Weighing up the Dead and Missing

Reflections on Inverse-probability Weighting and Principal Stratification to Address Truncation by Death

Basile Chaix,[a,b] *David Evans,*[a,b,c] *Juan Merlo,*[d] *and Etsuji Suzuki*[e]

In the present issue of the Journal, Weuve et al[1] provide an innovative analysis of the effect of smoking on cognitive decline in an aged population followed during 5 successive waves in the Chicago Health and Aging Project. As the authors describe with a signed directed acyclic graph,[2,3] selection bias may distort the relationship between smoking and cognitive decline owing to the effect of both the exposure and the determinants of cognitive decline on attrition in the study.[4,5]

The authors recognize that adjustment for the baseline or intermediate values of the cognitive score could generate other biases,[6,7] and therefore implement inverse-probability weighting (IPW) methods.[8] Such an approach compensates for the underrepresentation of smoking and other determinants of cognitive decline associated with attrition, through weights that are inversely related to the probability of remaining alive and in the study. Key strengths of the authors' analytic strategy include the separate modeling of 2 sources of attrition (mortality and study drop-out) and the application of inverse-probability-of-continuation weights at a person-wave level to account for the increasing selection that distorts the sample over successive study waves.

A Highly Model-dependent Estimate of the Smoking–cognitive Decline Relationship

Although technically innovative, the present application of IPW is particularly courageous, given that only 20% of the baseline participants were still present in the study at wave 5, and that the models for attrition showed acceptable, but limited, performance. The relatively broad range of stabilized weights (from 0.2 to 14.7) shows that weighting led to a dramatic recomposition of the sample.

In replacing dead participants by cloning the living, IPW generates a sample in which participants are not allowed to die. Moreover, IPW attributes particularly high weights to the participants most likely to die, ie, to people with poor health characteristics associated with death in the attrition model. In doing so, IPW not only prevents people from dying but also artificially maintains the lives of people in very poor health—arguably a form of statistical cruelty.

This structure of the IPW-based pseudo-population likely explains the most striking consequence of weighting in this study, namely that the rate of cognitive decline increased for both never-smokers and smokers (and only slightly more so for the latter). The up-weighting of the person-wave observations with poor health characteristics, increasing over the waves of the study, leads to an estimate of the rate of cognitive decline, as if this elderly population was allowed to get sicker and sicker without being allowed to die. This

From the [a]Inserm, U707, Paris, France; [b]Université Pierre et Marie Curie-Paris6, UMR-S 707, Paris, France; [c]EHESP School of Public Health, Rennes, France; [d]Unit for Social Epidemiology, CRC, Faculty of Medicine, Lund University, Malmö, Sweden; and [e]Department of Epidemiology, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama, Japan.

Correspondence: Basile Chaix, UMR-S 707, Faculté de Médecine Saint-Antoine, 27 rue Chaligny, 75012 Paris, France. E-mail: chaix@u707.jussieu.fr.

Editors' note: Related articles appear on pages 119 and 132.

ISSN: 1044-3983/12/2301-0129

DOI: 10.1097/EDE.0b013e3182319159

sharper cognitive decline observed after weighting reflects the effect of age after correcting for the selective removal (through death) of people with particularly low cognitive scores.

The overall effect of smoking on cognitive decline is a combination of effects in the various sociodemographic and health substrata of the sample. The observed effect across these substrata is likely to vary, with the variability being amplified by sampling error in each stratum. Despite the use of stabilized weights, heavily changing the relative importance of the various substrata can be a risky undertaking, as it attributes higher importance to sample subgroups with the fewest participants owing to attrition, in which the estimated effect is particularly vulnerable to sampling error.

Truncation by Death as a Missing Data Problem?

Of the 2 sources of attrition investigated, mortality was of particular importance for the expected bias of the smoking effect because most of that attrition was due to mortality (68%), and also because smoking was more clearly associated with mortality than with study drop-out.

Interestingly, in another thought-provoking article, the same authors indicated that "no amount of aggressive follow up can permit researchers to observe the events that would have occurred in those who died,"[7 p. 208] but they argued that "this problem can also be conceptualized as a problem of missing data, where subjects who die prematurely are missing data on the outcome."[7 p. 210]

Defining death as a competing risk of the ability to observe cognitive status, it is a matter of debate whether truncation by death can be treated as a missing-data problem and so addressed through IPW. It has been emphasized that if a subject dies before the outcome is assessed (cognitive status at a given wave), then the outcome is not simply missing but undefined.[9] It thus seems legitimate to ask whether it is conceptually relevant to generate pseudo-data for person-wave observations for which the outcome is undefined because of prior truncation by death.[10] It is important to emphasize that in the case of study drop-out, IPW attempts to reconstruct an unobserved but real target population for which the outcome is defined. By contrast, in the case of truncation by death, IPW generates a pseudo-population that has no real existence. Even if study drop-out and truncation by death have an identical relationship with smoking and cognitive decline in a directed acyclic graph, some investigators may be reluctant to overcome the undefinability of the outcome after truncation by death by creating a fictional statistical population, with no correspondence to the prevalence of poor health status in the real world (where sick people are usually allowed to die). Public-health decision makers in particular may be interested in knowing the effective impact of smoking on cognition in the real population, given that elderly smoking survivors may be few, rather than in a pseudo-population that exists only statistically.

The Principal-stratification Approach to Truncation by Death

The undefinability of the outcome for observations truncated by death has implications for formulating the causal effect of interest. Methodologists have emphasized that the potential outcomes needed to define the causal effect of interest cannot be defined among participants who would have died in at least one of the alternative exposure groups (eg, the counterfactual cognitive score would be undefinable for a nonsmoking survivor who would have died under the smoking scenario).[9,10] Assuming that people who survive as smokers would also survive as nonsmokers, the so-called principal-stratification approach[11,12] invites us to distinguish 3 different "principal strata" in this study: (i) people who would survive until outcome measurement, whether they smoke or not ("always-survivors"); (ii) those who would die before measurement, whatever their smoking status; and (iii) those who would survive as nonsmokers but would die as smokers. As long as we are reluctant to impute outcomes truncated by death, the causal effect of smoking on cognitive status is definable only in the always-survivor stratum where the 2 potential outcomes under the smoking and nonsmoking scenarios are defined.

However, a different subset of participants is included in the 2 exposure groups when comparing cognitive decline between smokers and nonsmokers among the effective survivors: under the hypothesis stated earlier, the smoker group comprises only always-survivors while the nonsmoker group comprises both always-survivors and people who would have died before the cognitive assessment had they smoked. The principal-stratification approach therefore suggests that the only fair definition of the causal effect of interest is the "survivor average causal effect," that is, the effect defined among always-survivors,[10] which is a different target population from that considered through IPW. The important point is that conditioning on the principal stratum through restriction to always-survivors is not as problematic as conditioning on "effective survival" because the principal stratum, as a fixed and predetermined characteristic of individuals, is not a postexposure variable—unlike effective survival status.[13]

Obviously, the critical problem is that investigators do not know the principal stratum to which each participant belongs. Related empirical approaches are based either on simple sensitivity analyses,[9] or on more complex estimation strategies that attempt to gain information on the underlying principal strata by considering additional covariates or the distributional shape of the outcome.[10,14] The latter invites investigators to speculate on the magnitude of bias resulting from the nonidentifiability of the always-survivors. These empirical applications often do not provide a single estimate

of effect, but allow one to derive bounds for the effect or to derive various estimates according to the value of a sensitivity-analysis parameter.[13] Although principal stratification provides an appealing conceptual framework for causal effects in the case of truncation by death, its empirical usefulness to clinical and public health practice is debatable. Its problems include the inherent impossibility to sort out participants' principal strata, the restriction of the target population to always-survivors, and the limited informativeness of the range of estimates provided.

Overall, the thoughtful application of IPW in the present article is of interest as a method for correcting mortality-related attrition bias, especially in the absence of an empirically credible alternative. However, we find it useful to emphasize the strong model-dependence of the IPW estimate provided, and the conceptual concerns with the pseudo-population created to address truncation by death. In the absence of a consensus approach, researchers obviously need to weigh up the current alternatives in light of the particular research question they pose.

ABOUT THE AUTHORS

BASILE CHAIX, as a permanent researcher at Inserm, coordinates the RECORD Cohort Study (which has its own set of selection problems!) on neighborhood health effects. DAVID EVANS is a doctoral student interested in estimating effects useful for policy and clinical decisions from observational data (and also works in pharmacoepidemiology). JUAN MERLO, as a Professor, is leading the Unit for Social Epidemiology at the Lund University and working mainly on life course and contextual analyses. ETSUJI SUZUKI is Assistant Professor of Epidemiology at Okayama University. His primary research interest concerns improving causal interpretations of observational studies.

REFERENCES

1. Weuve J, Tchetgen Tchetgen EJ, Glymour MM, et al. Accounting for bias due to selective attrition: the example of smoking and cognitive decline in older adults. *Epidemiology.* 2012;23:119–128.
2. VanderWeele TJ, Hernán MA, Robins JM. Causal directed acyclic graphs and the direction of unmeasured confounding bias. *Epidemiology*. 2008;19:720–728.
3. Chaix B, Billaudeau N, Thomas F, et al. Neighborhood effects on health: correcting bias from neighborhood effects on participation. *Epidemiology*. 2011;22:18–26.
4. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15:615–625.
5. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*. 2003;14:300–306.
6. Glymour MM, Weuve J, Berkman LF, Kawachi I, Robins JM. When is baseline adjustment useful in analyses of change? An example with education and cognitive change. *Am J Epidemiol*. 2005;162:267–278.
7. Glymour MM, Weuve J, Chen JT. Methodological challenges in causal research on racial and ethnic patterns of cognitive trajectories: measurement, selection, and bias. *Neuropsychol Rev*. 2008;18:194–213.
8. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11:550–560.
9. Chiba Y, VanderWeele TJ. A simple method for principal strata effects when the outcome has been truncated due to death. *Am J Epidemiol*. 2011;173:745–751.
10. Rubin DB. Causal inference through potential outcomes and principal stratification: application to studies with "censoring" due to death. *Statist Sci*. 2006;21:299–309.
11. Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics*. 2002;58:21–29.
12. Pearl J. Principal stratification - a goal or a tool? *Int J Biostat*. 2011;7:20. doi: 10.2202/1557-4679.1322.
13. VanderWeele TJ. Principal stratification — uses and limitations. *Int J Biostat*. 2011;7:28. doi: 10.2202/1557-4679.1329.
14. Hayden D, Pauler DK, Schoenfeld D. An estimator for treatment comparisons among survivors in randomized trials. *Biometrics*. 2005; 61:305–310.