

Generalizing Causal Effects to a Target Population Without Individual-Level Data from the Target Population

Wen Wei Loh^a and Dongning Ren^b

^aDepartment of Methodology and Statistics, Maastricht University, Maastricht, The Netherlands; ^bDepartment of Work and Social Psychology, Maastricht University, Maastricht, The Netherlands

ABSTRACT

Generalizability is a perennial concern in randomized studies. While randomized studies are the gold standard for establishing causality, study samples are rarely representative of broader populations due to factors such as convenience sampling, participant self-selection, and researchers' inclusion or exclusion criteria. To strengthen the generalizability of randomized experiments, the framework of causal effect generalizability provides a solution. However, existing methods require accessing representative individual-level data from the target population, which is often unavailable due to limited resources, data access restrictions, or privacy concerns. In this paper, we develop a novel method to generalize causal effects using only summary statistics on covariates from the target population. We illustrate the estimator using a real-world study by generalizing the impact of a climate change behavioral intervention from the study sample to a broader population. By avoiding the need for individual-level data from the target population, our method offers a practical tool for generalizing causal findings from randomized studies. We hope that the proposed method helps build more accurate theories and enhance the policy relevance of behavioral and psychological research.

KEYWORDS

Causal inference; treatment effect heterogeneity or moderation; potential outcomes; selection bias; generalizability

Randomized experiments are the gold standard for answering causal questions. Ideally, causal effects obtained from a randomized study should generalize to broader populations (Dodge et al., 2024; Sears, 1986), allowing experimental findings to inform psychological theories and guide public policies. In practice, however, randomized experiments often suffer from limited generalizability. Concerns about this limitation are longstanding, pervasive, and widely acknowledged across the behavioral, psychological, and social sciences (Dodge et al., 2024; Sears, 1986). Following influential critiques of so-called Western, Educated, Industrialized, Rich, and Democratic (WEIRD) study samples and decades of empirical audits of the literature, researchers and journal editors have repeatedly called for greater attention to the generalizability of research findings (Arnett, 2008; Bauer, 2022; Clarke et al., 2023; Henrich et al., 2010; Rad et al., 2018; Thalmayer et al., 2021; Vazire, 2024). Despite this sustained focus, practical solutions remain limited. Study samples are rarely representative of the intended target populations, due to factors such as convenience sampling, participant self-selection, and researchers' use of

inclusion or exclusion criteria (Diener et al., 2022; IJzerman et al., 2020; Loh & Ren, 2025b; Mumford & Schisterman, 2019).

Causal effect generalizability (Bareinboim & Pearl, 2016; Cole & Stuart, 2010; Pearl & Bareinboim, 2014) offers a framework for extending causal (or treatment) effects from randomized experiments to broader target populations. This framework was first developed in the causal inference and epidemiological literature, and has recently been applied to generalize causal effects to population-based cohorts (see, e.g., Hutcheon and Liao (2023); Lao et al. (2024)). Loh and Ren (2025a) offer an accessible introduction for a behavioral and psychological sciences audience. However, existing methods require accessing individual-level data from the target population (Dahabreh et al., 2019; Degtiar & Rose, 2023; Lesko et al., 2017). For example, generalizing a climate change behavioral intervention's effects from an experimental study sample to the entire adult population would necessitate individual-level records (e.g., each person's gender, age, education, and income) in the entire population,

or a representative (e.g., simple random) sample from the population. But such fine-grained individual-level data often remains inaccessible due to limited resources, data access restrictions, or privacy concerns, especially for hard-to-reach or vulnerable subgroups. Instead, what may be practically available and accessible to individual researchers are summary statistics (e.g., national averages or medians) of the target population, such as those commonly recorded in government census or administrative surveys.

Can researchers generalize causal effects without relying on individual-level data from, but using only the summary statistics of, the target population? Here, we propose a method that enables this. In the remainder of this paper, we will introduce the developed method, evaluate the proposed estimator's empirical performance under various scenarios using a simulation study, and provide a real-world application using publicly available data from a climate change behavioral intervention study. We have developed simple-to-use R scripts (R Core Team, 2024) to aid researchers in psychological and behavioral sciences, as well as policy-makers, in implementing the estimation method. The proposed method for generalizing causal effects is simple, cost-efficient, and can readily enhance policy relevance and insights of experimental evidence.

Causal effects in the study sample and target population

We first distinguish between the causal effects in the study sample and the target population. Let A denote a treatment either received ($A = 1$) or not received ($A = 0$), and let Y denote the outcome. Following the *potential outcomes* framework, commonly called the Neyman-Rubin causal model (Holland, 1986; Rubin, 1990; Splawa-Neyman et al., 1990), let Y^1 and Y^0 denote an individual's potential outcomes had they received or not received the treatment, respectively. Throughout, we maintain the *stable unit treatment value assumption* (SUTVA; Rubin (1986)), which holds that: (i) among individuals who did (not) receive treatment, there are “no hidden (non-)treatment variations” (Imbens & Rubin, 2015); and (ii) each individual's potential outcomes cannot be affected by other individuals' (non-)treatments, so that there is “no (treatment) interference” (Hudgens & Halloran, 2008). The average causal effect in the target population is defined as:

$$E(Y^1 - Y^0) = E(Y^1) - E(Y^0), \quad (1)$$

where $E(\cdot)$ denotes the expectation of a random variable across the whole population.

Let S denote a binary selection indicator with $S = 1$ if an individual is in the study, or $S = 0$ otherwise. The corresponding average causal effect within the study ($S = 1$) is:

$$E(Y^1 - Y^0 | S = 1) = E(Y^1 | S = 1) - E(Y^0 | S = 1), \quad (2)$$

where $E(\cdot | S = 1)$ denotes the conditional expectation among the study subpopulation. Therefore, the crucial difference between (1) and (2) is that the former pertains to the whole population, whereas the latter is restricted to the distinct subpopulation delineated by $S = 1$. The causal effects defined in (1) and (2) are equivalent only under stringent conditions, such as a constant individual causal effect $Y^1 - Y^0$, or selection S is (marginally) independent of both potential outcomes Y^1 and Y^0 . In reality, individual causal effects may be modified by covariates that concurrently act as effect moderators and influence selection or study inclusion (S). In other words, the distribution of effect moderators differs between the study sample and the target population. When this is the case, an estimator that is unbiased and consistent for the causal effect (2) will nonetheless be biased for the causal effect (1). In the next section, we describe how to estimate (1) using only observed data from the study sample ($S = 1$) and summary statistics for the target population. No individual-level data from the target population is needed.

Estimation procedure

We first introduce notation for developing the proposed estimator. Let W denote a set of covariates that are concurrently putative moderators of the causal effect (1) and possible predictors of study selection S , and let $\mu = E(W)$ denote their known expectations in the target population. Crucially, μ can include not only means but also higher moments such as variances and covariances, as well as quantiles, such as the median or interquartile range, which better characterize the empirical distribution. As an example, suppose there are two covariates: a continuous X_1 with mean $\mu_1 = E(X_1)$, variance $\sigma_1^2 = \text{var}(X_1)$, and median $\mu_1[0.5]$ where $\Pr(X_1 \leq \mu_1[0.5]) = 0.5$; and a binary X_2 with mean $\mu_2 = E(X_2)$; and let their covariance be $\sigma_{12} = \text{cov}(X_1, X_2)$. Let $W = (1, X_1, X_1^2, 1(X_1 \leq \mu_1[0.5]), X_2, X_1 X_2)^T$, where $1(E)$ is the indicator function that returns a value of one if the event E is true or zero otherwise, and we include a constant one to ease the subsequent mathematical equations. The corresponding population quantities of these non-linear functions of the covariates in W are $\mu = (1, \mu_1, \sigma_1^2 + \mu_1^2, 0.5, \mu_2, \sigma_{12} + \mu_1 \mu_2)^T$. Hence, W may include quadratic (or product interaction) terms for the covariates if

their population (co)variances are known, or binary indicators comparing the individual covariates in the study sample to their known population quantiles (such as the median). Note that a covariate distribution can be completely determined for a binary covariate by its mean, and for a normally distributed covariate by its mean and variance.

Let $\Pr(S = 1|W; \beta) = \{1 + \exp(-\beta^T W)\}^{-1}$ denote the selection or study inclusion probability in terms of a logistic regression model with coefficients β for the covariates W . (The intercept corresponds to the constant term in W .) Because no individual-level data is available for those with $S = 0$ (i.e., not selected for the study), it is impossible to empirically estimate β using only observed data for those with $S = 1$. Instead, we will estimate β by finding the values that equate the weighted covariate sample means of W (with weights being the inverse of the selection probabilities) in the study to the known population expectations μ . Note that no model(s) for the outcome describing the functional forms of the covariate-outcome relationships, especially how these associations may be modified or moderated by the treatment, are needed.

Let $\theta^a = E(Y^a)$, $a = 0, 1$, denote the population average potential outcome under each treatment level $A = a$. Estimators of θ^0 and θ^1 can be obtained by solving the following set of estimating equations:

$$E \begin{bmatrix} S(1-A)(Y - \theta^0) \frac{1}{\Pr(S = 1|W; \beta)} \\ SA(Y - \theta^1) \frac{1}{\Pr(S = 1|W; \beta)} \\ \frac{SW}{\Pr(S = 1|W; \beta)} - \mu \end{bmatrix} = 0. \quad (3)$$

The first two equations correspond to the population average potential outcomes θ^0 and θ^1 , respectively, and the remaining equations correspond to the coefficients β in the assumed logistic regression model for the selection probability $\Pr(S = 1|W; \beta)$. We explain in the [Appendix](#) how, under certain assumptions, the estimating equations in (3) are unbiased. It follows from standard estimating equation theory (Stefanski & Boos, 2002) that estimators of θ^0 and θ^1 are consistent and asymptotically normal. Solving the combined estimation equations for $(\theta^0, \theta^1, \beta^T)$ simultaneously allows their covariances to be freely estimated and ensures that sampling uncertainty for β is accounted for in the estimators of θ^0 and θ^1 . Hence, the asymptotic covariance of θ^0 and θ^1 can be consistently estimated, without relying on normality or constant variance assumptions, using an empirical sandwich estimator (Stefanski & Boos, 2002). These estimators can then be used to calculate

estimators of, and Wald confidence intervals for, the causal effect $\theta^1 - \theta^0$ in (1).

Because we can only observe individual-level data for those in the study, the estimating equations in (3) must be solved using the corresponding sample moments restricted to the $S = 1$ study sample. Following the law of iterated expectations, the estimating equations conditional on $S = 1$ are:

$$E \begin{bmatrix} (1-A)(Y - \theta^0) \frac{\Pr(S = 1)}{\Pr(S = 1|W; \beta)} \\ A(Y - \theta^1) \frac{\Pr(S = 1)}{\Pr(S = 1|W; \beta)} \\ W \frac{\Pr(S = 1)}{\Pr(S = 1|W; \beta)} - \mu \end{bmatrix} \bigg| S = 1 = 0. \quad (4)$$

Hence, in addition to the population expectations in μ , the overall proportion in the study relative to the whole target population $\pi = \Pr(S = 1)$ must also be known. Define the individual moment functions for the $S = 1$ study sample as:

$$U_i(\theta^0, \theta^1, \beta^T) = \begin{bmatrix} (1-A_i)(Y_i - \theta^0) \frac{\pi}{\Pr(S = 1|W_i; \beta)} \\ A_i(Y_i - \theta^1) \frac{\pi}{\Pr(S = 1|W_i; \beta)} \\ W_i \frac{\pi}{\Pr(S = 1|W_i; \beta)} - \mu \end{bmatrix}, \quad (5)$$

where the i subscript indexes the i -th individual for $i = 1, \dots, n$ in the study sample. The estimators of θ^0 and θ^1 , and thus $\theta^1 - \theta^0$, are then obtained by solving for the parameter values that yield a zero sample average; i.e.,

$$(\hat{\theta}^0, \hat{\theta}^1, \hat{\beta}^T) = \arg_{(\theta^0, \theta^1, \beta^T)} \frac{1}{n} \sum_{i=1}^n U_i(\theta^0, \theta^1, \beta^T) = 0. \quad (6)$$

We empirically evaluate the operating characteristics of this estimator using a simulation study later.

Conditions for unbiased estimation and practical recommendations

We first formalize the assumptions that must be satisfied to obtain unbiased estimating equations in (3) for generalizing causal effects back to the full target population, and offer practical recommendations to bolster their defensibility.

(A1) Study inclusion must be conditionally independent of the potential outcomes.

$$Y^a \perp\!\!\!\perp S | W; \text{ for all } a.$$

In other words, among individuals with the same values of W , selection or study inclusion

(S) is pseudo-randomized and conditionally independent of their potential outcomes (Y^a). This causal assumption, which is empirically untestable and must be justified by subject matter knowledge, is violated when unrecorded covariates act as effect moderators and also influence selection or study inclusion (Loh & Ren, 2025a; Ren & Loh, 2024). In practice, we encourage researchers to record a comprehensive set of pretreatment baseline covariates, even in randomized experiments, particularly when the study sample is nonrepresentative (e.g., due to researchers' resource constraints or participants' self-selection). We recommend visually representing the causal structure with a causal diagram (Grosz et al., 2020; Loh & Ren, 2023) and including a selection node (Bareinboim & Pearl, 2016; Pearl & Bareinboim, 2014; Poppe et al., 2025; Ren & Loh, 2024).

(A2) Study inclusion cannot be ruled out based on W . Under the assumed logistic model, everyone must have a non-zero selection probability conditional on W :

$$\Pr(S = 1|W; \beta) > 0.$$

Note that it is permissible to have a unit probability of selection or inclusion in the study. This assumption can be violated when individuals in the target population outside the study sample ($S = 0$) are ruled out or discouraged from participating in the study because of their characteristics W . In practice, we encourage researchers to define

$$Y = Y^1 A + Y^0 (1 - A).$$

In other words, each individual's observed outcome (Y) is exactly one of the potential outcomes (Y^1 or Y^0) revealed under the observed treatment (A). This assumption is violated when SUTVA doesn't hold, such as when there are peer influence or spillover effects among causally connected individuals (e.g., in social networks; An et al. (2022); Loh and Ren (2022)). (A4) Treatment is randomized within the study.

$$A \perp\!\!\!\perp Y^0, Y^1, W | S = 1.$$

Assumption (A4) further implies the following assumptions within the study:

- a. There is no unmeasured treatment-outcome confounding.

$$Y^0, Y^1 \perp\!\!\!\perp A | W, S = 1.$$

- b. The chances of assignment to each treatment level $A = a$ do not depend on W .

$$\Pr(A = a | W, S = 1) = \Pr(A = a | S = 1).$$

Assumptions (A3) and (A4) can be satisfied by design using a properly-conducted (double-blind) randomized controlled study.

Simulation study

We conducted a simulation study using the following data-generating process:

$$X_1 \sim N(\mu_1, \mu_4)$$

$$X_2 \sim \text{Exp}(\mu_2)$$

$$X_3 \sim \text{Bernoulli}(\mu_3)$$

$$\Pr(S = 1 | X_1, X_2, X_3) = \text{expit}(b_0 + 0.7X_2 - 0.7X_3 + 0.35X_1^2 - 0.35X_2X_3)$$

$$S \sim \text{Bernoulli}\{\Pr(S = 1 | X_1, X_2, X_3)\}$$

$$Y_0 = \min(|X_1|, 1/\sqrt{X_2}) + X_2 \log(X_1^2) + (2X_3 - 1) \sin(X_1) + \epsilon,$$

$$\epsilon \sim N(0, (1 + X_3)^2)$$

$$Y_1 = Y_0 + 0.5X_1^2 + 2X_1X_2 - 2X_3.$$

the target population using the same exclusion or inclusion criteria as the study sample (Power et al., 2022).

(A3) Causal, or counterfactual, consistency (Pearl, 2010) must be met.

We included a quadratic term (X_1^2) and product interaction terms (X_2X_3 and X_1X_2) among the covariates to illustrate how such nonlinear functions can be included in W . The population parameters were $\mu_1 = 0, \mu_2 = 1, \mu_3 = 0.1, \mu_4 = 1$. The intercept b_0 was numerically determined to minimize the absolute difference between the overall population probability of

Table 1. Results of the simulation study.

N	π	Results	Effect estimator		
			(S = 1) only	Generalized with minimal W	Generalized with all W
40000	0.005	Average	1.22	0.32	0.32
40000	0.005	Bias	0.92	0.02	0.02
40000	0.005	ESE	1.47	1.70	0.71
40000	0.005	ASE		1.55	0.59
40000	0.005	Coverage		0.94	0.89
40000	0.005	Convergence errors	0.00	0.01	0.19
40000	0.100	Average	0.84	0.30	0.30
40000	0.100	Bias	0.54	0.00	−0.00
40000	0.100	ESE	0.22	0.25	0.13
40000	0.100	ASE		0.25	0.13
40000	0.100	Coverage		0.95	0.95
40000	0.100	Convergence errors	0.00	0.00	0.00
250000	0.005	Average	1.23	0.31	0.30
250000	0.005	Bias	0.93	0.01	0.00
250000	0.005	ESE	0.58	0.66	0.28
250000	0.005	ASE		0.65	0.27
250000	0.005	Coverage		0.95	0.94
250000	0.005	Convergence errors	0.00	0.00	0.00
250000	0.100	Average	0.84	0.30	0.30
250000	0.100	Bias	0.54	0.00	0.00
250000	0.100	ESE	0.09	0.10	0.05
250000	0.100	ASE		0.10	0.05
250000	0.100	Coverage		0.95	0.95
250000	0.100	Convergence errors	0.00	0.00	0.00

ESE = Empirical standard error; ASE = Asymptotic standard error. The ESE is the standard deviation of the point estimates, and the ASE is the average of the asymptotic standard errors. “Coverage” is the proportion of 95% confidence intervals that included the true effect, and “Convergence errors” is the proportion of simulated datasets that failed to find a solution. For the estimator using the study sample ($S = 1$ only), we did not estimate the standard errors or confidence intervals; hence, these were left blank. All results were rounded to two decimal places.

selection $\hat{\Pr}(S = 1; b_0) = \hat{E}[\Pr(S = 1|X_1, X_2, X_3)]$ and a given value of π ; i.e., $b_0 = \arg \min_b |\pi - \hat{\Pr}(S = 1; b)|$. The average causal effect was $E(Y_1 - Y_0) = 0.3$. Within each study sample ($S = 1$), treatment was randomized as $A \sim \text{Bernoulli}(0.5)$.

We compared three possible estimators of the average causal effect. First, an effect estimator using the study sample ($S = 1$ only) using the ordinary least squares estimator of the coefficient for A in a simple linear regression of Y on A . Second, the proposed estimator with only population summaries for the covariates $W = (1, X_3, X_1^2)^T$. This was the minimal set of covariates that influenced selection and moderated the effect. Hence $\mu = (1, \mu_3, \mu_4 + \mu_1^2)$. Third, we used the proposed estimator with population summaries for all the covariates $W = (1, X_1, X_2, X_3, X_1^2, X_1X_2, X_2X_3)^T$. This included all covariates used to either predict selection or moderate the effect. Hence $\mu = (1, \mu_1, \mu_2, \mu_3, \mu_4 + \mu_1^2, \mu_1\mu_2, \mu_2\mu_3)$.

We considered two possible population sizes $N = 40000$ or $N = 250000$, and two possible selection proportions $\pi = 0.005$ or $\pi = 0.1$, so that the smallest possible average study sample size was $n = 200$. The results from 5000 replications for each combination of N and π are presented in Table 1 below. The estimator using only the study sample ($S = 1$) was biased in all scenarios, with greater bias when the selected study sample was a smaller subset of the population. The biases empirically demonstrate that findings from

nonrepresentative samples need not generalize to broader populations. Both estimators of the generalized effect using either minimal W or all W were unbiased in all settings. Adjusting for the minimal W that influenced selection and moderated the effect, using their correctly specified functional forms in the selection model, was sufficient to satisfy assumption (A1) and obtain unbiased estimators. Note that the proposed estimator does not require or rely on correctly specifying an outcome model. Hence, even under unknown (non-linear or complex) data-generating processes for the outcome, as in this simulation study, the estimator does not require modeling the outcome as a function of the treatment and covariates, especially how the treatment effect may be moderated by the covariates. Using more covariates yielded more precise estimates, but in small study samples, risked higher chances of convergence errors resulting from a failure to find a solution to the estimating equations. The Wald confidence intervals for both generalized effect estimators achieved their nominal coverage levels for large values of N or π .

Illustration using the effect of a climate change behavioral intervention among participants who already believed in climate change

Sinclair et al. (2025) compared various climate change behavioral interventions aimed at driving climate

action. As the authors rightfully point out, the study is highly policy-relevant for driving climate action in broader populations. However, the authors applied a strict inclusion criterion, restricting their study sample to comprise only participants who already believed in climate change. Climate change skeptics who “reported denying the existence or anthropogenic causes of climate change” (Sinclair et al., 2025) were thus excluded.

This choice of study sample undermined the generalizability of the causal conclusions regarding the effectiveness of the tested interventions. When predictors of climate change beliefs moderate an intervention’s effect, and the distributions of these predictors differ between climate change believers and skeptics, misleading conclusions about the intervention’s effectiveness in broader populations can ensue. Here, we will use this motivating example to illustrate how to generalize the causal effect of one such intervention from a narrowly selected study sample to a broader population, such as adults in the United States. Let $S = 1$ if an individual is selected for the study, or $S = 0$ otherwise. In particular, Sinclair et al. (2025) excluded anyone who (i) did not respond to the study being advertised as being about climate change; (ii) responded no to a single-item Prolific prescreening question, “Do you believe in climate change?” or; (iii) if they answered yes to that question, subsequently indicated that they denied or were uncertain or skeptical that anthropogenic climate change is occurring; see the “Exclusions” section on page 3 of the online supporting materials of Sinclair et al. (2025) for the specific criteria.

Following Goldberg et al. (2021); Hornsey and Pearson (2024); Todorova et al. (2025), and given the publicly available study data of Sinclair et al. (2025) on OSF¹, we will consider the following covariates in W : age, gender, education, and income. We set the population summaries μ at the following levels based on the available national figures. The median age was 39.1 years², so we defined the covariate to be a binary indicator taking value 1 if an individual was older than 39.1 years, or 0 otherwise. 50.9% were women³, so we defined the gender covariate to be 1 if a woman, or 0 otherwise. For educational attainment: 9% had less than a high school diploma or equivalent; 28% had high school as their highest level of school

completed; 15% had completed some college but not a degree; 10% had an associate degree as their highest level of school completed; 23% had a bachelor’s degree as their highest degree; and 14% had completed advanced education such as a master’s degree, professional degree, or doctorate⁴. We dummy-coded educational attainment (“edu1” to “edu5”) using a bachelor’s degree as the reference level. The median household income was about \$75000⁵, so we defined the covariate to be 1 if household income was below \$75000, or 0 otherwise.

For expository purposes, we used the “Individual Action Planning” intervention ($A = 1$) where participants selected a personal climate individual action goal, such as flying less or driving less, and developed a detailed plan to achieve it, including the steps involved, potential obstacles, and outcomes. The control group ($A = 0$) did not complete any intervention task. For the outcome, we considered participants’ intentions to engage in 12 pro-environmental behaviors (1 = a lot less, 7 = a lot more), including individual actions (eating vegan meals, flying by airplane, paying for renewable energy to power one’s home) and collective actions (e.g., donating, volunteering, or contacting representatives). We used the average of the 12 items as the outcome Y . We retained the $n = 1217$ observations in the study sample with complete data on the covariates, treatment, and outcome.

To estimate the average effect among the study sample ($S = 1$), we fitted a simple linear regression model of Y on A and used the coefficient estimator for A . The coefficient estimate was 0.11; 95% CI = (0.05, 0.18), in line with the results for this intervention shown in Figure 3 of Sinclair et al. (2025), reaffirming its effectiveness within the study sample. We then generalized the effect to a broader population with summaries μ as described above. To help researchers implement this in practice, we developed a single R function (“GEN.POP.IPSW”) which requires users to specify only four arguments: (i) the population summaries π and μ (“POP.summs”); (ii) the outcome name (“Y.name”); (iii) the treatment name (“A.name”); and (iv) the dataset name (“Data”). An example of the study dataset, the covariate sample means, the population summaries used, and the effect estimates are shown below in Box Illustration using the effect of a climate change behavioral intervention among participants who already believed in climate

¹<https://doi.org/10.17605/OSF.IO/X9C6J>.

²<https://www.census.gov/library/stories/2025/06/metro-areas-median-age.html>.

³<https://www2.census.gov/library/publications/decennial/2020/census-briefs/c2020br-06.pdf>.

⁴<https://www.census.gov/newsroom/press-releases/2023/educational-attainment-data.html>.

⁵<https://www.census.gov/library/publications/2023/demo/p60-279.html>.

change. The effect estimate among this broader population was 0.02; 95% CI = (−0.09, 0.13), suggesting that this intervention, which was estimated to increase

climate action by 0.11 among the study sample, would have a weaker, or even null, effect if implemented more widely beyond the study sample.

Box 1 Sample analysis code to generalize a causal effect using only population summaries in the climate change behavioral intervention study

Box 1: Sample analysis code to generalize a causal effect using only population summaries in the climate change behavioral intervention study

```
set.seed(12345)
DATA[sample(nrow(DATA), 6), ]

#      S age woman edu1 edu2 edu3 edu4 edu5 income MCII_individual  action
# 142 1   0     0    0    0    0    0    1     0              0 4.333333
#  51 1   0     0    0    0    0    0    0     0              1 5.000000
# 720 1   1     0    0    0    0    0    0     1              0 3.750000
# 730 1   1     0    0    0    0    0    0     1              0 4.166667
# 664 1   1     0    0    0    0    0    0     0              0 4.083333
# 826 1   1     0    0    0    0    1    0     0              1 4.333333

round(colMeans(DATA[, C_names]), 3)

#   age  woman  edu1  edu2  edu3  edu4  edu5 income
# 0.468 0.500 0.009 0.107 0.200 0.104 0.177 0.555

pop.summs

#      S      age  woman  edu1  edu2  edu3  edu4  edu5 income
# 0.650 0.500 0.509 0.090 0.280 0.150 0.100 0.140 0.500

ATE.hat <- GEN.POP.IPSW(
  POP.summs=pop.summs,
  Y.name="action",
  A.name="MCII_individual",
  Data=DATA)

round(ATE.hat, 2)

#  EST      SE  CI.L  CI.U
# 0.02  0.06 -0.09  0.13
```

Discussion

The central methodological contribution of our paper is the development of an estimator that generalizes the average causal effect from a single study to a known target population characterized solely by given covariate summaries, without any observed outcomes or individual-level data. We developed a weighting estimator that solves the estimating equations in (4) by equating the inverse probability of selection-weighted study-sample covariate moments with the given population moments. Hence, our estimator can be considered as belonging to a wider family of moment-based estimators (Chausse, 2023; Saul & Hudgens, 2020) that solve various estimating equations by matching weighted covariate moments across groups in different contexts, such as balancing treated and untreated groups within the same observational study (Imai & Ratkovic, 2014; Primbs et al., 2025; Thoemmes & Ong, 2016), trial-to-trial comparisons of observed outcomes from separate trials with different active treatments (Signorovitch et al., 2012), or meta-analysis that combines or pools trial-specific causal effects from different completed trials for evidence synthesis (Le et al., 2024; Vo et al., 2025).⁶ A key distinction of our paper is that we generalize a causal effect from a single study to a known, broader, and possibly nonrepresentative target population neither recruited nor selected for any study; whereas Le et al. (2024); Signorovitch et al. (2012); Vo et al. (2025) *transport* outcomes or effects across multiple, distinct trials to a population representative of participants in prior studied trials. Because generalizability extends causal effects from the study sample back to the full target population in which the study sample is nested, whereas transportability extends causal effects to a separate target population distinct from the study sample, the target causal estimands, and subsequently the weights, for generalizing and transporting causal effects differ (Lesko et al., 2017; Mumford & Schisterman, 2019).

We developed an estimator that intentionally avoids specifying an outcome model. This is advantageous when there are many putative moderators or limited guidance on the complex or nonlinear functional forms for the moderators. However, our estimator relies on correctly specifying a logistic regression model for selection or study inclusion predicted by the covariates. A potential direction for future work is to augment the weighting estimator with a model for potential outcomes so that the estimator is consistent if either the selection

model or the outcome model is correctly specified. Such a doubly robust estimator has been developed when individual-level population data are available (Dahabreh et al., 2019). In a different context, Campbell and Remiro-Azócar (2025) also developed a doubly robust estimator, although for cross-trial comparisons of outcomes from different single-arm trials, each with only treated participants, relative to a control population fully derived using external data sources with only covariate summaries. Because augmenting our estimator requires carefully considering how the treatment effect is moderated by covariates (e.g., treatment-covariate product interactions) in the study, but only their summaries (e.g., marginals unconditional on treatment) are available in the population, we defer this to future work.

In this paper, we focused on generalizing causal effects to broader, known target populations, as in a design-based approach (Rosenbaum, 2010). Hence, the population covariate summary statistics (μ, π) should be computed from a complete population frame (e.g., a census) or an enumeration (e.g., an administrative registry). These are finite-population quantities (not estimates) that are fixed by design and thus contribute no sampling variability. Hence, uncertainty arises only from the randomized study and the estimation of the weights, so the sandwich variance, conditional on fixed (μ, π) , is correct. We therefore encourage researchers to seek the true population values (μ, π) from a comprehensive or high-quality source. But if this is unfeasible, estimating (μ, π) introduces sampling error. We propose three options to account for this additional uncertainty in inference for the generalized causal effect, depending on the granularity of the available population information for (μ, π) . The first is to stack the score equations for estimating (μ, π) with our estimating equations for the weights to compute a joint sandwich variance. However, this likely requires individual-level population data, which obviates the need for our developed method. The second is a two-step procedure in which values of (μ, π) are first randomly drawn from their given distribution and then held fixed when applying the estimator to each draw. This approach has also been used in other contexts (Can & Rosseel, 2025; Vo et al., 2025). However, it requires estimating at least the full covariance matrix, if not the complete joint distribution, for (μ, π) . These first two options maintain the point estimator and affect only variance estimation. Published sources of population statistics often report only margins of error for the separate marginals (without their covariances). A third option is to conduct a sensitivity analysis by systematically varying (μ, π) jointly over the set of plausible values compatible with the reported uncertainty information

⁶We thank anonymous reviewer 2 for pointing us to both of these recent preprints in the last two references and for prompting us to clarify their distinctions from our work.

and summarizing the resulting range of estimates (Vansteelandt et al., 2006).

Conclusion

Causal effect generalizability enables the formal and rigorous extension of causal findings from randomized studies to broader populations. However, existing methods require access to representative individual-level data from the target population, which is often unavailable. In this paper, we developed a novel method to generalize causal effects using only summary statistics (e.g., means, variances, covariances, or quantiles such as the median) on covariates from the target population. Our method requires modeling study selection using logistic regression, but it avoids making any assumptions about the functional relationships between covariates and outcomes, especially how those associations may be moderated by treatment. Our method provides a practical tool to enhance the generalizability of experimental evidence in behavioral and psychological research.

Article information

Conflict of Interest Disclosures: Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical Principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: This work was partially supported by Grant 10430382410014 awarded to the authors from the Netherlands Organization for Health Research and Development (ZonMw), and by Grant 101220305 awarded to DR from the European Research Council (ERC) under the European Union's Horizon Europe research and innovation programme.

Role of the Funders/Sponsors: None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Acknowledgments: The authors thank the Editor-in-Chief, Dr. Emilio Ferrer, and the Associate Editor, Dr. Felix Thoemmes, for their advice and for guiding the manuscript through the review process. The authors also thank the two

anonymous reviewers for their constructive, detailed comments, which greatly improved the manuscript. The ideas and opinions expressed herein are those of the authors alone, and endorsement by the authors' institution, the Netherlands Organization for Health Research and Development (ZonMw), or the European Research Council (ERC) is not intended and should not be inferred.

Funding

This work was partially supported by Grant 10430382410014 awarded to the authors from the Netherlands Organisation for Health Research and Development (ZonMw), and by Grant 101220305 awarded to DR from the European Research Council (ERC) under the European Union's Horizon Europe research and innovation programme. The author(s) reported there is no funding associated with the work featured in this article.

References

- An, W., Beauville, R., & Rosche, B. (2022). Causal network analysis. *Annual Review of Sociology*, 48(1), 23–41. <https://doi.org/10.1146/annurev-soc-030320-102100>
- Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *The American Psychologist*, 63(7), 602–614. <https://doi.org/10.1037/0003-066X.63.7.602>
- Bareinboim, E., & Pearl, J. (2016). Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences of the United States of America*, 113(27), 7345–7352. <https://doi.org/10.1073/pnas.1510507113>
- Bauer, P. J. (2022). Psychological science stepping up a level. *Psychological Science*, 33(2), 179–183. <https://doi.org/10.1177/09567976221078527>
- Campbell, H., & Remiro-Azócar, A. (2025). Doubly robust augmented weighting estimators for the analysis of externally controlled single-arm trials and unanchored indirect treatment comparisons [arXiv preprint]. <https://doi.org/10.48550/arXiv.2505.00113>
- Can, S., & Rosseel, Y. (2025). Evaluating the standard error estimation of the local structural-after-measurement (LSAM) approach in structural equation modeling. *Methodology*, 21(4), 249–285. <https://doi.org/10.5964/meth.16517>
- Chausse, P. (2023). *momentfit: Methods of moments* [R package version 0.5]. <https://CRAN.R-project.org/package=momentfit>
- Clarke, B., Schiavone, S., & Vazire, S. (2023). What limitations are reported in short articles in social and personality psychology? *Journal of Personality and Social Psychology*, 125(4), 874–901. <https://doi.org/10.1037/pspp0000458>
- Cole, S. R., & Stuart, E. A. (2010). Generalizing evidence from randomized clinical trials to target populations: The ACTG 320 trial. *American Journal of Epidemiology*, 172(1), 107–115. <https://doi.org/10.1093/aje/kwq084>
- Dahabreh, I. J., Robertson, S. E., Tchetgen, E. J., Stuart, E. A., & Hernán, M. A. (2019). Generalizing causal inferences from individuals in randomized trials to all trial-eligible individuals. *Biometrics*, 75(2), 685–694. <https://doi.org/10.1111/biom.13009>

- Degtiar, I., & Rose, S. (2023). A review of generalizability and transportability. *Annual Review of Statistics and Its Application*, 10(1), 501–524. <https://doi.org/10.1146/annurev-statistics-042522-103837>
- Diener, E., Northcott, R., Zyphur, M. J., & West, S. G. (2022). Beyond experiments. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 17(4), 1101–1119. <https://doi.org/10.1177/17456916211037670>
- Dodge, K. A., Prinstein, M. J., Evans, A. C., Ahuvia, I. L., Alvarez, K., Beidas, R. S., Brown, A. J., Cuijpers, P., Denton, E.-g., Hoagwood, K. E., Johnson, C., Kazdin, A. E., McDanal, R., Metzger, I. W., Rowley, S. N., Schleider, J., & Shaw, D. S. (2024). Population mental health science: Guiding principles and initial agenda. *The American Psychologist*, 79(6), 805–823. <https://doi.org/10.1037/amp0001334>
- Goldberg, M. H., Gustafson, A., Ballew, M. T., Rosenthal, S. A., & Leiserowitz, A. (2021). Identifying the most important predictors of support for climate policy in the united states. *Behavioural Public Policy*, 5(4), 480–502. <https://doi.org/10.1017/bpp.2020.39>
- Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (2020). The taboo against explicit causal inference in nonexperimental psychology. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 15(5), 1243–1255. <https://doi.org/10.1177/1745691620921521>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? (2010/06/15). *The Behavioral and Brain Sciences*, 33(2–3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396), 945–960. <https://doi.org/10.1080/01621459.1986.10478354>
- Hornsey, M. J., & Pearson, S. (2024). Perceptions of climate change threat across 121 nations: The role of individual and national wealth. *Journal of Environmental Psychology*, 96, 102338. <https://doi.org/10.1016/j.jenvp.2024.102338>
- Hudgens, M. G., & Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482), 832–842. <https://doi.org/10.1198/016214508000000292>
- Hutcheon, J. A., & Liauw, J. (2023). Improving the external validity of antenatal late preterm steroids trial findings. *Paediatric and Perinatal Epidemiology*, 37(1), 1–8. <https://doi.org/10.1111/ppe.12856>
- Ijzerman, H., Lewis, N. A., Przybylski, A. K., Weinstein, N., DeBruine, L., Ritchie, S. J., Vazire, S., Forscher, P. S., Morey, R. D., Ivory, J. D., & Anvari, F. (2020). Use caution when applying behavioural science to policy. *Nature Human Behaviour*, 4(11), 1092–1094. <https://doi.org/10.1038/s41562-020-00990-w>
- Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(1), 243–263. <https://doi.org/10.1111/rssb.12027>
- Vo, T.-T., Le, T. T. K., Afach, S., & Vansteelandt, S. (2025). Integration of aggregated data in causally interpretable meta-analysis by inverse weighting. arXiv preprint arXiv: 2503.05634. <https://doi.org/10.48550/arXiv.2503.05634>
- Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.
- Lao, P., Young, C. B., Ezech, C., Lacayo, B., Seblova, D., Andrews, R. M., Gibbons, L., Kraal, A. Z., Turney, I., Deters, K. D., Dotson, V., Manly, J. J., Barnes, L. L., & Zahodne, L. B. (2024). Loneliness, cerebrovascular and alzheimer's disease pathology, and cognition. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 20(10), 7113–7123. <https://doi.org/10.1002/alz.14196>
- Le, T. T. K., Afach, S., & Vo, T.-T. (2024). A propensity score weighting approach to integrate aggregated data in random-effect individual-level data meta-analysis. arXiv preprint arXiv:2408.04854. <https://doi.org/10.48550/arXiv.2408.04854>
- Lesko, C. R., Buchanan, A. L., Westreich, D., Edwards, J. K., Hudgens, M. G., & Cole, S. R. (2017). Generalizing study results: A potential outcomes perspective. *Epidemiology (Cambridge, Mass.)*, 28(4), 553–561. <https://doi.org/10.1097/EDE.0000000000000664>
- Loh, W. W., & Ren, D. (2022). Estimating social influence in a social network using potential outcomes. *Psychological Methods*, 27(5), 841–855. <https://doi.org/10.1037/met0000356>
- Loh, W. W., & Ren, D. (2023). The unfulfilled promise of longitudinal designs for causal inference. *Collabra: Psychology*, 9(1), 89142. <https://doi.org/10.1525/collabra.89142>
- Loh, W. W., & Ren, D. (2025a). From experiments to policy insights: Generalizing causal effects from study samples to target populations. *Advances in Methods and Practices in Psychological Science*, 8(3), 25152459251359623. <https://doi.org/10.1177/25152459251359623>
- Loh, W. W., & Ren, D. (2025b). Upscaling behavioral interventions requires addressing selection bias. *Proceedings of the National Academy of Sciences of the United States of America*, 122(36), e2513887122. <https://doi.org/10.1073/pnas.2513887122>
- Mumford, S. L., & Schisterman, E. F. (2019). New methods for generalizability and transportability: The new norm. *European Journal of Epidemiology*, 34(8), 723–724. <https://doi.org/10.1007/s10654-019-00532-3>
- Pearl, J. (2010). On the consistency rule in causal inference: Axiom, definition, assumption, or theorem? *Epidemiology (Cambridge, Mass.)*, 21(6), 872–875. <https://doi.org/10.1097/EDE.0b013e3181f5d3fd>
- Pearl, J., & Bareinboim, E. (2014). External validity: From do-calculus to transportability across populations. *Statistical Science*, 29(4), 579–595. <https://doi.org/10.1214/14-STS486>
- Poppe, L., Steen, J., Loh, W. W., Crombez, G., De Block, F., Jacobs, N., Tennant, P. W. G., Cauwenberg, J. V., & De Paepe, A. L. (2025). How to develop. *Health Psychology Review*, 19(1), 45–65. <https://doi.org/10.1080/17437199.2024.2402809>
- Power, M. C., Engelman, B. C., Wei, J., & Glymour, M. M. (2022). Closing the gap between observational research and randomized controlled trials for prevention of alzheimer disease and dementia. *Epidemiologic Reviews*, 44(1), 17–28. <https://doi.org/10.1093/epirev/mxac002>
- Primbs, M. A., Bijlstra, G., Holland, R. W., & Thoemmes, F. (2025). Causal inference for dummies: A tutorial on directed acyclic graphs and balancing weights. *Social*

- Cognition*, 43(3), 217–237. <https://doi.org/10.1521/soco.2025.43.3.217>
- R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of homo sapiens: Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences*, 115(45), 11401–11405. <https://doi.org/10.1073/pnas.1721165115>
- Ren, D., & Loh, W. W. (2024). Advancing group-based disparities research and beyond: A cautionary note on selection bias. *Advances in Methods and Practices in Psychological Science*, 7(3), 25152459241260256. <https://doi.org/10.1177/25152459241260256>
- Rosenbaum, P. R. (2010). *Design of observational studies*. Springer-Verlag. <https://doi.org/10.1007/978-1-4419-1213-8>
- Rubin, D. B. (1986). Comment: Which ifs have causal answers. *Journal of the American Statistical Association*, 81(396), 961–962. <https://doi.org/10.1080/01621459.1986.10478355>
- Rubin, D. B. (1990). [On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.] Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies. *Statistical Science*, 5(4), 472–480. <https://doi.org/10.1214/ss/1177012032>
- Saul, B. C., & Hudgens, M. G. (2020). The calculus of M-estimation in R with geex. *Journal of Statistical Software*, 92(2), 1–15. <https://doi.org/10.18637/jss.v092.i02>
- Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, 51(3), 515–530. <https://doi.org/10.1037/0022-3514.51.3.515>
- Signorovitch, J. E., Sikirica, V., Erder, M. H., Xie, J., Lu, M., Hodgkins, P. S., Betts, K. A., & Wu, E. Q. (2012). Matching-adjusted indirect comparisons: A new tool for timely comparative effectiveness research. *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research*, 15(6), 940–947. <https://doi.org/10.1016/j.jval.2012.05.004>
- Sinclair, A. H., Cosme, D., Lydic, K., Reinero, D. A., Carreras-Tartak, J., Mann, M. E., & Falk, E. B. (2025). Behavioral interventions motivate action to address climate change. *Proceedings of the National Academy of Sciences of the United States of America*, 122(20), e2426768122. <https://doi.org/10.1073/pnas.2426768122>
- Splawa-Neyman, J., Dabrowska, D. M., & Speed, T. P. (1990). On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science*, 5(4), 465–472. <https://doi.org/10.1214/ss/1177012031>
- Stefanski, L. A., & Boos, D. D. (2002). The calculus of m-estimation. *The American Statistician*, 56(1), 29–38. <https://doi.org/10.1198/000313002753631330>
- Thalmayer, A. G., Toscanelli, C., & Arnett, J. J. (2021). The neglected 95% revisited: Is American psychology becoming less American? *The American Psychologist*, 76(1), 116–129. <https://doi.org/10.1037/amp0000622>
- Thoemmes, F., & Ong, A. D. (2016). A primer on inverse probability of treatment weighting and marginal structural models. *Emerging Adulthood*, 4(1), 40–59. <https://doi.org/10.1177/2167696815621645>
- Todorova, B., Steyerl, D., Hornsey, M. J., Pearson, S., Brick, C., Lange, F., Van Bavel, J. J., Vlasceanu, M., Lamm, C., & Doell, K. C. (2025). Machine learning identifies key individual and nation-level factors predicting climate-relevant beliefs and behaviors. *Npj Climate Action*, 4(1), 46. <https://doi.org/10.1038/s44168-025-00251-4>
- Vansteelandt, S., Goetghebuer, E., Kenward, M., & Molenberghs, G. (2006). Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statistica sinica*, 16(3), 953–979.
- Vazire, S. (2024). The next chapter for psychological science. *Psychological Science*, 35(7), 703–707. <https://doi.org/10.1177/09567976231221558>

Appendix

Unbiasedness of estimating equations

We formally demonstrate the unbiasedness of the estimating equations. For each treatment level $A = a$, the estimation equation is:

$$\begin{aligned}
 & E \left[S1(A = a)(Y - \theta^a) \frac{1}{\Pr(S = 1|W; \beta)} \right] \\
 &= E \left[S1(A = a)(Y^a - \theta^a) \frac{1}{\Pr(S = 1|W; \beta)} \right] \\
 &= E \left[E \left[S1(A = a)(Y^a - \theta^a) \frac{1}{\Pr(S = 1|W; \beta)} \middle| W \right] \right] \\
 &= E \left[E[1(A = a)(Y^a - \theta^a) | W, S = 1] \frac{E(S|W)}{\Pr(S = 1|W; \beta)} \right] \\
 &= E[E[1(A = a)(Y^a - \theta^a) | W, S = 1]] \\
 &= E[\Pr(A = a | W, S = 1)E(Y^a - \theta^a | W, S = 1)] \\
 &= E[\Pr(A = a | S = 1)E(Y^a - \theta^a | W)] \\
 &= \Pr(A = a | S = 1)E(Y^a - \theta^a) \\
 &= 0.
 \end{aligned}$$

The first equality follows from (A3); the second and third equalities follow from the law of iterated expectations; the fourth equality follows from correctly specifying the logistic model so that $E(S|W) = \Pr(S = 1|W; \beta)$; the fifth equality follows from (A4) a; the sixth equality follows from (A4)b and (A1); the penultimate equality follows from the law of iterated expectations; and the final equality follows from the definition of θ^a . For the estimating equation to be well-defined, (A2) must be met. The first two estimating equations in (3) correspond to $a = 0$ and $a = 1$, respectively. The remaining estimating equations in (3) for the logistic model coefficients are:

$$\begin{aligned}
 E \left[\frac{SW}{\Pr(S = 1|W; \beta)} - \mu \right] &= E \left[\frac{W}{\Pr(S = 1|W; \beta)} E(S|W) - \mu \right] \\
 &= E(W - \mu) = 0.
 \end{aligned}$$

The first equality follows from the law of iterated expectations; the second equality follows from correctly specifying the logistic model so that $E(S|W) = \Pr(S = 1|W; \beta)$; and the final equality follows from the definition of μ .