Response to Invited Commentary

Rose and van der Laan Respond to “Some Advantages of the Relative Excess Risk due to Interaction”

Sherri Rose* and Mark van der Laan

*Correspondence to Dr. Sherri Rose, Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115 (e-mail: rose@hcp.med.harvard.edu).

Initially submitted November 11, 2013; accepted for publication November 20, 2013.

We appreciate VanderWeele and Vansteelandt’s perspective (1) on our article (2). Our commentary largely focused on a discussion of marginal estimators for case-control study designs not mentioned in VanderWeele and Vansteelandt’s original article (3). In our presentation (2), we highlighted the case-control-weighted targeted maximum likelihood estimator (TMLE) (4–7) and Robins’ “approximately valid” inverse-probability-weighted estimator for case-control data (8). We appreciate VanderWeele and Vansteelandt’s continued dialogue on methods for case-control study designs, as well as their inclusion of a new double robust estimator in their commentary (1), since there is a strong need for more work in this area. In this response, we precisely frame the efficiency properties of the case-control-weighted TMLE, which have been discussed elsewhere (2, 4–7) but were not completely presented in VanderWeele and Vansteelandt’s commentary and Web Appendix (available at http://aje.oxfordjournals.org/) (1) or in our original commentary (2).

We also emphasize the need for flexible nonparametric estimators that incorporate machine learning in the modern “big data” era of epidemiology in large databases.

When defining our research question, we must be explicit about the model we are specifying. We wish to consider either a nonparametric model or a semiparametric model, thereby making fewer restrictive assumptions on our data-generating distribution than when imposing a parametric model. We are not limited to nonparametric statistical models, and we can make additional assumptions based on investigator knowledge in a semiparametric model. The efficiency claims made for the case-control-weighted TMLE are based on this nonparametric or semiparametric model (4–7). Before comparing the efficiency of estimators, it is important to agree on the model. Comparing parametric model efficiency with nonparametric or semiparametric model efficiency is not an apt comparison. Our case-control weighting effectively maps a function of the full-data sampled observations into a function for the biased case-control sampled observations. It has been demonstrated that case-control weighting of the efficient TMLE for the full-data model leads to an efficient TMLE for the case-control model. The required regularity conditions have been described previously (5). The case-control-weighted TMLE with known prevalence probability is consistent if either the outcome regression or the exposure mechanism is consistently estimated, and it is efficient if both are consistently estimated. Notably, the estimator is not defined as the solution to an estimating equation, although it does solve the efficient influence curve estimating equation.

We also wish to underscore that using a nonparametric or semiparametric model is not a limitation; in fact, we consider it a compelling advantage. Especially when considering the advent of large data sets in epidemiology, researchers are increasingly interested in more flexible procedures that do not rely on restrictive parametric models. Since the goal is to have a statistical model that contains the true data distribution, assuming a nonparametric or semiparametric model may be preferable, as will using an estimator that allows for the incorporation of machine learning or ensembling methods (9, 10). This avoids the problems of 1) having more parameters than observations in a parametric model, 2) committing to a specific functional form of the data, and 3) attempting to represent complex relationships with a parametric regression. Integrating machine learning methods and causal inference is a burgeoning field in statistical science, one with promising potential for new methodological innovation in epidemiology.

Novel robust estimators for case-control studies are an important area of methodological work, and we look forward to future contributions from VanderWeele, Vansteelandt, and other investigators.

ACKNOWLEDGMENTS

Author affiliations: Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts (Sherri Rose); and Division of Biostatistics, School of Public Health,
University of California, Berkeley, Berkeley, California (Mark van der Laan).
This work was supported by the National Institutes of Health (grant R01 A1074345-01 to M.v.d.L.).
Conflict of interest: none declared.

REFERENCES