COMMENTARY

Extending inferences from a randomized trial to a target population

Issa J. Dahabreh[1,2,3] · Miguel A. Hernán[3,4,5]

In this issue, Weiss discusses "generalizing" inferences from randomized trials to other populations [1]. However, he does not explicitly define what "generalizing" means, assumes that "generalizing" the results of a randomized trial has a single goal, and reduces generalizability to a binary subjective judgment—findings are either generalizable or not generalizable. A growing literature (e.g., [1–13]) precisely defines the several meanings and goals of extending inferences from randomized trials to another population, and describes analyses whose findings go beyond simple binary judgements. Here, we provide a non-technical overview of this literature. First, we briefly review the main concepts, then we outline the available study designs and statistical approaches.

What do we mean by extending inferences from randomized trials?

We can summarize the goals of extending inferences from randomized trials as learning about *counterfactual quantities* in a *target population* under *joint interventions* to scale-up trial engagement and the treatment strategies assessed in the trial [14]. To unpack this description, we consider its three key components one-at-a-time.

✉ Issa J. Dahabreh
  issa_dahabreh@brown.edu

1   Department of Health Services Policy and Practice, Center
    for Evidence Synthesis in Health, School of Public Health,
    Brown University, Box G-121-8, Providence, RI 02912,
    USA

2   Department of Epidemiology, School of Public Health,
    Brown University, Providence, RI, USA

3   Department of Epidemiology, Harvard T.H. Chan School
    of Public Health, Harvard University, Boston, MA, USA

4   Department of Biostatistics, Harvard T.H. Chan School
    of Public Health, Harvard University, Boston, MA, USA

5   Harvard-MIT Division of Health Sciences and Technology,
    Boston, MA, USA

Target population: generalizability versus transportability

Suppose that a randomized trial has compared several treatment strategies. Like Weiss, we would like to use the trial's data to learn about the effects of these treatments in another population: the target population. Treatment and outcome data from non-randomized individuals in the target population cannot be used to reliably estimate treatment effects because of, say, confounding by unmeasured variables or gross measurement error, but reliable baseline covariate data are available.

An explicit specification of the target population allows us to distinguish between two possible goals: generalizability and transportability. To explain this distinction, consider the steps in the selection of the population in a randomized trial. First, the investigators specify eligibility criteria that define the *eligible population*. Second, the actions to invite some eligible individuals to participate in the trial define the *invited population*. Last, the decision of some invited individuals to participate in the trial defines the *participant population*. These three populations are nested: the participant population is a subset of the invited population, which in turn is a subset of the eligible population. The composition of the participant population is largely outside of the investigators' control because the participation decision rests with the invited individuals.

We define *generalizability* as the extension of inferences from the trial to a target population that coincides, or is a subset of, the trial-eligible population [15]. We define *transportability* as the extension of inferences from the trial to a target population that includes individuals who are not part of the trial-eligible population (others [11] have proposed different definitions). In this context, we collectively refer to generalizability and transportability as *extending inferences from trial participants to a target population*. In our terminology, the commentary by Weiss is mostly about transportability, not generalizability, because the target populations he considers are "broader" than the trial-eligible populations.

Joint interventions on trial engagement and treatment

We define *trial engagement* as all actions related to the invitation to participate in the trial or trial participation itself (but not treatment). Trial engagement may have direct effects on the outcome that are not mediated through treatment. These effects include the psychological impact of participating in an experiment and the health effects of non-protocol-mandated care that is delivered in the trial and may not be part of usual practice. For example, in a trial comparing surgical weight-loss interventions, participants may change their dietary habits because they are being observed systematically and may be exposed to regular reminders to improve health behaviors (e.g., to quit smoking) at a higher intensity than in usual practice, even if such reminders are not part of the trial protocol.

In most randomized trials, the effects of trial engagement and the effects of treatment cannot be disentangled [16]. Therefore, when engagement has direct effects on the outcome, we can only learn about the effect of joint interventions to scale-up both trial engagement and treatment to the target population [14].

Counterfactual quantities

Causal questions about the effects of interventions are questions about the distribution of counterfactual (potential) outcomes that would be observed under the interventions. Most of the literature on extending inferences to a target population has focused on identifying the mean of the counterfactual distribution under each intervention and differences of the means under different interventions ("average treatment effects"), typically making an exclusion restriction assumption of no effects of trial engagement on the outcome (e.g., as reviewed in [4] and [17]).

Informally, the counterfactual quantities of interest are identifiable (that is, can be expressed in terms of the observed data) when two conditions hold: some sort of *exchangeability* between the trial-participant population and the target population (or its non-randomized subset) and *positive probability of trial participation* conditional on the variables needed to ensure exchangeability. These conditions cannot be verified empirically so we need to decide how plausible the conditions are based on their subject-matter knowledge. Additional conditions are required when the identification of the counterfactual quantities requires addressing drop-out or censoring, and non-adherence in the trial [14, 18].

How do we extend inferences from randomized trials?

When extending inferences from randomized trials to a target population, the choice of target population, interventions, and counterfactual quantities should guide the study design and statistical analysis.

Study designs for generalizability or transportability can be classified as either *nested trial designs,* if the randomized trial is embedded in a sample from the target population, or *non-nested trial designs,* if the trial is combined with a separately obtained random sample of non-randomized individuals [19, 20]. An example of a non-nested trial design involves constructing a composite dataset by appending a dataset from a completed trial (one that includes data on baseline covariates, treatments, and outcomes) to a dataset with baseline covariates from non-randomized individuals (e.g., trial-eligible individuals identified in routinely collected data). In non-nested trial designs, the sampling probability of non-randomized individuals is unknown and, thus, inferences are only possible about the non-randomized population represented by the sampled non-randomized individuals; inferences about the entire target population are not possible. In contrast, nested trial designs allow inference on the entire target population as well as the non-randomized subset. Different study designs have different implications for the plausibility of the identifiability conditions, our ability to study drivers of trial participation, and the statistical analysis [20].

When the identifiability conditions hold, various statistical methods can be used to learn about the counterfactual quantities of interest. A detailed description of the available statistical methods for generalizability and transportability is beyond the scope of our commentary. Much of the literature on the topic has considered g-formula (outcome model-based) and inverse probability/odds of participation weighting (participation model-based) approaches to estimate population means of counterfactual outcomes or average treatment effects (e.g., [7, 9, 12, 21] report simulation comparisons of different estimators). The correspondence of these approaches to well-known methods for handling confounding by measured variables in observational studies [22] also highlights that trial engagement is naturally viewed as part of a joint intervention. As usual, combining multiple models can help gain robustness without sacrificing efficiency [7, 9, 10, 12].

In practice, however, it is unlikely that the identifiability conditions hold precisely—this is particularly true regarding the exchangeability condition. Thus, judgements about generalizability and transportability can best be viewed as falling on a spectrum ranging from "not possible" (i.e., gross assumption violations are deemed likely) to "likely

to be a reasonable approximation" (i.e., any assumption violations are deemed negligible). These judgements can be informed by formal sensitivity analyses that quantify the impact of assumption violations on study conclusions. Different approaches to sensitivity analysis include the calculation of bounds (rather than point estimates) for the causal quantities, nonparametric just-identified models that produce a range of results over different magnitudes of assumption violations [23], and analyses that incorporate external information in the form of modeling assumptions, additional data, or prior beliefs [13, 24–26]. The preferred approach will vary depending on problem specifics and philosophy of inference, but all approaches can be useful in injecting a measure of inferential modesty to the interpretation of generalizability and transportability analyses.

Generalizability and transportability analyses: another tool in the epidemiologist's toolbox

Weiss suggests that, when inferences from a trial are deemed "non-generalizable" to the target population, we may be able to conduct an observational study in the target population. As a way to operationalize the formal counterfactual approach to causal inference, such observational study would emulate the trial as close as possible [27]. However, the effect estimates of any observational study are susceptible to confounding due to unmeasured variables.

As discussed in this commentary, an alternative approach to conducting an observational study is extending inferences to the target population from available randomized trials. Another alternative is conducting pragmatic trials with broad eligibility criteria in the target population, though these trials may suffer from low adherence and drop-out to such an extent that they are best thought as observational studies with baseline randomization [28]. In many cases, no single approach will be clearly preferred, and we might do better by pursuing multiple strategies in parallel. Because different approaches rely on different identifiability conditions and use different data in service of the same goal (or closely related goals), agreement in their results provides mutual support; large disagreements suggest that some of the approaches are not producing valid answers to the research question (and, with careful study, we may be able to develop hypotheses about what went wrong).

To sum up, provided the research question is clearly defined, we can combine our beliefs about the underlying causal structure with study design and data analysis choices that best allow us to answer it. Thus, as in other causal inference endeavors, in analyses extending inferences from randomized trials to a target population it is best to start by clearly articulating the research question.

References

1. Weiss NS. Generalizing from the results of randomized studies of treatment: Can non-randomized studies be of help? Eur J Epidemiol 2019. https://doi.org/10.1007/s10654-019-00516-3.
2. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. Am J Epidemiol. 2010;172(1):107–15.
3. Stuart EA, Cole SR, Bradshaw CP, Leaf PJ. The use of propensity scores to assess the generalizability of results from randomized trials. J R Stat Soc Ser A. 2011;174(2):369–86.
4. Tipton E. Improving generalizations from experiments using propensity score subclassification: assumptions, properties, and contexts. J Educ Behav Stat. 2013;38(3):239–66.
5. O'Muircheartaigh C, Hedges LV. Generalizing from unrepresentative experiments: a stratified propensity score approach. J R Stat Soc Ser C. 2014;63(2):195–210.
6. Hartman E, Grieve R, Ramsahai R, Sekhon JS. From sample average treatment effect to population average treatment effect on the treated: combining experimental with observational studies to estimate population treatment effects. J R Stat Soc Ser A. 2015;178(3):757–78.
7. Dahabreh IJ, Robertson SE, Tchetgen EJT, Stuart EA, Hernán MA. Generalizing causal inferences from individuals in randomized trials to all trial-eligible individuals. Biometrics. 2018;10:1–12. https://doi.org/10.1111/biom.13009.
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):1193–209.
9. Zhang Z, Nie L, Soon G, Hu Z. New methods for treatment effect calibration, with applications to non-inferiority trials. Biometrics. 2016;72(1):20–9.
10. Rudolph KE, van der Laan MJ. Robust estimation of encouragement design intervention effects transported across sites. J R Stat Soc Ser B. 2017;79(5):1509–25.
11. Westreich D, Edwards JK, Lesko CR, Stuart E, Cole SR. Transportability of trial results using inverse odds of sampling weights. Am J Epidemiol. 2017;186(8):1010–4.
12. Dahabreh IJ, Robertson SE, Stuart EA, Hernán MA. Transporting inferences from a randomized trial to a new target population. arXiv preprint arXiv:1805.00550. 2018.
13. Chan W. Partially identified treatment effects for generalizability. J. Res. Educ. Effect. 2017;10(3):646–69.
14. Dahabreh IJ, Robins JM, Haneuse SJ, Hernán MA. Generalizing causal inferences from randomized trials: counterfactual and graphical identification. 2019 (forthcoming).
15. Hernán MA. Discussion of "Perils and potentials of self-selected entry to epidemiological studies and surveys" by N Keiding and TA Louis. J R Stat Soc Ser A Stat Soc. 2016;179(2):346–7.
16. Heckman JJ. Randomization and social policy evaluation. Cambridge: National Bureau of Economic Research; 1991.
17. Lesko CR, Buchanan AL, Westreich D, Edwards JK, Hudgens MG, Cole SR. Generalizing study results: a potential outcomes perspective. Epidemiol. 2017;28(4):553–61.

18. Lu H, Cole SR, Hall HI, et al. Generalizing the per-protocol treatment effect: the case of ACTG A5095. Clin Trials. 2019;16(1):52–62. https://doi.org/10.1177/1740774518806311.

19. Dahabreh IJ, Hernán MA, Robertson SE, Buchanan A, Steingrimsson JA. Generalizing trial findings in nested trial designs with sub-sampling of non-randomized individuals. arXiv preprint arXiv:1902.06080. 2019.

20. Dahabreh IJ, Haneuse SJPA, Robins JM, Robertson SE, Buchanan AL, Stuart EA, et al. Study designs for extending causal inferences from a randomized trial to a target population. 2019. arXiv preprint arXiv:1905.07764.

21. Kern HL, Stuart EA, Hill J, Green DP. Assessing methods for generalizing experimental impact estimates to target populations. J Res Educ Effect. 2016;9(1):103–27.

22. Hernán MA, Robins JM. Causal inference. Boca Raton: Chapman & Hall/CRC; 2019, forthcoming.

23. Robins JM, Rotnitzky A, Scharfstein DO. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. Statistical models in epidemiology, the environment, and clinical trials: Springer; 2000. p. 1–94.

24. Nguyen TQ, Ebnesajjad C, Cole SR, Stuart EA. Sensitivity analysis for an unobserved moderator in RCT-to-target-population generalization of treatment effects. Ann Appl Stat. 2017;11(1):225–47.

25. Nguyen TQ, Ackerman B, Schmid I, Cole SR, Stuart EA. 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. https://doi.org/10.1371/journal.pone.0208795.

26. Dahabreh IJ, Robins JM, Haneuse SJ, et al. Sensitivity analysis using bias functions for studies extending inferences from a randomized trial to a target population. arXiv preprint arXiv:1905.10684. 2019.

27. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol. 2016;183(8):758–64. https://doi.org/10.1093/aje/kwv254.

28. Toh S, Hernán MA. Causal inference from longitudinal studies with baseline randomization. Int J Biostat. 2008. https://doi.org/10.2202/1557-4679.1117.