American Journal of Epidemiology

Vol. 188, No. 2
DOI: 10.1093/aje/kwy228
Advance Access publication:
October 9, 2018

Practice of Epidemiology

Target Validity and the Hierarchy of Study Designs

Daniel Westreich*, Jessie K. Edwards, Catherine R. Lesko, Stephen R. Cole, and Elizabeth A. Stuart

* Correspondence to Dr. Daniel Westreich, CB 7435, McGavran-Greenberg Hall, Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 (e-mail: djw@unc.edu).

Initially submitted November 16, 2017; accepted for publication September 27, 2018.

In recent years, increasing attention has been paid to problems of external validity, specifically to methodological approaches for both quantitative generalizability and transportability of study results. However, most approaches to these issues have considered external validity separately from internal validity. Here we argue that considering either internal or external validity in isolation may be problematic. Further, we argue that a joint measure of the validity of an effect estimate with respect to a specific population of interest may be more useful: We call this proposed measure *target validity*. In this work, we introduce and formally define target bias as the total difference between the true causal effect in the target population and the estimated causal effect in the study sample, and target validity as target bias = 0. We illustrate this measure with a series of examples and show how this measure may help us to think more clearly about comparisons between experimental and nonexperimental research results. Specifically, we show that even perfect internal validity does not ensure that a causal effect will be unbiased in a specific target population.

causal inference; external validity; generalizability; internal validity; study design; target population; target validity; transportability

In recent years, increasing attention has been paid to problems of external validity: both the *generalizability* of study results to the population from which the study sample was drawn and the *transportability* of results from a study sample to an external target population (1–12). In either case, a result is externally valid if the true effect in the study sample is unbiased for the true effect in the target population. (This usage is formalized in Web Appendix 1, available at https://academic.oup.com/aje; it is consistent with that of Shadish et al., who regard external validity as the question of whether "the causal relationship holds over variation in persons" (13, p. 472) and other factors, as well as that of Imbens (11).) In contrast, a result is internally valid when the effect *estimated* in the study sample is unbiased for the *true* effect in that sample (and not necessarily in some target population; see Web Appendix 1).

Both generalizability and transportability can be addressed by a combination of assumptions and statistical approaches (1, 9). However, most discussions of external validity to date have considered external validity separately from questions of internal validity and have typically been predicated on an assumption of perfect internal validity. Such an approach can create substantial issues for a public health decision-maker. For example, suppose a decision-maker wants to know whether to intervene in a particular target population and has 2 sources of information: 1) results from a well-conducted trial of the intervention in a population that is unlike the target population and 2) results (possibly still confounded) from a well-conducted observational study of the intervention in a population that is representative of the target population. (Here and hereafter, the reader can take "representative" to mean "a simple random sample of," though the reader can consider an expanded usage including situations in which we have a nonrandom sample of the target population with known sampling probabilities and we use those probabilities properly in analysis to weight back to the target population.)

How should a decision-maker decide on a next step in such a situation? Conventional approaches would suggest that the internally valid evidence—in this case, that from the trial—should be privileged, with only informal attention being paid to the external validity of those results. In this work, we argue that such questions are more subtle, and particularly that considering either internal validity or external validity in isolation may be problematic. Further, we argue that a joint measure of the validity of an effect estimate with respect to a specific population of interest (target population), a measure we here call *target validity*, may be more useful.

In this paper, we first discuss the necessity of a target population to inference and decision-making and the near-certainty that causal effects derived from classical randomized clinical trials cannot be unconditionally generalized to arbitrary target populations when studying humans or other complex animals. We define target validity and illustrate the measure with an example. We conclude with discussion of the implications of a target validity approach to causal inference for attempts to describe a "hierarchy" of study designs.

For convenience and conceptual clarity, we initially concentrate on randomized studies. This is because it is widely understood that an intent-to-treat analysis of a randomized clinical trial (with perfect measurement and no missing data, including no loss to follow-up) will yield an internally valid estimate of the causal effect of treatment (or intervention) assignment. Additionally, under perfect adherence, this will also be an estimate of the effect of the treatment itself. We later expand to discuss observational studies as well, where internal validity is not guaranteed. We also focus primarily on the problem of generalizability (and not transportability) to simplify our discussion; again, in a generalizability framework, the study sample is a proper subset of the target population, whereas in transportability the study sample is at least partially external to the target population (1, 9). While superficially similar, current methods address these 2 problems separately; in particular, presently only generalizability can be addressed using familiar causal diagrams (14), while transportability requires a somewhat different graphical approach (7). We consider differences between these concepts further in the Discussion section.

TARGET POPULATIONS AND EXTERNAL VALIDITY

It is near-universally overlooked that estimates of causal effects obtained from a study sample are only well-defined if they include specific reference to a target population in which they are said to apply (15). Indeed, in few, if any, studies—randomized trials or otherwise—do authors report the target population for their causal effect, much less attempt to generalize quantitatively to that target. As an illustration, we reviewed all randomized trials published in the *New England Journal of Medicine* between January 1, 2016, and December 31, 2016, and found that only 1 in 5 (21%) of these papers discussed generalizability or external validity concerns in describing the results; of those that did mention such issues, relatively few described statistical issues related to generalizability, and none attempted to standardize trial results to a target population (16).

Why is external validity a problem? When a study sample is not formally representative of (or alternately, the same as) the target population, we cannot assume that the true causal effect in a study sample will be the same as the true effect in the population, and thus that the (even unbiased) estimate of the study sample effect will also be an accurate estimate of the target population effect. Differences in these effects may occur for 2 main reasons: nonexchangeability or mathematical necessity. Nonexchangeability for generalizability (in particular) can be thought of as a strong analogy to confounding, and it can be illustrated using causal diagrams (17). Figure 1A shows confounding of the effect of X on Y due to Z through the open path $X \leftarrow Z \rightarrow Y$; Figure 1B shows nonexchangeability due to sampling into the study, which is seen in the open path from $S = 1$ (which is boxed, to indicate that we analyze only those who are in our study sample) to the outcome Y, $[S = 1] \leftarrow Z \rightarrow Y$ (17, 18). In addition, regardless of whether or not there is this sort of nonexchangeability, it is a mathematical certainty that nonnull causal effects will always exhibit heterogeneity on either the risk difference scale or the risk ratio scale if the baseline risk of the outcome changes (see Web Appendix 2) (19).

In the absence of discussion of the target population, it is typically implied or assumed that the target population is either 1) exactly the study sample (in which case generalizability is a nonissue) or 2) the population of which the study sample is a simple random sample (in which case generalizability is assured in expectation). In both cases, it is assumed that the target population is in some way implicitly defined by the inclusion and exclusion criteria of the study.

These are questionable assumptions, however. Regarding the first case noted above, rarely do we desire inference only in the study sample itself, without any interest in principles generalizable to other groups. Indeed, the idea that no inference is desired outside of the study sample may be ethically as well as scientifically problematic: Few human subjects review boards would approve a randomized trial that actively sought to create knowledge which was of no use beyond the study sample (though we can imagine this being the case in large-scale pragmatic settings). In addition, if the goal is to take action in the target population, then the study sample will always differ from the target population in time: We will only apply the results of the study to the target population in the future (20, 21).

The second case is likewise questionable. The fact that randomized trials require informed individual consent from study participants effectively eliminates the possibility of a true simple random sample from a target population (again, this can sometimes be avoided in pragmatic or community-randomized settings). More importantly, perhaps, randomized trials routinely overenroll persons at high risk of the outcome under study in order to enhance precision; for example, human immunodeficiency virus prevention trials will frequently overenroll persons who report high-risk sexual behavior. Such risk enrichment ensures that the study sample is not a simple random sample of the target population and can easily lead to lack of generalizability (see Web Appendix 2). Furthermore, it is common for inclusion and exclusion criteria of a randomized trial to walk a thin line between ensuring high rates of the outcome of interest and ensuring low rates of adverse events, further increasing the likelihood that the study sample is not a simple random sample of the target population of interest (22).

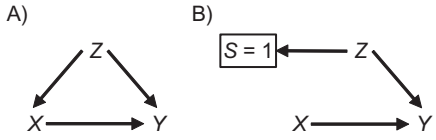

Figure 1. Causal diagrams for nonexchangeability for internal validity due to confounding (A) and nonexchangeability for external validity due to sampling bias (B).

Irrespective of ongoing debates on the *value* of representativeness (23–27), it is rare for the study sample in a randomized trial to *actually be* representative of the target population. When the sample is nonrepresentative of a target population, claims of external validity in that target population cannot be made without assumptions about similarities between the study sample and the target population or about homogeneity of effects.

In addition, even if a study sample is sampled at random from the target population of interest, it will not be representative of *other* target populations. For example, consider a randomized trial for which the investigators' target population is all residents of North Carolina. If the trial is conducted in a randomly selected subgroup of North Carolina residents, it might be representative of (and thus externally valid for) the state of North Carolina—but we would have no such guarantees for the United States as a whole (or the state of South Carolina, though that is a problem of transportability rather than generalizability). Because of this, to claim generally that a trial has "good external validity" is a category error: Since the prevalence of effect-measure modifiers differs from one target population to another, the degree to which a causal effect estimate from a randomized trial is externally valid will (in general) differ from one target population to another as well (9). For example, if biological sex is an effect-measure modifier and a randomized trial population is composed of 75% women, the generalizability of the results will differ between 3 target populations containing 75%, 50%, and 25% women, respectively. If external validity changes depending on the choice of target population, then any claim of "external validity" or "generalizability" which does not specify the target population is not a meaningful scientific claim. Ultimately, it is useful to remember that generalizability is a relationship between a study sample and a target population for a particular question—rather than a single inherent characteristic of a study (1).

IDENTIFICATION OF EXTERNALLY VALID EFFECTS IN A TARGET POPULATION

Pearl, Bareinboim, and others give conditions for both generalizability and transportability (1, 3, 5, 7), key among them being independence (conditional or otherwise) of study participation (or sampling) from the outcome under study. In generalizability, this can be seen as a missing-data problem under complete-case analysis, in which the outcomes in the target population are missing except in the study sample (18). It may often be more intuitive (though less formal) to think of generalizability in terms of effect-measure heterogeneity: If there is an effect-measure modifier of a causal effect that has a different prevalence in the study sample than in the target population, generalizability from the study sample to the target population is not guaranteed (see Web Appendix 2) (16, 28).

If all such effect-measure modifiers are measured in the randomized trial and the target population, then (under additional assumptions or conditions, often including external positivity (1), external consistency (5), and external interference equivalence (5)) it is possible to analytically generalize a causal effect from a study sample to a specific target population (9, 16). As with confounding, the possibility of unmeasured effect-measure modifiers' leading to nonexchangeability between the study sample and the target population is a potential issue. As noted above, this issue can be solved in expectation with random sampling of study participants from a target population. In the absence of such random sampling, we can assert generalizability to a specific target population only under an assumption similar to the typical "no unmeasured confounding" assumption in observational studies (1). (Recently developed methods provide an analysis of sensitivity of target population effect estimates to unobserved effect-measure modification (29).)

Few population health scientists would accept the assertion—implicit or explicit—that the crude (unadjusted) results of an observational study were valid without attention to any possible confounding, but population health science has paid less attention to the highly analogous claim that the results of a randomized trial are immediately, or crudely, generalizable to a particular target population. The latter claim is neither more nor less valid than the former: Both depend on an assumption of exchangeability. An exposure-outcome relationship in the presence of unexamined, possibly uncontrolled confounding is not considered more than an association: This is uncontroversial. We suggest that a result from a randomized trial in the presence of unexamined nongeneralizability—while interpretable as a causal effect in the study sample—should likewise be considered merely an "association" *with respect to estimation of the causal parameter in the target population*. That is, while the result of a randomized trial is indeed a valid estimate of a causal effect in the study sample, the potential for nonexchangeability with the target populations should give us pause before we interpret the effect as a valid estimate of the causal effect in the target population (30).

TOWARD TARGET VALIDITY

Above, we discussed the idea that results that are identified as "internally valid" are not as useful (or as complete) as sometimes thought, because no causal effect estimate is well-specified without a target population, and the study sample is rarely if ever equal to, and often is not representative of, the target population. We also noted that a simple and unelaborated claim that a result is "externally valid" (without reference to a specific and well-characterized target population) was poorly formed as a scientific statement. In sum, then, it makes little sense to speak of internal validity in isolation, and it makes little sense to speak of external validity in the abstract.

How then ought we discuss the validity of our studies? Here, we suggest that public health and medicine may benefit from a more integrated perspective than what is promoted by the current separation of internal and external validity. We suggest a combined metric that addresses the total validity of a causal effect estimate in the specified target population, a quantity we call "target validity."

This concept extends the framework presented by Imai et al. (30), which decomposed the overall bias in a treatment effect estimate for a well-defined target population into internal and external components. Those authors showed that the bias present when one is using a sample of treated and control subjects to estimate the average causal effect in a target population can be decomposed into 4 pieces: internal validity bias due to

Table 1. Relationships Between Random Sampling/Treatment Assignments and Internal and External Validity[a]

Is the Treatment Randomized or Observed?	Is the Study Sample Representative of the Target?		Internal Validity
	Representative	Not Representative	
Randomized	Web Appendix 3, example a	Web Appendix 3, example b	High internal validity
Observed	Web Appendix 3, example c	Web Appendix 3, example d	Unknown internal validity
External validity	High external validity	Unknown external validity	

[a] Interior cells refer to numerical examples from Web Appendices 3 and 4. Representativeness of the study sample may be achieved most simply through random sampling of the study sample from the target population; it may also arguably be achieved by sampling of the study sample from the target population with known sampling proportions and then applying those weights correctly during analysis.

1) observed and 2) unobserved factors and external validity bias due to 3) observed and 4) unobserved factors. As they documented in their paper (30), different study designs have different trade-offs in terms of these 4 components. For example, a "typical" nonrepresentative randomized trial may have (in expectation) higher internal validity because randomization provides exchangeability between study arms in expectation, but lower external validity for the target population of interest. In contrast, a "typical" nonexperimental study conducted in a large-scale data set may have lower internal validity (especially due to unobserved confounders) but larger external validity.

Starting with these ideas, we give a formal definition (in terms of potential outcomes) of target validity for the risk difference in Web Appendix 1 for both the generalizability and transportability cases. Here, we informally note that target bias for the risk difference is simply the sum of internal bias and external bias for a specific target population, and can be thought of therefore as the difference between the *true causal risk difference in the target population* and the *estimated risk difference in the study sample*. We illustrate the utility of the target validity concept in Web Appendix 3 with an example, in several parts; data for this example are given in an accompanying Excel (Microsoft Corporation, Redmond, Washington) spreadsheet (Web Appendix 4).

The example from Web Appendix 3 helps us fill out a 2×2 table to guide our understanding of target validity, addressing 2 questions: First, is the study sample representative of the target population (the columns)? And second, is the treatment randomized or observed (the rows)? We offer such a 2×2 table as Table 1. In this table, the top-left cell represents a doubly randomized experiment (31), while the contents of the cells correspond to examples a–d in Web Appendix 3. Changing our orientation to target validity helps us recognize that *only* when a study sample is representative of the target (randomly sampled from the target; alternately, sampled from the target with known sampling proportions and then weighted properly to represent the target) and treatment is randomized in the study sample can we count on high validity with respect to the target population (Web Appendix 3, example a)—in all other situations (Web Appendix 3, examples b–d), the balance of internal and external validity is a priori unknown.

DISCUSSION

Here we have introduced the concept of target validity, as the overall validity of a causal effect estimate in the specific target population of interest. This idea, while simple, has potentially significant implications for the way we generate and evaluate evidence. Several points are worth discussing.

We argue that the hierarchy in which internal validity is of primary importance to overall validity and external validity is considered only secondarily, while sometimes a useful perspective, may also be misleading. While is it true that, lacking internal validity, perfect external validity (or perfect representativeness of the target population) will not help an investigator obtain an estimate of effect that is unbiased with respect to the target population, it is equally true that even *with* perfect internal validity, a lack of external validity leads to bias with respect to the target population (Web Appendix 3, example b). Integrating our thinking about internal and external validity to a greater degree may be more useful in improving our approach to causal effect estimation.

Once we accept that validity should be measured with respect to a specific target population, the symmetry of internal and external validity becomes apparent: Internal validity can be threatened by confounding and selection bias, which can be dealt with by randomizing treatment. External validity, in a similar way, can be threatened by lack of exchangeability between the study sample and the target population, which can be dealt with by random sampling from the target population into the study. In both cases we can still sometimes (though not always (17, 32)) obtain valid results through quantitative adjustment for confounding factors and selection bias (internal validity) or for causal pathways linking sampling into the study and the outcome (external validity)—if we have correctly chosen, measured, and modeled the variables for which we need to account.

It may alarm some scientists (and perhaps some policymakers) to confront quantitatively the idea that—unless a study population is randomly sampled from the target population of interest (or sampled with known probabilities and then reweighted)—randomization of treatment alone does not *guarantee* an unbiased estimate of effect in the target population (as we illustrate in Web Appendix 3, example c). Indeed, the idea that the result of a randomized trial will be valid in the target population rests on the same kind of (potentially invalid) assumptions as those in any observational analysis. It follows from all of the above that the gold standard of evidence for assessing a causal effect estimate in a given target population is not simply a randomized trial but rather *a randomized trial conducted among a random sample of the target population*.

This, in turn, puts in question the many rankings of evidence—hierarchies of study designs—that near-universally place randomized trials above observational studies. Such rankings seem to us likely to be based on one of 3 points: 1) a belief that the target population is equal to, or perfectly represented by, the study sample; 2) a belief that internal validity alone is a proper basis for decision-making or scientific evidence (possibly related to underlying beliefs that randomized trials uncover scientific truths or laws about the universe, a belief that seemingly discounts the possibility of effect heterogeneity); or 3) a belief that randomized trials have greater target validity (in general) than observational studies (a counterpoint to which was given above). Regardless of underlying motivation, the fact of such hierarchies is questionable: Specifically, given a choice between an observational but possibly confounded estimate and a randomized but poorly generalizable estimate, it does not seem obvious which should inform public health or clinical decision-making to a greater degree.

In this work we have focused on generalizability (in which the study sample is a subset of the target population) rather than transportability (the study sample is not a subset of the target population). In Web Appendix 1, we derive a formal definition of target bias and thus target validity for transportability; but here we wish to draw the reader's attention to 2 key differences between generalizability and transportability. First, at present, problems of transportability cannot be expressed using causal diagrams but require (similar, but not identical) selection diagrams instead (7). Thus, direct graphical analogies of transportability to generalizability should proceed cautiously. Second, we argued above that the only way to guarantee external validity in expectation is to randomly select the study sample from the target population. In the transportability case—in which the study sample is not a subset of the target population—such sampling is simply not possible, and thus analytical approaches are probably necessary.

We mention several additional points briefly. First, some study designs, including pragmatic and community-randomized trials and certain pseudoexperiments and nonexperimental studies, may have extremely high target validity under some circumstances and are attractive study designs in part for exactly that reason. More work is needed to characterize the conditions under which target validity is high and exactly what data are required when quantitative approaches to generalizability or transportability are to be pursued (9, 16), as well as power calculations for such analyses. Second, consideration of target validity may have applications to the way we conduct and consider meta-analyses, especially meta-analyses of observational studies (33). We should first ask, what is the target population of each study in the meta-analysis? And then, what is the final target population for the result of the meta-analysis? Third, we have assumed throughout this work that the target population is broadly relevant to the intervention under study; clearly, a smoking cessation intervention is unlikely to prevent mortality in a population of nonsmokers. Thus, the role of commonsense contextual information should not be overlooked. Fourth, investigators may wish to estimate effects only for a subset of a study sample, as in the effect of the treatment in the treated; but the external validity of those effects must be evaluated just as for sample average causal effects.

Finally, it is possible that the components of target validity are not of equal importance: Perhaps (consistent with widely held views) internal validity is more important than external validity in many or even most real-world cases. This might be particularly true if issues of internal validity, such as confounding or selection bias, lead to reversal of sign (e.g., make a harmful treatment look helpful) more often than external validity. Evidence on whether internal validity is more important than external validity in real-world settings would be welcome, but such broad empirical investigation of this issue is beyond the scope of the present work.

In conclusion, we propose that the prioritization of internal validity over external validity, and the related consideration of internal validity and external validity separately, is a perspective that has weakened, rather than strengthened, public health evidence writ large. We encourage other investigators to generalize their thinking about validity and study hierarchies and to embrace the idea of target validity.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina (Daniel Westreich, Jessie K. Edwards, Stephen R. Cole); Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland (Catherine R. Lesko); and Departments of Mental Health, Biostatistics, and Health Policy and Management, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland (Elizabeth A. Stuart).

This research was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development and the Office of the Director of the National Institutes of Health (award DP2-HD084070) and the National Institute of Allergy and Infectious Diseases (grant R01 AI100654).

The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest: none declared.

REFERENCES

1. Lesko CR, Buchanan AL, Westreich D, et al. Generalizing study results: a potential outcomes perspective. *Epidemiology*. 2017;28(4):553–561.
2. Bareinboim E, Lee S, Honavar V, et al. Transportability from multiple environments with limited experiments. In: Burges CJC, Bottou L, Welling M, et al., eds. *Advances in Neural Information Processing Systems 26*. Red Hook, NY: Curran Associates, Inc.; 2014:136–144.
3. Bareinboim E, Pearl J. A general algorithm for deciding transportability of experimental results. *J Causal Infer*. 2013; 1(1):107–134.
4. Bareinboim E, Pearl J. Transportability from multiple environments with limited experiments: completeness results. In: Ghahramani Z, Welling M, Cortes C, et al., eds. *Advances*

in Neural Information Processing Systems 27. Red Hook, NY: Curran Associates, Inc.; 2015:280–288.
5. Hernán MA, VanderWeele TJ. Compound treatments and transportability of causal inference. *Epidemiology*. 2011;22(3):368–377.
6. Pearl J, Bareinboim E. External validity and transportability: a formal approach. In: *2011 JSM Proceedings: Papers Presented at the Joint Statistical Meetings, Miami Beach, Florida, July 30–August 4, 2011, and Other ASA-Sponsored Conferences*. Alexandria, VA: American Statistical Association; 2011:157–171.
7. Pearl J, Bareinboim E. External validity: from do-calculus to transportability across populations. *Stat Sci*. 2014;29(4):579–595.
8. Petersen ML. Compound treatments, transportability, and the structural causal model: the power and simplicity of causal graphs. *Epidemiology*. 2011;22(3):378–381.
9. Westreich D, Edwards JK, Lesko CR, et al. Transportability of trial results using inverse odds of sampling weights. *Am J Epidemiol*. 2017;186(8):1010–1014.
10. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med*. 2018;210:2–21.
11. Imbens G. Understanding and misunderstanding randomized controlled trials: a commentary on Deaton and Cartwright. *Soc Sci Med*. 2018;210:50–52.
12. Pearl J. Challenging the hegemony of randomized controlled trials: a commentary on Deaton and Cartwright. *Soc Sci Med*. 2018;210:60–62.
13. Shadish W, Cook T, Campbell D. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. New York, NY: Houghton Mifflin Company; 2002.
14. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10(1):37–48.
15. Maldonado G, Greenland S. Estimating causal effects. *Int J Epidemiol*. 2002;31(2):422–429.
16. 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.
17. Bareinboim E, Tian J, Pearl J. Recovering from selection bias in causal and statistical inference. In: *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence and the Twenty-Sixth Innovative Applications of Artificial Intelligence Conference*. Palo Alto, CA: AAAI Press; 2014:2410–2416.
18. Daniel RM, Kenward MG, Cousens SN, et al. Using causal diagrams to guide analysis in missing data problems. *Stat Methods Med Res*. 2012;21(3):243–256.
19. Hernán MA. Invited commentary: selection bias without colliders. *Am J Epidemiol*. 2017;185(11):1048–1050.
20. Hoggatt KJ, Greenland S. Commentary: extending organizational schema for causal effects. *Epidemiology*. 2014;25(1):98–102.
21. Rogawski ET, Gray CL, Poole C. An argument for renewed focus on epidemiology for public health. *Ann Epidemiol*. 2016;26(10):729–733.
22. 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.
23. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. *Int J Epidemiol*. 2013;42(4):1012–1014.
24. Elwood JM. Commentary: on representativeness. *Int J Epidemiol*. 2013;42(4):1014–1015.
25. Nohr EA, Olsen J. Commentary: epidemiologists have debated representativeness for more than 40 years—has the time come to move on? *Int J Epidemiol*. 2013;42(4):1016–1017.
26. Richiardi L, Pizzi C, Pearce N. Commentary: representativeness is usually not necessary and often should be avoided. *Int J Epidemiol*. 2013;42(4):1018–1022.
27. Schooling CM, Jones HE. Is representativeness the right question? *Int J Epidemiol*. 2014;43(2):631–632.
28. Olsen RB, Orr LL, Bell SH, et al. External validity in policy evaluations that choose sites purposively. *J Policy Anal Manage*. 2013;32(1):107–121.
29. Nguyen TQ, Ebnesajjad C, Cole SR, et al. Sensitivity analysis for an unobserved moderator in RCT-to-target-population generalization of treatment effects. *Ann Appl Stat*. 2017;11(1):225–247.
30. Imai K, King G, Stuart EA. Misunderstandings between experimentalists and observationalists about causal inference. *J R Stat Soc Ser A Stat Soc*. 2008;171(2):481–502.
31. Cole SR. Nondogmatism. *Ann Epidemiol*. 2016;26(4):231–233.
32. Hernán MA, Alonso A, Logroscino G. Cigarette smoking and dementia: potential selection bias in the elderly. *Epidemiology*. 2008;19(3):448–450.
33. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med*. 2000;342(25):1887–1892.