RESOLVING AN APPARENT PARADOX IN DOUBLY ROBUST ESTIMATORS

Doubly robust estimators are an approach used for estimating causal effects, usually based on fitting 2 statistical models (1). As the initial motivating example, Scharfstein et al. defined a robust estimator of the causal effect of some exposure X on outcome Y using models for both X and Y; they demonstrated that such an estimator is consistent if "at least one of the [fitted] models… is correct" (2, p. 1142). Such estimators were later termed "doubly robust" (3, p. 6). Here, we demonstrate that using common (but incorrect) intuition about what makes a model "correct" or "incorrect" can turn doubly robust estimators into estimators that are inconsistent if at least one of the fitted models is wrong. We introduce and resolve this double-robust paradox, demonstrating what must be meant by "correct model."

We illustrate this double-robust paradox using 2 scenarios for binary, random variables X, Y, and Z measured in a target population. We wish to estimate the average causal effect of X on Y, given by $E(Y(1) - Y(0))$, where the potential outcome $Y(x)$ is the value of Y we would have observed had X been set to x. Doubly robust estimation proceeds with a statistical model for Y, given by $f_{Y|X,C}(x, C, \beta)$, and a statistical model for X, given by $f_{X|C'}(C', \alpha)$, where covariate sets C and C' are both either empty ($\varnothing$) or contain Z, and where C may (or may not) equal C'; α and β are the parameters that index those models. If Z or $\varnothing$ is sufficient to control confounding between X and Y, a doubly robust approach fitted in this way can consistently estimate the average causal risk difference.

Given uncertainty about whether Z or $\varnothing$ is sufficient to control confounding, the doubly robust property seems advantageous. In such cases, we may be tempted to use Z in one model and $\varnothing$ in the other. For example, we could fit regression models $f_{Y|X,Z}(x, z, xz, \beta)$ and $f_X(\alpha)$. Our regression model for Y contains X, Z, and their product as regressors, while the "model" for X contains a single parameter α for $E(X)$.

First, suppose Z is a confounder. The model Y is "correct" because 1) it adjusts for Z and 2) it is saturated. The model for X is "incorrect" because it omits the confounder Z. (We emphasize that "correct" and "incorrect" have common-use definitions rather than technical definitions, a key distinction explained below.) Due to the doubly robust property, we might anticipate that our estimator would consistently estimate the average causal effect of X on Y.

Now suppose Z is instead a mediator. In this case, the model for X is "correct" because, to identify the average causal effect of X on Y, we should not generally adjust for variables affected by exposure (4). The model for Y is "incorrect" for this same reason. Under the common-use definition of correct model specification, our doubly robust estimator will (again) appear to be consistent for the average causal effect. Therein lies the problem; for a doubly robust estimator to give us a correct answer, regardless of whether Z is a mediator or a confounder, the *statistical* model would necessarily yield *causal* information about X and Z.

If the common-use definition were correct, we could, in principle, compare doubly robust estimators with other estimators to test whether Z was a mediator or a confounder.

Seen another way, if Z is a mediator and is not a modifier on the additive scale, one model is "correct" for the controlled, direct risk difference $E(Y(1, z) - Y(0, z))$, the effect we would observe if we could manipulate X while *also* setting the mediator to $Z = z$. The doubly robust estimator in the fully saturated case estimates the quantity $\mu = \sum_z [\{E(Y|X = 1, Z = z) - E(Y|X = 0, Z = z)\}\Pr(Z = z)]$; this holds regardless of whether Z is a confounder or a mediator. Under the common-use definition of correct model specification, our doubly robust estimator would be consistent for *both* $E(Y(1, z) - Y(0, z))$ *and* $E(Y(1) - Y(0))$ *under the same model specification*. If Z is truly a mediator, these estimands *cannot*, in general, be equal, meaning that the doubly robust property has failed. We term this contradiction the "double-robust paradox."

This apparent paradox is resolved by observing that the common-use definition of correct model specification is, itself, incorrect. The error results from bestowing a causal interpretation on the doubly robust estimate without considering causal identification conditions. Through the lens of causal directed acyclic graphs, we cannot infer the direction of the edges that go between X and Z without external information (5). That is, numerical values in a single data set do not tell us how variables affect each other (6). Only with external information, such as temporal ordering or biological plausibility, can we make accurate causal inference.

Under appropriate causal identification conditions, this identifies *an* effect of X under the observed distribution of Z. Thus, μ can be a marginal effect if Z is a confounder; a controlled, direct effect if Z is a mediator and is not a modifier on the additive scale; or something else that may not have a causal interpretation (7). For our scenario, a doubly robust method will estimate μ if Z is in either the model for Y or the model for X, meaning that we can make incorrect causal inferences by misspecifying (common usage) either model. Thus, if we mistake a mediator for a confounder, what was a doubly robust estimator becomes a "double-jeopardy" estimator: Estimators are inconsistent if at least one of the fitted models is wrong.

The paradox is resolved by noting that a "correctly specified" model refers to a statistical model that contains the true density of the data; to estimate causal effects, we also must understand identification conditions for the potential outcome of interest, be it $Y(x)$ or $Y(x, z)$. We propose that definitions of "model specification" should avoid conflation of statistical and causal concepts by clearly separating the two sets of concepts. Pearl notes that causal assumptions reflect concepts "that cannot be realized by imposing statistical assumptions" (8, p. 39). Further clarification of what distinguishes causal concepts from statistical concepts is needed: We are not aware of a general set of rules that

distinguish between causal mistakes that are amenable to double-robustness and causal mistakes that result in double jeopardy.

REFERENCES

1. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *J Am Stat Assoc*. 1994;89(427):846–866.
2. Scharfstein DO, Rotnitzky A, Robins JM. Rejoinder. [Re: "Adjusting for nonignorable drop-out using semiparametric nonresponse models"]. *J Am Stat Assoc*. 1999;94(448): 1135–1146.
3. Robins JM. Robust estimation in sequentially ignorable missing data and causal inference models. In: *Proceedings of the American Statistical Association Section on Bayesian Statistical Science*. (Vol. 1999). Alexandria, VA: American Statistical Association; 1999:6–10.
4. Rosenbaum PR. The consquences of adjustment for a concomitant variable that has been affected by the treatment. *J R Stat Soc Ser A*. 1984;147(5):656–666.
5. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10(1):37–48.
6. Robins JM. Data, design, and background knowledge in etiologic inference. *Epidemiology*. 2001;12(3):313–320.
7. Hubbard AE. Direct effects and effect among the treated. In: Van der Laan MJ, Rose S, eds. *Targeted Learning: Causal Inference for Observational and Experimental Data*. 1st ed. (Springer Series in Statistics). New York, NY: Springer Science & Business Media; 2011:133–143.
8. Pearl J. *Causality: Models, Reasoning, and Inference*. New York, NY: Cambridge University Press; 2000.

Alexander P. Keil[1], Stephen J. Mooney[2], Michele Jonsson Funk[1], Stephen R. Cole[1], Jessie K. Edwards[1], and Daniel Westreich[1] (e-mail: akeil@unc.edu)
[1] *Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC*
[2] *Harborview Injury Prevention and Research Center, University of Washington, Seattle, WA*