American Journal of Epidemiology

Vol. 192, No. 3
https://doi.org/10.1093/aje/kwac200
Advance Access publication:
November 8, 2022

Practice of Epidemiology

Proxy Variables and the Generalizability of Study Results

Anton Nilsson*, Jonas Björk, and Carl Bonander

* Correspondence to Dr. Anton Nilsson, Tornblad Building, Occupational and Environmental Medicine, Lund University, Box 117, SE-22100 Lund, Sweden (e-mail: anton.nilsson@med.lu.se).

Initially submitted September 22, 2021; accepted for publication November 4, 2022.

When individuals self-select (or are selected) into a study based on factors that influence the outcome, conclusions may not generalize to the full population. To compensate for this, results may be adjusted, for example, by standardization on the set of common causes of participation and outcome. Although such standardization is useful in some contexts, the common causes of participation and outcome may in practice not be fully observed. Instead, the researcher may have access to one or several variables related to the common causes, that is, to proxies for the common causes. This article defines and examines different types of proxy variables and shows how these can be used to obtain generalizable study results. First of all, the researcher may exploit proxies that influence only participation or outcome but which still allow for perfect generalizability by rendering participation and outcome conditionally independent. Further, generalizability can be achieved by leveraging 2 proxies, one of which is allowed to influence participation and one of which is allowed to influence the outcome, even if participation and outcome do not become independent conditional on these. Finally, approximate generalizability may be obtained by exploiting a single proxy that does not itself influence participation or outcome.

directed acyclic graphs; external validity; generalizability; proxy variables

Abbreviations: ATE, average treatment effect; CRR, causal risk ratio; DAG, directed acyclic graph.

Epidemiologists usually place great emphasis on internal validity, striving to obtain estimates that reflect causal effects among the individuals present in the sample. There has traditionally been less focus on issues of external validity, such as whether estimates of causal effects obtained from a sample may generalize to the full population from which the sample participants were obtained. However, participants in clinical trials, cohorts, and other epidemiologic studies are rarely fully representative of the population as a whole (1, 2), meaning that results may not generalize, even if they are valid internally. Self-selection is one common reason for lack of representativeness, but the phenomenon can also arise because of decisions by researchers or data collectors, who for convenience or reasons of scientific interest may choose to focus on certain groups of individuals.

There is a recent and growing literature on the generalizability of experimental and cohort findings, showing that lack of generalizability has similarities with confounding, and providing methods to correct for it (3–10), such as by the use of standardization or reweighting. These methods work by balancing the sample with respect to some set of characteristics that are "fully observed," that is, that are observed both in the study sample and in (a representative sample of) the population.

Conceptually, the resemblance between confounding and lack of generalizability is quite straightforward: Whereas confounding is due to common causes of treatment and outcome, lack of generalizability is, essentially, due to common causes of study participation and outcome (10–13). A simple strategy to determine which variables to account for to achieve generalizability is therefore to identify all factors that one believes influence both participation and outcome. In practice, some of these factors may be unobserved in the sample and/or in the population, raising the question of what alternative adjustments can be made to restore generalizability, either perfectly (i.e., without bias) or approximately.

Drawing on the similarities between confounding and lack of generalizability, this article elaborates on different proxy variable adjustments that can be made to achieve generalizable results in cases where (some of) the common causes of

study participation and the outcome are not fully observed. We define and examine 5 different types of variables that can act as proxies for common causes of study participation and outcome. First, we define type 1 proxy variables, which influence participation but not the outcome, but allow for generalizability by rendering participation and outcome conditionally independent. Correspondingly, type 2 proxy variables influence the outcome but not participation, but allow for generalizability by rendering participation and outcome conditionally independent. Further, we define type 1B and 2B proxy variables, which do not render participation and outcome conditionally independent but can together nevertheless be used to restore generalizability. Finally, we define type 0 proxy variables, which do not render participation and outcome conditionally independent but can be used to mitigate bias due to lack of representativeness. We use directed acyclic graphs (DAGs) (14, 15) to illustrate the concepts and we provide realistic examples of the different types of proxy variables.

COMMON CAUSES AND CONDITIONAL INDEPENDENCE

Before turning to proxy variables, we will in this section review some basic concepts and conditions under which study results can be generalized to a target population from which the study participants were obtained. We assume that we have a study sample that contains data on an outcome, a treatment, and some background characteristics (i.e., characteristics that do not depend on the treatment), which may be fully observed. When we say that a finding can be "generalized," we mean that it can be extended from the sample to the population (16), either directly (i.e., the unadjusted sample estimate is, in expectation, identical to the corresponding population quantity), or with proper adjustment for some background characteristics.

The core scenario of interest to this article is illustrated by the DAG in Figure 1, where we use A to denote a randomized treatment. The outcome is represented by Y, whereas study participation is represented by S (i.e., the sample is restricted to those with $S = 1$). There is also a background variable, denoted by C. As always in DAGs, directed arrows represent direct causal effects. Hence, Figure 1 encodes the assumptions that A influences Y and that C is a common cause of S and Y. In Figure 1 as throughout our examples, we assume that S does not influence anything (other than treatment assignment, which does not need to be indicated in the DAG (5)) and that S is not influenced by Y. The latter implies that we do not consider case-control studies, where the inclusion probability depends on the outcome.

In Figure 1, we can imagine that C contains aspects of health or personality characteristics, such as conscientiousness, which is a characteristic that is related to both study participation (17, 18) and disease outcomes such as mortality (19, 20). Since study participation and mortality are linked via this underlying factor, unadjusted sample-based calculations pertaining to mortality, including how it is influenced by a treatment A, may fail to generalize (3, 11, 13).

In studies of causal effects, researchers are often interested in estimating the average treatment effect (ATE) of a

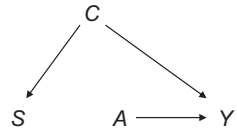

Figure 1. A causal diagram with common cause(s) (C) of study participation (S) and outcome (Y), posing a threat to generalizability of the causal effect of treatment A on Y unless adjusted for, such as by standardization.

binary treatment. (If the outcome is also binary, the ATE is the same as the causal risk difference.) For the unadjusted ATE to not generalize, the scenario in Figure 1 needs to be complemented with an additional condition: C needs to be an effect measure modifier of the effect of A on Y on an additive scale (21–25). Expressed differently, the ATE needs to vary across levels of C for generalizability to be compromised.

As an alternative to the ATE, researchers may be interested in estimating a measure of relative effect, such as the causal risk ratio (CRR). Whereas the unadjusted ATE fails to generalize if C in Figure 1 is an effect measure modifier on an additive scale, the unadjusted CRR fails to do so if C is an effect measure modifier on a multiplicative scale (24, 25). In general, given that both C and A influence Y, effect measure modification is present on at least one of these two scales (26). At least one of the unadjusted ATE and the unadjusted CRR would therefore fail to generalize in a scenario like Figure 1.

While unadjusted results in a scenario like Figure 1 may not generalize, generalizability can typically be obtained by adjustment for C, provided that C is fully observed. Additional adjustments may be needed if the treatment was not randomized in the sample, as causal effects may otherwise not even be identified in the sample. Below, we provide a comprehensive list of conditions that allows population causal effects to be identified.

A sufficient set of conditions for identifying and generalizing a causal effect. Let $A \perp\!\!\!\perp B \mid C$ denote that A and B are independent conditional on some covariate(s) C, and Y^a denote the potential outcome that an individual would have if, possibly contrary to fact, their treatment was $A = a$. Assume that the following conditions hold for all a:

1. Treatment assignment ignorability: $Y^a \perp\!\!\!\perp A \mid (C, S = 1)$.
2. S-ignorability: $Y^a \perp\!\!\!\perp S \mid C$.
3. A and Y are measured in the study sample.
4. C is fully observed, that is, is measured in both the study sample and in (a representative sample of) the population.
5. Positivity, meaning that for each value of C that exists in the population, there is a positive probability of observing individuals with treatment $A = a$ in the study sample.
6. Consistency of potential outcomes: $Y = Y^a$ whenever $A = a$, meaning that treatment does not exist in different versions that are relevant for producing the outcome.

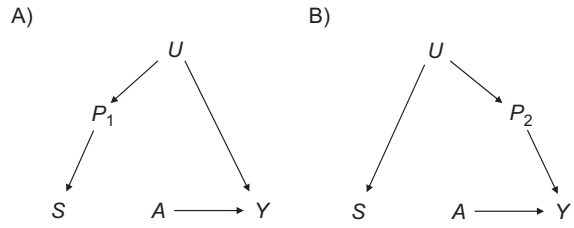

Figure 2. Causal diagrams with proxy variables of type 1 and 2: A) type 1 (P_1); B) type 2 (P_2). As before, S represents sample participation, Y the outcome, and A the treatment. U represents the unobserved common cause(s) of sample participation and the outcome. Generalizability can be achieved by standardizing on the relevant proxy.

If, given some background factors C, the above conditions hold, the population expected value of Y^a is identified by the standardization formula:

$$E\left(Y^a\right) = \sum_c E\left(Y|A=a, C=c, S=1\right) \Pr\left(C=c\right). \quad (1)$$

The proof of equation 1, which can also be found in previous literature (7), is provided in Web Appendix 1 (available at https://doi.org/10.1093/aje/kwac200). If one assumes that conditions 3–6 hold by default, condition 1 is the critical condition needed for identification of causal effects in the sample, and condition 2 is the critical condition needed for generalization of such effects. For simplicity, we focus on the standardization approach in this article, but note that other methods for identifying and generalizing causal effects exist as well, including inverse probability weighting (7–9, 27).

Using the standardization formula (equation 1), population causal effects can be obtained by contrasting $E(Y^a)$ for different choices of a. For example, the population ATE is given by $E(Y^1) - E(Y^0)$, whereas the population CRR is given by $E(Y^1)/E(Y^0)$. Particular choices of effect measure may allow condition 2 to be replaced by some weaker condition (i.e., by a condition that may be fulfilled with fewer adjustments than what condition 2 requires). When the aim is to generalize an ATE, it can be replaced by $Y^1 - Y^0 \perp\!\!\!\perp S \mid C$ (21–23).

Sufficient conditions for identifying and generalizing causal effects can be formally expressed in graph-based terminology (3, 5, 11). Given the presumptions that the outcome does not influence participation and that participation does not influence the outcome (other than by influencing treatment), the question of whether a crude sample association between A and Y corresponds to a population causal effect boils down to whether there are common causes of any of the following:

- A and Y; or
- S and Y, where one should ignore any connection that may operate via A.

The absence of common causes of any of these two types corresponds to conditions 1 and 2 being fulfilled unconditionally. If common causes of any of the two types are present, adjustment by formulas such as equation 1 may be necessary—either adjustment for the common causes themselves or for some proxy variable(s), the topic that we return to in subsequent sections.

In Web Appendixes 1–4 (Web Figures 1–5), we consider adjustments in the general scenario, where neither condition 1 nor condition 2 may hold unconditionally, so that the investigator must simultaneously deal with identification and generalization of causal effects. Throughout the rest of the main text of this article, we will, however, assume that condition 1 holds unconditionally, allowing us to focus on generalization. We achieve this by assuming, as in Figure 1, that: 1) treatment is randomized in the sample, so that there are no common causes of A and Y; and that 2) A does not influence S, thus avoiding issues of collider bias (28). This means that measures of causal effects can be identified in the sample even without any adjustment, and the issue of whether adjustments may be needed to determine a population causal effect will depend only on whether there are any common causes S and Y (particularly common causes that are effect measure modifiers on a relevant scale). Throughout both the main text and the appendices, we assume by default that conditions 3 and 6 hold, as well as that conditions 4 and 5 hold for certain (possibly proxy) variables, so that the terms in standardization formula (equation 1)—and in other formulas, to be discussed—can be calculated.

PERFECT GENERALIZABILITY WITH A PROXY VARIABLE

In many cases, some common causes of S and Y may not be fully observed. Assume that X represents the set of common causes that are fully observed, whereas U represents those that are not. For simplicity, we henceforth refer to U just as "unobserved." Since U is unobserved, (X, U) cannot be standardized on, which poses a potential threat to the generalizability of results from the study. Particularly, the generalizability of the ATE (CRR) is threatened if U contains variables that are effect measure modifiers on an additive (multiplicative) scale.

Whereas (X, U) cannot be standardized on, the researcher might be able to achieve generalizability by exploiting some fully observed "proxy variables," P, that is, by standardizing on (X, P) instead of (X, U). In graphical terms (3, 5, 11), the researcher is looking for some fully observed background variables P such that (X, P) blocks all paths between S and Y. We refer the reader to previous literature on DAGs (14, 15) for more comprehensive discussions of the concept of blocking, but note that a common cause C of two variables A and B gives rise to a path between the two (e.g., $A \leftarrow P_1 \leftarrow C \rightarrow P_2 \rightarrow B$), and that this path said to be open (or "unblocked") if no variable along the path is being adjusted for. In this scenario, A and B will typically be statistically dependent. If some variable along the path, either C or some other variable P_k, is being adjusted for, the path is instead said to be blocked. If all paths between two variables are

blocked, the variables are said to be d-separated, which implies that the 2 variables are conditionally independent (14, 15). In our scenario, with X and U being common causes of S and Y and where we want to block all the paths between S and Y, we may hence either adjust for X and U themselves or, if this is not possible (as with U), adjust for some other factors along the paths generated by the common causes.

Figure 2 defines 2 simple types of proxy variables that can be used to generalize study results. (Here, as in the subsequent figures, we for simplicity omit the possible X variables.) First, in Figure 2A, we display a variable that we refer to as a type 1 proxy, P_1. The characteristic feature of this variable is that it depends on the common cause(s) U, influences S but not Y, and blocks all paths between S and Y, rendering them conditionally independent.

As for a concrete example, suppose that the target population consists of individuals with breast cancer, and we want to evaluate survival, Y, and how it depends on a randomized treatment, A. In their study, a team of researchers has primarily included breast cancer patients with a high value on a severity score, P_1, a score that depends on the true but unobserved severity of the condition, U. The unobserved severity in turn influences Y. Other than via P_1, study participation is assumed to not depend on U. While unadjusted effect estimates from the study may not generalize to the target population, it is possible to achieve generalizability by standardizing on the population distribution of P_1.

In Figure 2B, we display a variable that we refer to as a type 2 proxy, P_2. The characteristic feature of this variable is that it depends on the common cause(s) U, influences Y but not S, and blocks all paths between S and Y, rendering them conditionally independent. This time, suppose that the target population is the full US population, and a research team wants to examine the effect of a randomized dietary supplement, A, on the risk of lung cancer, Y. We assume that participants in the study volunteered based on a personality characteristic U, such as conscientiousness. The characteristic influences smoking behavior P_2, which in turn influences the risk of lung cancer. Other than via P_2, the outcome is assumed to not depend on U. While unadjusted effect estimates from the study may not generalize to the target population, it is possible to obtain generalizable results by standardizing on the population distribution of P_2.

PERFECT GENERALIZABILITY WITH 2 PROXY VARIABLES

Next, we define a scenario with 2 (sets of) proxy variables that can together be used to generalize study results. The setting, which combines and generalizes the two settings in Figure 2, is displayed in Figure 3 (the effects indicated by the dashed lines are allowed for but not necessary). Since U represents some unobserved common cause(s) of participation and outcome, it poses a potential threat to generalizability, particularly so for the ATE (CRR) if U modifies the effect of A on Y on an additive (multiplicative) scale.

There are two proxy variables present at the same time, one that is allowed to influence S but not Y, and one that is allowed to influence Y but not S. Unlike before, U also has direct impacts on S and Y, not mediated by these proxies. Because of this feature, we refer to the proxies as type 1B and type 2B proxies instead of just type 1 and type 2, and we use the symbols P_{1B} and P_{2B} in the graph. Due to the direct effects of U on S and Y, exploiting P_{1B} and/or P_{2B} to standardize with equation 1 may not provide generalizable results. Given that P_{1B} and P_{2B} are observed in the sample and that the distribution of P_{2B} is known in the population, however, an alternative adjustment formula, the proof of which we provide in Web Appendix 3, may be applied. In the case of categorical or binary variables, a crucial requirement to be able to apply the formula is that both proxy variables have at least as many levels or components as U (otherwise, the proxies would not fully capture the variation in U).

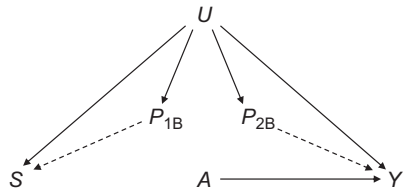

Figure 3. A causal diagram with proxy variables of type 1B (P_{1B}) and type 2B (P_{2B}). As before, S represents sample participation, Y the outcome, A the treatment, and U the unobserved common cause(s) of sample participation and the outcome. Generalizability can be achieved by exploiting P_{1B} and P_{2B} simultaneously. The dashed lines indicate causal effects that are allowed for but are not necessary for the conclusions to hold.

With a binary U, two binary type 1B and type 2B proxy variables P_{1B} and P_{2B}, and, as in Figure 3, a randomized treatment that does not influence participation and with no other fully observed background variable X involved, the adjustment formula becomes:

$$E\left(Y^a\right) = \frac{E(P_{2B})-E(P_{2B}|P_{1B}=0,S=1)}{E(P_{2B}|P_{1B}=1,S=1)-E(P_{2B}|P_{1B}=0,S=1)} * E\left(Y|P_{1B}=1,A=a,S=1\right) + \frac{E(P_{2B}|P_{1B}=1,S=1)-E(P_{2B})}{E(P_{2B}|P_{1B}=1,S=1)-E(P_{2B}|P_{1B}=0,S=1)} * E\left(Y|P_{1B}=0,A=a,S=1\right). \quad (2)$$

As for a concrete example, suppose that a team of researchers is interested in examining emotional well-being in a target population of individuals having experienced a natural disaster, and decide to conduct a randomized experiment to decide whether a counseling program may be helpful to alleviate symptoms. Again, a challenging issue may be that conscientious individuals are more prone to participate in the study. If such individuals are also better at avoiding emotional issues, such as via coping strategies, effect estimates may not generalize. Assume that P_{2B} represents the use of coping strategies such as mindfulness, which might partly explain why conscientious individuals fare better. Further assume that individuals could submit

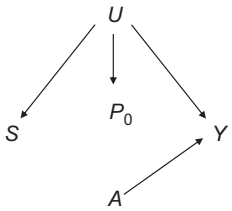

Figure 4. A causal diagram with a proxy of type 0 (P_0). As before, *S* represents sample participation, *Y* the outcome, *A* the treatment, and *U* the unobserved common cause(s) of sample participation and outcome. Given that *A*, *U*, and P_0 are binary, approximate generalizability of the average treatment effect may be achieved by standardizing on P_0.

an application of interest to participate in the study. The action of submitting the application is denoted by P_{1B}, and its likelihood depends on conscientiousness. Submitting the application might in turn increase the likelihood of participating (e.g., because applicants receive a reminder to participate in the study), but conscientious individuals are also more likely to enroll regardless of whether they submitted the application. Given this setup—and assuming that using coping strategies does not influence study participation, that submitting the application of interest does not influence well-being, that the prevalence of the coping strategy is known in the target population of individuals affected by the natural disaster, and that U, P_{1B}, and P_{2B} are all binary—equation 2 could be exploited to generalize causal effects from the study.

IMPERFECT GENERALIZABILITY WITH A PROXY VARIABLE

Finally, we define a proxy variable that does not influence any variable of interest but which represents a misclassified version of the underlying common cause(s) of participation and outcome. The scenario is depicted in Figure 4, where P_0 is the proxy variable and U represents the underlying, unobserved, common cause(s) of participation and outcome. We refer to the proxy variable as a type 0 proxy, because it does not influence anything, and because it corresponds to what is probably the most common way in which epidemiologists use the term "proxy" in the context of confounding—that is, as referring to a misclassified version of a confounder.

As for a concrete example, we again assume that U represents conscientiousness and that this characteristic influences both study participation and a disease outcome, such as cardiovascular disease. Since conscientiousness is difficult to measure, it is unobserved. However, the proxy P_0, which represents a misclassified version of conscientiousness, has via questionnaires been determined in the study sample as well as in another sample that is representative of the population. Hence, standardization on P_0 can be done. By carrying out such a standardization with equation 1, one would expect to obtain approximately generalizable results—or, at least, results that are closer to the true population effect as compared to crude sample estimate. In Web Appendix 4, we show that this is indeed the case, at least in scenarios where U, P_0, and A are all binary and where interest lies in the ATE.

Briefly returning to the concepts of type 1B and type 2B proxies, we note that the conclusions regarding these proxies did not hinge on any assumption that the proxies had to influence participation and the outcome, as the effects indicated by the dashed arrows in Figure 3 did not need to be present. In a scenario without these dashed arrows, P_{1B} and P_{2B} can practically be seen as 2 type 0 proxies, as the setup is the same as in Figure 4, except that there are 2 proxy variables involved, rather than just one. Two imperfect measurements of an underlying cause of participation and outcome can hence be exploited to achieve perfect generalizability.

DISCUSSION

When study participation and outcome have common causes, results may not generalize from study to target population, at least not without adjustments. One way of restoring generalizability is to standardize on the full set of common causes of participation and outcome; however, this approach is often not feasible, as not all common causes are fully observed. Borrowing from the literature on proxy variables in the context of confounding (29–31), we have defined and examined several types of proxy variables that can instead be utilized, and provided a simple classification system.

The fact that proxies of type 1 and type 2 allow for perfect generalizability should come as no surprise to readers familiar with the DAG framework. Indeed, these proxies block the paths between participation and outcome, rendering these conditionally independent. The opportunities to obtain imperfectly generalizable results with a proxy of type 0 may also be unsurprising, but our study is, to our knowledge, the first to establish conditions under which this presumption holds.

The finding that a combination of a type 1B and type 2B proxy can be used to obtain perfectly generalizable results is perhaps the most remarkable one. In contrast to when exploiting proxies of type 1 or 2, participation and outcome are here not independent conditional on the proxies, as U directly influences both participation and the outcome.

The degree of generalizability that can be achieved with a type 0 proxy depends in part on the relationship between U and P_0. A nearly perfect relationship between these two means that close to perfect generalizability can be obtained by standardizing on P_0, whereas a negligible relationship means that standardization on P_0 will hardly make any difference. A formula for the bias that remains after standardization on P_0 in the case of binary U, P_0, and A and where interest lies in the ATE is given by equation A13 in Web Appendix 4. Specifically, in a scenario without fully observed background variables X, the bias is represented as a product of two terms, of which the first measures the effect of U on the ATE, and the second is a weighted average of P_0-conditional relationships between U and S. An investigator

may try out different scenarios for the relationships between U, P_0, S, and the ATE to conduct a sensitivity analysis based on this formula, similarly to how sensitivity analyses can be conducted in scenarios where no proxy variable is available (32–34).

As we have emphasized, although unobserved common causes of sample participation and outcome threaten the generalizability of study results in general, the generalizability of the ATE or CRR is threatened only if the common causes act as effect measure modifiers on a relevant scale. Standard DAGs are nonparametric and can therefore not be used to illustrate whether effect measure modification on a particular scale is present (12). However, variants of DAGs that allow for this have been proposed (25, 35). In Web Appendix 5 (Web Figures 6–9), we replicate the figures in the main text of this article, using such a DAG variant (25).

Issues of generalizability are often neglected in empirical research, perhaps in part due to a limited understanding of why these issues arise and how they can be dealt with. However, just like confounding, lack of generalizability may be corrected for by adjustment for certain background variables, which may be identified with DAGs. For readers familiar with confounding and DAGs, our article should serve as a useful guide to how generalizability can be improved.

ACKNOWLEDGMENTS

Author affiliations: Epidemiology, Population Studies and Infrastructures (EPI@LUND), Department of Laboratory Medicine, Lund University, Lund, Sweden (Anton Nilsson, Jonas Björk); Clinical Studies Sweden, Forum South, Skåne University Hospital, Lund, Sweden (Jonas Björk); Centre for Societal Risk Research, Karlstad University, Karlstad, Sweden (Carl Bonander); and Health Economics and Policy, School of Public Health and Community Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden (Carl Bonander).

This work was funded by the Swedish Research Council for Health, Working Life and Welfare (FORTE; grant 2020-00962).

The article does not make use of any data.

The views expressed in this article are those of the authors and do not reflect those of FORTE.

Conflict of interest: none declared.

REFERENCES

1. Galea S, Tracy M. Participation rates in epidemiologic studies. *Ann Epidemiol.* 2007;9(9):643–653.
2. Keiding N, Louis TA. Perils and potentials of self-selected entry to epidemiological studies and surveys. *J R Stat Soc Ser A Stat Soc.* 2016;179(2):319–376.
3. Bareinboim E, Tian J, Pearl J. Recovering from selection bias in causal and statistical inference. In: Brodley C, Stone P, eds. *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*. Palo Alto, CA: AAAI Press; 2014: 2410–2416.
4. Rudolph KE, Díaz I, Rosenblum M, et al. Estimating population treatment effects from a survey subsample. *Am J Epidemiol.* 2014;180(7):737–748.
5. Pearl J. Generalizing experimental findings. *J Causal Inference.* 2015;3(2):259–266.
6. Stuart EA, Bradshaw CP, Leaf PJ. Assessing the generalizability of randomized trial results to target populations. *Prev Sci.* 2015;16(3):475–485.
7. Lesko CR, Buchanan AL, Westreich D, et al. Generalizing study results: a potential outcomes perspective. *Epidemiology.* 2017;28(4):553–561.
8. Buchanan AL, Hudgens MG, Cole SR, et al. Generalizing evidence from randomized trials using inverse probability of sampling weights. *J R Stat Soc Ser A Stat Soc.* 2018;181(4): 1193–1209.
9. Dahabreh IJ, Roberts SE, Hernán MA. On the relation between g-formula and inverse probability weighting estimators for generalizing trial results. *Epidemiology.* 2019; 30(6):807–812.
10. Westreich D, Edwards JK, Lesko CR, et al. Target validity and the hierarchy of study designs. *Am J Epidemiol.* 2019; 188(2):438–443.
11. Bareinboim E, Pearl J. Causal inference and the data-fusion problem. *PNAS.* 2016;113(27):7345–7352.
12. Hernán MA. Invited commentary: selection bias without colliders. *Am J Epidemiol.* 2017;185(11):1048–1050.
13. Infante-Rivard C, Cusson A. Reflection on modern methods: selection bias—a review of recent developments. *Int J Epidemiol.* 2018;47(5):1714–1722.
14. Pearl J. Causal diagrams for empirical research. *Biometrika.* 1995;82(4):669–710.
15. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology.* 1999;10(1):37–48.
16. Dahabreh IJ, Hernán MA. Extending inferences from a randomized trial to a target population. *Eur J Epidemiol.* 2019;34(8):719–722.
17. Lönnqvist J-E, Paunonen S, Verkasalo M, et al. Personality characteristics of research volunteers. *Eur J Pers.* 2007;21: 1017–1030.
18. Thielsch M, Vasilev E, Back MD. Will they stay or will they go? Personality predictors of dropout in an online study. *Int J Internet Sci.* 2015;10(1):37–48.
19. Hudek-Knežević J, Kardum I. Five-factor personality dimensions and 3 health-related personality constructs as predictors of health. *Croat Med J.* 2009;50(4):394–402.
20. Taylor MD, Whiteman MC, Fowkes GR, et al. Five factor model personality traits and all-cause mortality in the Edinburgh Artery Study cohort. *Psychosom Med.* 2009;71(6): 631–641.
21. Tipton E. Improving generalizations from experiments using propensity score subclassification: assumptions, properties, and contexts. *J Educ Behav Stat.* 2013;38(3):239–266.
22. Kern HL, Stuart EA, Hill J, et al. Assessing methods for generalizing experimental impact estimates to target populations. *J Res Educ Eff.* 2016;9(1):103–127.
23. Egami N, Hartman E. Covariate selection for generalizing experimental results: application to a large-scale development program in Uganda. *J R Stat Soc Ser A.* 2021;184(4): 1524–1548.
24. Huitfeldt A, Swanson SA, Stensrud MJ, et al. Effect heterogeneity and variable selection for standardizing causal effects to a target population. *Eur J Epidemiol.* 2019;34(12): 1119–1129.

25. Nilsson A, Bonander C, Strömberg U, et al. A directed acyclic graph for interactions. *Int J Epidemiol.* 2021;50(2): 613–619.
26. VanderWeele TJ, Knol MJ. A tutorial on interaction. *Epidemiol Methods.* 2014;3:33–72.
27. Li F, Lock Morgan K, Zaslavsky AM. Balancing covariates via propensity score weighting. *J Am Stat Assoc.* 2018;113: 390–400.
28. Hernán MA, Hernández-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology.* 2004;15(5): 615–625.
29. de Luna X, Fowler P, Johansson P. Proxy variables and nonparametric identification of causal effects. *Econ Lett.* 2017;150:152–154.
30. Miao W, Geng Z, Tchetgen Tchetgen ET. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika.* 2018;105(4):987–993.
31. Ogburn EL, VanderWeele TJ. On the nondifferential misclassification of a binary confounder. *Epidemiology.* 2012;23(3):433–439.
32. Nguyen TQ, Ackerman B, Schmid I, et al. Sensitivity analyses for effect modifiers not observed in the target population when generalizing treatment effects from a randomized controlled trial: assumptions, models, effect scales, data scenarios, and implementation details. *PLoS One.* 2018;13(12):e0208795.
33. Chan W. An evaluation of bounding approaches for generalization. *J Exp Educ.* 2019;89(4):690–720.
34. Colnet B, Josse J, Scornet E, et al. Causal effect on a target population: a sensitivity analysis to handle missing covariates. *J Causal Inference*. 2022;10: 372–414.
35. Weinberg CR. Can DAGs clarify effect modification? *Epidemiology.* 2007;18(5):569–572.