Title
Generalizability of Subgroup Effects.

Journal
Epidemiology, 32(3)

Authors
Seamans, Marissa
Hong, Hwanhee
Ackerman, Benjamin
et al.

Publication Date
2021-05-01

DOI
10.1097/EDE.0000000000001329

Peer reviewed

HHS Public Access

Author manuscript

Epidemiology. Author manuscript; available in PMC 2022 May 01.

Published in final edited form as:

Epidemiology. 2021 May 01; 32(3): 389–392. doi:10.1097/EDE.0000000000001329.

Generalizability of subgroup effects

Marissa J. Seamans[a], **Hwanhee Hong**[b], **Benjamin Ackerman**[c], **Ian Schmid**[d], **Elizabeth A. Stuart**[c,d,e]

[a]Department of Epidemiology, Fielding School of Public Health, University of California, Los Angeles

[b]Department of Biostatistics and Bioinformatics, Duke University, Durham

[c]Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore

[d]Department of Mental Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore

[e]Department of Health Policy and Management, Bloomberg School of Public Health, Johns Hopkins University, Baltimore

Abstract

Generalizability methods are increasingly used to make inferences about the effect of interventions in target populations using a study sample. Most existing methods to generalize effects from sample to population rely on the assumption that subgroup-specific effects generalize directly. However, researchers may be concerned that in fact subgroup-specific effects differ between sample and population. In this brief report, we explore the generalizability of subgroup effects. First, we derive the bias in the sample average treatment effect estimator as an estimate of the population average treatment effect when subgroup effects in the sample do not directly generalize. Next, we present a Monte Carlo simulation to explore bias due to unmeasured heterogeneity of subgroup effects across sample and population. Finally, we examine the potential for bias in an illustrative data example. Understanding the generalizability of subgroup effects may lead to increased use of these methods for making externally valid inferences of treatment effects using a study sample.

Keywords

generalizability; treatment effect heterogeneity; subgroup; bias; causality; effect modifier

Corresponding author: Marissa J. Seamans, Department of Epidemiology, UCLA Fielding School of Public Health, 650 Charles E. Young Dr. S., Los Angeles, CA 90095. mseamans@ph.ucla.edu.

Conflict of interest: The authors have no conflicts of interest to report.

Code for the simulation presented in this article is available at https://github.com/MarissaSeamans/GeneralizabilitySubgroupEffects

Introduction

Generalizability methods are increasingly used to make population inferences of treatment effects using a study sample.[1–3] These methods account for differences in distributions of treatment effect modifiers (e.g., sex) in the study sample and target population using weighting or outcoming modeling approaches[4] under the assumption that effects within subgroups defined by those treatment effect modifiers (e.g., men) transport from the study sample to the target population. However, researchers may be concerned that the subgroup effects themselves may differ between sample and population. For example, men in the study sample may differ from the men in the target population by another covariate (e.g., smoking status) that modifies the treatment effect among men, and may or may not be observed. Concerns about the potential for subgroup effects to vary between the sample and population may arise, but to our knowledge there has been no formal investigation of how large these differences must be to cause substantial bias in population average treatment effect estimates. In this brief report, we provide the formula for bias and use simulations and a data example to illustrate the impact of unmeasured treatment effect heterogeneity within subgroups on inferences about population average treatment effect (ATE) estimates using a study sample.

Cole and Stuart[5] derived the bias in the sample ATE estimator as an estimate of the population ATE when sample selection (S = 1 for individuals in the sample; S = 0 for those in the target population) depends on a single binary covariate Z and there is heterogeneity in the effect of treatment A due to Z when the outcome follows a simple linear model: $E(Y_i) = b_0 + b_aA_i + b_{az}A_iZ_i$ as

$$b_{az}\left[\frac{P(Z=1)}{P(S=1)}[P(S=1 \mid Z=1) - P(S=1)]\right]$$

If this model is correct, accounting for the sample selection dependency with Z using inverse probability of selection weighting or direct standardization provides an unbiased estimate of the population ATE.

Now, suppose the true outcome model includes additional covariates and interaction terms: $E(Y_i) = b_0 + b_aA_i + b_zZ_i + b_uU_i + b_{az}A_iZ_i + b_{au}A_iU_i + b_{zu}Z_iU_i + b_{azu}A_iZ_iU_i$ where U is a single binary covariate that is unmeasured in the study sample, and there is heterogeneity in the treatment effect across Z and U and within subgroups of Z due to U.

Borrowing notation from Cole and Stuart, two estimands are defined as follows: $PATE = E(Y^1) - E(Y^0)$, the mean difference in potential outcomes in the population, and $\text{SATE} = E\left(Y^1 \mid S=1\right) - E\left(Y^0 \mid S=1\right)$, the mean difference in outcomes in the study sample.

When sample selection depends on Z and U, one can derive the bias in the SATE estimator as an estimate of the PATE as:

$$b_{az}\left[\frac{P(Z=1)}{P(S=1)}[P(S=1 \mid Z=1) - P(S=1)]\right] + b_{au}\left[\frac{P(U=1)}{P(S=1)}[P(S=1 \mid U=1) - P(S=1)]\right] + b_{azu}\left[\frac{P(Z=1,U=1)}{P(S=1)}[P(S=1 \mid Z=1,U=1) - P(S=1)]\right]$$

Thus, if effects within subgroups of Z are homogenous (i.e., $b_{azu} = 0$), then bias also depends positively on the heterogeneity of treatment effects across groups defined by U, the prevalence of the heterogeneity characteristic U, the proportion of the target population not sampled, and the extent to which sample selection depends on U. If subgroup effects are not constant (i.e., $b_{azu} \neq 0$), then bias also depends positively on the subgroup effect heterogeneity, the prevalence of Z and U, the proportion of the target population not sampled, and the extent to which sample selection depends on Z and U. The derivation of the bias for a simple linear outcome model is given in the eAppendix. Expressions for the bias will differ for a log–link model or non-additive effects.

Monte Carlo Simulation

To explore the generalizability of subgroup effects for finite sample sizes, we adapted the data generating mechanism from Lesko et al.[2] and generated large target populations where Z and U were two independent Bernoulli random variables with expectations 0.15 and 0.20, A was a Bernoulli random variable with expectation 0.5 and independent of Z, U, and potential outcomes Y(1) and Y(0). Potential outcomes were generated as Bernoulli random variables where $P(Y_i) = 0.1073 + 0.05A_i + 0.2Z_i + 0.2U_i + 0Z_iU_i + b_{az}A_iZ_i + b_{au}A_iU_i$ for 125 scenarios that combined b_{az}, b_{au}, and b_{azu} ranging from −0.15 to 0.15 by increments of 0.1 including 0. A main effect for ZU can be added to the potential outcome model but adds additional complication without introducing bias in treatment effect estimates because it changes P(Y) by the same amount for those with A=1 and A=0. For each scenario, we drew a study sample of $n = 2000$ individuals from the target population, where sample selection depended on strata defined by Z and U: (Z=0, U=0) = 320, (Z=1, U=0) = 480, (Z=0, U=1) = 480, (Z=1, U=1) = 720.

We used an outcome modeling approach (also known as “G-computation”[6]) to estimate the population ATE, using generalized linear models to model the outcome in the study sample, then used the model coefficients to predict outcomes under treatment and control in the target population. For each of the 125 scenarios, we first modeled the outcome “correctly” in the study sample using a saturated model with A, Z, U, and all two- and three-way interactions. Next, we misspecified the outcome model to omit U, AU, and AZU, reflecting a scenario where U is unknown (e.g., an unobserved factor that differs between males in the sample and population). For each of the 125 scenarios, we ran 5000 simulations and calculated absolute bias and mean squared error (MSE) of the misspecified model relative to the correctly specified model. This analysis using synthetic data did not require ethical review.

Simulation results

Results for bias and MSE were similar across values of b_{az} because Z, the measured treatment effect modifier, was always included in the outcome modeling approach for all scenarios. Thus, we present results from the 25 scenarios when $b_{az} = 0.15$ in the Figure. For a given value of b_{au} or b_{az}, varying three-way interaction effect sizes for b_{azu} did not add appreciably to bias in the population ATE estimate. For example, when there was no treatment effect heterogeneity independently by U (i.e., $b_{au} = 0$), even large three-way interactions did not increase bias appreciably (i.e., moving across the x-axis). In contrast, for a given value of b_{azu}, large two-way interactions between treatment and an unmeasured covariate (b_{au}) resulted in large increases in bias of the population ATE estimate (i.e., moving up and down the y-axis). Results for MSE showed a similar pattern indicating that variance is not a major component of MSE in this setting.

Applied example

We explored the generalizability of subgroup effects using data from the PREMIER Study, which randomized 810 adults at risk for hypertension to simultaneous lifestyle changes.[7] For this illustrative example, we combined participants in the two lifestyle intervention programs as the treatment group. The main outcome was the difference in systolic blood pressure (SBP) between 6 months and baseline. The International Population Study on Macronutrients and Blood Pressure (INTERMAP) study was used to represent the target population.[8] INTERMAP is a cross-sectional epidemiologic study of men and women ages 40–59 years in the US, UK, China, and Japan. For this illustrative example, we restricted to participants in the US who met the eligibility criteria of PREMIER to represent individuals who would receive the lifestyle intervention programs. The final sample size was 478. This secondary analysis of the PREMIER and INTERMAP studies was approved by the Duke University Medical Center Institutional Review Board (Pro00101084).

Across the trial and target population, we consider race (Black/Other), smoker (yes/no), and systolic blood pressure (SBP) at baseline as common covariates. The distributions of these covariates are shown in Table 1. Compared to the target population, more participants in the trial were Black and non-smokers. There was evidence that race and smoking status modified the treatment effect independently and jointly. For this illustrative example, we assume that race is unmeasured in the target population, thus omitting the variable from the g-computation models.

The results of g-computation models to estimate population mean differences in SBP for the effect of lifestyle intervention in INTERMAP are presented in Table 2. A model that included treatment, race, smoking, and all two- and three-way interactions resulted in an estimate of the population ATE = −4.2 (95% CI: −6.0, −2.5). In contrast, a model that included only treatment, smoking, and their two-way interaction resulted in an estimate of the population ATE = −4.0 (95% CI: −5.7, −2.2). A model that included treatment, race, smoking, and two-way interactions but excluded a three-way interaction (i.e., assuming homogeneity of subgroup effects) resulted in an estimate of the PATE =−4.1 (95% CI: −5.9,

−2.4), which did not differ substantially from the estimate when we adjust for heterogeneity of subgroup effects.

Discussion

In this brief report, we explored the issue of generalizability of subgroup effects. Though all-important effect modifiers should be measured in the study sample and target population, the implications of unmeasured heterogeneity of main effects is substantially larger than unmeasured heterogeneity of subgroup effects. Although we did not find an appreciable increase in bias due to the omission of a three-way interaction relative to the omission of a two-way interaction in the g-computation model, future work could explore flexible approaches such as machine learning methods to account for all possible interactions between measured variables, as well as consider sensitivity analyses if subgroup effects vary substantially by a covariate that is not measured in the target population. We hope that by examining the generalizability of subgroup effects in relation to main effects, epidemiologists will increasingly use these important methods for making population inferences using a study sample.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

We thank Jay Xu and anonymous reviewers for helpful comments on earlier drafts of the manuscript.

Financial Support: Funding for this study was provided in part by the National Institutes of Health (T32DA007292 (PIs: Maher/Johnson), R00MH111807 (PI: Hong), P50MH115842 (PI: Daumit), T32MH109436 (PI: Barry/Stuart)) and the US Department of Education R305D150003 (PI: Stuart).

References

1. Bengtson AM, Pence BW, Gaynes BN, Quinlivan EB, Heine AD, O'Donnell JK, Crane HM, Mathews WC, Moore RD, Westreich D, O'Cleirigh C, Christopoulos K, Mimiaga MJ, Mugavero MJ. Improving Depression Among HIV-Infected Adults: Transporting the Effect of a Depression Treatment Intervention to Routine Care. J Acquir Immune Defic Syndr 2016;73(4):482–488. [PubMed: 27668804]
2. Lesko CR, Buchanan AL, Westreich D, Edwards JK, Hudgens MG, Cole SR. Generalizing Study Results: A Potential Outcomes Perspective. Epidemiology 2017;28(4):553–561. [PubMed: 28346267]
3. Susukida R, Crum RM, Ebnesajjad C, Stuart EA, Mojtabai R. Generalizability of findings from randomized controlled trials: application to the National Institute of Drug Abuse Clinical Trials Network. Addiction 2017;112(7):1210–1219. [PubMed: 28191694]
4. Kern HL, Stuart EA, Hill J, Green DP. Assessing methods for generalizing experimental impact estimates to target populations. J Res Educ Eff 2016;9(1):103–127. [PubMed: 27668031]
5. 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. [PubMed: 20547574]
6. Robins J A New Approach to Causal Inference in Mortality Studies with a Sustained Exposure Period - Application to Control of the Healthy Worker Survivor Effect. Mathematical Modelling 1986;7(9–12):1393–1512.

7. Appel LJ, Champagne CM, Harsha DW, Cooper LS, Obarzanek E, Elmer PJ, Stevens VJ, Vollmer WM, Lin PH, Svetkey LP, Stedman SW, Young DR, Writing Group of the PCRG. Effects of comprehensive lifestyle modification on blood pressure control: main results of the PREMIER clinical trial. JAMA 2003;289(16):2083–93. [PubMed: 12709466]
8. Stamler J, Elliott P, Dennis B, Dyer AR, Kesteloot H, Liu K, Ueshima H, Zhou BF, Group IR. INTERMAP: background, aims, design, methods, and descriptive statistics (nondietary). J Hum Hypertens 2003;17(9):591–608. [PubMed: 13679950]

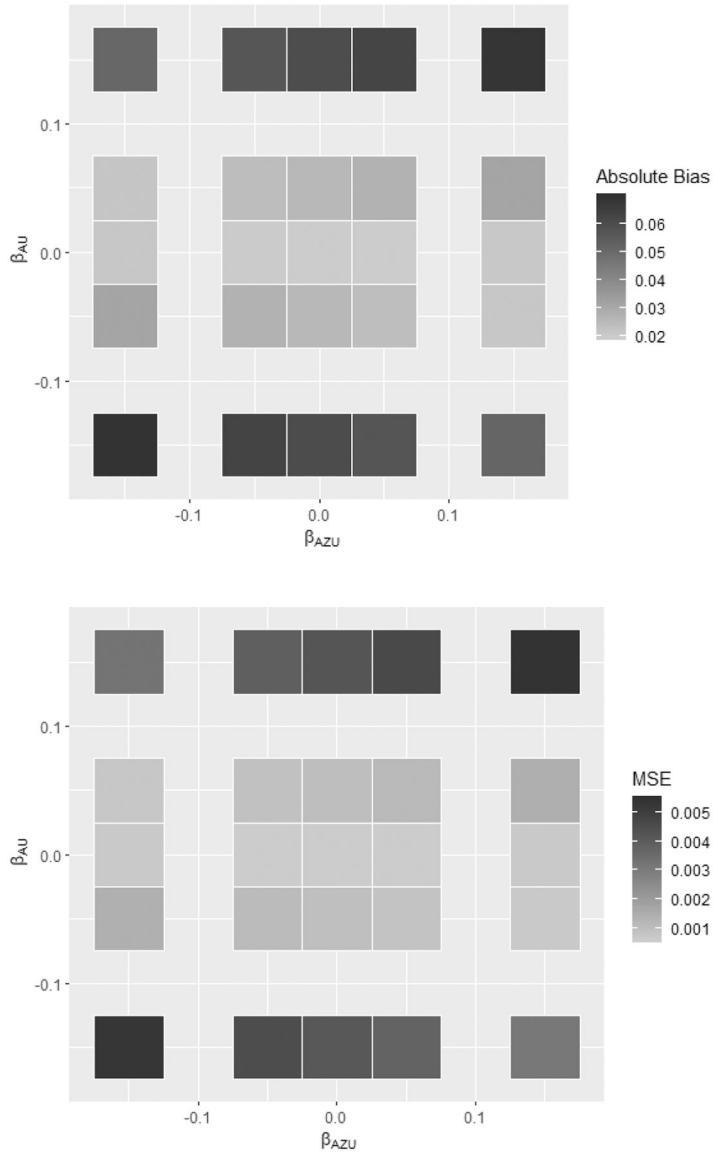

Figure.

Absolute value of bias and mean squared error (MSE) of the population average treatment effect estimates across simulation scenarios when $b_{az} = 0.15$.

Table 1.

Select baseline characteristics of participants in the PREMIER Study and the INTERMAP Study in the United States.

	Trial	Population		
	PREMIER (N=810)	**INTERMAP (N=478)**	**SMD[a]**	**95% CI**
Black (N (%))	279 (34.4)	116 (24.3)	0.22	(0.11, 0.34)
Smoke (N (%))	39 (4.8)	83 (17.4)	−0.41	(−0.52, −0.30)
Black × Smoke (N (%))			0.49	(0.37, 0.60)
1. Black & smoker	17 (2.1)	29 (6.1)		
2. Other race & smoker	22 (2.7)	54 (11.3)		
3. Black & non-smoker	262 (32.3)	87 (18.2)		
4. Other race & non-smoker	509 (62.8)	308 (64.4)		
Systolic blood pressure at baseline (mmHg, Mean (SD))	134.89 (9.57)	134.57 (9.94)	0.03	(−0.08, 0.15)

[a]Standardized mean difference, SMD. Imbalance defined as SMD absolute value > 0.20. CI=confidence interval.

Table 2.

Estimates of the population average treatment effect of a lifestyle intervention on systolic blood pressure (SBP in mmHg) at 6 months among INTERMAP participants in the United States.

	Target population	
Outcome model specification	**Mean difference in SBP (mmHg)**	**95% CI**
Naïve SATE		
Average treatment effect in the trial	−4.2	−5.7, −2.8
Assuming race as unmeasured[a]		
A × smoking	−4.0	−5.7, −2.2
A × smoking + A × race	−4.1	−5.9, −2.4
A × smoking + A × race + A × race × smoking	−4.2	−6.0, −2.5

[a]All models include SBP at baseline as a covariate; A denotes treatment. CI=confidence interval, SATE=sample average treatment effect, SBP=systolic blood pressure.