Practice of Epidemiology

Transportability of Trial Results Using Inverse Odds of Sampling Weights

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

* Correspondence to Dr. Daniel Westreich, Campus Box 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).

Increasingly, the statistical and epidemiologic literature is focusing beyond issues of internal validity and turning its attention to questions of external validity. Here, we discuss some of the challenges of transporting a causal effect from a randomized trial to a specific target population. We present an inverse odds weighting approach that can easily operationalize transportability. We derive these weights in closed form and illustrate their use with a simple numerical example. We discuss how the conditions required for the identification of internally valid causal effects are translated to apply to the identification of externally valid causal effects. Estimating effects in target populations is an important goal, especially for policy or clinical decisions. Researchers and policy-makers should therefore consider use of statistical techniques such as inverse odds of sampling weights, which under careful assumptions can transport effect estimates from study samples to target populations.

causal inference; epidemiologic methods; external validity; generalizability; transportability

Abbreviation: HIV, human immunodeficiency virus.

Large randomized trials with complete compliance and no missing data provide internal validity in expectation as a matter of design (1). However, *external* validity with respect to a specific, investigator-defined target population is not similarly provided (2–7). Unless the study sample (P_S) was sampled at random from the target population (P_T), there is no expectation of exchangeability of the study sample and the (again, investigator-defined) target population. Yet nearly all trials are conducted among study samples that are not sampled at random from the target population, for reasons of either design (e.g., to maximize statistical power, a trial is conducted among those at highest risk of an outcome) or happenstance (e.g., if persons who exhibit health-seeking behaviors participate in the trial at higher frequencies than others). In these cases, despite having an internally unbiased sample average treatment effect, that sample average treatment effect may differ from the average treatment effect in the target population.

Given that we have internally valid trial results, we often wish to ask the question: What would happen had this trial been conducted in another, external population—the target? Above we suggested that we might want to ask about the causal effect of the treatment in the population *from which the study population was sampled* (albeit perhaps nonrandomly); we also might wish to address the causal effect of the treatment in a target population *distinct from the study sample*, that is, one which is partially or completely nonoverlapping with the study sample. In this latter case—such as when we have a randomized trial and wish to infer a causal effect in a target population—the question can be framed as one of direct standardization to the external target population. As a distinction of language, and to be consistent with the evolving literature on this topic, we refer to the former case (where the study sample is a subset of the target population) as a problem of "generalizability," and to the latter case (where the study sample is *not* a subset of the target population) as a problem of "transportability."

In either case, when externally valid estimates of effect are desired but not guaranteed by design, quantitative approaches are needed. In general, these approaches rely on assumptions which parallel the identification conditions necessary for internally valid causal effect estimation, particularly conditional exchangeability (8) with positivity (9), treatment variation irrelevance (10), no measurement error, and no misspecification of relevant parametric or semiparametric models. The last point is not necessary if nonparametric inference is possible, but in most cases the relevant space of covariates is high-dimensional, and thus a robust approach to quantitative generalizability or transportability requires some degree of modeling.

Am J Epidemiol. 2017;186(8):1010–1014

Previously, Cole and Stuart (4) introduced inverse probability weights for quantitative generalization of trial results, but they did not explain how to operationalize this approach or whether their approach was applicable to problems of both generalizability and transportability. More recently, Bareinboim and Pearl (3) highlighted several key distinctions between generalizability and transportability and introduced a method for deriving a transport formula, which relies on a detailed understanding of the causal relationships among all relevant variables. Here we integrate these 2 methods to introduce an approach to quantitative transportability which may be simpler to implement than the transport formula.

A brief additional note on terminology: Where Cole and Stuart refer to inverse probability of *selection* weights (4), we refer to inverse probability of *sampling* weights. We note, however, that in many (perhaps most) cases the study subjects were probably not formally *sampled*, and we do not wish to imply so with the use of that term; rather, we simply obtain a study sample through some (perhaps unclear) mechanism. For simplicity, we assume that once a study sample has been enumerated, treatment is randomized and follow-up is complete, so that there is no confounding bias in expectation and no additional missing data or selection into the analytical sample (and therefore no selection bias as a problem of internal validity).

METHODS

Preliminary issues and notation

Sampling (S) might relate to covariates (Z) in several ways, including S causing Z (or S indicating differences in distributions of Z), Z causing S, and both S and Z being caused by some additional variable U; here we restrict our attention to the first of these cases, which is closely related to Bareinboim and Pearl's term "transportability" (3). We assume that the epidemiologist has conducted a study and wishes to transport the effect estimate from that study sample to an external target population. For convenience, we assume that information on the same set of covariates has been collected in the study sample and target data, and that the epidemiologist has concatenated the 2 data sets.

In the following, i indicates a participant index $i = 1, 2, \ldots n, n + 1, \ldots N$ such that the study sample comprises n participants and the target population $N - n$ participants; study participants are designated $S_i = 1$, while individuals in the target population are designated $S_i = 0$; and Z_i is a vector of pretreatment covariates for participant i (see the Discussion section for comments on components of Z). Y_i^a indicates the potential outcome under some specific treatment $A = a$ for participant i.

Method

Our goal is estimation of $P(Y_i^a = 1 \mid S_i = 0)$, the risk of the outcome under a particular treatment (a) in the target population. In the Appendix, we use the transport formula to derive a set of weights, which when applied to estimates of observed quantities in the study sample yield an estimate of this estimand. Specifically, we derive the following expression for inverse odds of sampling weights:

$$W_i = \begin{cases} \dfrac{P(S_i = 0 \mid Z_i)}{P(S_i = 1 \mid Z_i)} \times \dfrac{P(S_i = 1)}{P(S_i = 0)}, & S_i = 1, \\ 0, & S_i = 0, \end{cases}$$

where S_i, Z_i, and i are as described above. The weight for individual i is 0 if they did not participate in the study. Otherwise, the first term of the weight is the inverse of the ratio of an individual's probability of being in the study sample as opposed to the target population (hereafter "being sampled"), conditional on Z_i divided by their Z_i-conditional probability of *not* being sampled—that is, the inverse of their Z_i-conditional sampling odds. The second part of the weight is the ratio of the *unconditional* sampling probability to the *unconditional* nonsampling probability—that is, the unconditional sampling odds.

We note that this approach differs from the inverse probability (rather than odds) of selection weights; the latter method, described by Cole and Stuart (4), is appropriate when the study sample is a subset of the target population (i.e., for generalizability rather than transportability). Inverse odds weights are appropriate when the study sample and target population are nonoverlapping; if we consider "being in the study sample" to be a kind of treatment, this method is analogous to weighting for the average treatment effect in the untreated in nonexperimental studies (11, 12).

NUMERICAL EXAMPLE

To aid intuition around this method, consider a hypothetical trial of assignment to a new antiretroviral therapy regimen for human immunodeficiency virus (HIV) compared with assignment to a reference regimen, for the outcome of virological failure at 1 year, conducted in HIV-positive people living in the United States. Suppose the *study sample* for the trial comprises 2,000 participants, 1,000 with single covariate $Z = 1$ and 1,000 with $Z = 0$. Among participants with $Z = 1$, the risk difference is −0.2 (novel treatment is protective against failure); among participants with $Z = 0$, the risk difference is 0.0 (no effect of intervention). The crude sample average causal risk difference is therefore −0.1, a simple average of the 2 strata.

Our *target population* (alternately, a random sample from our target population) comprises 2,000 persons living with HIV in the United States, of whom 80% have $Z = 1$ and 20% have $Z = 0$. In this very simple case, we can hand-calculate the (target) population average causal effect in our external setting as $0.8 \times (-0.20) + 0.2 \times (0.00) = -0.16$. In real data, we would need to use model-based approaches to account for the joint distribution of multiple continuous and categorical variables.

We concatenate the trial data ($n = 2,000$; 50% with $Z = 1$) with the target population ($n = 2,000$; 80% with $Z = 1$), obtaining a combined population of size 4,000, including 2,600 (65%) with $Z = 1$ and 1,400 (35%) with $Z = 0$. We proceed by estimating $O(S = 1 \mid Z) = P(S = 1 \mid Z)/(1 - P(S = 1 \mid Z))$, which in this case is $(1,000/2,600)/(1 - 1,000/2,600) = 1,000/1,600$ where $Z = 1$ and $(1,000/1,400)/(1 - 1,000/1,400) = 400/1,000$ where $Z = 0$. We would use these odds to calculate a weighted pseudopopulation of 1,600 persons $((1,000 \times 1,600/1,000) = 1,600)$ for $Z = 1$ and 400 persons $((1,000 \times 400/1,000) = 400)$

for $Z = 0$; we would then calculate the weighted risk difference as $(1{,}600 \times -0.2 + 400 \times 0.0)/(1{,}600 + 400) = -0.16$.

This estimate coincides with the common-sense simple weighted average we derived immediately above. In addition, these inverse odds weights coincide with the intuitive explanation of how individuals from the study sample *ought* to be weighted so as to represent individuals in the target population, as shown in Figure 1. Notably, increasing or decreasing the size of the target population has no impact on the final estimate, which is not necessarily the case in the method proposed by Cole and Stuart (4).

DISCUSSION

Typical epidemiologic and biostatistical analyses emphasize internal validity of causal effects, but (as others have noted (13, 14)) a causal effect without a specified target population is poorly defined. In practice, study samples for randomized trials are rarely sampled at random directly from the target population; indeed, because consent is an ethical necessity for enrollment in a clinical trial, trial participants are effectively never a random sample of the target population. Yet this is the premise that underlies the assumption of unconditional generalizability or transportability between the study sample and the target population—a frequent (if informal) claim in randomized trials. In contrast, the methods discussed and presented here allow us to relax this questionable premise: We no longer assume that the results from the study population are unconditionally transportable from the study

sample to an arbitrary target population; rather, we assume that they are transportable conditional on variables in our model.
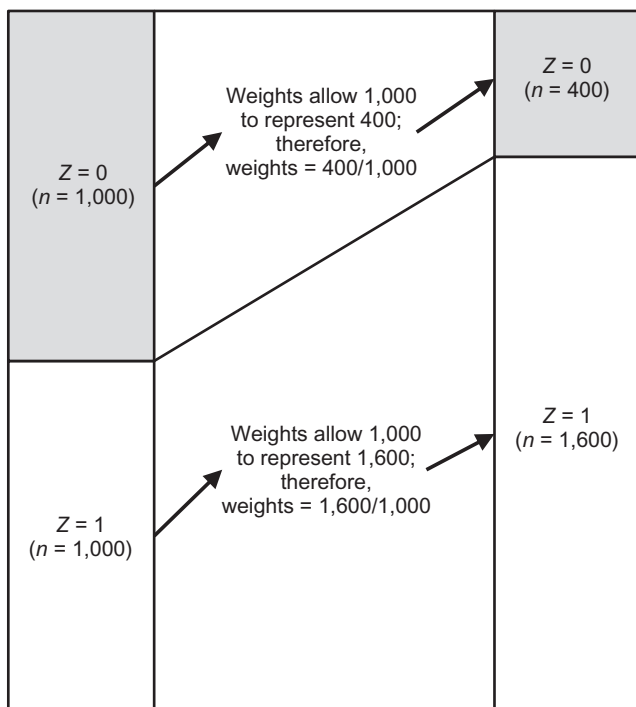
Some people may be uncomfortable with our assumption of conditional transportability, perhaps because herein we are explicit about assumptions that are typically hidden within vague statements about how "representative" the study sample is without addressing the questions 1) representative of what target population? and 2) representative according to which characteristics Z? There is a useful conceptual parallel here with the assumption of exchangeability between treated and untreated subjects for internal validity in an observational setting. The assumption of unconditional transportability is similar (but not identical) to the assumption of unconditional exchangeability (e.g., the causal effect is unconfounded), while the assumption of transportability conditional on variables in the model is similar (but again not identical) to the notion of conditional exchangeability (e.g., the causal effect is unconfounded conditional on a set of confounders).

These parallels are useful in considering the contents of Z. In earlier work, investigators have variously described Z as comprising all effect-measure modifiers (5) or as having components identifiable from causal diagrams (3, 7, 15). The clearest guideline is that Z should be S-admissible (16)—that is, that Z should include pretreatment covariates sufficient to d-separate sampling and the outcome variable (3). This guideline is analogous to that of selecting variables for d-separation of the exposure and outcome variables for internal validity.

As with conditional exchangeability for internal validity, conditional transportability of external validity carries with it additional assumptions: namely, positivity (9) and correct model specification. For transport-positivity to hold, the probability of being included in the sample must be greater than 0 for participants in all strata defined by Z in the target population. This assumption is necessary so that the Z-specific probability of the outcome estimated in the study can "stand in" for the Z-specific probability of the outcome in the target population (see Appendix). Of course, as with positivity for internal validity, transport-positivity may be replaced by making additional assumptions (e.g., smoothing under a parametric model). As noted elsewhere, additional conditions are necessary for transportability—specifically, similar patterns of interference and similar versions of treatment between the study sample and the target population (5).

The concerns about external validity discussed here are highly relevant to observational studies as well as to trials. The results of a trial are frequently, naively assumed to be transportable to a target population. Just as often, however, an epidemiologist assumes that observational cohorts are more representative of the target population and thus that there is less need to evaluate transportability directly (much less to identify the target population explicitly). In fact, the transportability of observational studies to a particular target population of interest is not guaranteed and must be evaluated carefully.

In many clinical trials, external validity is considered only an afterthought; however, consideration of both internal validity in the study sample and external validity in a target population is crucial to providing evidence which will best improve medicine and public health (13, 17). Quantitative approaches to transportability, such as the one described here, are straightforward and should be applied more widely.

Figure 1. Concepts of weights to map from a study sample with oversampled $Z = 0$ (on left) to a target population (on right).

REFERENCES

1. Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials*. 2012;9(1): 48–55.
2. Frangakis C. The calibration of treatment effects from clinical trials to target populations. *Clin Trials*. 2009;6(2):136–140.
3. Bareinboim E, Pearl J. A general algorithm for deciding transportability of experimental results. *J Causal Inference*. 2013;1(1):107–134.
4. 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.
5. Hernán MA, Vanderweele TJ. Compound treatments and transportability of causal inference. *Epidemiology*. 2011;22(3): 368–377.
6. Weisberg HI, Hayden VC, Pontes VP. Selection criteria and generalizability within the counterfactual framework: explaining the paradox of antidepressant-induced suicidality? *Clin Trials*. 2009;6(2):109–118.
7. Bareinboim E, Lee S, Honavar V, et al. Transportability from multiple environments with limited experiments. In: Burges CJ, Bottou L, Welling M, et al., eds. *Advances in Neural Information Processing Systems 26*. (Proceedings of the 26th Annual Conference on Neural Information Processing Systems (NIPS 2013)). La Jolla, CA: Neural Information Processing Systems Foundation; 2013:136–144.
8. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. 2006; 60(7):578–586.
9. Westreich D, Cole SR. Invited commentary: positivity in practice. *Am J Epidemiol* 2010;171(6):674–677.
10. VanderWeele TJ. Concerning the consistency assumption in causal inference. *Epidemiology*. 2009;20(6):880–883.
11. Sato T, Matsuyama Y. Marginal structural models as a tool for standardization. *Epidemiology*. 2003;14(6):680–686.
12. 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.
13. Maldonado G, Greenland S. Estimating causal effects. *Int J Epidemiol*. 2002;31(2):422–429.
14. Hoggatt KJ, Greenland S. Commentary: extending organizational schema for causal effects. *Epidemiology*. 2014; 25(1):98–102.
15. 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.
16. Pearl J, Bareinboim E. External validity and transportability: a formal approach. In: *2011 JSM Proceedings*. Alexandria, VA: Statistical Computing Section, American Statistical Association; 2011:157–171.
17. 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.

(Appendix follows)

APPENDIX

Inverse Odds of Sampling Weights

Weights with which to estimate the expected value of a binary potential outcome in a target population, $P(Y^a = 1 | S = 0)$, using data from a study sample ($S = 1$) and with covariates Z, can be derived from the g-formula.

By the law of total probability,

$$P(Y^a = 1 | S = 0)$$
$$= \sum_z P(Y^a = 1 | S = 0, Z = z) P(Z = z | S = 0).$$

Assuming exchangeability between treatment arms conditional on Z (i.e., the independence of exposure A and the potential outcome Y^a), we can substitute $P(Y^a = 1 | S = 0, Z = z, A = a)$ for $P(Y^a = 1 | S = 0, Z = z)$ in the above expression, such that

$$P(Y^a = 1 | S = 0)$$
$$= \sum_z P(Y^a = 1 | S = 0, Z = z, A = a) P(Z = z | S = 0).$$

Note that we must also assume exposure positivity, or $P(A = a | Z = z) > 0 \; \forall \; z$.

By counterfactual consistency, we can replace the potential outcome Y^a with Y, where $A = a$,

$$P(Y^a = 1 | S = 0)$$
$$= \sum_z P(Y = 1 | S = 0, Z = z, A = a) P(Z = z | S = 0).$$

Assuming exchangeability (see note at end of Appendix) between the study sample $S = 1$ and the target population $S = 0$, conditional on Z (i.e., independence of the outcome and sampling), we allow the conditional outcome distribution in the sample, $P(Y = 1 | S = 1, Z = z, A = a)$, to stand in for the conditional outcome distribution in the target, $P(Y = 1 | S = 0, Z = z, A = a)$, such that

$$P(Y^a = 1 | S = 0)$$
$$= \sum_z P(Y = 1 | S = 1, Z = z, A = a) P(Z = z | S = 0).$$

Note that we must also assume transport positivity, or $P(S = 1 | Z = z) > 0 \; \forall \; z$. The above equation is analogous to the transport formula described by Bareinboim and Pearl (3).

Next, we rewrite the conditional probability of the outcome $P(Y = 1 | S = 1, Z = z, A = a)$ in terms of the joint distribution of Y, Z, and A among sampled individuals, $P(Y = 1, A = a, Z = z | S = 1)$:

$$P(Y^a = 1 | S = 0)$$
$$= \sum_z \frac{P(Y = 1, A = a, Z = z | S = 1)}{P(A = a | Z = z, S = 1) P(Z = z | S = 1)}$$
$$\times P(Z = z | S = 0).$$

Then we rearrange the formula so that

$$P(Y^a = 1 | S = 0) = \sum_z \frac{P(Y = 1, A = a, Z = z | S = 1)}{P(A = a | Z = z, S = 1)}$$
$$\times \frac{P(Z = z | S = 0)}{P(Z = z | S = 1)}.$$

As written, the last term may be difficult to estimate when Z is high-dimensional. To ease implementation, we rearrange the last term using Bayes' theorem:

$$\frac{P(Z = z | S = 0)}{P(Z = z | S = 1)} = \frac{\dfrac{P(S = 0 | Z = z) P(Z = z)}{P(S = 0)}}{\dfrac{P(S = 1 | Z = z) P(Z = z)}{P(S = 1)}}$$
$$= \frac{P(S = 0 | Z = z)}{P(S = 1 | Z = z)} \times \frac{P(S = 1)}{P(S = 0)}.$$

Thus, the final expression reads

$$P(Y^a = 1 | S = 0) = \sum_z \frac{P(Y = 1, A = a, Z = z | S = 1)}{P(A = a | Z = z, S = 1)}$$
$$\times \frac{P(S = 0 | Z = z)}{P(S = 1 | Z = z)} \times \frac{P(S = 1)}{P(S = 0)},$$

where the last 2 terms,

$$\left(\frac{P(S = 0 | Z = z)}{P(S = 1 | Z = z)} \times \frac{P(S = 1)}{P(S = 0)} \right),$$

constitute the stabilized inverse odds of sampling weights.

Note that the 2 exchangeability assumptions above may require different sets of covariates; thus, for convenience, Z can be thought of as the union of the sets of covariates required for both exchangeability assumptions.