OXFORD
JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

American Journal of Epidemiology, 2025, **194**, 2150–2153
https://doi.org/10.1093/aje/kwaf090
Advance access publication date April 29, 2025
Invited Commentary

Descriptive social epidemiology: putting the question before the methods

S. Ghazaleh Dashti*,[1,2] and Margarita Moreno-Betancur[1,2]

[1]Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Melbourne, VIC, Australia
[2]Department of Pediatrics, University of Melbourne, Melbourne, VIC, Australia
*Corresponding author: S. Ghazaleh Dashti, Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, 50 Flemington Rd, Parkville, Melbourne, VIC 3052 (ghazaleh.dashti@unimelb.edu.au)

Abstract

In studies describing socioeconomic inequities in health outcomes, the choice of estimand and the planned analytic approach are central to the interpretability and policy relevance of findings. In this commentary, we aimed to highlight this by revisiting some of the choices made in the article by Eisenberg-Guyot and Renson (Am J Epidemiol. 2025;194(8):2440-2444) and presenting a discussion on how these choices impact the meaning of the inequity estimates obtained, in particular what they tell us about the world. These choices concern (1) the estimand in the presence of competing events (ie, the measure of inequity to be estimated), (2) the timescale with time-to-event outcomes, and (3) covariate adjustment. When describing inequities in health outcomes in the presence of competing events, it is indispensable to start with a clear research question and choosing the most relevant estimand to address it. This should then be followed by a study design and data analytic approaches that appropriately target that estimand. Following these steps will help avoid findings with obscure or misleading interpretation.

This article is part of a Special Collection on Methods in Social Epidemiology.

Key words: descriptive epidemiology; health disparities; health inequities; mortality; competing risks; social epidemiology.

Editor's note: *The opinions expressed in this article are those of the authors and do not necessarily reflect the views of the* American Journal of Epidemiology.

Introduction

In a recent article in the *American Journal of Epidemiology*,[1] Eisenberg-Guyot and Renson sought to illustrate the application of g-computation for describing inequities by social class (defined based on employment status) in cause-specific mortality, where death from other causes are competing events. The study used data from the National Health Interview Survey (NHIS), a repeated cross-sectional survey of a nationally representative sample of the noninstitutionalized U.S. population that was linked to the National Death Index. For each cause of death and social class, the authors estimated the age- and gender-adjusted cause-specific cumulative incidence function among individuals aged 18 to 64 years at the time of survey data collection (baseline). Follow-up time, starting at baseline and ending at year of death or 2020, whichever occurred first, was used as the underlying timescale.

Describing socioeconomic inequities in mortality and other health outcomes across population subgroups is key to informing public health and policy decision-making. In this context, clarity in the question asked and the associated estimand (ie, the measure of inequity to be estimated), and a study design and data analytic approach that are appropriately targeted to that estimand, are essential ingredients to ensuring that study findings have clear and policy-relevant interpretations so they can be used effectively for promoting equity. In this commentary, we revisit three interrelated estimand and analytic choices made in the article by Eisenberg-Guyot and Renson concerning (1) the estimand in the presence of competing events, (2) the timescale with time-to-event outcomes, and (3) covariate adjustment. Our commentary is not intended as a criticism of all these choices. Rather, our aim is to present an expanded discussion regarding the implications of these decisions for the interpretation and policy relevance of findings in this as well as other studies describing socioeconomic inequities in health outcomes.

Estimand in the presence of competing events

In a seminal paper, Young et al. used a formal counterfactual framework to clarify the interpretation, identifiability, and estimation of common causal estimands in the survival analysis literature in the presence of competing events.[2] The authors highlighted that there does not seem to be an ideal or one-size-fits-all estimand choice in the presence of competing risks in the causal inference setting, and emphasized the importance of the research question and context in guiding the choice between different estimands. Much of that discussion is also relevant for descriptive questions. In this section, we briefly review the descriptive analogs of the causal estimands described by Young

Received: November 21, 2024. **Accepted:** April 16, 2025

et al. in the context of describing socioeconomic inequities in cause-specific mortality, where death from other causes is a competing event. We focus on interpretation, policy relevance, and comparability across different settings of these estimands. We begin by introducing some notation.

Notation

Let X denote the socioeconomic grouping variable, measured at time 0 (ie, start of follow-up) and assumed binary for simplicity, where $X = 1$ if individual is in the disadvantaged group and $X = 0$ otherwise. Let Y denote the outcome (cause-specific death), where $Y = 1$ if the cause-specific death occurs during follow-up and $Y = 0$ if not; and D the competing event (death from other causes), which upon occurring during follow up ($D = 1$) prevents Y ($Y = 0$). Let $Y^{D=0}$ denote the potential outcome under elimination of death from other causes. For simplicity, we assume no loss to follow-up.

Estimand (i): Direct disparity

Direct disparity is the analog of the so-called *direct effect* causal estimand,[2] and is defined as:

$$\mathbf{P}\left(Y^{D=0} = 1|X = 1\right) - \mathbf{P}\left(Y^{D=0} = 1|X = 0\right)$$

This estimand treats the competing event as a censoring event and can be interpreted as the difference in the risk of cause-specific death in those disadvantaged versus not disadvantaged, under a hypothetical intervention that eliminates death from other causes.

Eisenberg-Guyot and Renson refer to direct disparity as the "conditional risk." As pointed out by them, the policy relevance of this estimand is questionable because such a scenario of "elimination of death by other causes" is implausible. Formally, the counterfactual outcome $Y^{D=0}$ is ill-defined as there is no single well-defined intervention by which death from other causes could be eliminated. On the other hand, considering the scenario under elimination of competing events renders the estimand comparable across populations with different risks of competing events, which could be a desirable feature in public health monitoring settings.

Estimand (ii): Total disparity

Total disparity is the analog of the so-called *total effect* causal estimand,[2] and is defined as:

$$\mathbf{P}(Y = 1|X = 1) - \mathbf{P}(Y = 1|X = 0)$$

This estimand can be interpreted as the difference in risk of cause-specific death in those disadvantaged versus not disadvantaged, in the presence of competing events, without their hypothetical elimination. Of note, $\mathbf{P}(Y = 1|X = x)$ is equal to the cause-specific cumulative incidence function for those with $X = x$.

This estimand is what Eisenberg-Guyot and Renson called the "unconditional risk" and chose as estimand in their study. Contrary to the direct disparity, the total disparity does not invoke an implausible intervention and may, therefore, be of greater relevance for policy. Indeed, the estimand "sticks to this world,"[3] describing it as is rather than under a hypothetical intervention, which is generally the goal in descriptive research. However, its interpretation is complicated by the fact that it depends on the risk of death from other causes, both overall in the population and by socioeconomic disadvantage. Importantly, socioeconomic inequities in the risk of the event of interest and competing events may differ and even exhibit opposite directions.[2] The authors allude to this issue in their speculation that the reason for observing a lower risk of Alzheimer's disease mortality among those not in the labor force (the disadvantaged group) compared with incorporated business owners might be that the former group was more likely to die from other causes at an earlier age, as shown by their estimates. This feature of total disparity makes it difficult to compare estimates across populations (eg, defined by time-period or geographical region) with differing distributions of the competing event, which may hinder its usefulness in some cases.

Estimand (iii): Disparity in survivors

Disparity in survivors is the analog of the so-called *survivor average causal effect,*[2] and is defined as:

$$\mathbf{P}(Y = 1|X = 1, D = 0) - \mathbf{P}(Y = 1|X = 0, D = 0)$$

The estimand can be interpreted as the difference in risk of cause-specific death in those disadvantaged versus not disadvantaged, but only among those who survive death from other causes during follow-up.

Like the total disparity, the disparity in survivors does not invoke an implausible intervention. However, the policy relevance of this estimand is questionable. Firstly, the subpopulation it describes is not identifiable at time 0. Secondly, the estimand may paint a distorted picture of inequities when there are also socioeconomic inequities in death from other causes, and there exist common risk factors between cause-specific death and death from other causes. For example, smoking is a common risk factor for both death from other causes and cancer-related deaths. Individuals in the disadvantaged group are more likely to die from other causes, and those who survive tend to have some protective factors, such as not being smokers. As a result, comparing risk of cancer-related deaths between the disadvantaged and advantaged groups among those who survive other causes of death inherently compares a disadvantaged population with more protective factors (eg, nonsmokers) to an advantaged population where this so-called "depletion of susceptibles" has occurred to a lesser extent. If there are socioeconomic inequities in the risk factor, as is likely for smoking, then the difference will be even more pronounced. Either way, the resulting picture of disparities captured by this estimand may be misleading.[4] Finally, even in the unlikely absence of inequities in death from other causes, the disparity estimate in the subpopulation of those who survive death from other causes is unlikely to be generalizable to or comparable across other population groups with differing distributions of the common risk factors for death from other causes and cause-specific death.[5,6]

In summary, of the three descriptive estimands described above, we agree with Eisenberg-Guyot and Renson that the total disparity was the likely preferred estimand in their setting. But, given the issues described, we wish to caution the readers against reporting and interpreting this estimand in silo in similar settings, without concurrent consideration of inequities in risk of death from other causes.[7]

Choice of the timescale with time-to-event outcomes

Using follow-up time as the timescale, Eisenberg-Guyot and Renson estimated cause-specific cumulative incidence function by social class over 34 years from baseline (the time of survey data collection) among individuals aged 18-64 years old.[1] Pooling age groups in this way may sometimes be motivated by statistical efficiency concerns, although in this case the sample sizes were quite large. On the other hand, this choice raises a challenge for the meaning of results because the risk of death over 34 years would inherently be different for an 18-year-old compared with a 64-year-old individual. Therefore, even with age adjustment (which, as described below, raises its own issues), pooling the age groups in this way makes it hard to interpret and translate the findings. An alternative would be to use attained age as the timescale, which would arguably lead to more readily interpretable estimates (eg, cumulative risk of heart disease death over lifetime since age 18 years).[8] Stratifying the analyses by age group at baseline would have been another way of making the results more interpretable. Additionally, if inequities by social class were expected to have changed over time, further stratifying the analyses by time period would have also helped make estimates more meaningful.

In general, we encourage researchers to think explicitly about the interpretation of the descriptive estimands chosen, and what they tell us about the world, to ensure that findings can be used in a meaningful way to inform public health and policy decision-making.

Covariate adjustment

The rationale of the paper by Eisenberg-Guyot and Renson is to present a methodological approach (g-computation) that is useful when adjustment is necessary, especially high-dimensional adjustment, as the authors mention. Yet, we note that in the context of describing socioeconomic inequities in health outcomes, high-dimensional adjustment may rarely be defensible.[6,9-11]

In this setting, adjustment for certain covariates might be warranted to address selection bias, for example due to sampling design or missing data, and thus enable unbiased estimation of the chosen inequity estimand.[6,12] Adjustment might also be justified to enable comparisons between socioeconomic groups that differ in the distribution of certain covariates but where those distributional differences are considered "fair" and therefore not contributors to inequities (the so-called "outcome-allowable" covariates).[4,13,14] In this case, the estimand may be defined as pertaining to a target population where those distributional differences are absent (for example, where the covariate distribution in each group is the same as in the total population, ie, the pooled population across all socioeconomic groups).[15] Carefully considering such covariates, and explicitly stating the value judgments underpinning this choice, are essential for clearly defining the inequity estimand.[13]

Eisenberg-Guyot and Renson present all estimates adjusted for age and gender without clearly outlining the justification. Implicitly, this adjustment defines an estimand pertaining to a target population where the age and gender distributions within each socioeconomic group align with those of the total population. Whether this estimand definition is appropriate is based on value judgments, and ultimately context-dependent.[4,11,13,14] For example, suppose those in the disadvantaged group are, on average, younger than those in the advantaged groups. This younger age distribution may partly be responsible for their higher risk of certain causes of death that are more common in younger people, such as motor vehicle accidents. In this case, adjusting for age estimates disparities in a scenario where all socioeconomic groups share the same age distribution as the total population, which could mask inequities in deaths from motor vehicle accidents.

Age and gender are of course key factors to consider when studying mortality and other health outcomes, but considering them explicitly when defining the estimand (eg, treating them as stratification variables, or considering their distribution in the target population) before determining the need for adjustment may help produce more meaningful results in settings like this one. Similar considerations would apply to other covariates.

Conclusion

In recent years, the field of epidemiology has benefited from a growing body of literature advocating for and providing guidance for more rigorous descriptive epidemiology (see for example[4,6,10,11,16-18]). In this commentary, we draw upon this literature, as well as the causal inference literature, to provide an expanded discussion on how estimand and analytic choices like those made in the study by Eisenberg-Guyot and Renson could influence the interpretation of findings in studies describing inequities in health outcomes subject to competing events. An overall principle for these studies is ensuring that the research question and the associated estimand are clear and aligning the data analytic approaches with the estimand, so that the findings provide a meaningful description of the world.

Acknowledgments

Study investigators, conference presentations, preprint publication information, thanks.

Funding

The authors were supported by National Health and Medical Research Council (NHMRC) Investigator Grants: S.G.D. (ID 2027171) and M.M.-B. (ID 2009572).

Conflict of interest

The authors have no conflicts of interest to disclose.

References

1. Eisenberg-Guyot J, Renson A. Competing classes confront competing risks: unraveling mortality inequities with parametric g-computation. *Am J Epidemiol*. 2025;194(8):2440-2444. https://doi.org/10.1093/aje/kwae417
2. Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, et al. A causal framework for classical statistical estimands in failure-time settings with competing events. *Stat Med*. 2020;39(8):1199-1236. https://doi.org/10.1002/sim.8471
3. Andersen PK, Geskus RB, de Witte T, et al. Competing risks in epidemiology: possibilities and pitfalls. *Int J Epidemiol*. 2012;41(3):861-870. https://doi.org/10.1093/ije/dyr213
4. Jackson JW, Hsu Y-J, Greer RC, et al. The target study: a conceptual model and framework for measuring disparity. *Sociol Methods Res*. 2025. https://doi.org/10.1177/00491241251314037
5. Hernan MA. Invited commentary: selection bias without colliders. *Am J Epidemiol*. 2017;185(11):1048-1050. https://doi.org/10.1093/aje/kwx077
6. Lesko CR, Fox MP, Edwards JK. A framework for descriptive epidemiology. *Am J Epidemiol*. 2022;191(12):2063-2070. https://doi.org/10.1093/aje/kwac115
7. Latouche A, Allignol A, Beyersmann J, et al. A competing risks analysis should report results on all cause-specific hazards

and cumulative incidence functions. *J Clin Epidemiol*. 2013;66(6): 648-653. https://doi.org/10.1016/j.jclinepi.2012.09.017
8. Vyas MV, Fang J, Kapral MK, et al. Choice of time-scale in time-to-event analysis: evaluating age-dependent associations. *Ann Epidemiol*. 2021;62:69-76. https://doi.org/10.1016/j.annepidem.2021.06.006
9. Kaufman JS. Statistics, adjusted statistics, and maladjusted statistics. *Am J Law Med*. 2017;43(2-3):193-208. https://doi.org/10.1177/0098858817723659
10. Lesko CR, Zalla LC. Rigorous descriptive epidemiology for health justice. *Epidemiology*. 2023;34(6):838-840. https://doi.org/10.1097/EDE.0000000000001658
11. Zalla LC, Lesko CR. Re: "eliminating ambiguous treatment effects using estimands". *Am J Epidemiol*. 2023;192(6):1029-1030. https://doi.org/10.1093/aje/kwad070
12. Lu H, Howe CJ, Zivich PN, et al. The evolution of selection bias in the recent epidemiologic literature-a selective overview. *Am J Epidemiol*. 2025;194(3):580-584. https://doi.org/10.1093/aje/kwae282
13. Chang TH, Nguyen TQ, Jackson JW. The importance of equity value judgments and estimator-estimand alignment in measuring disparity and identifying targets to reduce disparity. *Am J Epidemiol*. 2024;193(3):536-547. https://doi.org/10.1093/aje/kwad209
14. Jackson JW. Meaningful causal decompositions in health equity research: definition, identification, and estimation through a weighting framework. *Epidemiology*. 2021;32(2):282-290. https://doi.org/10.1097/EDE.0000000000001319
15. Moreno-Betancur M. The target trial: a powerful device beyond well-defined interventions. *Epidemiology*. 2021;32(2):291-294. https://doi.org/10.1097/EDE.0000000000001318
16. Conroy S, Murray EJ. Let the question determine the methods: descriptive epidemiology done right. *Br J Cancer*. 2020;123(9): 1351-1352. https://doi.org/10.1038/s41416-020-1019-z
17. Fox MP, Murray EJ, Lesko CR, et al. On the need to revitalize descriptive epidemiology. *Am J Epidemiol*. 2022;191(7):1174-1179. https://doi.org/10.1093/aje/kwac056
18. Thurber KA, Thandrayen J, Maddox R, et al. Reflection on modern methods: statistical, policy and ethical implications of using age-standardized health indicators to quantify inequities. *Int J Epidemiol*. 2022;51(1):324-333. https://doi.org/10.1093/ije/dyab132

American Journal of Epidemiology, 2025, **194**, 2150–2153
https://doi.org/10.1093/aje/kwaf090
Invited Commentary