The Epidemiologic Toolbox: Identifying, Honing, and Using the Right Tools for the Job

Catherine R. Lesko*, Alexander P. Keil, and Jessie K. Edwards

* Correspondence to Dr. Catherine R. Lesko, Department of Epidemiology, Johns Hopkins School of Public Health, 615 N. Wolfe Street, Baltimore, MD 21205 (e-mail: clesko2@jhu.edu).

Initially submitted October 21, 2019; accepted for publication March 2, 2020.

There has been much debate about the relative emphasis of the field of epidemiology on causal inference. We believe this debate does short shrift to the breadth of the field. Epidemiologists answer myriad questions that are not causal and hypothesize about and investigate causal relationships without estimating causal effects. Descriptive studies face significant and often overlooked inferential and interpretational challenges; we briefly articulate some of them and argue that a more detailed treatment of biases that affect single-sample estimation problems would benefit all types of epidemiologic studies. Lumping all questions about causality creates ambiguity about the utility of different conceptual models and causal frameworks; 2 distinct types of causal questions include 1) hypothesis generation and theorization about causal structures and 2) hypothesis-driven causal effect estimation. The potential outcomes framework and causal graph theory help efficiently and reliably guide epidemiologic studies designed to estimate a causal effect to best leverage prior data, avoid cognitive fallacies, minimize biases, and understand heterogeneity in treatment effects. Appropriate matching of theoretical frameworks to research questions can increase the rigor of epidemiologic research and increase the utility of such research to improve public health.

bias; causality; descriptive studies; epidemiologic methods; inference

The future of epidemiologic norms has recently been the subject of much discussion. In particular, there has been a debate about the degree of emphasis placed on the use of explicit and specific causal frameworks and advanced statistical models for elucidating causes and effects (1–6). This debate is not new (7–9) and is not constrained to the literature. In our own institutions, discussions about the content of epidemiologic methods courses and the focus of doctoral theses seem imbued with existential angst among faculty who trained before potential outcomes permeated many curricula and who decry this new framework as reductionist and harmful to the advancement of public health and among students who are clamoring to apply the latest methods. This angst is not entirely misplaced. As early-career researchers who were exposed to a "causal" curriculum and who are now engaged in training new epidemiology students to refine their research questions and design sound studies, we have often found ourselves torn between promoting and defending the use of formal causal frameworks and cautioning against the indiscriminate application of causal methods as spice for otherwise bland research questions. We speak here to the skeptics of causal inference and to its devoted practitioners.

Herein, we give examples of the breadth of the field of epidemiology to give context to questions deriving from causal effect estimation; estimation of causal effects is only 1 type of epidemiologic question out of many valuable lines of epidemiologic inquiry (5). We consider how epidemiologic studies contribute to scientific knowledge about threats and opportunities in public health. We disambiguate the term "causal inference." Specifically, we differentiate between the act of hypothesizing and theorizing about causes and the act of estimating a causal effect and, in doing so, suggest instances in which an explicit causal framework is useful and when it may not be necessary (3, 10). We finish by briefly discussing the role of context in epidemiologic studies and outlining what we believe are the implications of these arguments for current and future epidemiologists.

THE SCOPE OF EPIDEMIOLOGIC INFERENCE

Epidemiology has been defined as "the study of the occurrence and distribution of health-related states and events in specified populations, including the study of the determinants influencing such states, and the application of this knowledge to control the health problems." (11, p. 95). Under the umbrellas of distribution and determinants live many research questions, the answers to which can improve public health (12). Loosely speaking, these research goals fall along a spectrum with purely descriptive epidemiology at 1 end; hypothesis generation, prediction, and outbreak investigation somewhere in the middle; and causal effect estimation and program evaluation at the other end. Here, we envision the spectrum signifying the approximate strength of assumptions required to obtain useful results and perhaps the temporal order in which investigations tend to unfold (e.g., describe the problem, hypothesize about the problem, intervene on the problem, evaluate the intervention). However, there are always exceptions to this ordering and feedback loops between related questions. Others have proposed other frameworks (12, 13).

A comprehensive framework for organizing epidemiologic questions is beyond the scope of this commentary, but it is evident that 1) epidemiologic principles and methods are applicable to many questions beyond causal effect estimation, and 2) epidemiologic curricula and journals have prioritized analytic epidemiology and questions related to identifying (causal) determinants of disease over descriptive epidemiology and questions related to accurately characterizing the health of populations (14, 15). Descriptive epidemiologic studies are frequently excluded from peer-reviewed journals for not being generalizable enough. We contextualize this particular criticism in a later section of this paper.

Descriptive epidemiology addresses many important public health questions and accurate answers to these questions are crucial for prioritizing, targeting, and staffing interventions. Furthermore, characterizing emerging public health problems, including estimating crude associations (e.g., "risk factor analysis") can help generate hypotheses for further study.

Collecting, analyzing, and interpreting data from a descriptive study, such as public health surveillance, is challenging. Yet methods for doing so are given cursory coverage in most epidemiology curricula or are relegated to elective courses. An epidemiology curriculum that emphasized descriptive epidemiology might spend the entirety of the first term on single-sample estimation problems and describing the natural course of disease (i.e., the course of disease in the absence of any interventions). This could be framed in terms of designing a target study, or an idealized study that would accurately estimate the descriptive parameter of interest in the absence of real-world constraints like missing data and measurement error (foreshadowing introduction of the target trial as a heuristic for study design for causal effect estimation but encompassing a broader set of questions). This task is already challenging. Loss to follow-up, measurement error, selection processes, and other more traditional sources of missing data blind us from some persons' characteristics and we are forced to make assumptions about the value of those missing characteristics (16). This novel epidemiology curriculum would spend more time defining target populations and describing how changing population compositions may affect disease occurrence; sampling methods and selection bias in single-sample estimation problems; properties of screening and diagnostic tests and case definitions; methods for handling information bias; and missing data mechanisms and techniques for dealing with nonresponse.

A solid understanding of biases that plague single-sample estimation problems is made even more urgent as our modes of communication and transportation and our expectations of privacy evolve. Staple sampling methodologies such as random-digit dialing or using Department of Motor Vehicle registries are becoming less reliable. New sampling methods (e.g., Internet sampling, respondent-driven sampling) can reach hidden populations or return large samples quickly (17–21), but drawing population-level inference may be challenging. Furthermore, "big data" has made available massive administrative data sets that may lull us into a false sense of security about the potential for systematic error by effectively eliminating considerations of random error (22, 23). Given the reliance of analytic epidemiologic studies on these data sources, all other study designs would arguably benefit from such a renewed focus on biases in descriptive epidemiology.

Ultimately, a fundamental goal of epidemiologic studies is to inform interventions that could improve public health. The frequency, distribution, and impact of diseases should influence where we invest our time and energy. The frequency, distribution, and manipulability of determinants of disease influence the impact that interventions on those determinants will have on improvements to health and realizations of justice, given limited resources. A novel epidemiology curriculum would include transparent discussions about the metrics we use to measure improvement in health, whose health is improved through various interventions, and whether that improvement comes at the detriment or neglect of others' health.

WEIGHING THE SCIENCE ON CAUSES AND EFFECTS: INTEGRATING EVIDENCE ACROSS FIELDS

Ultimately, epidemiologic studies should strive to inform specific decisions about how and in which populations we might intervene to achieve such control. Descriptive epidemiologic studies are crucial for informing the "in which populations" part of the decision. With respect to the "how" part of the decision, we can (and should) distinguish between best practice for evaluating the scientific evidence to inform those decisions and best practice for conducting an individual epidemiologic study such that it contributes maximally to the scientific evidence. A single epidemiologic study should not determine the policy most likely to result in improved public health as a matter of course.

To inform a public health decision, the scientific evidence need not be represented by an axiom or law of nature but rather a reasonably accurate, consistent association (within

some target population, perhaps conditional on modifiers) between some stimuli (e.g., exposure, intervention) and some health state (e.g., outcome). The degree of accuracy and consistency required is likely to vary depending on the consequences of action or inaction. Ideally, the scientific evidence about a public health challenge could be weighed after numerous studies are conducted on a subject across domains of knowledge, spanning from abstract mathematics and human or animal laboratory studies to ecological studies, contextual case studies, cohort studies, and, sometimes, randomized trials. Yet the process through which we typically integrate evidence is often informal and iterative. Most public health decisions are necessarily made with insufficient information. Early studies often provide inspiration for subsequent studies specifically designed to address the limitations of prior work. For example, policy makers have used evidence from uranium miners to set guidelines for residential radon exposure while the evidence on the health effects of low-level radon exposure accumulates from ongoing residential-exposure studies (24). We conduct postmarketing surveillance for the effects of drugs because rare adverse events or long-term effects of drugs are not detected in randomized trials.

It seems obvious that the optimal strategy for integrating evidence across studies depends on the research question and the quantity and quality of evidence available. For questions about the effect of an intervention on an outcome, where results from multiple randomized trials are available, meta-analytic methods may be appropriate. In contrast, when there are few studies or gaps in the available data, Bayesian reasoning and data fusion methods to integrate multiple knowledge domains may be best. Triangulation of evidence from imperfect yet complementary study designs may best address questions in the presence of multiple sources of bias (25).

EVIDENCE ON WHAT?

Epidemiologic studies for which the aim is to inform actions for improved public health are inherently causal. However, there are at least 2 types of causal questions and ambiguity about the type of causal question being asked may derail the investigation (26). One type of question involves looking back in time to try to identify the causes of some health state, and another involves quantifying the changes in a health state we might expect to see as a result of toggling 1 or more of its causes (3). These 2 types of questions have not always been clearly delineated and would both seem to be causal inference. Yet we argue that identifying the causes of prevalent health states is much more of a hypothesis-generating exercise and best guided by conceptual models, whereas estimating causal effects is more amenable to hypothesis testing and best guided by potential outcomes/graphical causal models. There is some circularity in this statement: If we determine that an intervention has an effect on some outcome, we may deduce the intervention is a cause of that outcome for at least some subset of individuals. But in general, we contend the most reliable, reproducible conclusions about causal relationships result from studies designed to investigate the effect of a limited subset of interventions (i.e., causal effect estimation), rather than studies designed to identify causes of an outcome (27, 28).

That said, some have argued that reliance on the potential outcomes framework for causal inference inhibits other forms of epidemiologic inquiry. We find sympathy with these concerns, but we would separate the framework from its practitioners and learners. We hope this commentary allays some of these concerns about the framework while addressing how we might produce more well-rounded practitioners that select a framework based on a question, rather than the reverse. As we have outlined, causal-effect estimation is only 1 of many possible goals of epidemiologic studies and only 1 possible goal of causal inference. The potential outcomes framework is often unnecessary for investigating a causal relationship (29), particularly in the early stages of investigation when hypothesis generation rather than hypothesis testing is more efficient (30–32). Other times, the link between association and causation is strong, immediate, and obvious: We can do a case-control study after a foodborne outbreak, estimate the increased odds of gastrointestinal illness due to the potato salad, and interpret this odds ratio causally because the induction period is short, the biologic mechanism is known, and, perhaps, we have confirmatory laboratory evidence of the offending pathogen. Other causal relationships may not lend themselves to causal effect estimation, such as when an intervention cannot be clearly defined; in these instances, researchers may justifiably seek to demonstrate a persistent association between 2 variables despite accounting for other possible explanations.

ESTIMATION OF CAUSAL EFFECTS IN EPIDEMIOLOGY

Undertaking any epidemiologic study requires making many decisions about study design and analyses, among other things. Crucial study design features include inclusion/exclusion criteria, the time origin, and follow-up time; estimating causal effects requires specifying interventions or treatment regimens of interest, as well (33). These features are likely to determine the magnitude and perhaps even direction of the effect we estimate. Without a framework for organizing lessons learned from prior missteps (34–36), how to do we efficiently and reliably design our studies to best leverage prior data, avoid cognitive fallacies, and minimize bias in our results (37)?

Despite the many theoretical, statistical, and computing advances in the field, epidemiology is still a relatively new science. This may explain the paucity of comprehensive conceptual frameworks to organize the lessons learned and best practices. We have multiple frameworks for thinking about how causal relationships might manifest in an observed association (e.g., Koch's postulates, Bradford-Hill's causal considerations, Rothman's causal pies, Neyman's potential outcomes) (38–40). One can find multipage lists of biases that have been identified over the years as threats to inferring causation (though often not expressed in terms of causality) (41). There has been a recent push to classify these more than 200 specific biases into 1 of 3 classes: confounding bias, selection bias, and information bias (14, 41, 42), but

even that classification is not always clear. Is inappropriately including a variable in a regression model, such that it induces bias between exposure and outcome (i.e., conditioning on a collider) a form of confounding bias, selection bias, or model misspecification bias, and to that end, where would model misspecification bias fall (43–45)?

We, like many others, see utility in the potential outcomes framework and causal graph theory for conceptualizing questions about causal effect estimation, particularly when the interventions or exposures under study have delayed effect, or are complex, time varying, or dynamic. A full introduction to potential outcomes can be found in articles by Little and Rubin (46) and Hernán and Robins (47), and an introduction to causal graph theory presented by Greenland et al. (48) and Pearl (49); the 2 theories share many underlying principles and assumptions, and have been formally integrated in Single-World Intervention Graphs in an article by Richardson and Robins (50). A key benefit of any of these frameworks is that they allow us to formally define "effect" and write down the estimand of interest. Then, we can evaluate the data and consider the causal assumptions that would allow us link the observed data to the causal estimand (16, 51, 52). In being so explicit about the inferential goal, we may avoid logical pitfalls (37). We can imagine a target trial or public health intervention we would conduct, absent ethical or logistic constraints, and attempt to emulate it with observational data (51–54). Distinguishing the causal estimand from the statistical estimator makes explicit the assumptions we must make for valid causal and statistical inference, which, in turn, allows for scientific debate on, and revision of, those assumptions (51, 52). In most cases, the potential outcomes framework or causal graph theory does not imply a particular statistical methodology is necessary and there will be many instances where generalized linear models are sufficient to execute the analysis. However, for certain questions (particularly when there is time-varying confounding affected by prior exposure), so-called g-methods give us a better chance at getting the right answer than do traditional methods (4, 55, 56). Our advocacy for teaching and using the potential outcomes framework and causal graph theory is not an argument for more complex methods; rather, it is an argument for employing useful tools that can help us ask more clearly defined research questions and to have a clearer understanding of the assumptions needed to answer such questions with data using the least complex approach necessary.

A PRACTICAL APPROACH TO EXPLICIT CAUSAL FRAMEWORKS

Conceptual models and theoretical frameworks are useful for guiding research programs and discussing bodies of evidence. These frameworks may benefit from including on them the big, nebulous social determinants of disease, such as poverty or racism. However, the same constructs are not sufficiently well defined to be provide useful guidance in a study designed to estimate their causal effects. Arguably, a strength of the potential-outcomes framework and causal graph theory are that they force us to imagine the interventions that might alter the system, linking our research more closely to policy decisions. An intervention on poverty might involve giving housing vouchers or increasing wages or permitting collective bargaining, for example (26). If we do not have the data to identify the effects of interest, that is a practical, rather than a theoretical, problem. There are exposure-outcome pairs for which confounding is intractable and exposures of importance that are not well defined. In such instances, 1 reaction has been to answer a question about a different exposure that *can* be answered, rather than the question of initial interest. This reaction certainly contributes to the perception that the potential outcomes framework and causal graph theory are somehow inherently limiting with respect to the exposures that are "valid" to study. Although we sympathize with this sentiment, this is not a limitation of the potential outcomes theory itself but rather an uncomfortable reality (4). Alternatively, we can attempt to answer the original questions, acknowledging that the exposures we study are inherently poorly defined, interpreting our resultant estimates of effect accordingly, and attempting to refine our exposure definition in subsequent investigations.

A strength of the potential-outcomes framework and causal graph theory is that they allow us to articulate clearly the assumptions sufficient to identify a causal effect from data. A limitation, and caution to researchers new to the field, is that there has been a misconception that by clearly stating the causal effect of interest, researchers can then interpret the association they estimate as a causal effect. Reciting identification assumptions like catechism or an incantation does not make them true. We must consider carefully whether they are met in each circumstance and design better studies to address instances in which they are not. Formalizing statistical assumptions using potential outcomes and probability logic, or drawing causal diagrams, can help determine whether an effect is identifiable from data. Causal diagrams can help identify a sufficient adjustment set—under the assumption that the diagrams are correct enough for the question at hand. We should be honest about the possible errors in our causal models, which are expressed in the diagrams. The assumptions sufficient for drawing causal conclusions from observational data are heroic and we should not lose sight of that. As reviewers of our peers' work, we should embrace the open and precise statements about uncertainty when it exists rather than letting uncertainty diminish the importance of a result or method. Otherwise, we risk over-certainty in our literature, and we also risk the trust of the public when we are inevitably both certain and wrong. Furthermore, we should continue our work to educate nonscientists on the importance of acknowledging uncertainty in decision-making.

ADDITIONAL CONSIDERATIONS FOR USING EPIDEMIOLOGIC STUDIES TO IMPROVE POPULATION HEALTH

The potential-outcomes framework can help us define a causal estimand in a specific population and also help us understand why the causal effect of a well-defined exposure

may differ across populations (10, 57–65). Briefly, an internally valid estimate of effect may not generalize or transport from 1 population to another because 1) there is effect modification and the distribution of modifiers differs across populations; 2) the way in which treatment or exposure occurs (the version of treatment) makes a difference in the magnitude of the effect observed, and the treatment is not dispensed similarly across populations; 3) there is interference, spillover, or disseminated effects and the distribution of exposures differs across populations; or 4) the exposure or outcome is measured with different degrees of accuracy across populations. It is useful to remember that we are not working in a field that studies fundamental laws of the universe, rather epidemiology studies the mechanisms or effects of exposures in specific populations, contexts, and times; the distribution of other determinants of disease shift with those contexts and thus the relative influence of the exposure under investigation will surely change. The idea that epidemiologists should aim to identify and intervene in determinants of disease that are the largest levers for improving health and supporting justice implies that the largest levers may differ across and within populations. When we find different effect estimates from different studies in different populations, the explanation is not always bias. The importance of sample composition and context when drawing epidemiologic inference again underscores the need for good descriptive epidemiology. Quantitative solutions to estimating effects relevant to target populations of interest depend on having high-quality descriptions of those target populations (58, 59, 61, 65–67).

IMPLICATIONS FOR THE FIELD OF EPIDEMIOLOGY

What are the implications of our assertions for the future of epidemiology and, in particular, for how we train the next generation of epidemiologists? First, let us agree that epidemiologists can and should bring unique skill sets, knowledge, perspectives, and critical frameworks to scientific inquiry. If this were not the case, there would be no need for epidemiologic methods courses; epidemiology students would just enroll in some combination of biostatistics, ecology, microbiology, and physiology courses, to name a few. Certainly, 1 strength of epidemiology is its focus on interdisciplinary collaboration. Yet, if we epidemiologists abandon our emphasis on well-defined study questions (causal or not) that are linked to appropriate study design and analysis, what do we contribute to interdisciplinary collaborations?

We are not suggesting that expertise in causal diagrams, potential outcomes, and g-methods are necessary for all epidemiologists to do good epidemiologic research. However, we argue that our ability to answer epidemiologic questions of interest is improved by the formalization of theory and language. Interdisciplinary collaborations can help identify important questions, draw a conceptual model (even if one does not choose to use a causal diagram to formalize that process), and provide context for interpreting our results. But the epidemiologic perspective and skill set is necessary to refine the research question, link the estimand of interest with a statistical estimation strategy (e.g., choose the model form), determine what goes into the model, choose which associations to report, and appropriately interpret those associations. Only if we understand how the numbers underlying associations relate to possible change we could observe in the world can we begin to link observations to causes, and epidemiology to actions. Again, we may not always, or even typically, need advanced methods to make those links. But ignorance of such methods guarantees that, unless we are lucky, we will always be worse off when those methods are needed.

CONCLUSIONS

Epidemiology is an applied science—one aimed at improving population health. We need to re-emphasize the importance and impact of asking and answering descriptive questions. Yet some important public health questions are causal. Indeed, frequently, causal and descriptive questions go hand-in-hand: Descriptive analyses are necessary to identify public health problems and targets for intervention; causal analyses are necessary to inform how we should improve the health of populations; and descriptive analyses are necessary to track the implementation of our interventions. When asking causal questions, if we cannot ultimately produce actionable information, we have failed in our charge. If we focus on associations because we cannot reliably estimate causal effects, we abdicate responsibility and obfuscate our intent (68).

When designing studies to answer causal questions, the potential-outcomes framework and causal graph theory allow new epidemiologists to build a foundation for causal inference in which sources of bias can be identified from first principles, such that the burden of memorizing lists of biases and rules of thumb (which often have important exceptions) is removed. We cannot simply try to teach intuition or many rules of thumb for identifying potential sources of bias; this intuition is not reliably transferable across institutions and instructors.

Ultimately, we cannot abdicate responsibility to identify causes or attempt to estimate effects of interventions from imperfect data. We should arm ourselves with the most reliable tools to accomplish this goal. As summarized by Jenicek, "Epidemiology suffers often from its 'yes – but' approach. Let us not forget, that ultimately . . . we must always make a 'yes or no' decision . . . We are in our positions for this kind of decision making and it is expected from us" (69, p. 192). Let us take the best frameworks and ideas on offer and not be afraid to make our science more rigorous. Not only will the field benefit, the public's health will, too.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland (Catherine R. Lesko); Department of Epidemiology, Gillings School of

Global Public Health, University of North Carolina, Chapel Hill, Chapel Hill, North Carolina (Alexander P. Keil, Jessie K. Edwards); and National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina (Alexander P. Keil).

This work was funded by National Institutes of Health grants K01 AA028193, R01 ES029531, and K01 AI125087.

Conflict of interest: none declared.

REFERENCES

1. Vandenbroucke JP, Broadbent A, Pearce N. Causality and causal inference in epidemiology: the need for a pluralistic approach. *Int J Epidemiol*. 2016;45(6):1776–1786.
2. Krieger N, Davey Smith G. The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology. *Int J Epidemiol*. 2016;45(6): 1787–1808.
3. VanderWeele TJ. Commentary: on causes, causal inference, and potential outcomes. *Int J Epidemiol*. 2016;45(6): 1809–1816.
4. Robins JM, Weissman MB. Commentary: counterfactual causation and streetlamps: what is to be done? *Int J Epidemiol*. 2016;45(6):1830–1835.
5. Greenland S. For and against methodologies: some perspectives on recent causal and statistical inference debates. *Eur J Epidemiol*. 2017;32(1):3–20.
6. Daniel RM, De Stavola BL, Vansteelandt S. Commentary: the formal approach to quantitative causal inference in epidemiology: misguided or misrepresented? *Int J Epidemiol*. 2016;45(6):1817–1829.
7. Taubes G. Epidemiology faces its limits. *Science*. 1995; 269(5221):164–169.
8. Wynder EL. Invited commentary: response to science article, "Epidemiology faces its limits". *Am J Epidemiol*. 1996; 143(8):747–749.
9. Wing S. Limits of epidemiology. *Medicine and Global Survival*. 1994;1(2):74–86.
10. Gelman A. Causality and statistical learning. *Am J Sociol*. 2011;117(3):955–966.
11. Porta MS, Greenland S, Hernán M, et al. *A Dictionary of Epidemiology*. 6th ed. New York, NY: Oxford University Press; 2014.
12. Lau B, Duggal P, Ehrhardt S. Epidemiology at a time for unity. *Int J Epidemiol*. 2018;47(5):1366–1371.
13. Vittinghoff E. *Regression Methods in Biostatistics : Linear, Logistic, Survival, and Repeated Measures Models*. 2nd ed. New York, NY: Springer; 2012.
14. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008.
15. Gordis L. *Epidemiology*. 5th ed. Philadelphia, PA: Elsevier/Saunders; 2014.
16. Edwards JK, Cole SR, Westreich D. All your data are always missing: incorporating bias due to measurement error into the potential outcomes framework. *Int J Epidemiol*. 2015;44(4): 1452–1459.
17. Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. *Soc Probl*. 1997; 44(2):174–199.
18. Goel S, Salganik MJ. Assessing respondent-driven sampling. *Proc Natl Acad Sci*. 2010;107(15):6743–6747.
19. Wise LA, Rothman KJ, Mikkelsen EM, et al. Design and conduct of an Internet-based preconception cohort study in North America: Pregnancy Study Online. *Paediatr Perinat Epidemiol*. 2015;29(4):360–371.
20. Sinclair M, O'Toole J, Malawaraarachchi M, et al. Comparison of response rates and cost-effectiveness for a community-based survey: postal, Internet and telephone modes with generic or personalised recruitment approaches. *BMC Med Res Methodol*. 2012;12(1):132.
21. Ross MW, Månsson S-A, Daneback K, et al. Biases in Internet sexual health samples: comparison of an Internet sexuality survey and a national sexual health survey in Sweden. *Soc Sci Med*. 2005;61(1):245–252.
22. Khoury MJ, Ioannidis JP. Medicine. Big data meets public health. *Science*. 2014;346(6213):1054–1055.
23. Toh S, Platt R. Is size the next big thing in epidemiology? *Epidemiology*. 2013;24(3):349–351.
24. National Research Council. *Health Effects of Exposure to Radon: BEIR VI.* Vol 6. Washington, DC: National Academies Press; 1999.
25. Lawlor DA, Tilling K, Davey SG. Triangulation in aetiological epidemiology. *Int J Epidemiol*. 2016;45(6): 1866–1886.
26. Galea S, Hernán MA. Win-win: reconciling social epidemiology and causal inference [published online ahead of print October 3, 2019. *Am J Epidemiol*. doi: 10.1093/aje/kwz158.
27. VanderWeele TJ. Outcome-wide epidemiology. *Epidemiology*. 2017;28(3):399–402.
28. Cole P. The hypothesis generating machine. *Epidemiology*. 1993;4(3):271–273.
29. Hernán MA. Does water kill? A call for less casual causal inferences. *Ann Epidemiol*. 2016;26(10):674–680.
30. Greenland S, Gago-Dominguez M, Castelao JE. The value of risk-factor ("black-box") epidemiology. *Epidemiology*. 2004; 15(5):529–535.
31. Altman N, Krzywinski M. Association, correlation and causation. *Nat Methods*. 2015;12(10):899–900.
32. Savitz DA. In defense of black box epidemiology. *Epidemiology*. 1994;5(5):550–552.
33. Maldonado G, Greenland S. Estimating causal effects. *Int J Epidemiol*. 2002;31(2):422–429.
34. Hernández-Díaz S, Schisterman EF, Hernán MA. The birth weight "paradox" uncovered? *Am J Epidemiol*. 2006;164(11): 1115–1120.
35. Hernán MA, Robins JM, Garcia Rodriguez LA. Discussion on "statistical issues arising in the Women's Health Initiative". *Biometrics*. 2005;61(4):922–930.
36. Greenhouse JB, Kaizar EE, Kelleher K, et al. Generalizing from clinical trial data: a case study. The risk of suicidality among pediatric antidepressant users. *Stat Med*. 2008;27(11): 1801–1813.
37. Lash TL. Heuristic thinking and inference from observational epidemiology. *Epidemiology*. 2007;18(1):67–72.
38. Hill AB. The environment and disease: association or causation? *Proc R Soc Med*. 1965;58(5):295–300.
39. Rothman KJ. Causes. *Am J Epidemiol*. 1976;104(6): 587–592.
40. Splawa-Neyman J, Dabrowska DM, Speed T. On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Stat Sci*. 1990;465–472.
41. Delgado-Rodriguez M, Llorca J. Bias. *J Epidemiol Community Health*. 2004;58(8):635–641.

42. Hernán MA, Hernández-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15(5): 615–625.
43. Lederer DJ, Bell SC, Branson RD, et al. Control of confounding and reporting of results in causal inference studies. Guidance for authors from editors of respiratory, sleep, and critical care journals. *Ann Am Thorac Soc*. 2019; 16(1):22–28.
44. Keil AP, Mooney SJ, Jonsson Funk M, et al. Resolving an apparent paradox in doubly robust estimators. *Am J Epidemiol*. 2018;187(4):891–892.
45. Cole SR, Platt RW, Schisterman EF, et al. Illustrating bias due to conditioning on a collider. *Int J Epidemiol*. 2010; 39(2):417–420.
46. Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annu Rev Public Health*. 2000;21(1): 121–145.
47. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. 2006; 60(7):578–586.
48. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10(1): 37–48.
49. Pearl J. *Causality*. New York, NY: Cambridge University Press; 2009.
50. Richardson TS, Robins JM. Single world intervention graphs (SWIGs): a unification of the counterfactual and graphical approaches to causality. *Cent Stat Soc Sci Univ Washingt Ser Work Pap*. 2013;128(30):2013.
51. Petersen ML. Commentary: applying a causal road map in settings with time-dependent confounding. *Epidemiology*. 2014;25(6):898–901.
52. Petersen ML, van der Laan MJ. Causal models and learning from data: integrating causal modeling and statistical estimation. *Epidemiology*. 2014;25(3): 418–426.
53. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol*. 2016;183(8):758–764.
54. Westreich D, Edwards JK, Rogawski ET, et al. Causal impact: epidemiological approaches for a public health of consequence. *Am J Public Health*. 2016;106(6): 1011–1012.
55. Buckley JP, Keil AP, McGrath LJ, et al. Evolving methods for inference in the presence of healthy worker survivor bias. *Epidemiology*. 2015;26(2):204–212.
56. Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. *Int J Epidemiol*. 2017;46(2):756–762.
57. Balzer LB. "All generalizations are dangerous, even this one."—Alexandre Dumas. *Epidemiology*. 2017;28(4):562–566.
58. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. *Am J Epidemiol*. 2010;172(1):107–115.
59. Lesko CR, Buchanan AL, Westreich D, et al. Generalizing study results: a potential outcomes perspective. *Epidemiology*. 2017;28(4):553–561.
60. Pearl J. Generalizing experimental findings. *J Causal Infer*. 2015;3(2):259–266.
61. Stuart EA, Ackerman B, Westreich D. Generalizability of randomized trial results to target populations: design and analysis possibilities. *Res Soc Work Pract*. 2018; 28(5):532–537.
62. Bareinboim E, Pearl J. A general algorithm for deciding transportability of experimental results. *J Causal Infer*. 2013; 1(1):107–134.
63. Hernán MA, VanderWeele TJ. Compound treatments and transportability of causal inference. *Epidemiology*. 2011; 22(3):368–377.
64. Pearl J, Bareinboim E. External validity and transportability: a formal approach. In: *Paper presented at the Joint Statistical Meetings*. Miami Beach, FL; 2011.
65. Dahabreh IJ, Robertson SE, Tchetgen EJ, et al. Generalizing causal inferences from individuals in randomized trials to all trial-eligible individuals. *Biometrics*. 2019;75(2):685–694.
66. Lesko CR, Cole SR, Hall HI, et al. The effect of antiretroviral therapy on all-cause mortality, generalized to persons diagnosed with HIV in the USA, 2009-11. *Int J Epidemiol*. 2016;45(1):140–150.
67. Westreich D, Edwards JK, Lesko CR, et al. Target validity and the hierarchy of study designs. *Am J Epidemiol*. 2019; 188(2):438–443.
68. Hernán MA. The C-word: scientific euphemisms do not improve causal inference from observational data. *Am J Public Health*. 2018;108(5):616–619.
69. Jenicek M. Epidemiology, evidenced-based medicine, and evidence-based public health. *J Epidemiol*. 1997;7(4): 187–197.