## Invited Commentary

# Invited Commentary: Estimating Population Impact in the Presence of Competing Events

## Ashley I. Naimi* and Eric J. Tchetgen Tchetgen

* Correspondence to Dr. Ashley I. Naimi, Department of Obstetrics and Gynecology, McGill University, 687 Pine Avenue West, Room F 432, Montreal, Quebec H3A 1A1, Canada (e-mail: ashley.naimi@mcgill.ca).

The formal approach in the field of causal inference has enabled epidemiologists to clarify several complications that arise when estimating the effect of an intervention on a health outcome of interest. When the outcome is a failure time or longitudinal process, researchers must often deal with competing events. In this issue of the *Journal*, Picciotto et al. (*Am J Epidemiol*. 2015;181(8):563–570) use structural nested failure time models to assess potential population effects of hypothetical interventions and censor competing events. In the present commentary, we discuss 2 interpretations that result from treating competing events as censored observations and how they relate to measures of public health impact. We also comment on 2 alternative approaches for handling competing events: an inverse probability weighting estimator of the survivor average causal effect and the parametric g-formula, which can be used to estimate a functional of the subdistribution of the event of interest. We argue that careful consideration of the tradeoff between the interpretation of the parameters from each approach and the assumptions required to estimate these parameters should guide researchers on the various ways to handle competing events in epidemiologic research.

causal inference; comparative effectiveness research; competing risks; implementation science; intervention; principal strata

The formal approach in causal inference enables researchers to identify whether (and how) a given data set and modeling strategy can be used to estimate the effect of an intervention that would yield some change in the exposure's value. This approach clarifies several complications that arise when aiming to infer causation. When the outcome is a failure time or longitudinal process, researchers must often deal with 2 complications that result in incomplete data for a subset of the sample: right censoring, which occurs due to study termination or loss to follow-up, and competing events, which arise when alternative events preclude the event of interest from occurring altogether.

Though very distinct, competing events are often handled in the same way as right-censored observations. This handling leads to a "cause-specific" (1) or "death-blocking" (2) interpretation involving what would be observed in a world in which competing events were prevented from occurring. However, when the aim of an analysis is to quantify public health effects, removing competing events from the equation can lead to misleading population average inferences.

In this issue of the *Journal*, Picciotto et al. (3) use structural nested failure time models to assess the public health impacts of limiting exposure to oil-based metalworking fluids on death due to various cardiovascular causes. With characteristic insight and clarity, they highlight several strengths of their approach and give due diligence to some of its limitations. As they note, their handling of competing events leads to a "somewhat odd interpretation" (Web Appendix in reference 3) of their cause-specific mortality results. In the present commentary, we take the opportunity to build on why this is the case and discuss alternative approaches that have been used to make sense of intervention effects when competing events are present.

### CENSORING COMPETING EVENTS

Consider a failure time outcome $T_\delta$ that ends in 1 of 2 possible states: death due to an outcome of interest ($\delta = 1$) or death due to a competing risk ($\delta = 2$). Consider further a binary time-varying exposure $A(j)$ and a (scalar or vector-valued)

*Am J Epidemiol.* 2015;181(8):571–574

confounder $W(j)$ that can take values for each of the $j = 0$ to $t$ possible time points in a hypothetical cohort study, where $t$ represents the largest integer value just before $T_{\delta = 1}$. Often, the goal of an analysis is to estimate the parameter from a structural nested failure time model, as follows:

$$T^{\bar{0}} = \int_0^{T_{\delta=1}} \exp\{\psi_{CS} A(u)\} du,$$

where $T^{\bar{0}}$ is the outcome that would be observed under no exposure and $\psi_{CS}$ quantifies the relation between the time-varying exposure $A(j)$ and the cause-specific failure-time $T_{\delta = 1}$. As previously illustrated (4), g-estimation of a structural nested failure time model can be implemented by 1) choosing a set of candidate values for the true value of the structural nested model parameter; 2) using the observed failure time, observed exposure history, set of candidate values for $\psi_{CS}$ (denoted $\tilde{\psi}$), and the structural nested failure time model to impute a set of potential outcomes under no exposure (denoted $T(\tilde{\psi})$, as described by Naimi et al. (4)); and 3) testing whether, conditional on all measured confounders, the exposure is independent of the imputed potential outcomes for each unique value in the candidate set.

For a binary exposure, step 3 is usually implemented via a pooled logistic regression model, such as:

$$\text{logit}\{P[A(j) = 1 | \overline{W}(j), \overline{A}(j - 1), T(\tilde{\psi}), Y(j - 1) = 0]\}$$
$$= \alpha_j + \alpha_1 W(j) + \alpha_2 W(j - 1) + \alpha_3 A(j - 1) + \alpha_4 A(j - 2) + \alpha_5 T(\tilde{\psi}),$$

where $Y(j - 1)$ is an indicator of the event of interest and $T(\tilde{\psi})$ is the imputed potential outcome for the event of interest. Furthermore, overbars (e.g., $\overline{W}(j)$) denote variable histories that are often summarized by a few time points (e.g., $W(j)$, $W(j - 1)$). This logistic model is fit separately for each unique element in the set of candidate values $\tilde{\psi}$. The candidate value that renders a test statistic for $\alpha_5$ equal to 0 is taken as a point estimate for the exposure effect (5).

When competing events are present, this procedure will yield a biased estimate of $\psi_{CS}$ because it fails to account for the impact of individuals who die from the competing event. To address this problem, researchers often use inverse probability of censoring weights which, for g-estimation of a structural nested failure time model, are defined as:

$$sw_{CS}(j) = \begin{cases} \prod_{k=j}^{int(t)} \dfrac{P[C(k) = 0 | C(k - 1) = Y(k - 1) = 0, \overline{A}(k - 1)]}{P[C(k) = 0 | C(k - 1) = Y(k - 1) = 0, \overline{A}(k - 1), \overline{W}(k - 1)]}, & C(j) = 0 \\ 0, & C(j) = 1 \end{cases}$$

where $C(k)$ is an indicator that the competing event occurred at time $k$ (i.e., 1 if yes, 0 if no). One can then fit the pooled logistic model weighted by $sw_{CS}(j)$ and continue with step 3 of the g-estimation procedure, conditional on the added constraint that $C(j - 1) = 0$. When using these weights, the test statistic for step 3 must be based on the robust variance estimator described previously (6).

There are 2 possible interpretations of the parameter $\psi_{CS}$ when it is estimated using the censoring weights described above. The first (common) population average interpretation considers $sw_{CS}(j)$ as an inverse probability weight for missing data in that it is used to account for individuals not observed in the sample at time $j$. This interpretation corresponds to the typical "cause-specific" interpretation described in the survival literature (1). When $\delta = 2$ denotes being lost to follow-up (rather than being a competing event), this interpretation would be unproblematic, and one could thus easily translate the estimate of $\psi_{CS}$ into a measure of public health impact. This would not be problematic because $\delta = 2$ would represent a limitation of the study itself (e.g., not enough resources to track down individuals lost to follow-up), rather than an inherent characteristic of the system under study. Public health interventions aimed at reducing the exposure burden will have no impact on the distribution of censored observations because those observations do not exist outside of the context of the study. Hence, using inverse probability of censoring weights to address problems due to censored observations creates no complications for interpreting parameters.

However, when $\delta = 2$ denotes a competing risk, then this variable represents an inherent characteristic of the system under study. Any public health intervention aimed at reducing the exposure in a population will also affect the distribution of competing events, which will change the number of life-years saved (even though $\psi_{CS}$ can still be consistently estimated from the data). In fact, using inverse probability weights to deal with competing events often leads to an interpretation involving what would be observed under 2 interventions:

1. One that enables the modification of the exposure in the population, and
2. One that prevents competing events from occurring.

It will often be difficult to identify interventions that can altogether prevent the occurrence of competing events (such as death due to acute myocardial infarction or cerebrovascular disease). Therefore, relying on this interpretation can lead to ill-defined counterfactual outcomes and thus violations of counterfactual consistency (7–9). As a result, interpreting parameters causally when competing events are censored has long been cautioned against (10).

A second interpretation of $\psi_{CS}$ when inverse probability of censoring weights are used was derived by Tchetgen Tchetgen et al. (11) to avoid assuming that competing events could be prevented. Under additional no-confounding assumptions, this approach leads to a causal interpretation in which the parameter is a combination of 1) the survivor average causal effect (discussed below) of the exposure directly on the outcome and 2) a function of population average and survivor average causal effects of the exposure through intermediate time-varying covariates (11). This framework permits a causal interpretation that does not require identifying an intervention that prevents competing events from occurring. However, under this interpretation, the parameter no longer represents a population average effect. In fact, the public health implications of this interpretation are not easy to articulate without involving simplifying parametric assumptions (11) because of the complex combination of population average and survivor average causal effects that it entails.

## SURVIVOR AVERAGE CAUSAL EFFECTS

Instead of invoking either of these limited interpretations of $\psi_{CS}$, one can estimate the survivor average causal effect $\psi_{SACE}$, which is the exposure effect among a subgroup of the population that would not have died from a competing event irrespective of their exposure history. The survivor average causal effect has recently been extended to time-varying exposures via a generalization of standard inverse probability weighted marginal structural models (12). However, these results can be applied to structural nested failure time models. To estimate $\psi_{SACE}$ from a structural nested failure time model, one need only generate an alternative set of inverse probability weights defined as:

$$sw_{SACE}(j) = \prod_{k=j}^{int(t)} \frac{P[C(k)=0|C(k-1)=Y(k-1)=0,\overline{A}(k-1)=\overline{1},\overline{W}(k-1)]}{P[C(k)=0|C(k-1)=Y(k-1)=0,\overline{A}(k-1),\overline{W}(k-1)]}.$$

The numerator of each factor of $sw_{SACE}(j)$ can be computed from a pooled logistic regression model such as:

$$\text{logit}\{P[C(k)=0|C(k-1)=Y(k-1)=0,\overline{A}(k-1),\overline{W}(k-1)]\}$$
$$= \beta_k + \beta_1 A(k-1) + \beta_2 A(k-2) + \beta_3 W(k-1) + \beta_4 W(k-2),$$

where the numerator factors of $sw_{SACE}(j)$ correspond to predicted values from this logistic model in which both $A(t-1)$ and $A(t-2)$ are set to 1 for all individuals in the sample. The denominator of $sw_{SACE}(j)$ can be obtained from the same model, with predicted values obtained under each person's observed covariate values. One can then fit a pooled logistic model weighted by $sw_{SACE}(j)$ to implement step 3 of the g-estimation procedure described above. When using these weights, the test statistic for step 3 must be based on a bootstrap (13) rather than a standard or robust variance estimator, which does not appropriately acknowledge estimation of the weights. Furthermore, one must bootstrap both the model for the numerator and denominator of the weights and the model for the exposure to obtain a consistent variance estimator for $\widehat{\psi}_{SACE}$.

The logic behind weighting to obtain the survivor average causal effects is as follows: Assuming there is no person in the population for whom the exposure is protective with respect to the competing risk, unexposed individuals are overrepresented relative to exposed individuals in every risk set given the harmful effects of exposure on survival. The modified weight is guaranteed to be 1 for individuals who are always exposed, but it will be less than 1 for an individual who remains unexposed, thus down-weighting these individuals' contributions to the analysis. This balances the risk of the competing event across individuals with different exposure histories, accounting for any bias due to competing events and yielding an estimate of the exposure's effect among those would have survived irrespective of their exposure history.

Using $sw_{SACE}(j)$ to estimate the survivor average causal effect requires that a number of assumptions hold, including sequential monotonicity and the concordant survivorship assumption for time-varying exposures (12). These assumptions entail use of cross-world counterfactuals, which have been the source of some controversy in a mediation analysis setting (14). Moreover, although survivor average causal effects are intuitively appealing (because causal effects are difficult to define among individuals who do not survive to experience the outcome), Joffe (2) has shown that they can sometimes be misleading when the exposure has no overall effect on the outcome.

## THE PARAMETRIC G-FORMULA

A second approach that can be used to estimate intervention-based effects for a failure time outcome is the parametric g-formula. This approach can be used to estimate a parameter $\psi_{SD}$ corresponding to a contrast of subdistribution functions under different exposure scenarios (15). Measures of occurrence based on subdistribution functions have often been use to deal with the complications introduced by competing events (1, 16).

Details on how to implement the parametric g-formula have been published previously (15, 17–19). Rather than estimating $\psi_{SD}$ and then translating the estimate into a measure of public health impact (such as the cumulative risk or number of life-years saved), the parametric g-formula can be used to obtain such an estimate directly (15, 20). Indeed, if the modeling assumptions are correct, the g-formula can provide parameter estimates that closely align with questions about public health impact (20). This close alignment, however, comes with the (sometimes hefty) cost of requiring correct specification of all models used to implement the approach (3, Web Appendix). Moreover, when the time-varying exposure of interest

has no effect on the outcome, the parametric g-formula may provide biased estimates due to the g-null paradox (21).

## BALANCING ASSUMPTIONS WITH INTERPRETATIONS

Several approaches in epidemiology are converging toward the goal of estimating parameters for population impact (20, 22, 23). These parameters are appealing because they more precisely correspond to some action that might be taken to improve population health. However, each approach comes with its own strengths ( public health relevance of the interpretation) and limitations (required assumptions). Competing events have long been recognized as difficult obstacles to overcome in pursuit of this goal (10). Ultimately, the choice of how to handle competing events should be based on thoughtful consideration of the tradeoff between the assumptions required and the interpretation permitted by each approach.

## REFERENCES

1. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol.* 2009;170(2):244–256.
2. Joffe M. Principal stratification and attribution prohibition: good ideas taken too far. *Int J Biostat.* 2011;7(1):Article 35.
3. Picciotto S, Peters A, Eisen EA. Hypothetical exposure limits for oil-based metalworking fluids and cardiovascular mortality in a cohort of autoworkers: structural accelerated failure time models in a public health framework. *Am J Epidemiol.* 2015; 181(8):563–570.
4. Naimi AI, Richardson DB, Cole SR. Causal inference in occupational epidemiology: accounting for the healthy worker effect by using structural nested models. *Am J Epidemiol.* 2013; 178(12):1681–1686.
5. Robins JM. Structural nested failure time models. In: Andersen P, Keiding N, eds. *The Encyclopedia of Biostatistics.* Chichester, UK: John Wiley and Sons; 1998:4372–4389.
6. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology.* 2000;11(5): 561–570.
7. Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? *Epidemiology.* 2009; 20(1):3–5.
8. VanderWeele TJ. Concerning the consistency assumption in causal inference. *Epidemiology.* 2009;20(6):880–883.
9. Hernán MA, VanderWeele TJ. Compound treatments and transportability of causal inference. *Epidemiology.* 2011;22(3): 368–377.
10. Prentice RL, Kalbfleisch JD, Peterson AV Jr, et al. The analysis of failure times in the presence of competing risks. *Biometrics.* 1978;34(4):541–554.
11. Tchetgen Tchetgen EJ, Glymour MM, Shpitser I, et al. Rejoinder: to weight or not to weight? On the relation between inverse-probability weighting and principal stratification for truncation by death. *Epidemiology.* 2012;23(1):132–137.
12. Tchetgen Tchetgen EJ. Identification and estimation of survivor average causal effects. *Stat Med.* 2014;33(21): 3601–3628.
13. Efron B, Tibshirani R. *Introduction to the Bootstrap.* Boca Raton, FL: Chapman & Hall/CRC; 1993.
14. Naimi AI, Kaufman JS, MacLehose RF. Mediation misgivings: ambiguous clinical and public health interpretations of natural direct and indirect effects. *Int J Epidemiol.* 2014;43(5): 1656–1661.
15. Cole SR, Richardson DB, Chu H, et al. Analysis of occupational asbestos exposure and lung cancer mortality using the g formula. *Am J Epidemiol.* 2013;177(9):989–996.
16. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc.* 1999; 94(446):496–509.
17. Daniel RM, Cousens SN, De Stavola BL, et al. Methods for dealing with time-dependent confounding. *Stat Med.* 2013; 32(9):1584–1618.
18. Robins J, Hernán M. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice G, Davidian M, Verbeke G, et al., eds. *Advances in Longitudinal Data Analysis.* Boca Raton, FL: Chapman & Hall; 2009:553–599.
19. Keil AP, Edwards JK, Richardson DB, et al. The parametric g-formula for time-to-event data: intuition and a worked example. *Epidemiology.* 2014;25(6):889–897.
20. Westreich D. From exposures to population interventions: pregnancy and response to HIV therapy. *Am J Epidemiol.* 2014; 179(7):797–806.
21. Robins JM, Wasserman L. Estimation of effects of sequential treatments by reparameterizing directed acyclic graphs. In: Geiger D, Shenoy P, eds. *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence.* San Francisco, CA: Morgan Kaufmann; 1997:409–420.
22. Naimi AI, Moodie EE, Auger N, et al. Stochastic mediation contrasts in epidemiologic research: interpregnancy interval and the educational disparity in preterm delivery. *Am J Epidemiol.* 2014;180(4):436–445.
23. Young JG, Hernán MA, Robins JM. Identification, estimation and approximation of risk under interventions that depend on the natural value of treatment using observational data. *Epidemiol Methods.* 2014;3(1):1–19.