THE UNIVERSITY OF RHODE ISLAND

University of Rhode Island
DigitalCommons@URI

Pharmacy Practice and Clinical Research Faculty Publications

Pharmacy Practice and Clinical Research

2017

Practical considerations when generalizing study results: a potential outcomes perspective

Catherine R. Lesko

L. Buchanan
University of Rhode Island, buchanan@uri.edu

Daniel Westreich

Jessie K. Edwards

Michael G. Hudgens

See next page for additional authors

Follow this and additional works at: https://digitalcommons.uri.edu/php_facpubs

Citation/Publisher Attribution

Lesko, C. R., Buchanan, A. L., Westreich, D., Edwards, J. K., Hudgens, M. G., & Cole, S. R. (2017). Practical considerations when generalizing study results: a potential outcomes perspective. *Epidemiology, 28*(4), 553-561. doi: 10.1097/EDE.0000000000000664
Available at: http://dx.doi.org/10.1097/EDE.0000000000000664

This Article is brought to you by the University of Rhode Island. It has been accepted for inclusion in Pharmacy Practice and Clinical Research Faculty Publications by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu. For permission to reuse copyrighted content, contact the author directly.

Practical considerations when generalizing study results: a potential outcomes perspective

Authors

Catherine R. Lesko, L. Buchanan, Daniel Westreich, Jessie K. Edwards, Michael G. Hudgens, and Stephen R. Cole

The University of Rhode Island Faculty have made this article openly available. Please let us know how Open Access to this research benefits you.

This is a pre-publication author manuscript of the final, published article.

Terms of Use

This article is made available under the terms and conditions applicable towards Open Access Policy Articles, as set forth in our Terms of Use.

This article is available at DigitalCommons@URI: https://digitalcommons.uri.edu/php_facpubs/92

Practical considerations when generalizing study results: a potential outcomes perspective

Catherine R. Lesko[1,2]

Ashley L. Buchanan[3,4]

Daniel Westreich[1]

Jessie K. Edwards[1]

Michael G. Hudgens[3]

Stephen R. Cole[1]

[1]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC

[2]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

[3]Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC

[4]Department of Epidemiology, TH Chan School of Public Health, Harvard University, Boston, MA

Keywords: Causality; Effect modifier, epidemiologic; Selection bias; Sampling studies

Abstract

Great care is generally taken in epidemiologic studies to ensure the internal validity of causal effect estimates; however, the *external* validity of effect estimates, has received considerably less attention. The causal effect in a given target population is the average of heterogeneous subgroup effects, weighted according to the prevalence of the subgroups in the target population. When the study sample is not a random sample of the target population, the sample average treatment effect, even if internally valid, cannot be expected to equal the average treatment effect in the target population. There are several categories of choices for the target population. The study sample may be a census of the target population; the population from which the study sample is a random sample or from which the study sample is not a random sample; or some other population of which, the study sample is not a subset of the target population. The identification conditions sufficient for external validity closely parallel the identification conditions for internal validity, namely: conditional exchangeability; positivity; similar distributions of the versions of; similar patterns of interference; no measurement error; and correct model specification. The value of an effect estimate for planning purposes and decision making will depend on the degree of departure from both internal *and* external validity. If the study sample is not a random sample of the target population, direct standardization (the g-formula or transport formula) or inverse probability weighting can be used to estimate a causal effect in the target population.

Epidemiology as a discipline is distinguished by its efforts to identify causes of disease for the purpose of intervening to improve public health. Great care is generally taken in epidemiologic studies to ensure the internal validity of causal effect estimates,[1] including the application of methods to minimize the potential for bias due to measurement error, confounding, selection (specifically, due to missing data, including censoring and truncation), and model misspecification. However, the *external* validity of effect estimates, has received considerably less attention. For the purposes of this discussion, we use the term external validity to refer to the potential for an internally valid treatment (or exposure or intervention) effect measured in a study sample to differ from the treatment effect that would have been estimated in the population of interest[2] (henceforth, the target population). External validity encompasses generalizability and transportability, which we distinguish below. We advance the discussion of external validity herein using a potential outcomes framework. We enumerate a set of identification assumptions sufficient to estimate an externally valid effect, and note the parallel between these and the identification assumptions sufficient to estimate an internally valid effect. Finally, we illustrate some issues regarding generalizability with a simple example and discuss practical considerations for addressing generalizability in epidemiological study design.

Definitions and causal framework

A well-defined causal question states the outcome(s) of interest, denoted by Y; the treatments of interest, denoted by A; and the target population, of size N

described by a set of characteristics denoted by W. Here, we assume variables in W are discrete, however, all concepts are easily extended to the case of continuous W. The causal effect of interest is a contrast (i.e., a difference or ratio) of the distribution of potential outcomes, $Y(a)$, in the target population under two different interventions, treatments, or policies of interest, for example:

$$E_T\{E[Y(a)|\mathrm{W}]P_T(\mathrm{W})\} - E_T\{E[Y(a')|\mathrm{W}]P_T(\mathrm{W})\} \qquad (1)$$

where $Y(a)$ denotes the outcome that a participant would have if he or she received treatment a, and the subscript T denotes that the set of characteristics W takes the distribution seen in target population. This notation makes it clear that the overall causal effect in the target population is the average of effects that are heterogeneous according to W, with weights defined by the distribution of W in the target population, $P_T(\mathrm{W})$. $P_T(\mathrm{W})$ is often omitted when the causal effect of interest is written (i.e. $\mathrm{E}[Y(a)] - \mathrm{E}[\mathrm{Y}(a')]$) under the implicit assumption that either 1) the study sample is a census of the target population; 2) the study sample is a simple random sample from the target population, i.e., the distribution of W in the study sample, denoted $P_S(W)$, is equal to $P_T(\mathrm{W})$ in expectation; 3) there is a distinct target population, M, of which the study sample is not a proper subset, for which $P_M(\mathrm{W})$ is equal to $P_S(\mathrm{W})$; or 4) the causal effect is homogeneous across all W for the causal contrast of interest.

It is helpful to distinguish threats to validity that arise after enumeration of the study sample, which we define as threats to *internal* validity, from threats to validity due to eligibility and enrollment of study subjects, which we define as threats to *external* validity. We define an estimator as *internally* valid when the

estimator of association in the study sample is a consistent estimator of the treatment effect in the source population from which the study participants are randomly sampled (the sample average treatment effect). The distribution of W in the study sample, $P_S(W)$, may differ from $P_T(W)$ without threatening internal validity. We define a causal estimator to be *externally* valid when the sample average treatment effect is a consistent estimator of the average treatment effect in the target population (population average treatment effect). An estimate will be externally valid under one of the first three scenarios described above, all of which result in the distribution of W in the study sample being equal to the distribution of W in the target population.

Assumptions

Given we cannot observe all potential outcomes for subjects in our target population or our study sample,[3,4] we can rely on a sufficient set of identification assumptions under which (with a consistent estimator) an estimate of association could be interpreted causally.

These assumptions are well described in the literature for estimating an internally valid causal estimate. They include: 1) the unexposed are a good substitute for the experience of the exposed in the absence of exposure and vice versa (exchangeability), perhaps conditional on a set of covariates, Z (conditional exchangeability); 2) a non-zero probability of exposure within every stratum defined by Z (positivity); 3) treatment variation irrelevance or no versions of treatment (sometimes referred to as consistency);[5-8]4) no interference or partial

interference (or some other restriction on the interference structure);[9,10] 5) no measurement error; and 6) if Z is high dimensional and non-parametric inference is unfeasible, correct specification of parametric models.

The identification conditions sufficient for external validity closely parallel the identification conditions for internal validity.[18,27] First, we assume that the participants included in the sample are exchangeable with members of the target population who were not sampled, perhaps conditional on W (conditional exchangeability):

$$S \perp [Y(a), Y(a')]|\mathrm{W}$$

where S is an indicator of enrollment into the study sample. Enrollment into the sample is both under the control of the researcher (in designing a recruitment strategy) and under the control of the participants (in deciding whether to provide consent to participate). For the purposes of identification of the causal contrast in the target population, the set of characteristics, W, is sufficient if it includes all causes of (or proxies for causes of) sampling and the outcome. If the researcher is willing to specify a single causal contrast of interest, then the set of characteristics W may be restricted to a subset of W that are effect measure modifiers of that contrast.[2] Second, we assume that, within strata of W, all subjects in the target population have some probability of being selected into the sample ($S = 1$) (positivity):

$$0 < \Pr(\mathrm{S} = 1|\mathrm{W}) < 1, \text{for all W, such that } P_T(\mathrm{W}) > 0$$

Third, we assume similar distribution of the versions of treatment in the study sample and the target population (of which treatment variation irrelevance is a

special case). This may be a strong assumption when the delivery mechanism for treatment differs dramatically between the study sample and the target population (e.g., treatment given to trial participants may have been accompanied by more adherence education and supportive services, as well as Hawthorne effects due to trial participation).[11] Fourth, we assume similar patterns of interference in the target population and the study sample (of which no interference is a special case).[11-13] Finally, we assume no measurement error for exposure, outcome, and all variables W, and correct model specification (if we use a parametric model).

Defining the target population

Failure to specify the target population explicitly precludes comparisons between the target population and the study sample of patient characteristics (exchangeability), details of the implementation of the intervention (treatment versions), or the comparison of patterns of interference between the target population and the study sample. In failing to make these aspects explicit in study design and planning, the generalizability of study results to some unspecified target population is nebulous.[14-18]

There are several categories of choices for the target population.[19] First, the study sample may be a census of the target population. This is almost never the case, because in nearly all instances, we have done research to inform decisions about a population at least somewhat different than that under study. A second choice for the target population may be the population from which the study sample was sampled. When the sampling is random, $P_S(\mathrm{W})$ is the same as $P_T(\mathrm{W})$ in expectation

and sample average treatment effect equals the population average treatment effect in expectation (Figure 1a). When the study sample is not a random sample from the target population and $P_S(\mathrm{W})$ differs from $P_T(\mathrm{W})$, the sample average treatment effect can be expected to differ from the population average treatment effect (Figure 1b). A third choice for the target population is some other population that does not include the study sample (Figure 1c). In both of the latter cases (study sample not a random sample of the target; external population), additional information is needed to estimate the effect of interest in the target population, namely the distribution of $P_T(\mathrm{W})$.

A distinction between *generalizing* results to a target population that includes as members those persons included in the study sample (Figures 1a and 1b) and *transporting* results to a target population that is non-overlapping with the study sample (Figure 1c) has been made previously, but not discussed in depth.[20] In the former case, a physical probability of sampling can be envisioned;[21] in the latter, the probability of sampling is not physical.[21,22] This should not alarm us; transporting results to a target population that is non-overlapping with the study sample is simply direct standardization. Graphical criteria can assist in determining whether an estimate of effect is directly transportable and if not, can help identify the appropriate "transport formula" for estimating an effect for the target population.[23,24] Despite these distinctions between target populations are overlapping or non-overlapping with the study sample, the same set of generalizability assumptions described above holds in both scenarios.

Example

To demonstrate how the sample average treatment differs from the population average treatment effect when the study sample is not a random sample of the target population, consider the following example. Imagine the target population comprises N = 50,000 individuals, in whom the prevalence of two dichotomous causes of the outcome, W_1 and W_2, is 0.15 and 0.20, respectively. In the study sample (n=1,000), as in many trials, participants at greater risk of the outcome ($W_1 = 1$ or $W_2 = 1$) were oversampled, with $P(W_1) = 0.5$ and $P(W_2) = 0.5$. In both the study sample and in in the target population, the exposure, A, is randomly assigned with probability 0.5. The 1-year risk for the outcome is defined by the function: $P(Y = 1) = 0.1073 - 0.05A + 0.20W_1 + 0.20W_2 - 0.15AW_1W_2$. The data from one realization of a target population and study sample generated under these conditions appears in Table 1.

Given this realization of the data, the estimate of the risk difference due to A in the target population is -5.3%. The estimate of the risk difference due to A in the study sample is -9.7%.

The g-formula or inverse probability (IP) of sampling weights can be used to use the study sample data to estimate the population average treatment effect. The g-formula (equivalent to Bareinboim & Pearl's "transport formula")[23] and equation (1) above is:

$$P[Y(a)] = \sum_{W} P[Y(a)|W = w]P_T(W = w)$$

Which, if we assume consistency and conditional exchangeability given W, can be estimated by:

$$P[Y = 1|A = a] = \sum_{W} P[Y = 1|A = a, W = w]P_T(W = w)$$

where the quantity $P[Y = 1|A = a, W = w]$ is estimated in the study sample and $P_T(W = w)$ is the probability that $W = w$ in the target population. The estimate of effect in the target population using the non-parametric g-formula was -5.4%.

Scaled IP of sampling weights were defined:

$$\frac{P(S = 1)}{P(S = 1|W_1 = w_1, W_2 = w_2)}$$

If the set of covariates, W, that should be included in calculating IP of sampling weights is high dimensional, the denominator of the weights can be modeled parametrically. In this example, we used a fully saturated model for the denominator of the weights. The IP of sampling weighted estimate of effect in the target population based on data from the study sample was -5.4%.

Note that the results from the IP of sampling approach and the g-formula/transport formula will be equivalent when both are estimated non-parametrically. In practice, if the dimension of W is large and $P(S = 1)$ or $P(Y = 1|A = a, W = w)$ is modeled, the two approaches may give different results due to different modelling assumptions or model misspecification. Doubly robust estimation of the population average treatment effect is also possible.[25]

Practical considerations for study design

The only way to ensure an estimate is directly generalizable (in expectation) to a particular target population would be draw the study sample as a random sample from that target population as described above.[12] However, beyond the logistical, financial, and ethical challenges to conducting such a study, in certain circumstances, a study sample that is representative of the target population may be undesirable.[18,26,27] When first exploring the existence of a causal effect, non-random sample selection may be purposefully undertaken. For example, an investigator might enroll a sample for a trial that has a higher than average risk of disease to increase statistical efficiency, or restrict enrollment into an observational study to control for an important confounder. Oversampling to avoid sparse numbers of patients within subgroups improves precision during confounder control and also allows estimation of subgroup effects,[18,26] although trials are rarely powered to estimate such subgroup effects.

Epidemiologists are typically primarily concerned with the internal validity of effect estimates. However, the value of an effect estimate for planning purposes and decision making will depend on the degree of departure from both internal *and* external validity. External validity will be threatened to the degree that 1) the prevalence of other causes of the outcome (also versions of treatment, patterns of interference) differs in the study sample and the target population, and 2) the exposure or intervention causal contrast is modified by those other causes of the outcome that differ in the study sample and the target population.[2] For example, Greenhouse et al., describe a trial of antidepressants in adolescents that pointed to an increased risk of suicide, but which excluded participants with the most severe

depression who would have experienced the greatest benefits from the therapy.[15] In this case, while the trial effects were internally valid, the lack of external validity had serious implications for policy and removed potentially beneficial treatment options from depressed adolescent patients. A second example is, the discrepancies between the conclusions about the effects of combined estrogen/progestin menopausal hormone replacement therapy (HRT) on coronary heart disease based on the Women's Health Initiative trial and the Nurses' Health Study can be recast as a generalizability problem if we consider that the age- and time-on-exposure stratum specific effects of HRT estimated in both studies were similar,[28] but the distribution of women by age and time-on-exposure in the target population (young women with no prior exposure) did not match the distribution of women in the study sample from the Nurses' Health Study (older women with lots of prior exposure). In this instance, while internal validity (confounding by some unmeasured factor) was initially blamed for the discrepancy, and while generalizability is typically only thought of as an issue for clinical trials, the generalizability of this observational study undermined policy recommendations based on its results. Such examples highlight the importance of balancing study design decisions to maximize both internal and external validity.

If our study sample is not a random sample of the target population, we can estimate causal effects in a specified target population using direct standardization or inverse probability weighting (a semiparametric extension of direct standardization)[29] if all predictors of both selection into the study sample and the outcome are measured in both the study sample and the target population. This

methodological solution would allow us to use a single study sample to calculate generalized estimates of effect for multiple different specified target populations[2,12,13] as long as the distribution of W was available for the target population.[15,30]

Discussion

Commentaries on the lack of generalizability of randomized trials typically implore the reader to evaluate a lengthy check list of potential determinants of external validity,[16,17,31] or further classify the list into categories to distinguish "external validity" from "applicability."[32] We argue that such exercises could be more efficient if considered quantitatively, within the potential outcomes framework. This approach would quickly narrow the scope of future research needed to ascertain the effect estimate of interest because specific threats to the external validity of the estimate could be identified and used to guide future study design. Furthermore, understanding the mechanism by which differences between the sample and the target population influence the generalizability of a sample estimate would help identify the most appropriate methods to account for those differences. Understanding differences in the distribution of risk factors for the outcome has implications for the selection of the study sample for future research.

Arguments about the generalizability of study results are not well-formed until the relevant characteristics of the target population are explicitly stated (i.e., patient characteristics are designated or interference patterns are specified or the types of therapy defined). This is analogous to the estimation of controlled direct

effects; such direct effects are not well-defined until the researcher specifies that the estimator is the direct effect not through M, where M is some possible mediator.[33]

Finally, distinguishing internal and external threats to validity is useful for determining which parameters in the study sample or target population are estimable. When collider stratification bias due to selection is present in a study, it may threaten causal inference being made for any population,[34] even the study sample, and preclude attempts to generalize results to either the source population or any specified target population (associational estimate is biased for the causal effect in the study sample; generalizability is irrelevant). In contrast, if a study can be determined to be free of selection bias (and confounding) (associational estimate is an unbiased estimate of the causal effect in the study sample) and differences in average treatment effects attributed to nonrandom sampling of the study population, methods exist to generalize results to a the target population (contingent on all assumptions outlined above).[2,35] Standardizing effect estimates to the appropriate target population will improve their utility to clinicians and public health practitioners, and better inform implementation of interventions in target populations.

Funding

Dr. Westreich was supported in part by NIH grant DP2 HD084070. Dr. Cole and Dr. Lesko were supported in part by NIH grant R01 AI100654. Dr. Cole, Dr. Westreich and Dr. Hudgens were supported in part by NIH grant P30 AI50410. Dr. Buchanan was supported in part by NIH grants U01 AI103390-02, and R24 AI067039. The

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

Acknowledgements

References

1. Hernan MA, Robins JM. Estimating causal effects from epidemiological data. *Journal of epidemiology and community health.* Jul 2006;60(7):578-586.
2. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: The ACTG 320 trial. *Am J Epidemiol.* Jul 1 2010;172(1):107-115.
3. Holland PW. Statistics and Causal Inference. *Journal of the American Statistical Association.* Dec 1986;81(396):945-960.
4. Edwards JK, Cole SR, Westreich D. All your data are always missing: incorporating bias due to measurement error into the potential outcomes framework. *Int J Epidemiol.* Apr 28 2015.
5. Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? *Epidemiology.* Jan 2009;20(1):3-5.
6. Pearl J. On the consistency rule in causal inference: axiom, definition, assumption, or theorem? *Epidemiology.* Nov 2010;21(6):872-875.
7. VanderWeele TJ. Concerning the consistency assumption in causal inference. *Epidemiology.* Nov 2009;20(6):880-883.
8. Hernan MA, Taubman SL. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *International journal of obesity.* Aug 2008;32 Suppl 3:S8-14.
9. Hudgens MG, Halloran ME. Toward Causal Inference With Interference. *J Am Stat Assoc.* Jun 2008;103(482):832-842.
10. Tchetgen EJ, VanderWeele TJ. On causal inference in the presence of interference. *Stat Methods Med Res.* Feb 2012;21(1):55-75.
11. Hernan MA, VanderWeele TJ. Compound treatments and transportability of causal inference. *Epidemiology.* May 2011;22(3):368-377.
12. Stuart EA, Cole SR, Bradshaw CP, Leaf P. *The use of propensity scores to assess the generalizability of results from randomized trials.* Johns Hopkins Department of Biostatistics, Working Papers;2010.
13. Tipton E. Improving Generalizations From Experiments Using Propensity Score Subclassification: Assumptions, Properties, and Contexts. *Journal of Educational and Behavioral Statistics.* 2013;38(3):239-266.
14. Gandhi M, Ameli N, Bacchetti P, et al. Eligibility criteria for HIV clinical trials and generalizability of results: the gap between published reports and study protocols. *AIDS.* Nov 4 2005;19(16):1885-1896.
15. Greenhouse JB, Kaizar EE, Kelleher K, Seltman H, Gardner W. Generalizing from clinical trial data: a case study. The risk of suicidality among pediatric antidepressant users. *Stat Med.* May 20 2008;27(11):1801-1813.
16. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet.* Jan 1-7 2005;365(9453):82-93.
17. Rothwell PM. Commentary: External validity of results of randomized trials: disentangling a complex concept. *Int J Epidemiol.* Feb 2010;39(1):94-96.
18. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. *Int J Epidemiol.* Aug 2013;42(4):1012-1014.

19. Maldonado G, Greenland S. Estimating causal effects. *International journal of epidemiology.* 2002;31(2):422-429.
20. Hoggatt KJ, Greenland S. Commentary: extending organizational schema for causal effects. *Epidemiology.* Jan 2014;25(1):98-102.
21. Good IJ. Kinds of Probability: Although there are at least five kinds of probability, we can get along with just one kind. *Science.* Feb 20 1959;129(3347):443-447.
22. Lindley DV. *Understanding uncertainty.* Revised edition. ed. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2014.
23. Bareinboim E, Pearl J. A general algorithm for deciding transportability of experimental results. *J Causal Inference.* 2013;1(1):107-134.
24. Petersen ML. Compound treatments, transportability, and the structural causal model: the power and simplicity of causal graphs. *Epidemiology.* May 2011;22(3):378-381.
25. Rudolph KE, Diaz I, Rosenblum M, Stuart EA. Estimating population treatment effects from a survey sub-sample. *American Journal of Epidemiology.* 2014.
26. Rothman K, Hatch E, Gallacher J. Representativeness is not helpful in studying heterogeneity of effects across subgroups. *Int J Epidemiol.* Apr 2014;43(2):633-634.
27. Rothman KJ, Gallacher JE, Hatch EE. Rebuttal: When it comes to scientific inference, sometimes a cigar is just a cigar. *Int J Epidemiol.* Aug 2013;42(4):1026-1028.
28. Hernan MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology.* Nov 2008;19(6):766-779.
29. Sato T, Matsuyama Y. Marginal structural models as a tool for standardization. *Epidemiology.* Nov 2003;14(6):680-686.
30. Weisberg HI, Hayden VC, Pontes VP. Selection criteria and generalizability within the counterfactual framework: explaining the paradox of antidepressant-induced suicidality? *Clinical trials.* Apr 2009;6(2):109-118.
31. Szklo M. Population-based cohort studies. *Epidemiologic reviews.* 1998;20(1):81-90.
32. Dekkers OM, von Elm E, Algra A, Romijn JA, Vandenbroucke JP. How to assess the external validity of therapeutic trials: a conceptual approach. *Int J Epidemiol.* Feb 2010;39(1):89-94.
33. Cole SR, Hernan MA. Fallibility in estimating direct effects. *Int J Epidemiol.* Feb 2002;31(1):163-165.
34. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology.* Sep 2004;15(5):615-625.
35. Stuart EA, Cole SR, Bradshaw CP, Leaf PJ. The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society. Series A.* Apr 1 2011;174(2):369-386.

Table 1. Data from a hypothetical target population (N=50,000) and nonrandom study sample (n=1,000)

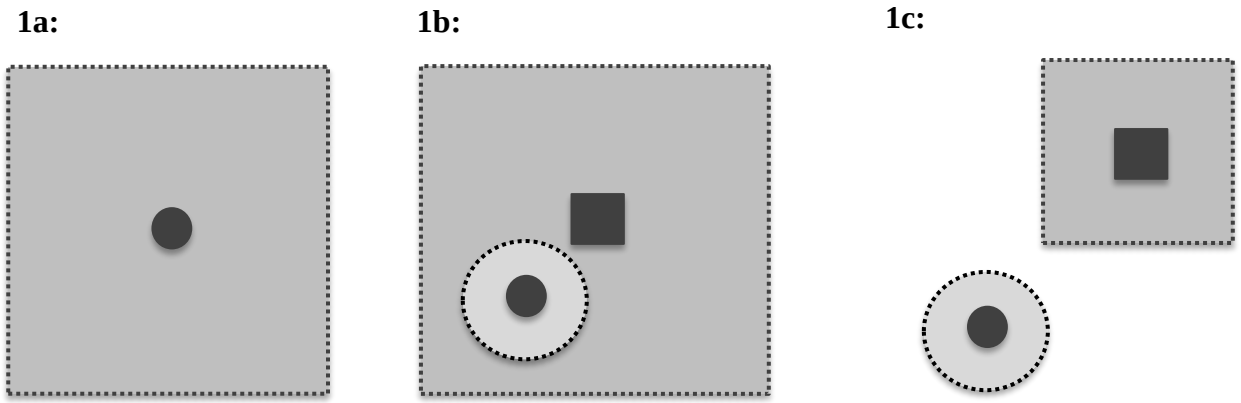
Z_1	Z_1	A	Y	N (target)	n (sample)	Scaled IP sampling weight[a]	Weighted sample
0	0	0	0	15,023	68	4.239	288.2
0	0	0	1	1807	11	4.239	46.6
0	0	1	0	16,046	75	4.239	317.9
0	0	1	1	1,035	6	4.239	25.4
0	1	0	0	2,954	85	0.712	60.5
0	1	0	1	1,285	35	0.712	24.9
0	1	1	0	3,221	91	0.712	64.7
0	1	1	1	1,078	29	0.712	20.6
1	0	0	0	2,087	88	0.506	44.5
1	0	0	1	914	28	0.506	14.2
1	0	1	0	2,285	88	0.506	44.5
1	0	1	1	787	36	0.506	18.2
1	1	0	0	360	84	0.082	6.9
1	1	0	1	378	97	0.082	8.0
1	1	1	0	517	125	0.082	10.3
1	1	1	1	223	54	0.082	4.4

[a] Scaled inverse probability of sampling weight

Table 2. Summary data from a hypothetical target population (N=50,000) and crude and inverse probability of sampling weighted nonrandom study sample (n=1,000)

Target population				Study sample				Weighted sample			
	Y=0	Y=1	Risk		Y=0	Y=1	Risk		Y=0	Y=1	Risk
X=0	20,424	4,384	0.177	X=0	325	171	0.345	X=0	400	94	0.190
X=1	22,069	3,123	0.124	X=1	379	125	0.248	X=1	437	69	0.136
RD:	-0.053			RD:	-0.097			RD:	-0.054		
RR:	0.702			RR:	0.719			RR:	0.716		

Figure 1. Possible choices for the target population of interest



The target population (large, light gray square) can relate to the study sample (small, dark gray circle) in multiple ways. (a) The study sample is a random sample from the target population. (b) The study sample is not a random sample from the target population. (c) The target population is external to the study sample. The target population can be described by either a census of the target population or a random sample of the target population (small, dark gray square). In situation (b) or (c), the study sample can be thought of as a random sample from a larger super-population (larger, light gray circle) that differs from the target population; if the analysis does not standardize or transport results to the target population, inference is restricted to this super-population.