American Journal of Epidemiology

Vol. 182, No. 10
DOI: 10.1093/aje/kwv215
Advance Access publication:
October 26, 2015

Point-Counterpoint

Counterpoint: Epidemiology to Guide Decision-Making: Moving Away From Practice-Free Research

Miguel A. Hernán*

* Correspondence to Dr. Miguel A. Hernán, Harvard T. H. Chan School of Public Health, Department of Epidemiology, 677 Huntington Avenue, Boston, MA 02115 (e-mail: miguel_hernan@post.harvard.edu).

Initially submitted April 1, 2015; accepted for publication July 2, 2015.

Analyses of observational data aimed at supporting decision-making are ideally framed as a contrast between well-defined treatment strategies. These analyses compare individuals' outcomes from the start of the treatment strategies under consideration. Exceptions to this synchronizing of the start of follow-up and the treatment strategies may be justified on a case-by-case basis.

causal inference; decision-making; left truncation; survival curves

Editor's note: *Counterpoints to this article appear on pages 826 and 840, and a response appears on page 846.*

Suppose you and your colleagues are reading an epidemiology blog. The first post says, "When analyzing a randomized trial, the first 5 years of follow-up can be safely deleted from the data set." You all think this is nonsensical. Excluding the early follow-up seems a bad idea for multiple reasons, not the least because individuals who develop the outcome, die, or are lost to follow-up during the first 5 years would be automatically excluded from the analyses.

The second post says, "The previous post is silly. Why delete part of your data set? Now, if a computer malfunction erased the first 5 years of follow-up of the randomized trial, then it would be safe to analyze the remaining data without trying to recover the erased data." You and your colleagues smile. There is no difference between actively deleting your data and having them missing for other reasons.

Your colleagues write the third post: "The previous posts reflect a lack of understanding of randomized trials. You can ignore the first 5 years of follow-up only in observational studies." Now *you* are perplexed. If your colleagues agreed that ignoring the first 5 years of follow-up was dangerous in randomized trials, how can they think that it is safe in observational studies?

You might then be perplexed by Vandenbroucke and Pearce's (1) message to epidemiologists: Let us not refrain from conducting observational studies in which the first part of the follow-up is missing. You would not be alone in your confusion. A strong warning against such studies was already voiced by Feinstein (2, 3) in the 1970s. A nontruncated follow-up was taken for granted by the causal inference methods proposed by Rubin (4, 5) for time-fixed treatments and by Robins (6–8) for time-varying treatments. In 2003, Ray (9) provided compelling examples of potential biases that might arise when the first part of the follow-up is missing. More recently, some of us (10) have also privileged studies with an intact early follow-up.

All of these authors argue for the synchronization of the start of follow-up and the start of treatment of exposure, that is, for the use of incident exposures in observational research. Vandenbroucke and Pearce point out that "studying incident exposures may be necessary in some situations, but it is not always necessary and is not the preferred option in many instances" (1, p. 000). It is hard to disagree with the last sentence. (In fact, one can replace "studying incident exposures" with almost anything. Try "starting a war" or "being polite.") The question is, then, "Under which circumstances will omission of the first part of the follow-up be appropriate in both observational studies and randomized trials?"

This counterpoint article discusses Vandenbroucke and Pearce's response to that question. Before doing so, let us define the class of studies for which this discussion is relevant: observational studies that try to answer the same causal questions as (hypothetical) randomized experiments.

EPIDEMIOLOGIC RESEARCH TO SUPPORT PRACTICE

Most randomized trials are designed to assist decision-making. The goal is to help decision-makers—patients, clinicians, policy-makers, public health officers, regulators—decide among several possible strategies. This counterpoint article is concerned with observational studies that, like randomized trials, are designed to guide clinical, policy, public health, regulatory, or other decisions. For some exposures (e.g., complex clinical strategies, long-term dietary and lifestyle changes, air pollution), randomized experiments are impractical so, realistically, we can only conduct observational studies. For other exposures (e.g., some medical interventions, short-term dietary and lifestyle changes, some components of the health system), we have a choice between randomized experiments and observational studies.

Regardless of whether the studies are randomized or observational, the design and analysis of studies to assist decision-making are determined by practical needs (11), with the starting point: "A decision needs to be made." Should patients with human immunodeficiency virus be treated as soon as they are diagnosed or later in the course of the disease? Should I start eating more fish? Should we recommend a radiographic examination of the breasts (a mammogram) annually or every other year? Does a change in reimbursement systems affect patients' outcomes? The most helpful answers to these questions are provided by studies that compare well-defined strategies and compare their outcomes since the strategy was initiated.

For example, consider a trial in which human immunodeficiency virus-positive individuals are randomly assigned to immediate initiation of 1 of 2 treatment strategies. To compare the 5-year risk of death under each strategy, trial researchers will typically construct cumulative incidence (risk) curves or their complementary survival curves. Time zero in these curves is the initiation of the treatment strategy. In addition, trial researchers often estimate the average hazard or rate ratio during the follow-up, which is a convenient summary (it is a single number and therefore well suited for the abstract of the article). The average hazard ratio, however, is not the most informative measure for decision-making, as it does not inform us about the absolute risk at a particular time point, about the time when the survival curves started to diverge, or whether the curves ever crossed. These are all key pieces of information for decision-making.

Survival curves are more informative than hazard ratios in observational studies designed to assist decision-makers. The primary goal of these observational studies is, then, the estimation of survival (or risk) over time under each treatment strategy. This goal is hard to accomplish in analyses of observational data that lack the first part of the follow-up, that is, data with left truncation.

THE PROBLEM OF LEFT TRUNCATION

Consider again a randomized trial that compares the survival between groups who initiated different treatments. An individual's follow-up is left truncated if a period after initiation of her treatment strategy is omitted from the analysis. An individual with a left-truncated follow-up will appear in the analytical data set only if she did not develop the outcome during the missing period (or, for nonfatal outcomes, if the analyst is unaware that she developed the outcome).

A more precise definition of left truncation requires a more precise characterization of the treatment strategies that are being compared. In a randomized trial, we may say that we compare the survival curves under "treatment A" and "treatment B," but treatment A is shorthand for something like "initiate treatment A at baseline and take it continuously unless toxicity arises, in which case you will discontinue treatment A and initiate an alternative therapy to be decided by you and your physician according to your clinical status and preferences at that time," and similarly for treatment B. Assignment to an exposure or treatment really means assignment to an intervention, exposure regimen, or treatment strategy that will often be sustained over time. The start of follow-up is the time of assignment to the intervention or strategy of interest. Data from a randomized trial are left truncated if the period after strategy assignment is missing.

The same logic applies to observational studies. When we use the term "exposure," we gain clarity by mapping it to a hypothetical intervention, exposure regimen, or treatment strategy. For decision-making purposes, we cannot simply compare exposed and unexposed person-years but persons whose data are consistent with different treatment strategies with a clearly defined starting point. As an example, suppose we are interested in making dietary recommendations to healthy adults who eat less than 1 serving of fish per week. We could then compare individuals who, at the time of eligibility, increased their fish intake to 2 servings per week (a hypothetical intervention) versus individuals who increased their fish intake to 4 servings per week (another hypothetical intervention). Comparing person-years currently exposed to 2 servings per week versus person-years currently exposed to 4 servings per week does not directly inform decisions about dietary changes, because such comparison does not specify the time of onset of the strategies of interest, that is, the time when a decision about a change in fish intake would be made. (As a consequence, the times at which confounder history should be measured remain unspecified too.)

Vandenbroucke and Pearce (1) review another example in which failure to specify the treatment strategies of interest and their starting point may be misleading. Both a large randomized trial (12) and a large observational study (13) found no benefit of estrogen-plus-progestin therapy on the coronary heart disease risk of postmenopausal women aged 50 years or more. This was the conclusion of data analyses that compared women following the strategies "initiate hormone therapy at baseline and stop it whenever you want after baseline" versus "do not initiate hormone therapy at baseline, but start it whenever you want after baseline" (with extensive confounding adjustment for baseline initiation in the observational study). In contrast, if the data from either the randomized trial or the observational study had been analyzed by simply comparing "exposed" and "unexposed" person-years after excluding the first few years of follow-up (when women who already developed coronary heart disease have stopped contributing person-time), the incidence rate of coronary heart disease would have been lower in the "exposed" than in the "unexposed" person-time.

The left truncation of the latter analysis results in misleading estimates for women who need to make a decision about hormone therapy initiation at a particular time based on what happens from that time onward. Omitting the first period of follow-up misses the early increase in risk of coronary heart disease and results in person-time that is partially depleted of susceptible individuals. Visualizing this left truncation is easy when the contrast of interest is conceptualized as a contrast between women following the treatment strategies "initiate hormone therapy" versus "do not initiate hormone therapy" at baseline. However, the left truncation remains hidden when we compare "exposed" versus "unexposed" person-years because this contrast does not map into any treatment strategies with a precise onset.

NO OBVIOUS SOLUTION FOR LEFT TRUNCATION

When *all* individuals in the data are missing their early follow-up, it is generally impossible to reconstruct the cohort of individuals who initiated each strategy. These studies with so-called prevalent users may yield misleading estimates, as discussed above. When not all individuals have a left-truncated follow-up, Vandenbroucke and Pearce (1) review 2 possible approaches: stratification on time since initiation and reconstruction of the full cohort via life tables.

Stratifying the analysis on time since initiation of the strategies is helpful to illustrate how the hazard ratio changes over the follow-up. In the hormone therapy example, the hazard ratio is greater than 1 in the early follow-up and less than 1 in the late follow-up. However, changes in the period-specific hazard ratio cannot be generally interpreted as changes in actual risk (notwithstanding Vandenbroucke and Pearce's use of the word "risk" when discussing their Figure 3), because the numerical value of period-specific hazard ratios is a function of selection in previous periods (14). In the extreme, a hazard ratio may be less than 1 later in the follow-up if treatment depleted most susceptible individuals earlier in the follow-up, even if treatment had no beneficial effect on any single individual. From a decision-making standpoint, it is therefore unclear what is gained by presenting hazard ratios stratified by time since initiation as opposed to just presenting the hazard ratio (or even better, the survival curve) since time zero.

Life tables appear more promising. The idea is to estimate the time-varying hazards in order to reconstruct the survival curves under each strategy of interest. Because the hazard at each time is based only on those who survived without developing the outcome, Vandenbroucke and Pearce claim that all individuals' data can now be used in the analysis: each individual would only contribute to the hazards model from the time (since initiation) when her data are first available. Unfortunately, we still need to know each individual's complete treatment history (otherwise we could not determine whether she has been following the strategy of interest from time zero) and confounder history (otherwise we could not appropriately weight her contribution to the hazards model). If, however, we could ascertain every individual's treatment and confounder history from time zero, including that of individuals who died or were lost to follow-up, then there would be no left truncation.

A related point concerning hazards and risks: Let us not conflate the ends with the means. From a decision-making standpoint, estimating the time-varying hazards is an intermediate step toward constructing survival (or cumulative incidence) curves (15). The hazards themselves, and the corresponding hazard ratios, may be hard to interpret causally as the hormone therapy example illustrates (14). Specifically, precise estimation of hazards in late periods may be of little help if one cannot combine that information with hazard estimates in early periods in order to calculate the absolute risk.

THE EXTENT OF THE PROBLEM IN OBSERVATIONAL RESEARCH

Vandenbroucke and Pearce believe that serious misunderstandings due to left truncation are rare. For them, the case of estrogen-plus-progestin hormone therapy and coronary heart disease is unusual and, they say, "strong warnings about the inherent biasedness of studies involving prevalent exposures . . . seem inconsistent with the practice and experience of epidemiology" (1, p. 000). After all, left truncation did not create confusion about the direction of the effect of estrogen-plus-progestin hormone therapy on stroke and several other outcomes.

This brings seat belts to mind. Strong warnings about the inherent danger of riding a car also seem inconsistent with the practice and experience of driving. After all, we get involved in a car accident during less than, perhaps, 1% of our car trips. And yet, despite the low frequency of accidents, the use of seat belts is recommended. Our societies have determined that the cost of designing, installing, wearing, and even policing the use of seat belts is acceptable in order to prevent rare but potentially catastrophic injuries.

Analogously to cars without seat belts, studies with left truncation cannot receive a free pass simply because they may not lead to bias frequently. In fact, Vandenbroucke and Pearce do not give them such a free pass. Rather, they advise us to use our expert knowledge to postulate the expected hazard function in our study. If the hazard is expected to have certain shapes (refer to their Figure 3 and accompanying discussion), they claim that left truncation would not lead to bias and that the use of prevalent users would not only be acceptable but preferable. There are 2 distinct problems with their argument.

First, our expert knowledge is sadly incomplete. In the hormone therapy example, there was an unexpected early increase in the risk of coronary heart disease but not of stroke and other outcomes. Who knew that before the studies were conducted? What we believe about the hazard function may be inaccurate. For similar reasons, we recommend that everybody wear a seat belt rather than try to guess who will not get into a car accident.

Second, Vandenbroucke and Pearce (1) do not express much concern about studies in which left truncation distorts the magnitude but not the direction of the average hazard ratio during the follow-up (which is a weighted average of the time-varying hazard ratios). For example, the mortality hazard ratio for statin versus no statin in persons with prior heart disease is closer to the null in observational studies of incident users than in observational studies of prevalent users (16),

possibly because the hazard ratio for prevalent users does not include the null hazard ratio during the early follow-up. If observational studies with left truncation (prevalent users) are exaggerating the benefits of statins, benefit-risk considerations may be incorrect and decisions may be suboptimal. Estimating the absolute risks from time zero under each strategy would be preferable.

When Vandenbroucke and Pearce (1) argue that subject matter knowledge about hazards relieves us from synchronizing start of follow-up and start of exposure, their logic is sound. Unfortunately, our subject matter knowledge is fallible. Further, even if the assumptions about hazards were correct, left truncation generally prevents the estimation of absolute risks, a crucial piece of information for decision-making.

Yet Vandenbroucke and Pearce raise an important question: How often do studies with left truncation yield seriously distorted association measures? There are many pharmacoepidemiologic applications in which left truncation is potentially problematic. Here we have mentioned the cases of hormone therapy and coronary heart disease and of statin therapy and mortality (also reviewed by Vandenbroucke and Pearce). Ray (9) discussed several other examples. The failure to include all incident users in a contrast is also responsible for immortal time biases, as reviewed by Suissa (17), and for biases in the comparison of dynamic treatment strategies (18, 19).

Outside of drug epidemiology, we have previously called attention to problems deriving from left truncation in pregnancy research (e.g., studies of the effect of prenatal exposures that include liveborns only) (20), obesity research (e.g., the so-called obesity paradox) (21), and aging research (e.g., studies of the effect of cigarette smoking on dementia that excluded the first 70 years of life) (22). There are probably many other examples, but no systematic review seems to be available. Vandenbroucke and Pearce have identified an interesting line of methodological research.

DISCUSSION

We complain when decisions are not supported by sound research. We call it evidence-free practice. Conversely, we need to be careful not to engage in practice-free research, that is, research that is conducted without adequate thought given to the needs of those who will make decisions in practice—patients, physicians, policy-makers, public health officials, regulators, and others. One way to move toward practice-free research is the comparison of exposure groups that do not correspond to well-defined interventions (23, 24). Another one is to conduct analyses of observational data with left truncation that renounce estimation of absolute risks and yield estimates that are hard to interpret causally.

Starting the follow-up when the treatment strategies start, however, raises some challenges.

First, only individuals with treatment, confounder, and outcome information since the start of follow-up can be included in the analysis. As Vandenbroucke and Pearce remind us, this may result in a small sample size. The sample size can be further reduced when the definition of the strategies requires information on the individuals' history before the start of follow-up. For example, an analysis of electronic medical records to compare the strategies "initiate hormone therapy" versus "do not initiate hormone therapy" among women 50 years of age who have not used hormone therapy for at least 2 years needs to be restricted to women in the database since at least age 48 years. Another example: An analysis of an epidemiologic cohort to compare the hypothetical interventions "increase fish intake to 2 servings per week" versus "increase fish intake to 4 servings per week" in the year 2000 needs to be restricted to individuals with data on fish intake before 2000. Otherwise we would not be able to assign the strategies and adjust for prior history of fish intake.

Second, the assessment of long-term effects requires that individuals be followed for long periods. Individuals will be right censored when their data on treatment, confounders, or outcome becomes unavailable. Vandenbroucke and Pearce see right censoring as a problem of studies of incident users. On the contrary, being able to quantify and possibly adjust for right censoring is another advantage of using incident users. In studies of prevalent users, right censoring tends to be ignored because investigators do not know who among the incident users was lost to follow-up or died between the initiation of the treatment strategy and the start of the truncated follow-up.

Third, synchronizing start of follow-up and treatment strategies requires that confounders be measured and adjusted for during the entire follow-up period. In particular, the analysis requires confounding adjustment for the effect of treatment initiation to prevent so-called "protopathic bias" (25). For example, in a second example of statins reviewed by Vandenbroucke and Pearce (this time on the effect of statins for primary prevention of coronary heart disease), an estimated increase in the 1-year risk after statin initiation may well reflect incomplete confounding adjustment, because unrecorded symptoms of heart disease may encourage treatment initiation, rather than an early harmful effect of statins.

Finally, the elimination of left truncation requires a rethinking of traditional epidemiologic studies based on the comparison of incidence rates under the implicit assumption that all units of person-time are exchangeable. It is precisely this assumption that creates the illusion of *the* hazard ratio as an intrinsic property of the relation between exposure and outcome (as opposed to a weighted average of time-varying hazard ratios that varies with the duration of follow-up) in a particular population and that impedes clear thinking about confounding adjustment (26, 27). The limitations of analyses based on "a pool of person-time" rather than on "a group of persons" apply to the analyses of both longitudinal data sets and case-control samples from those data sets.

As Vandenbroucke and Pearce remind us, this discussion has implications for teaching. Prominent among these is that teachers ought to be explicit about the goal of the methods they are teaching. If the goal is to provide evidence to support decision-making, then the natural starting point is teaching methods to estimate risks from the time the decision needs to be made. We can then teach that deviations from this paradigm must be justified on a case-by-case basis. The burden of the proof is on those who decide not to start follow-up at the same time as the treatment strategies.

In addition, teachers have to be careful when using the term "prevalent exposure." The follow-up need not start when individuals are first exposed but when individuals would

need to decide whether to follow a particular course of action. For example, when estimating the effect of fish intake, we do not require that the follow-up start when individuals first eat fish during their lifetimes but when a particular strategy concerning fish intake starts (28). This starting point may be middle age if the goal of the analysis is to help middle-aged individuals decide whether to increase their fish intake. That is, prevalent exposures like fish intake, when used to define precise interventions, are allowed in causal analyses of observational data.

One way of reducing misunderstandings is teaching the counterfactual theory of causal effects of time-varying treatments by Robins (6, 7). In this theory, causal effects are defined in terms of contrasts between the treatment strategy-specific survival curves of individuals with a particular treatment and covariate history before the start of the strategies of interest. Causal graphs (29, 30) facilitate the representation of counterfactual concepts and the assumptions required by a particular analytical method in each context. Counterfactuals and graphs have become the common language in which methodologically sophisticated practitioners from disciplines concerned with causal inference communicate. Vandenbroucke and Pearce, on the other hand, seem to argue that the old ways suffice and that an understanding of causal inference theory is largely unnecessary.

This discussion focused on causal questions for which we have enough observational data to attempt a direct emulation of a randomized trial. Many decisions involve systems so complex that even an approximate direct emulation of a trial is impossible. Consider contrasts of the effects of different strategies to deal with climate change or of alternative transformations of the health-care system on health outcomes. The reliance on subject-matter knowledge and modeling assumptions required by this approach (31) may make left truncation and other issues discussed here look like just minor inconveniences.

In summary, analyses of observational data aimed at supporting decision-making are ideally framed as a contrast between different hypothetical courses of action—or interventions or exposure regimes or treatment strategies—and compare individuals from the decision time, that is, the time when the change under consideration could have taken place. Exceptions to this synchronizing of the start of follow-up and the treatment strategies may be considered when the only available data (or the only data that we can afford) are left truncated. If we believe that analyzing those data will improve the existing evidence for decision-making, we must defend the use of left-truncated data explicitly, rather than defaulting into using the data without any justification. Vandenbroucke and Pearce (1) have provided us with solid arguments to mount such defense.

ACKNOWLEDGMENTS

Author affiliations: Departments of Epidemiology and Biostatistics, Harvard T. H. Chan School of Public Health, Boston, Massachusetts (Miguel A. Hernán); and Harvard-MIT Division of Health Sciences and Technology, Boston, Massachusetts (Miguel A. Hernán).

Supported by grant NIH R01 AI102634.

I thank Drs. Sander Greenland and James Robins for their comments.

Conflict of interest: none declared.

REFERENCES

1. Vandenbroucke J, Pearce N. Point: Incident exposures, prevalent exposures, and causal inference: Does limiting studies to persons who are followed from first exposure onward damage epidemiology? *Am J Epidemiol*. 2015;182(10):826–833.
2. Feinstein AR. Clinical biostatistics. XI. Sources of 'chronology bias' in cohort statistics. *Clin Pharmacol Ther*. 1971;12(5):864–879.
3. Feinstein AR. Sources of 'chronology bias'. In: *Clinical Biostatistics*. St. Louis, MO: The C. V. Mosby Company; 1977:89–104.
4. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol*. 1974;66(5):688–701.
5. Rubin DB. Bayesian inference for causal effects: the role of randomization. *Ann Statist*. 1978;6(1):34–58.
6. Robins JM. Addendum to "a new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect" [published errata appear in *Comput Math Appl*. 1989;18:477]. *Comput Math Appl*. 1987;14:923–945.
7. Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect [published errata appear in *Math Model*. 1987;14:917–921]. *Math Model*. 1986;7:1393–1512.
8. Robins JM. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *J Chronic Dis*. 1987;40(suppl 2):139S–161S.
9. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol*. 2003;158(9):915–920.
10. Hernán MA. With great data comes great responsibility: publishing comparative effectiveness research in epidemiology. *Epidemiology*. 2011;22(3):290–291.
11. Galea S. An argument for a consequentialist epidemiology. *Am J Epidemiol*. 2013;178(8):1185–1191.
12. Manson JE, Hsia J, Johnson KC, et al. Estrogen plus progestin and the risk of coronary heart disease. *N Engl J Med*. 2003;349(6):523–534.
13. Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology*. 2008;19(6):766–779.
14. Hernán MA. The hazards of hazard ratios. *Epidemiology*. 2010;21(1):13–15.
15. Cole SR, Hudgens MG, Brookhart MA, et al. Risk. *Am J Epidemiol*. 2015;181(4):246–250.
16. Danaei G, Tavakkoli M, Hernán MA. Bias in observational studies of prevalent users: lessons for comparative effectiveness research from a meta-analysis of statins. *Am J Epidemiol*. 2012;175(4):250–262.
17. Suissa S. Immortal time bias in pharmaco-epidemiology. *Am J Epidemiol*. 2008;167(4):492–499.

18. Kitahata MM, Gange SJ, Abraham AG, et al. Effect of early versus deferred antiretroviral therapy for HIV on survival. *N Engl J Med*. 2009;360(18):1815–1826.
19. Hernán MA, Robins JM. Early versus deferred antiretroviral therapy for HIV. *N Engl J Med*. 2009;361(8):822–823; author reply 823–824.
20. Hernán MA, Hernández-Díaz S, Werler MM, et al. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol*. 2002;155(2):176–184.
21. Lajous M, Banack HR, Kaufman JS, et al. Should patients with chronic disease be told to gain weight? The obesity paradox and selection bias. *Am J Med*. 2015;128(4):334–336.
22. Hernán MA, Alonso A, Logroscino G. Cigarette smoking and dementia: potential selection bias in the elderly. *Epidemiology*. 2008;19(3):448–450.
23. Hernán MA, Taubman SL. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *Int J Obes (Lond)*. 2008;32(suppl 3):S8–S14.
24. Glass TA, Goodman SN, Hernán MA, et al. Causal inference in public health. *Annu Rev Public Health*. 2013;34:61–75.
25. Feinstein AR, Horwitz RI. A critique of the statistical evidence associating estrogens with endometrial cancer. *Cancer Res*. 1978;38(11 pt 2):4001–4005.
26. Greenland S, Rothman KJ, Lash TL. Measures of effect and measures of association. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*. 3rd ed. Philadelphi, PA: Lippincott, Williams & Wilkins; 2008:51–70.
27. Greenland S, Rothman KJ. Measures of occurrence. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott, Williams & Wilkins; 2008:32–50.
28. Lajous M, Willett WC, Robins J, et al. Changes in fish consumption in midlife and the risk of coronary heart disease in men and women. *Am J Epidemiol*. 2013;178(3):382–391.
29. Pearl J. Causal diagrams for empirical research. *Biometrika*. 1995;82(4):669–710.
30. Spirtes P, Glymour C, Scheines R. *Causation, Prediction and Search*. 2nd ed. Cambridge, MA: MIT Press; 2000.
31. Hernán MA. Invited commentary: Agent-based models for causal inference—reweighting data and theory in epidemiology. *Am J Epidemiol*. 2015;181(2):103–105.