as an additional covariate. If it were not possible to adjust for age, one could simply state as a caveat to interpretation that the estimated effect of EGFR will only be unbiased if the effects of age are entirely mediated through EGFR; in this case, a DAG could be helpful for communicating the assumption underlying the caveat (b). When multiple studies are represented in the dataset, age and EGFR may be causal determinants of dose (because of inclusion/exclusion differences in the various studies). In such a case, failure to adjust for either age or EGFR will result in biased estimates for the effect of dose (represented in yellow to indicate that this is now the explantory variable of interest) on plasma concentration, signaled in this case by the red pathway passing through age when this variable is not in the outcome model (c). When both age and EGFR are included in the model, the direct (not EGFR-mediated) effect of age (now represented as the explanatory variable of interest) can still be estimated (depicted with thick black arrow), but adjustment for EGFR prevents estimation of the total effect of age (d).

What about questions related to Q2, for example, what is required if we want to estimate the effect of dose level on exposure in a subpopulation defined by an EGR threshold? In the simplest of cases, if our analysis were based only on a single study in which dose level had been randomized (thereby removing the possibility of confounding the effects of dose), it would suffice to include only EGFR in the model and then simulate exposures in the EGFR-restricted population. More commonly, however, pharmacometric datasets will include multiple studies, each with potentially different inclusion/exclusion criteria. In such a multistudy context, both age and EGFR could easily be causes of dose level, for example, if higher doses were only given in earlier studies that enrolled subjects with age and EGFR restrictions (confounding of this nature is a hallmark of analyses based on integrated evidence). The resulting DAG in that case would include backdoor paths through age and EGFR connecting dose to exposure, indicated in red in Figure 3c. As such, closure of backdoor paths would require conditioning on both age and EGFR and so would require the inclusion of both covariates in the model. Given an outcome model with both of these covariates included, the causal effect of dose level may be computed by *g*-formula, as illustrated in Supplementary Material S1.

In summary, adequate answers to questions Q1 and Q2 will in most cases require inclusion of both age and EGFR, regardless of the correlation between these two variables (so long as this correlation is less than one to avoid strict problems of nonidentifiability).

Our previous discussion has neglected questions related to the causal effect of age. Given the DAG in Figure 3d, the inclusion of EGFR as a covariate permits the unbiased estimation of the direct causal effect of age, but prevents the unbiased estimation of the total causal effect of age. Given space constraints for this article, we simply invite the reader to consider which (if any) of these two estimands is likely to be of interest.

CONCLUDING REMARKS

As suggested by the citations in our introductory section, causal concepts are being leveraged with increasing frequency in a number of disciplines that intersect with pharmacometrics, including epidemiology and AI/ML. This observation gave rise to one of our goals in writing this tutorial: to foster cross-disciplinary dialogue by way of helping pharmacometricians to become more conversant in the language and notation of causal inference.

Although we are indeed very hopeful that causal inference may provide a basis for new interdisciplinary directions in pharmacometric research, we wish to emphasize that causal inference is not merely peripheral to our field of study but, rather, is central and even foundational to it. In his highly influential 1989 article, Sheiner called attention to the importance of addressing "what if" questions[4] (notably, *What If* is now the title of one of the definitive references on causal inference that has been cited throughout this article[37]). Soon thereafter, Sheiner came to articulate his advocacy for "theory" (as distinct from mere "empiricism") and his related critique of the excessive emphasis on intention-to-treat estimands[49] using Rubin's potential outcomes framework.[5] Current beneficiaries of these efforts within pharmacometrics have perhaps varying degrees of awareness as to the role played by formal causal inference in their own intellectual pedigree. Nonetheless, the enduring and far-reaching nature of that work (extending beyond the pharmacometric community, as evidenced by its recent permeation into statistical regulatory guidance,[22,50] for example) seems at least partly attributable to the decision of Sheiner and Rubin to phrase their arguments in the formal language of causal inference. Our hope is that by fostering a greater awareness of this formal language, we will enable pharmacometricians to more comfortably navigate both the periphery and the deepest roots of their discipline.

CONFLICT OF INTEREST

The authors declared no competing interests for this work.

FUNDING

No public funding was received for this work.

REFERENCES

1. Bies RR, Gastonguay MR, Schwartz SL. Mathematics for understanding disease. *Clin Pharmacol Ther*. 2008;83:904-908.
2. Food and Drug Administration. Guidance for industry exposure-response relationships -study design, data analysis, and regulatory applications; 2003. Accessed June 14, 2022. https://www.fda.gov/media/71277/download
3. Food and Drug Administration. Demonstrating substantial evidence of effectiveness for human drug and biological products; 2019. Accessed June 14, 2022. https://www.fda.gov/media/133660/download
4. Sheiner LB. Clinical pharmacology and the choice between theory and empiricism. *Clin Pharmacol Ther*. 1989;46:605-615.
5. Sheiner LB, Rubin DB. Intention-to-treat analysis and the goals of clinical trials. *Clin Pharmacol Ther*. 1995;57:6-15.
6. Sheiner LB. Is intent-to-treat analysis always (ever) enough? *Br J Clin Pharmacol*. 2002;54:203-211.
7. Nedelman JR, Rubin DB, Sheiner LB. Diagnostics for confounding in PK/PD models for oxcarbazepine. *Stat Med*. 2007;26:290-308.
8. Yang J, Zhao H, Garnett C, et al. The combination of exposure-response and case-control analyses in regulatory decision making. *J Clin Pharmacol*. 2013;53:160-166.
9. Dai HI, Vugmeyster Y, Mangal N. Characterizing exposure-response relationship for therapeutic monoclonal antibodies in Immuno-oncology and beyond: challenges, perspectives, and prospects. *Clin Pharmacol Ther*. 2020;108:1156-1170.
10. Staplin N, Herrington WG, Judge PK, et al. Use of causal diagrams to inform the design and interpretation of observational studies: an example from the study of heart and renal protection (SHARP). *Clin J Am Soc Nephrol*. 2017;12:546-552.
11. Spakowicz D, Hoyd R, Muniak M, et al. Inferring the role of the microbiome on survival in patients treated with immune checkpoint inhibitors: causal modeling, timing, and classes of concomitant medications. *BMC Cancer*. 2020;20:383.
12. Andreano A, Bergamaschi W, Russo AG. Immune checkpoint inhibitors at any treatment line in advanced NSCLC: real-world overall survival in a large Italian cohort. *Lung Cancer*. 2021;159:145-152.
13. Ogami C, Tsuji Y, Seki H, et al. An artificial neural network-pharmacokinetic model and its interpretation using Shapley additive explanations. *CPT Pharmacometrics Syst Pharmacol*. 2021;10:760-768.
14. Basu S, Munafo A, Ben-Amor AF, Roy S, Girard P, Terranova N. Predicting disease activity in patients with multiple sclerosis: an explainable machine-learning approach in the mavenclad trials. *CPT Pharmacometrics Syst Pharmacol*. 2022;11(7):843-853.
15. Heskes T, Sijben E, Bucur IG, Claassen T. Causal shapley values: Exploiting causal knowledge to explain individual predictions of complex models. *arXiv* 2020. doi:10.48550/ARXIV.2011.01625
16. Lecca P. Machine learning for causal inference in biological networks: perspectives of this challenge. *Frontiers in Bioinformatics*. 2021;1. doi:10.3389/fbinf.2021.746712

17. Crown WH. Real-world evidence, causal inference, and machine learning. *Value Health*. 2019;22:587-592.
18. Neal B. *Introduction to Causal Inference from a Machine Learning Perspective*. Accessed June 20, 2022. https://www.bradyneal.com/Introduction_to_Causal_Inference-Aug28_2020-Neal.pdf
19. Van Driest SL, Choi L. Real-world data for pediatric Pharmacometrics: can we upcycle clinical data for research use? *Clin Pharmacol Ther*. 2019;106:84-86.
20. Jaksa A, Wu J, Jónsson P, Eichler HG, Vititoe S, Gatto NM. Organized structure of real-world evidence best practices: moving from fragmented recommendations to comprehensive guidance. *J Comp Eff Res*. 2021;10:711-731.
21. Thorlund K, Dron L, Park JJH, Mills EJ. Synthetic and external controls in clinical trials – a primer for researchers. *Clin Epidemiol*. 2020;12:457-467.
22. Akacha M, Bartels C, Bornkamp B, et al. Estimands-what they are and why they are important for pharmacometricians. *CPT Pharmacometrics Syst Pharmacol*. 2021;10:279-282.
23. Lipkovich I, Ratitch B, Mallinckrodt CH. Causal inference and Estimands in clinical trials. *Stat Biopharm Res*. 2020;12:54-67.
24. Rubin D. Estimating causal effects of treatments in experimental and observational studies. *ETS Res Bull Series*. 1972;1972:i-31.
25. Pearl J, Glymour M, Jewell NP. *Causal Inference in Statistics: A Primer*. John Wiley & Sons; 2016.
26. Pearl J. Causal inference without counterfactuals: technical report. *J Comput Graph Stat*. 2000;95:428-431.
27. Pearl J, Mackenzie D. *The Book of why: the New Science of Cause and Effect*. Basic Books; 2018.
28. Plato. *Plato in Twelve Volumes*. Vol 12. Fowler HN, trans. Harvard University Press, William Heinemann Ltd.; 1921.
29. Hernán MA. Does water kill? A call for less casual causal inferences. *Ann Epidemiol*. 2016;26(10):674-680.
30. Rogers JA. Causa nostra: the potentially legitimate business of drawing causal inferences from observational data. *CPT Pharmacometrics Syst Pharmacol*. 2019;8:253-255.
31. Vansteelandt S, Keiding N. Invited commentary: G-computation–lost in translation? *Am J Epidemiol*. 2011;173:739-742.
32. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math Model*. 1986;7:1393-1512.
33. Snowden JM, Rose S, Mortimer KM. Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. *Am J Epidemiol*. 2011;173:731-738.
34. NHANES – National Health and Nutrition Examination Survey Homepage; 2018. Accessed August 30, 2018. http://www.cdc.gov/nchs/nhanes/
35. Tannenbaum SJ, Holford NH, Lee H, Peck CC, Mould DR. Simulation of correlated continuous and categorical variables using a single multivariate distribution. *J. Pharmacokinet. Pharmacodyn*. 2006;33:773-794.
36. Oganisian A, Roy JA. A practical introduction to Bayesian estimation of causal effects: parametric and nonparametric approaches. *Stat Med*. 2021;40:518-551.
37. Hernán MA, Robins JM. *Casual Inference: What if*. Chapman & Hall/CRC; 2020.
38. Greenland S, Pearl J. *Causal Diagrams*. Wiley; 2017.
39. Textor J, van der Zander B, Gilthorpe MS, Liskiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *Int J Epidemiol*. 2016;45:1887-1894.
40. Munafò MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: when selection bias can substantially influence observed associations. *Int J Epidemiol*. 2018;47:226-235.
41. Sackett DL. Bias in analytic research. *J Chronic Dis*. 1979;32:51-63.
42. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10:37-48.
43. Tennant PWG, Harrison WJ, Murray EJ, et al. Use of directed acyclic graphs (DAGs) in applied health research: review and recommendations. *Int J Epidemiol*. 2021;50(2):620-632.
44. Pearl J. Direct and Indirect Effects. Technical Report R-273-UAI, University of California, Los Angeles; 2001.
45. Verma R, Foster RE, Horgan K, et al. Lymphocyte depletion and repopulation after chemotherapy for primary breast cancer. *Breast Cancer Res*. 2016;18:10.
46. Galluzzi L, Buqué A, Kepp O, Zitvogel L, Kroemer G. Immunological effects of conventional chemotherapy and targeted anticancer agents. *Cancer Cell*. 2015;28:690-714.
47. Bonate PL. The effect of collinearity on parameter estimates in nonlinear mixed effect models. *Pharm Res*. 1999;16:709-717.
48. Center for Drug Evaluation and Research. Pharmacokinetics in Patients with Impaired Renal Function—Study Design, Data Analysis, and Impact on Dosing and Labeling; 2020. Accessed January 25, 2022. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/pharmacokinetics-patients-impaired-renal-function-study-design-data-analysis-and-impact-dosing-and
49. Sheiner LB. The intellectual health of clinical drug evaluation. *Clin Pharmacol Ther*. 1991;50:4-9.
50. Center for Drug Evaluation and Research. E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands; 2021. Accessed January 25, 2022. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e9r1-statistical-principles-clinical-trials-addendum-estimands-and-sensitivity-analysis-clinical

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Rogers JA, Maas H, Pitarch AP. An introduction to causal inference for pharmacometricians. *CPT Pharmacometrics Syst Pharmacol*. 2023;12:27-40. doi:10.1002/psp4.12894