Original Article

Sociological Methodology
2024, Vol. 54(1) 92–117

DOI: 10.1177/00811750231183711
http://sm.sagepub.com
Sage

Choosing an Optimal Method for Causal Decomposition Analysis with Continuous Outcomes: A Review and Simulation Study

Soojin Park[1], Suyeon Kang[1], and Chioun Lee[1]

Abstract

Causal decomposition analysis is among the rapidly growing number of tools for identifying factors ("mediators") that contribute to disparities in outcomes between social groups. An example of such mediators is college completion, which explains later health disparities between Black women and White men. The goal is to quantify how much a disparity would be reduced (or remain) if we hypothetically intervened to set the mediator distribution equal across social groups. Despite increasing interest in estimating disparity reduction and the disparity that remains, various estimation procedures are not straightforward, and researchers have scant guidance for choosing an optimal method. In this article, the authors evaluate the performance in terms of bias, variance, and coverage of three approaches that use different modeling strategies: (1) regression-based methods that impose restrictive modeling assumptions (e.g., linearity) and (2) weighting-based and (3) imputation-based methods that rely on the observed distribution of variables. The authors find a trade-off between the modeling assumptions required in the method and its performance. In terms of performance, regression-based methods operate best as long as the restrictive assumption of linearity is met. Methods relying on mediator models without imposing any modeling assumptions are sensitive to the ratio of the group-mediator association to the mediator-outcome association. These results highlight the importance of selecting an appropriate estimation procedure considering the data at hand.

Keywords

disparity reduction, disparity remaining, performance evaluation, estimators, sensitivity

A key objective of decomposition analysis is to identify risks or resources ("mediators") that contribute to disparities between groups of individuals defined by social characteristics, such as race, ethnicity, gender, class, and sexual orientation. Examples of such mediators include incarceration, which explains the racial earnings gap among men (Western and Pettit 2005); socioeconomic status (SES), which explains the cardiovascular health (CVH) gap across race-gender groups (Lee, Park, and Boylan 2021); and the opportunity to learn, which explains the math achievement gap across racial/ethnic groups (Schmidt, Guo, and Houang 2021). The key to this

[1]University of California, Riverside, Riverside, CA, USA

Corresponding Author:
Soojin Park, University of California, Riverside, 1215 Sproul Hall, 900 University Avenue, Riverside, CA 92521, USA
Email: soojinp@ucr.edu

approach is to single out contributing factors that could play a role in reducing such disparities across social groups.

Several studies have situated decomposition analysis within the counterfactual framework of causal inference. VanderWeele and Robinson (2014) first advanced the idea of focusing on observed social disparities rather than the causal effects of social groups such as race and gender. Jackson and VanderWeele (2018) further developed the approach and proposed various definitions of disparity reduction and disparity remaining using an interventional perspective (Nguyen, Schmid, and Stuart 2021). In this article, we use one of their definitions that quantifies the extent to which the observed disparity would be reduced or remain if we hypothetically intervened to set the mediator distributions equal between social groups among individuals with similar demographic backgrounds.

Another counterfactual approach to mediation is causal mediation analysis on the basis of natural direct and indirect effects (Pearl 2001; Robins 2003). We include a detailed discussion regarding the similarities and differences between these methods in the section "Relations to Causal Mediation Analysis," but one notable difference is that causal decomposition analysis requires fewer assumptions than causal mediation analysis. One critical assumption required in causal mediation analysis, but not in causal decomposition analysis, is "no intermediate confounding." Given that a myriad of factors contributes to social disparities in health outcomes, intermediate confounding (effects of social groups confounding the mediator-outcome relationship) is likely to occur in disparities research, so causal decomposition analysis has a substantial advantage over causal mediation analysis.

Despite this merit, allowing intermediate confounders in causal decomposition analysis adds a modeling burden, because the identification of disparity reduction and remaining depends on the conditional probability of intermediate confounders in addition to that of the mediator and outcome. To reduce the modeling burden, estimation methods for causal decomposition analysis use different strategies. In this study, we focus on six methods: two regression methods based on the difference-in-coefficients and the product-of-coefficients estimators (Jackson and VanderWeele 2018), two weighting methods based on ratio of mediator probability weighting (RMPW) and inverse odds ratio weighting (IORW) estimators (Jackson 2021), and two imputation methods based on the single-mediator imputation estimator (Lundberg 2022; Sudharsanan and Bijlsma 2021) and the multiple-mediator imputation estimator (Park, Qin, and Lee 2022). The two regression methods impose restrictive modeling assumptions such as linearity. The weighting and imputation methods rely on the observed distribution of variables instead of imposing a restrictive modeling assumption. Specifically, the weighting methods rely on the observed distribution of the outcome; the single-mediator imputation method relies on the observed distribution of the intermediate confounders; and the multiple-mediator imputation method relies on the observed distribution of the mediator. However, the consequences, in terms of performance, of relying on the modeling assumptions or observed distributions are poorly understood, particularly when combined with varying data conditions critical to mediation settings.

Therefore, the goal of this study is to review the modeling strategies of each method and assess their performance in various data conditions. Comparing the performance of methods that use different strategies to reduce the modeling burden is a focal point of our simulation study, which makes it distinct from other simulation studies that compare performance across traditional mediation methods (e.g., MacKinnon et al. 2002). To achieve this goal and for simplicity, this review and simulation study focuses on continuous outcomes.

To empirically ground our investigation, we examine disparities in CVH across race-gender groups using the Midlife Development in the United States (MIDUS) study. Specifically, we focus on the research question "To what extent would the CVH disparity between Black women and White men be reduced if the college completion rate was equal between the groups?" and we illustrate estimation methods in the context of this example.

CAUSAL DECOMPOSITION ANALYSIS: REVIEW

In this section, we review causal decomposition analysis and then discuss several issues that emerge when one uses traditional mediation analysis or causal mediation analysis to study health disparities across socially defined groups.

Using the motivating example, we consider the setting in which the groups are Black women ($R=1$) and White men ($R=0$), the mediator is college completion status (M), and the outcome is CVH (Y). Because the mediator is not randomized, the relationship between the mediator and the outcome could be confounded by many life-course factors. On the basis of previous literature (Suglia et al. 2018; Winkleby et al. 1992), we identified age (C_1), genetic vulnerability (C_2), childhood SES (X_1), and childhood abuse (X_2) as confounders in the mediator-outcome relationship. Among these confounders, we need to further distinguish baseline covariates ($\boldsymbol{C}=(C_1,C_2)$) from other confounders ($\boldsymbol{X}=(X_1,X_2)$). Baseline covariates characterize demographics through which the differences in the mediator or outcome are considered equitable (Jackson 2021); in our example, these are age and genetic vulnerability measured by parental history of cardiovascular and metabolic illness (heart problems, stroke, and diabetes). The rest of the variables are effects of social groups confounding the mediator-outcome relationship, which are early-life adversity (i.e., childhood SES and abuse) in the example. We refer to childhood SES and abuse as the intermediate confounders.

Initial Disparity

We are interested in the CVH disparity between Black women and White men, controlling for age and genetic vulnerability. Formally, the initial disparity between Black women and White men is defined as $\tau_c(1,0)\equiv E[Y|R=1,\boldsymbol{c}]-E[Y|R=0,\boldsymbol{c}]$, where $\boldsymbol{c}\in\mathcal{C}$. Note that the defined initial disparity is simply the observed mean difference in an outcome between social groups given baseline covariates. Causal decomposition analysis does not attempt to estimate the causal effect of social groups, thus avoiding the issue of assigning counterfactual outcomes to nonmanipulable factors such as race and gender (Lundberg 2022; VanderWeele and Robinson 2014). In line with this approach, we also focus on the observed disparity in CVH between Black women and White men. The observed disparity is conditional on a specific age and genetic vulnerability because the differences in CVH through these background characteristics are considered equitable.

Disparity Reduction and Remaining

Once we observe the disparity between Black women and White men, we would also want to identify how to reduce the disparity, for example, by increasing Black women's college completion rate to that of White men. We thus equalize the college completion rate between the groups *among those with the same covariate level*, because we consider potential differences in the college completion rate that arise through age (e.g., cohort differences) or genetic vulnerability as unrelated to disparities in the college completion rate across gender and racial groups. Then, the disparity reduction is defined as, conditional on baseline covariates, the difference between the average CVH of Black women and their counterfactual CVH after setting their college completion rate equal to that of White men among those with the same baseline covariate level. Formally, $\delta_c(1) \equiv E[Y|R=1,\boldsymbol{c}] - E[Y(G_{m|\boldsymbol{c}}(0))|R=1,\boldsymbol{c}]$, where $\boldsymbol{c} \in \mathcal{C}$ and $G_{m|\boldsymbol{c}}(0)$ is a random value drawn from the distribution of White men's college completion status given baseline covariates. Disparity remaining is defined as, conditional on baseline covariates, the difference between the average CVH of White men and the average counterfactual CVH after equalizing the college completion rate between the groups among those with the same baseline covariate level. Formally, $\zeta_c(0) \equiv E[Y(G_{m|\boldsymbol{c}}(0))|R=1,\boldsymbol{c}] - E[Y|R=0,\boldsymbol{c}]$. The initial disparity can be obtained by summing disparity reduction and remaining as $\tau_c(1,0) = \delta_c(1) + \zeta_c(0)$.

Interpretation of these conditional estimands would apply to a specific level of baseline covariates $\boldsymbol{C} = \boldsymbol{c}$. For example, the disparity reduction and remaining defined above are among individuals who have the same values of age and genetic vulnerability. Hence, it is critical to choose the values of the baseline covariates to be set when estimating these conditional estimands. However, some researchers may be more interested in marginal decomposition, in which the effects are averaged over distributions of baseline covariates. Formally, marginal disparity reduction is defined as $\delta(1) \equiv E[Y|R=1,\boldsymbol{c}]P(\boldsymbol{c}) - E[Y(G_{m|\boldsymbol{c}}(0))|R=1,\boldsymbol{c}]P(\boldsymbol{c})$; marginal disparity remaining is defined as $\zeta(0) \equiv E[Y(G_{m|\boldsymbol{c}}(0))|R=1,\boldsymbol{c}]P(\boldsymbol{c}) - E[Y|R=0,\boldsymbol{c}]P(\boldsymbol{c})$. Interpretation of these marginal estimands would pertain to a population across different levels of baseline covariates (for more details, see Jackson and VanderWeele 2018; Lundberg 2022). Marginal disparities can be obtained for the weighting and imputation methods but not for the regression methods. Therefore, throughout this article, we focus on disparities conditional on baseline covariates to ensure comparability across different estimation methods.

To give causal interpretations to disparity reduction and remaining, we need to make the following identification assumptions:

A1. Conditional Independence: $Y(m) \perp M | R = r, \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{C} = \boldsymbol{c}$ for $r \in \{0, 1\}$, all $m \in M, \boldsymbol{x} \in \mathcal{X}$, and all $\boldsymbol{c} \in \mathcal{C}$, where $Y(m)$ denotes the potential value of the outcome under $M = m$. There is no omitted confounding in the mediator-outcome relationship given the group status, intermediate confounders, and baseline covariates.

A2. Positivity: $0 < P(M = m | R = 1, \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{C} = \boldsymbol{c})$ for all $m \in M, \boldsymbol{x} \in \mathcal{X}$, and all $\boldsymbol{c} \in \mathcal{C}$. Positivity states a positive conditional probability among Black women ($R = 1$) of each observed value for the mediator given covariates. This implies that Black women should have a possibility of experiencing all levels of the mediator given covariates and intermediate confounders.

A3. Consistency: If $M_i = m$ then $Y_i = Y_i(m)$ for all $m \in \mathcal{M}$, where M_i and Y_i represent the mediator M and the outcome Y of the ith subject, respectively, and $Y_i(m)$ denotes the potential value of the outcome of the ith subject under $M_i = m$. The observed outcome under a particular exposure value is the same as the outcome after intervening to set the exposure to that value. This assumption would be violated, for example, if the CVH outcome for an individual is affected by another person's college completion status.

All these assumptions are strong, and whether the assumptions are met or not depends on a substantive example. Assessing the plausibility of the assumptions is essential but beyond the scope of this study. Given the assumptions, disparity reduction and remaining are nonparametrically identified as

$$\begin{aligned} \delta_c(1) &= E[Y|R=1, \boldsymbol{c}] - \sum_{\boldsymbol{x}, m} E[Y|R=1, \boldsymbol{x}, m, \boldsymbol{c}] P(\boldsymbol{x}|R=1, \boldsymbol{c}) P(m|R=0, \boldsymbol{c}) \text{ and} \\ \zeta_c(0) &= \sum_{\boldsymbol{x}, m} E[Y|R=1, \boldsymbol{x}, m, \boldsymbol{c}] P(\boldsymbol{x}|R=1, \boldsymbol{c}) P(m|R=0, \boldsymbol{c}) - E[Y|R=0, \boldsymbol{c}], \end{aligned} \tag{1}$$

where $\boldsymbol{x} \in \mathcal{X}, m \in \mathcal{M}$, and $\boldsymbol{c} \in \mathcal{C}$. We will review different estimation methods derived from this identification result in the section "Estimation Methods."

Relations to Traditional Mediation Analysis

Traditionally, decomposition analysis has been understood, formulated, and conducted within the linear framework on the basis of the difference-in-coefficients estimator (Freedman and Schatzkin 1992; Olkin and Finn 1995). For example, a seminal work by Fryer (2011) used this difference-in-coefficients approach to examine whether controlling for test scores reduced observed racial/ethnic disparities in wages, unemployment, incarceration, and health. In the context of our motivating example and on the basis of the traditional difference-in-coefficients estimator, disparity reduction is estimated by $\hat{\eta}_1 - \hat{\theta}_1$ and disparity remaining is estimated by $\hat{\theta}_1$ in the following linear regression models:

$$\begin{aligned} Y &= \eta_0 + \eta_1 R + \eta_2 X_1 + \eta_3 X_2 + \eta_4^T \boldsymbol{C} + e_3, \text{ and} \\ Y &= \theta_0 + \theta_1 R + \theta_2 X_1 + \theta_3 X_2 + \theta_4 M + \theta_5^T \boldsymbol{C} + e_4. \end{aligned} \tag{2}$$

The disparity reduction and remaining obtained from this traditional approach differs from those defined at the beginning of this section. Specifically, if identification assumptions A1, A2, and A3 are met, $\hat{\eta}_1 - \hat{\theta}_1$ is disparity reduction after controlling for college completion status (M) within levels of childhood SES and abuse (X_1 and X_2) given baseline covariates. For proof, see proposition 2 from Jackson and VanderWeele (2018). This traditional estimator is meaningful if investigators are interested in controlling for differential college completion status between the groups that cannot be attributed to childhood SES and abuse. However, Jackson and VanderWeele (2018) argued that this estimate within the same levels of intermediate confounders (childhood SES and abuse) may not be desirable, because reducing disparity for children who have the same childhood SES is suboptimal in equity perspectives. For example, if we only consider individuals with a high level of childhood SES, the disparity reduction after equalizing college completion rate between the groups is likely underestimated compared with people across all levels of childhood SES. In contrast, the defined quantity described earlier estimates disparity reduction across all levels of intermediate confounders.

Another widely used approach in decomposition is the Kitakawa-Oaxaca-Blinder (KOB) decomposition (Kitagawa 1955; Oaxaca 1973; Blinder 1973). The KOB is often used to decompose social disparities in an outcome to the explained (by the fact that groups have different means for the mediator) and unexplained portions. Hou (2014) extended the KOB decomposition to address intermediate confounders using a regression-based mediation analysis framework. His article shows the equivalence between product-of-coefficients and difference-in-coefficients estimators when no exposure-mediator interaction exists.

The explained and unexplained portions defined in the KOB decomposition correspond to disparity reduction and disparity remaining, respectively. The only difference is that the typical KOB decomposes a marginal disparity where it is not conditional on baseline covariates **C**. Jackson and VanderWeele (2018) pointed out that this difference is relatively minor but has important implications regarding causal interpretations if assumption A1 (conditional independence) is met after conditioning on baseline covariates. After conditioning on baseline covariates, causal decomposition analysis would be a causal implementation of the KOB decomposition (Jackson and VanderWeele 2018).

Relations to Causal Mediation Analysis

One popular counterfactual approach to mediation is causal mediation analysis on the basis of natural direct and indirect effects (Pearl 2001; Robins 2003). Bauer and Scheim (2019) adopted this approach and applied VanderWeele's (2013) three-way decomposition method for disparities research. However, in prior literature (e.g., Jackson and VanderWeele 2019; Lundberg 2022; Park et al. 2022) it is argued that causal decomposition analysis is preferred over causal mediation analysis when studying contributing factors to disparities for the following three reasons.

First, causal decomposition analysis adopts the framework of a descriptive disparity, focusing on estimating the causal effects of manipulable factors rather than social groups. In our example, we are interested in the causal effect of college completion in reducing a CVH disparity but are agnostic about the causal effect of social groups on CVH. This framework circumvents the issue of assigning counterfactual outcomes to nonmanipulable factors such as race and gender. In contrast, causal mediation analysis is often applied to settings in which a manipulated treatment affects an outcome. Hence, it focuses on estimating the causal effects of the treatment as well as the mediator.

Second, causal decomposition analysis is based on interventional effects (Didelez, Dawid, and Geneletti 2012), which provide a straightforward interpretation of direct and indirect effects defined in disparities research. If causal mediation analysis was applied to our example, natural indirect effects would compare each Black woman's CVH with the potential CVH outcome of each Black woman after setting their mediator (college completion status) to a value that would have naturally resulted had she been born a White man. Considering this potential outcome is somewhat strange because a Black woman cannot be reborn as a different race-gender status and experience the mediator. In contrast, disparity reduction computes the difference between the average Black woman's CVH and the average counterfactual outcome of Black women after hypothetically intervening to equalize the college completion rate between groups. Compared with natural indirect effects, disparity reduction is more straightforward to interpret.

Third, identifying disparity reduction requires a weaker assumption than natural indirect effects. Identifying natural indirect effects requires no omitted confounding in the (1) exposure-outcome, (2) exposure-mediator, and (3) mediator-outcome relationships. In identifying disparity reduction, no counterfactual outcome is assigned to social groups, so we do not need to assume there is no omitted confounding in the exposure-outcome and exposure-mediator relationships. Most importantly, natural indirect effects require an additional assumption, that is, no intermediate confounding (Pearl 2009) or no interaction in the group-mediator relationship at the individual level (Robins 2003). Each assumption is restrictive and unrealistic, and neither assumption is met in our example. Early-life adversity (childhood SES and abuse) can affect the risk for dropping out of college and CVH, and the effect of college completion on CVH might vary by social group. In contrast, disparity reduction on the basis of interventional indirect effects requires neither assumption. That is, observed intermediate confounders are allowed, and interaction in the group-mediator relationship is allowed.

However, estimating disparity reduction and remaining in causal decomposition analysis is challenging because of these added intermediate confounders. As shown in equation (1), the identification result for disparity reduction and remaining depends on the conditional probability of intermediate confounders given the group status and baseline covariates. This dependence implies that, unless we either make restrictive modeling assumptions or rely on the observed distribution of variables, then intermediate confounders need to be modeled to estimate disparity reduction and remaining.

ESTIMATION METHODS

The following section details the estimation procedure of each method and how each method addresses the modeling burden of intermediate confounders. To estimate disparity reduction and remaining conditional on $\boldsymbol{C}=\boldsymbol{c}$, we center each continuous covariate in $\boldsymbol{C}$ at a prespecified value (e.g., mean value) and set a reference group to be zero for each categorical covariate in $\boldsymbol{C}$ when fitting regression models. For the same purpose, we omit $\frac{P(r)}{P(r|\boldsymbol{c})}$ for weighting and imputation methods. Readers can refer to the original articles that we review for marginal effects.

Regression-Based Approaches

Difference-in-Coefficients Method. The estimation procedure requires modeling the following four successive outcome models regressed on group status and baseline covariates, and additionally intermediate confounding (childhood SES X_1 and abuse X_2), and finally the mediator (college completion status M) as

$$Y=\phi_0+\phi_1 R+\phi_2^T\boldsymbol{C}+e_1, \tag{3}$$

$$Y=\gamma_0+\gamma_1 R+\gamma_2 X_1+\gamma_3^T\boldsymbol{C}+e_2, \tag{4}$$

$$Y=\eta_0+\eta_1 R+\eta_2 X_1+\eta_3 X_2+\eta_4^T\boldsymbol{C}+e_3, \tag{5}$$

$$Y=\theta_0+\theta_1 R+\theta_2 X_1+\theta_3 X_2+\theta_4 M+\theta_5^T\boldsymbol{C}+e_4, \tag{6}$$

where ϕ_1 represents the CVH disparity between Black women and White men given baseline covariates, γ_1 represents the disparity within levels of childhood SES given baseline covariates, η_1 represents the disparity within levels of childhood SES and abuse given baseline covariates, and θ_1 represents the disparity within levels of the college completion rate, childhood SES, and abuse given baseline covariates.

Given equations (3) to (6), disparity reduction is estimated as $\hat{\delta}_c(1)=\hat{\eta}_1-\hat{\theta}_1+(\hat{\eta}_2-\hat{\theta}_2)\frac{\hat{\phi}_1-\hat{\gamma}_1}{\hat{\gamma}_2}+(\hat{\eta}_3-\hat{\theta}_3)\frac{(\hat{\gamma}_1-\hat{\eta}_1)+(1-\hat{\eta}_2/\hat{\gamma}_2)(\hat{\phi}_1-\hat{\gamma}_1)}{\hat{\eta}_3}$; disparity remaining is estimated as $\hat{\zeta}_c(0)=\hat{\theta}_1+\hat{\theta}_2\frac{\hat{\phi}_1-\hat{\gamma}_1}{\hat{\gamma}_2}+\hat{\theta}_3\frac{(\hat{\gamma}_1-\hat{\eta}_1)+(1-\hat{\eta}_2/\hat{\gamma}_2)(\hat{\phi}_1-\hat{\gamma}_1)}{\hat{\eta}_3}$, where $\hat{a}$ is the estimate of regression coefficient a in the equations above. Standard errors can be obtained by delta methods or bootstrap methods.

Note that $\hat{\eta}_1-\hat{\theta}_1$ is the disparity reduction estimate after equalizing the college completion rate within levels of childhood SES and abuse (X_1 and X_2) given baseline covariates. To obtain disparity reduction across all levels of childhood SES and abuse, we add $(\hat{\eta}_2-\hat{\theta}_2)\frac{\hat{\phi}_1-\hat{\gamma}_1}{\hat{\gamma}_2}+(\hat{\eta}_3-\hat{\theta}_3)\frac{(\hat{\gamma}_1-\hat{\eta}_1)+(1-\hat{\eta}_2/\hat{\gamma}_2)(\hat{\phi}_1-\hat{\gamma}_1)}{\hat{\eta}_3}$, which is the mediated effect of childhood SES and abuse scaled by the proportion of the mediated portion via the mediator. Likewise, we add $\hat{\theta}_2\frac{\hat{\phi}_1-\hat{\gamma}_1}{\hat{\gamma}_2}+\hat{\theta}_3\frac{(\hat{\gamma}_1-\hat{\eta}_1)+(1-\hat{\eta}_2/\hat{\gamma}_2)(\hat{\phi}_1-\hat{\gamma}_1)}{\hat{\eta}_3}$ to disparity remaining after equalizing the college completion rate within levels of childhood SES and abuse ($\hat{\theta}_1$). Because of this added term, disparity reduction and remaining estimators differ from the traditional difference-in-coefficients estimator (Freedman and Schatzkin

1992; Olkin and Finn 1995). We refer to this method as the *difference-in-coefficients* estimator to differentiate it from the other type of regression-based method we introduce later. The regression-based approach is generally efficient in terms of standard errors and is straightforward to use. However, this difference-in-coefficients estimator relies on the restrictive modeling assumption of no nonlinear relationships at the individual level (e.g., group-mediator interactions) as reflected in equations (3) to (6).

Product-of-Coefficients Method. A product-of-coefficients approach is obtained by posing a model for the causal KOB decomposition discussed by Jackson and VanderWeele (2018). There are different ways to pose a model to estimate disparity reduction and disparity remaining. For example, Jackson and VanderWeele require modeling intermediate confounders in addition to the mediator and the outcome (see page 17 of their appendix).

Here, we implement the causal KOB decomposition by only fitting the mediator and outcome models, which has an advantage in terms of reducing the modeling burden. For illustration, we assume the mediator is continuous (education) and we will show that the method can address a discrete mediator (college completion status). The estimation procedure requires modeling the following mediator and outcome models as

$$\begin{aligned} M &= \alpha_0 + \alpha_1 R + \alpha_2^T \boldsymbol{C} + e_m, \\ Y &= \beta_0 + \beta_1 R + \beta_2 X_1 + \beta_3 X_2 + \beta_4 M + \beta_5 RM + \beta_6^T \boldsymbol{C} + e_y \end{aligned} \tag{7}$$

where RM represents the group-mediator interaction. Here, α_1 is the average disparity in college completion status between Black women and White men given baseline covariates; β_4 and $\beta_4 + \beta_5$ are the effect of college completion on CVH for White men and Black women, respectively, given baseline covariates. Note that α_1 includes disparities in college completion status between the groups given baseline covariates, including differences attributable to childhood SES and abuse. Given equation (7), the disparity reduction is estimated as $\hat{\delta}_c(1) = \hat{\alpha}_1 \times (\hat{\beta}_4 + \hat{\beta}_5)$ and the disparity remaining is estimated as $\hat{\zeta}_c(0) = \hat{\phi}_1 - \hat{\alpha}_1 \times (\hat{\beta}_4 + \hat{\beta}_5)$, where $\hat{\alpha}_1$, $\hat{\beta}_4$, and $\hat{\beta}_5$ are the estimates of the regression coefficients in the equations, and $\hat{\phi}_1$ is the estimate of initial disparity given covariates from equation (3). This estimator for disparity remaining uses the fact that disparity reduction and remaining add to the initial disparity ($\hat{\zeta}_c(0) = \hat{\tau}_c(1, 0) - \hat{\delta}_c(1)$). The standard errors for disparity reduction and remaining are obtained by delta methods or bootstraps.

Alternatively, as shown in Jackson and VanderWeele's (2018) appendix, we can estimate the disparity remaining by modeling intermediate confounders as $\hat{\zeta}_{\text{alt}}(0) = \hat{\beta}_1 + \hat{\beta}_2 \hat{\kappa}_1 + \hat{\beta}_3 \hat{\kappa}_2 + \hat{\beta}_5 \hat{\alpha}_0$, where $\hat{\kappa}_1$ and $\hat{\kappa}_2$ are the average disparity in childhood SES (X_1) and childhood abuse (X_2), respectively, between Black women and White men given the covariates.

This product-of-coefficients estimator allows a group-mediator interaction, and it can be easily modified to address binary mediators. For instance, a logistic/probit regression should be fitted for a binary mediator as $P(M = 1) = logit^{-1}\left(\alpha_0 + \alpha_1 R + \alpha_2^T \boldsymbol{C}\right)$. Then, disparity reduction is estimated as

$\widehat{\delta_c}(1) = \left\{ \frac{\exp\left(\widehat{\alpha_0} + \widehat{\alpha_1} + \widehat{\alpha_2^T} \boldsymbol{c}\right)}{1 + \exp\left(\widehat{\alpha_0} + \widehat{\alpha_1} + \widehat{\alpha_2^T} \boldsymbol{c}\right)} - \frac{\exp\left(\widehat{\alpha_0} + \widehat{\alpha_2^T} \boldsymbol{c}\right)}{1 + \exp\left(\widehat{\alpha_0} + \widehat{\alpha_2^T} \boldsymbol{c}\right)} \right\} \times \left(\widehat{\beta_4} + \widehat{\beta_5}\right)$ and disparity remaining is estimated as $\hat{\zeta}_c(0) = \hat{\phi}_1 - \hat{\delta}_c(1)$. However, the estimator cannot address nonlinear relationships other than the group-mediator interaction, for which either a weighting or imputation method should be considered.

Weighting-Based Approaches

Jackson (2021) proposed two weighting-based estimators on the basis of adaptation of the RMPW and IORW estimation. These estimators were originally developed in the causal mediation literature by Hong, Deutsch, and Hill (2015) and Tchetgen Tchetgen (2013), respectively.

RMPW. The RMPW estimator can be applied to a single discrete mediator. The following estimation procedure relies on two mediator models in which any linear and nonlinear relationships are allowed (steps 1 and 2) while using the observed distribution of the outcome (step 3).

1. Fit a mediator model, regressing college completion status on baseline covariates among White men ($R = 0$). On the basis of this fitted model, compute the predicted probability of M_i given $\boldsymbol{C}_i$ for each subject (i.e., $P(M_i|R_i = 0, \boldsymbol{C}_i)$).
2. Fit another mediator model, regressing college completion status on baseline covariates and the intermediate confounders (childhood SES and abuse) among Black women ($R = 1$). On the basis of this fitted model, compute the predicted probability of M_i given $\boldsymbol{X}_i$ and $\boldsymbol{C}_i$ for each subject (i.e., $P(M_i|R_i = 1, \boldsymbol{X}_i, \boldsymbol{C}_i)$).
3. Calculate the average CVH (Y) among Black women given $\boldsymbol{C} = \boldsymbol{c}$, weighted by the ratio of the two predicted probabilities as $W_i = \frac{P(M_i|R_i = 0, \boldsymbol{C}_i)}{P(M_i|R_i = 1, \boldsymbol{X}_i, \boldsymbol{C}_i)}$. This estimates the average counterfactual outcome of Black women $E[Y(G_{m|\boldsymbol{c}}(0))|R = 1, \boldsymbol{c}] = \frac{1}{n_1}\sum_{i \in \Pi_1} W_i Y_i(\boldsymbol{c})$, where Π_r indicates the subjects (of size n_r) in group $R = r$ for $r \in \{0, 1\}$, and $Y_i(\boldsymbol{c})$ denotes the observed outcome value of the i th subject given $\boldsymbol{C}_i = \boldsymbol{c}$. This quantity is obtained as the intercept in a weighted regression of Y on $\boldsymbol{C}$ among individuals with $R = 1$.
4. The disparity reduction is estimated as $\hat{\delta}_c(1) = \frac{1}{n_1}\sum_{i \in \Pi_1} Y_i(\boldsymbol{c}) - \frac{1}{n_1}\sum_{i \in \Pi_1} W_i Y_i(\boldsymbol{c})$ and disparity remaining is estimated as $\hat{\zeta}_c(0) = \frac{1}{n_1}\sum_{i \in \Pi_1} W_i Y_i(\boldsymbol{c}) - \frac{1}{n_0}\sum_{i \in \Pi_0} Y_i(\boldsymbol{c})$. Standard errors are obtained from bootstraps.

IORW. IORW can also be applied to a single discrete mediator. The estimation procedure relies on two mediator models (step 1) and four exposure models (steps 2 and 3) while using the observed distribution of the outcome. The procedure is similar to RMPW, so we briefly describe the estimation procedure here.

1. Fit two mediator models regressing college completion status on baseline covariates and intermediate confounders. On the basis of the two fitted models, compute the predicted probabilities of M_i as $P(M_i|\boldsymbol{C}_i)$ and $P(M_i|\boldsymbol{X}_i, \boldsymbol{C}_i)$.
2. Fit two exposure models regressing group status on baseline covariates and intermediate confounders. On the basis of the two fitted models, compute the predicted probabilities of being in a specific group as $P(R_i = 0|\boldsymbol{C}_i)$ and $P(R_i = 1|\boldsymbol{X}_i, \boldsymbol{C}_i)$.
3. Fit two exposure models regressing group status on college completion status and baseline covariates and intermediate confounders. On the basis of the two fitted models, compute the predicted probabilities of being in a specific group as $P(R_i = 0|M_i, \boldsymbol{C}_i)$ and $P(R_i = 1|M_i, \boldsymbol{X}_i, \boldsymbol{C}_i)$
4. The remaining steps are the same as with the RMPW estimator, except the weight is given as $W_i = \frac{\frac{P(R_i=0|M_i,\boldsymbol{C}_i)}{P(R_i=1|M_i,\boldsymbol{X}_i,\boldsymbol{C}_i)}}{\frac{P(R_i=0|\boldsymbol{C}_i)}{P(R_i=1|\boldsymbol{X}_i,\boldsymbol{C}_i)}} \times \frac{P(M_i|\boldsymbol{C}_i)}{P(M_i|\boldsymbol{X}_i,\boldsymbol{C}_i)}$.

One advantage of these weighting estimators is their flexibility to accommodate linear and nonlinear relationships, as the estimators do not change regardless of the fitted models. However, the disadvantages of these estimators include addressing discrete mediators only, as most weighting-based approaches do not work very well with continuous variables. Also, weighting-based approaches are generally less efficient in terms of standard errors compared with regression-based approaches (VanderWeele 2010).

Imputation-Based Approaches

Single-Mediator Imputation Method. Sudharsanan and Bijlsma (2021) and Lundberg (2022) proposed an estimator on the basis of the parametric *g*-formula (Robins 1986). Their algorithm predicts potential outcomes by randomly drawing values from mediators and outcomes from probability distributions. To address the uncertainty associated with this procedure, random draws for mediators and outcomes are conducted hundreds or even thousands of times. Combined with bootstrapping, the algorithm requires substantial computational power and time. Here, we extend this approach by using a predicted mediator value for continuous mediators (and a randomly drawn value from the mediator distribution for binary mediators), and we directly address the uncertainty associated with predicting a mediator value by bootstrapping rather than randomly drawing from mediator probability distributions multiple times. Although it is a minor difference, it substantially reduces computational power and time. The following estimation procedure relies on modeling a mediator and an outcome in which any linear and nonlinear relationships are allowed (steps 1 and 2) while using the observed distribution of the intermediate confounder (step 2).

1. Fit a mediator model, regressing the mediator (college completion status) on group status and baseline covariates. Using the coefficients from the fitted model, we compute the predicted value of the mediator for each subject (denoted as $\tilde{m}_i$), after forcing $R = 0$.
2. Fit an outcome model, regressing CVH on group status, intermediate confounder, mediator, and baseline covariates as $\mu_{rxm}(\boldsymbol{c}) \equiv E(Y_i|R_i = r, \boldsymbol{X}_i = \boldsymbol{x}, M_i = m, \boldsymbol{C}_i = \boldsymbol{c})$. On the

basis of the fitted model, compute a predicted outcome value for each subject after imputing $\tilde{m}_i$ as $\mu_{R_i \boldsymbol{X}_i \tilde{m}_i}(\boldsymbol{c})$. Note that the observed value of the intermediate confounder ($\boldsymbol{X}_i$) is used in computing the predicted outcome value $\mu_{R_i \boldsymbol{X}_i \tilde{m}_i}$.

3. The predicted outcome values obtained from step 2 will be averaged over i among Black women conditional on $\boldsymbol{C} = \boldsymbol{c}$. This computes the average counterfactual outcome of Black women $E[Y(G_{m|\boldsymbol{c}}(0))|R = 1, \boldsymbol{c}] = \frac{1}{n_1}\sum_{i \in \Pi_1} \mu_{R_i \boldsymbol{X}_i \tilde{m}_i}(\boldsymbol{c})$, where Π_r indicates the subjects (of size n_r) in group $R = r$.
4. The disparity reduction is estimated as $\hat{\delta}_c(1) = \frac{1}{n_1}\sum_{i \in \Pi_1} Y_i(\boldsymbol{c}) - \frac{1}{n_1}\sum_{i \in \Pi_1} \mu_{R_i \boldsymbol{X}_i \tilde{m}_i}(\boldsymbol{c})$ and disparity remaining is estimated as $\hat{\zeta}_c(0) = \frac{1}{n_1}\sum_{i \in \Pi_1} \mu_{R_i \boldsymbol{X}_i \tilde{m}_i}(\boldsymbol{c}) - \frac{1}{n_0}\sum_{i \in \Pi_0} Y_i(\boldsymbol{c})$. Standard errors can be obtained via bootstraps.

This single-mediator imputation estimator is flexible in addressing linear and nonlinear relationships as well as discrete and continuous mediators and outcomes.

Multiple-Mediator Imputation Method. Park et al. (2022) proposed the multiple-mediator imputation estimator by adopting the result in VanderWeele and Vansteelandt (2014), which was originally developed for causal mediation analysis. Park et al. (2022) developed this estimator to address the case of intervening on multiple mediators simultaneously, which is useful when the causal ordering of the mediators cannot be easily determined. Although the method is for multiple mediators, it can also address a single mediator. The following estimation procedure relies on modeling the intermediate confounders and the outcome in which any linear and nonlinear relationships are allowed (steps 1 and 2) while using the observed distribution of the mediator (step 2).

1. Fit a confounder model, regressing each intermediate confounder (childhood SES and abuse) on group status and baseline covariates. Using the coefficients from the fitted model, we compute a predicted value of each confounder for each subject (denoted as $\tilde{\mathbf{x}}_i$) among White men ($R = 0$), after forcing $R = 1$.
2. Fit an outcome model, regressing CVH on social groups, intermediate confounders, mediator, and baseline covariates as $\mu_{rxm}(\boldsymbol{c}) \equiv E(Y_i|R = r, \boldsymbol{X}_i = \boldsymbol{x}, M_i = m, \boldsymbol{C}_i = \boldsymbol{c})$. On the basis of the fitted outcome model, compute a predicted outcome value for each subject, after forcing $R = 1$, and imputing $\tilde{\boldsymbol{x}}_i$ as $\mu_{1\tilde{x}_i M_i}(\boldsymbol{c})$. Note that the observed mediator value (M_i) is used in computing the predicted outcome value $\mu_{1\tilde{x}_i M_i}(\boldsymbol{c})$.
3. The predicted outcome values obtained from step 2 will be averaged over i among White men given $\boldsymbol{C} = \boldsymbol{c}$. This computes the average counterfactual outcome of Black women $E[Y(G_{m|\boldsymbol{c}}(0))|R = 1, \boldsymbol{c}] = \frac{1}{n_0}\sum_{i \in \Pi_0} \mu_{1\tilde{x}_i M_i}(\boldsymbol{c})$, where Π_r indicates the subjects (of size n_r) in group $R = r$.
4. The disparity reduction is estimated as $\hat{\delta}_c(1) = \frac{1}{n_1}\sum_{i \in \Pi_1} Y_i(\boldsymbol{c}) - \frac{1}{n_0}\sum_{i \in \Pi_0} \mu_{1\tilde{x}_i M_i}(\boldsymbol{c})$ and disparity remaining is estimated as $\hat{\zeta}_c(0) = \frac{1}{n_0}\sum_{i \in \Pi_0} \mu_{1\tilde{x}_i M_i}(\mathbf{c}) - \frac{1}{n_0}\sum_{i \in \Pi_0} Y_i(\boldsymbol{c})$. Standard errors can be obtained via bootstraps.

This multiple-mediator imputation estimator is highly flexible because it can address (1) any nonlinear terms, (2) multiple mediators and a single mediator, and (3) different variable types of mediators and outcomes.

Table 1. Summary of Available Methods

		Type of Mediator		Number of Mediators		Nonlinear Terms		
Approach	Estimator	Categorical	Continuous	Single	Multiple	No	$R \times M$	Others
Regression	1	✓	✓	✓		✓		
	2	✓	✓	✓		✓	✓	
Weighting	3	✓		✓		✓	✓	✓
	4	✓		✓		✓	✓	✓
Imputation	5	✓	✓	✓		✓	✓	✓
	6	✓	✓	✓	✓	✓	✓	✓

Note: Estimators are as follows: 1, difference-in-coefficients; 2, product-of-coefficients; 3, ratio of mediator probability weighting; 4, inverse odds ratio weighting; 5, single-mediator imputation; and 6, multiple-mediator imputation. $R \times M$ = differential effects of mediators by groups.

However, depending on the causal structure of variables, there could be more burden in correctly specifying models than in the single-mediator imputation method. This estimator requires modeling intermediate confounders instead of mediators. From a modeling perspective, this estimator is advantageous only when the number of mediators exceeds or equals the number of intermediate confounders.

SIMULATION STUDY

Weighting- or imputation-based methods are generally more flexible than regression-based methods because no restrictive modeling assumptions are required. However, this flexibility comes at the cost of relying on the observed distribution of variables. We conducted a simulation study to assess the performance of the methods, either relying on the observed distribution of variables or imposing restrictive modeling assumptions with various data conditions to help researchers choose an optimal method given the data at hand. For simplicity, we refer to difference-in-coefficients, product-of-coefficients, RMPW, IORW, single-mediator imputation, and multiple-mediator imputation as estimators 1, 2, 3, 4, 5, and 6, respectively, in this section. Table 1 shows the summary of available estimation methods depending on conditions.

Data Generation

To generate synthetic data that mimics real data, we use the distribution of each variable in the MIDUS data used for the motivating example, which contains the group status R, the outcome Y, baseline covariate C, and intermediate confounder X. For simplicity, we use a single intermediate confounder and baseline covariate. Because those variables are related to each other, we also analyze the relationship between variables using the data with the following regression models:

$$Y = a_0 + a_1 R + a_2 C + U_x \tag{8}$$

$$M = b_0 + b_1 R + b_2 X + b_3 C + U_m \tag{9}$$

Table 2. Coefficient Values for Each Scenario and Corresponding Parameters

Mediator Type	Ratio (r)	Effect sizes		Coefficients			True Effects	
		$R-M$	$M-Y$	b_1	c_1	c_3	$\delta_c(1)$	$\zeta_c(0)$
Continuous	0.3	.138 (small)	–.681 (large)	.244	–1.048	–1.322	–.262	–.611
	0.5	.177 (small)	–.617 (large)	.326	–1.048	–1.112	–.263	–.611
	1	.247 (small)	–.535 (large)	.477	–1.048	–.900	–.263	–.611
	2	.338 (medium)	–.467 (medium)	.689	–1.049	–.750	–.263	–.612
	3	.402 (medium)	–.434 (medium)	.852	–1.049	–.684	–.263	–.612
Binary	0.1	.157 (small)	–.551 (large)	.552	–.669	–1.932	–.249	–.583
	0.3	.267 (small)	–.390 (medium)	1.032	–.644	–1.244	–.239	–.558
	0.5	.355 (medium)	–.340 (medium)	1.463	–.674	–1.063	–.252	–.588
	0.7	.430 (medium)	–.312 (medium)	1.906	–.693	–.964	–.259	–.607
	0.9	.504 (large)	–.292 (small)	2.466	–.709	–.900	–.268	–.623

Note: b_1, c_1, and c_3 are regression coefficients from equations (9) and (10). For effect sizes, we used partial correlations for the $M-Y$ and $R-M$ associations and Cohen's (2013) rules, where 0.1, 0.3, and 0.5 are considered small, medium, and large, respectively.

$$Y = c_0 + c_1 R + c_2 X + c_3 M + c_4 RM + c_5 C + U_y \quad (10)$$

Specifically, we create a binary treatment R that takes the value of 1 or 0, with the probability 0.5. For observations with $R=1$, C is generated from a truncated normal distribution with mean 50 and standard deviation 12 within the interval $(25, 75)$; for observations with $R=0$, C is generated from the same distribution but with a shifted mean of 48. We then dichotomize the generated C at 50 to take the value of 0 (the reference group) and 1. We create X, M, and Y using equations (8), (9), and (10), respectively, with added error terms from a standard normal distribution, instead of the error terms U_x, U_m and U_y from the original data. For a fair comparison across methods, we do not include the group-mediator interaction effect ($R \times M$) in equation (10) for estimator 1.

For the binary mediator case, we use the same procedure, but dichotomize M at its median to fit the logistic regression of equation (9). We then generate a binary mediator for the synthetic data that takes the value of 0 or 1 with the probability of $logit^{-1}(b_0 + b_1 R + b_2 X + b_3 C)$ for being $M=1$.

On the basis of the synthetic data, we compute the true average outcome values for $R=1$ and $R=0$ conditional on $C=0$, which correspond to $E[Y|R=1, C=0]$ and $E[Y|R=0, C=0]$, respectively. To generate the true value of $E[Y(G_{m|c}(0))|R=1, C=0]$, we compute the predicted outcome value on the basis of equation (10) given $R=1$, $C=0$, and $X=X_i$. For M, we use random draws from the distribution of M for $R=0$ given $C=0$. Table 2 shows the true values of disparity reduction and disparity remaining and the regression coefficient values used to generate the synthetic data for each condition. Because we do not include the interaction term between R and M for estimator 1, different coefficient values are used in data-generation (these values can be found in part A of the online supplement). We fix the percentage of disparity reduction at 30 percent across different settings to ensure comparability.

Simulation Setting

We consider three conditions critical to mediation settings: type of mediator, sample size, and the ratio between the $R-M$ and $M-Y$ associations. First, we use a binary and continuous mediator because the performance of the estimation methods may depend on the variable type of the mediator. For each type of mediator, we then consider three sample sizes $n=\{100, 500, 1000\}$, which cover reasonably small, medium, and large sample sizes in observational studies to which causal decomposition analysis mainly applies.

Last, an important condition that we vary for each fixed sample size is the ratio between the $R-M$ and $M-Y$ association. If the variables are standardized, the effect size ratio can be used to assess the performance of methods, as shown in traditional mediation literature (e.g., Kelcey et al. 2017; MacKinnon et al. 2002). However, in some cases, keeping the original scale with a meaningful metric (e.g., income in thousands of dollars) is essential because it is more intuitive to interpret. In this case, we argue it is more beneficial to use the ratio expressed in the original scale, which depends on both effect sizes and variances of the variables (i.e., the conditional variance of R, M, and Y after removing the components explained by predictors). This is because parameter estimation and statistical inference are affected by the variances in addition to effect sizes. Moreover, the ratio is particularly relevant in evaluating the performance of causal decomposition methods that use different strategies to reduce the modeling burden. For example, some methods rely on modeling the mediator (weighting or single-mediation-imputation), and others rely on the observed distribution of mediators (multiple-mediator imputation). Given that the disparity reduction estimate has two sources (the $R-M$ association and the $M-Y$ association), we want to investigate whether one approach is better than another when the $R-M$ association is particularly small relative to the $M-Y$ association (or vice versa).

The ratio is defined as $r=\alpha_1/(\beta_4+\beta_5)$ for a continuous mediator and $r=\{\text{logit}^{-1}(\alpha_0+\alpha_1+\alpha_2 E[C])-\text{logit}^{-1}(\alpha_0+\alpha_2 E[C])\}/(\beta_4+\beta_5)$ for a binary mediator; where α s and β s are from equation (7). We consider $r=\{0.3, 0.5, 1, 2, 3\}$ for a continuous mediator, which covers cases where the $R-M$ association is smaller than ($r=0.3, 0.5$), equal to ($r=1$), or larger than ($r=2, 3$) the $M-Y$ association. We consider $r=\{0.1, 0.3, 0.5, 0.7, 0.9\}$ for a binary mediator, which covers cases where the probability difference in $M=1$ between $R=1$ and $R=0$ is less than ($r=0.1, 0.3$), equal to ($r=0.5$), or greater than ($r=0.7, 0.9$) half the size of the $M-Y$ association. For binary mediators, the $R-M$ association is bounded between 0 and 1; hence, finding a set of coefficients for b_1, c_1, and c_3 that makes the ratio greater than 1 was implausible. Therefore, we limited the ratio for binary mediators to less than 1. The midpoint (ratio of 0.5) indicates when the probability difference of $M=1$ equals half the size of the effect of M on Y.

Thus, we consider 15 scenarios with different n and r values for each type of mediator. To set the desired level of the ratio r, we change the coefficients in equations (9) and (10). Table 2 shows how to set the coefficients for each scenario. Other than the three coefficient values in the table, the remaining coefficient values are fixed as the coefficients from the MIDUS data. We also present the effect sizes of the $R-M$ and

$M - Y$ associations as information. The variance-covariance matrix for all scenarios is shown in part B of the online supplement.

In this study, we used the following metrics to compare the performance of each estimation method: relative bias, the normalized root mean squared errors (nRMSEs), and 95 percent confidence interval coverage using the percentile bootstrap method (Efron 1982) with the number of bootstrap replicates of 1,000. The relative bias measures the difference between the average of the estimates and the true value relative to the true value. The nRMSE measures the square root of the average squared difference between the estimate and the true value relative to the true value. For each scenario, we make 1,000 replicates of the sample from the population, and the performances are averaged over the 1,000 repetitions. The coverage rate for the 95 percent confidence interval is defined as the proportion of replications where the true value is covered by the 95 percent confidence interval out of 1,000 replications.

Simulation Results

The simulation results for a continuous and binary mediator are summarized in Figures 1 and 2, respectively. In the figures, we present the relative bias (first row), nRMSE (second row), and 95 percenet confidence interval coverage (third row) for disparity reduction and remaining. Each column represents a different estimator. The x-axis represents the ratio and the y-axis represents the performance metrics. The gray lines indicate no bias, no root mean squared error, or the nominal level (95 percent). The exact numerical values of the performance metrics in each scenario can be found in part C of the online supplement.

Figure 1 (continuous mediator) shows the performance of estimators 1, 2, 5, and 6. Estimators 3 and 4 are not considered because they are only available for a binary mediator. Estimators 1, 2, and 6 perform well with a medium or large sample size ($n \geq 500$) regardless of ratios. With a small sample size ($n = 100$), estimators 1 and 2 perform slightly better than estimator 6 in terms of variance. The nRMSEs for the disparity reduction estimate obtained from estimator 6 are larger than estimators 1 and 2 with a sample size of 100 regardless of ratios.

In contrast, estimator 5 does not perform well in bias and coverage when the ratio is less than 1. The coverage rate of estimator 5 exceeds 0.98 with ratios less than 1 even with the sample size of 1,000, which implies that estimator 5 is inefficient in standard errors (here, shown as wide confidence intervals). In addition, the bias for a small sample size ($n = 100$) is substantial. For instance, with a ratio of 0.3, the relative bias is 0.494 (49.4 percent of the true value) for disparity reduction.

Figure 2 (binary mediator) shows that estimators 1, 2, and 6 for disparity reduction perform well in terms of bias, variance, and coverage with a medium or large sample size ($n \geq 500$) regardless of ratios. Even with a small sample size ($n = 100$), estimators 1 and 2 demonstrate small biases (less than 5 percent of the true value) and coverage rates close to the nominal level, regardless of ratios. In contrast, the bias is somewhat larger for estimator 6 with a small sample size. For example, the relative bias for estimator 6 reaches 0.168 (16.8 percent of the true value) with a ratio of 0.7.

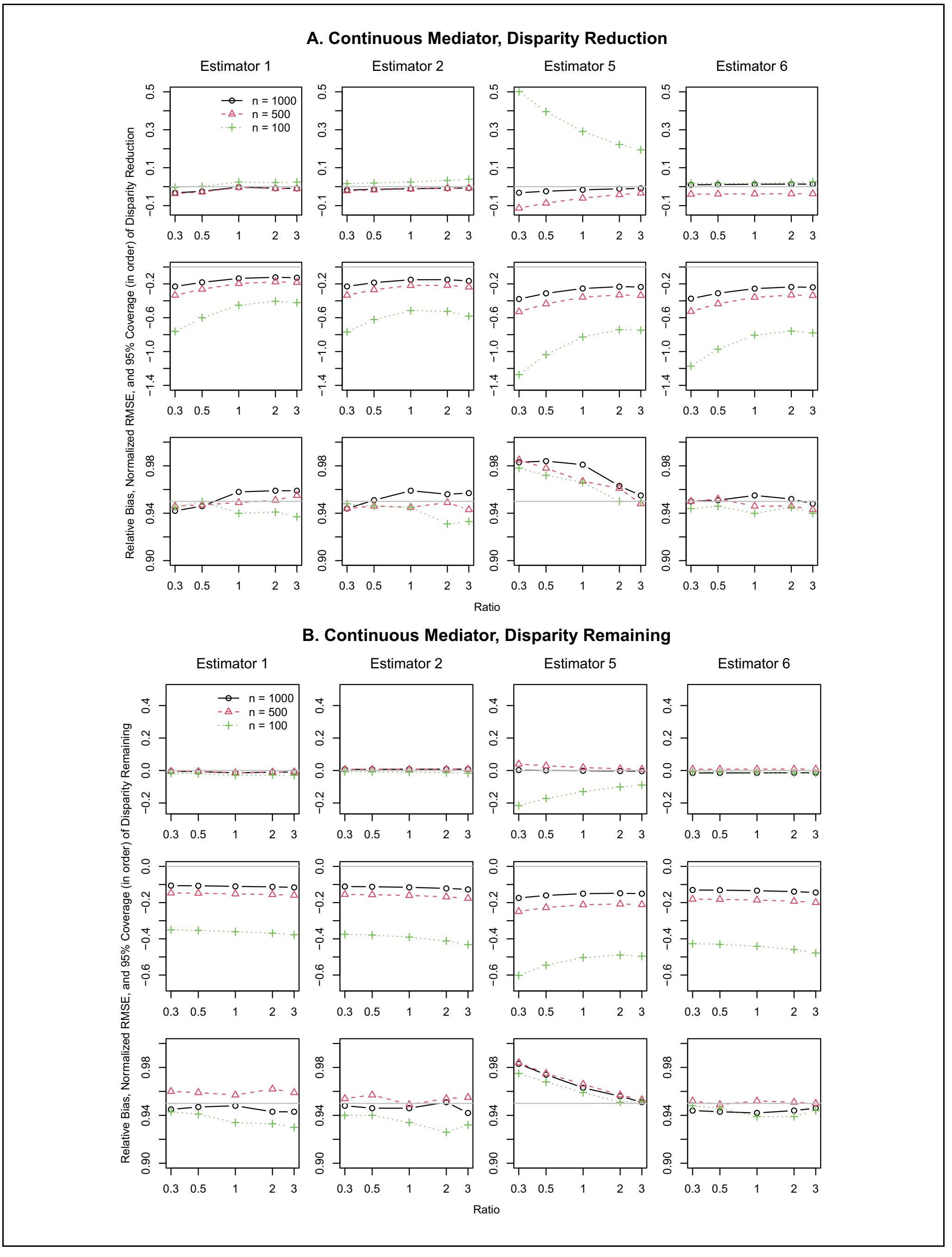

Figure 1. Performance of disparity reduction (A) and disparity remaining (B) with a continuous mediator.

Note: Estimators are as follows: 1, difference-in-coefficients estimator; 2, product-of-coefficients estimator; 5, single-mediator imputation estimator; and 6, multiple-mediator imputation estimator. Estimators 3 and 4 are not considered because they are only available for a binary mediator. In each panel, we used the log scale of ratio values such that the tick marks on the *x*-axis are equally spaced.

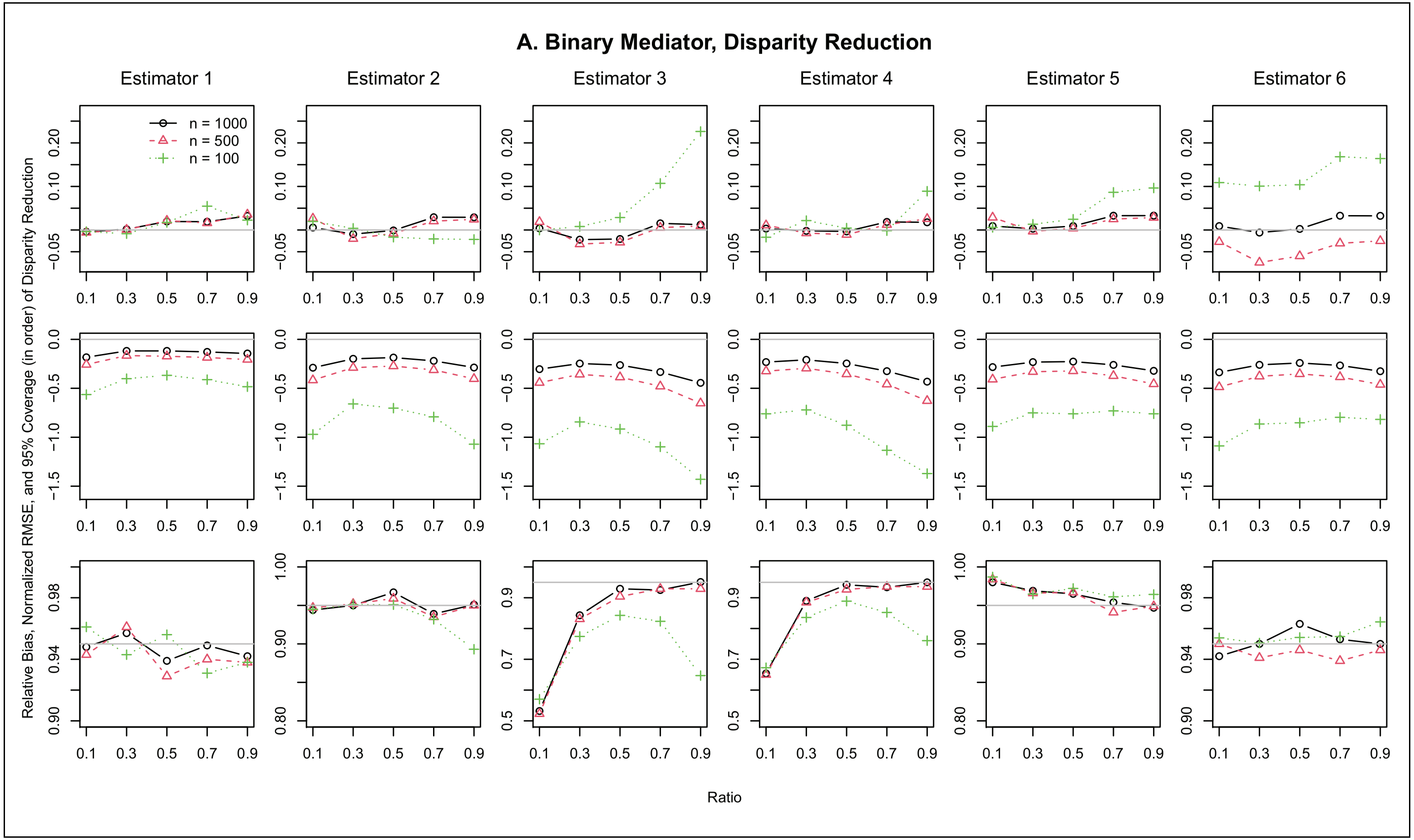

Figure 2. (continued)

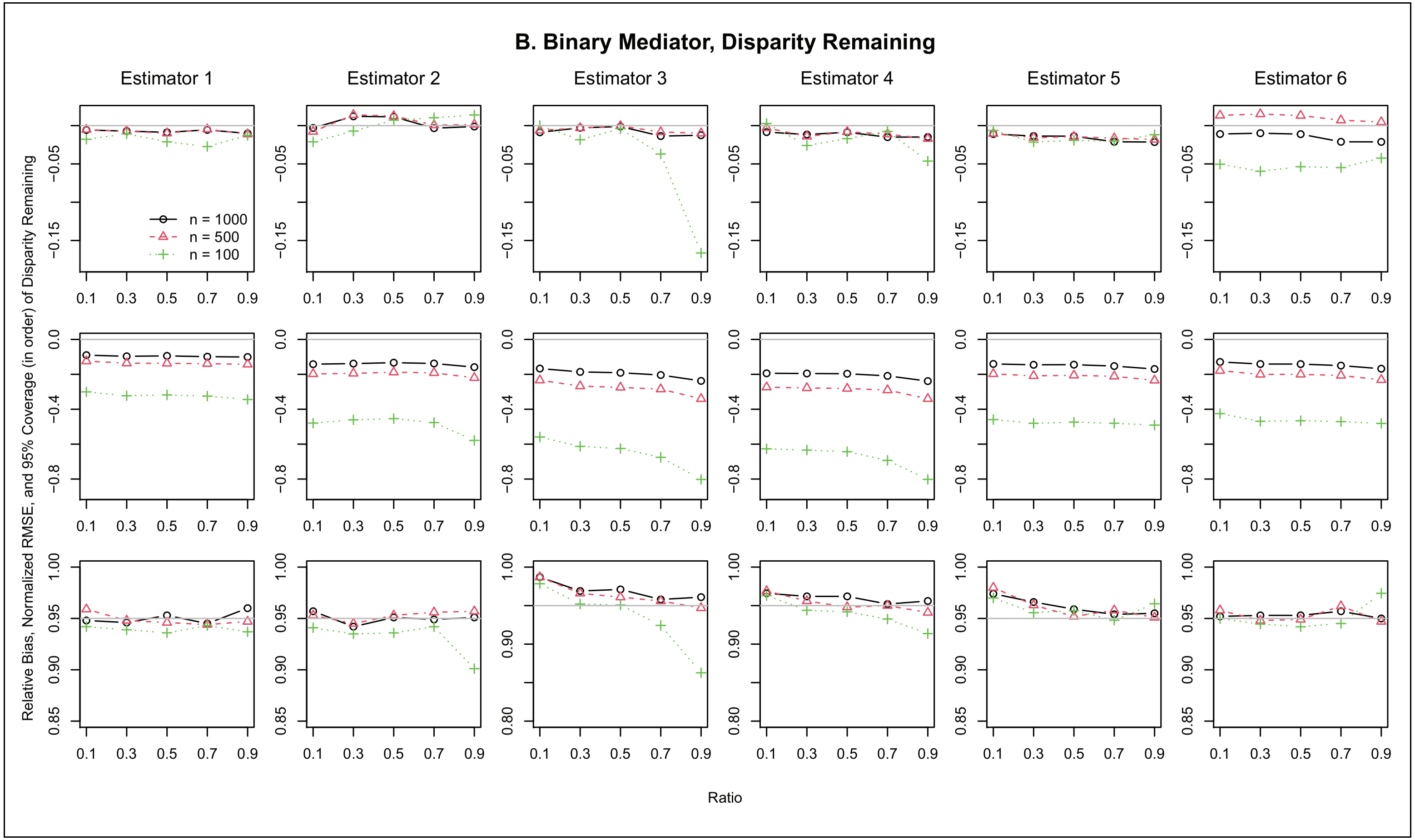

Figure 2. Performance of disparity reduction (A) and disparity remaining (B) with a binary mediator.
Note: Estimators are as follows: 1, difference-in-coefficients estimator; 2, product-of-coefficients estimator; 3, ratio of mediator probability weighting estimator; 4, inverse odds ratio weighting estimator; 5, single-mediator imputation estimator; and 6, multiple-mediator imputation estimator.

For estimators 3 and 4, we observe a low coverage rate with ratios less than 0.5. For example, with a sample size of 1,000 and a ratio of 0.1, the coverage rates of estimators 3 and 4 are only 0.53 and 0.65, respectively. These low coverage rates are due to narrow confidence intervals. Estimator 3 is advantageous in terms of modeling perspective because it only requires modeling two mediator models. However, the simulation result suggests this modeling advantage comes at the cost of low coverage when the ratio is small. We also observe a large bias (22.6 percent of the true value) and low coverage (0.64) for estimator 3 with a sample size of 100 and a ratio of 0.9.

For estimator 5, we observe a high coverage rate with ratios less than 0.5. For example, with a sample size of 1,000 and a ratio of 0.1, the coverage rate of estimator 5 is 0.98. These high coverage rates are due to the wide confidence intervals produced by the estimator, and this pattern remains consistent for continuous and binary mediators.

The pattern is similar with disparity remaining estimators, but there are two notable differences. First, estimator 4 has a better coverage rate for disparity remaining, achieving the nominal level across all sample sizes and ratios. Second, estimator 3 shows a high coverage rate for ratios less than 0.5. For example, with a sample size of 1,000 and a ratio of 0.1, estimator 3 has a coverage rate of 0.99. These high coverage rates are due to the wide confidence intervals produced by the estimator.

In addition to these 15 scenarios, we present another set of simulation studies under the same model specification but with standardized variables in part D of the online supplement. Although the relative bias and nRMSEs vary slightly, the low and high coverage issues of the weighting (estimators 3 and 4) and single-mediator imputation (estimator 5) methods persist, even after standardizing the variables.

In summary, we find a trade-off between the modeling assumptions required and performance in terms of bias, variance, and coverage. Methods that require a restrictive assumption perform best if the assumption is met (e.g., estimators 1 and 2). The performance of methods that do not require any restrictive modeling assumption but rely on modeling the mediator (estimators 3, 4, and 5) is sensitive to the ratio of the $R - M$ association to the $M - Y$ association. With these methods, ratios less than 1 for a continuous mediator could result in either a biased result or wider confidence intervals (estimator 5); ratios less than 0.5 for a binary mediator could result in narrower confidence intervals (estimators 3 and 4) or wider confidence intervals (estimator 5).

APPLICATION

Choosing between Methods

On the basis of our review of the methods and the simulation study, we provide recommendations for selecting an optimal method. We illustrate the practice of choosing an optimal method using the motivating example. Our research question is, to what extent would the CVH disparity be reduced if we increased the college completion rate of Black women to the level of White men among individuals with the same age and genetic vulnerability? The mediator is college completion status and the outcome is CVH, with higher values indicating better CVH (mean $= 8.09$, S.D. $= 2.12$). The mediator is binary and the outcome is continuous. We assume differential effects of college

Table 3. Estimates of the Disparity Reduction and Disparity Remaining for Black Women versus White Men

	Estimate				
Binary Mediator	Estimator 2	Estimator 3	Estimator 4	Estimator 5	Estimator 6
Initial disparity ($\tau_c(1,0)$)	–.913	–.965	–.965	–.965	–.965
(95% CI)	–1.236 to –.579	–1.250 to –.658	–1.250 to –.678	–1.262 to –.672	–1.258 to –.659
Disparity remaining ($\zeta_c(0)$)	–.62	–.563	–.563	–.615	–.623
(95% CI)	–.920 to –.331	–1.152 –.448	–1.069 to –.384	–.975 to –.253	–.944 to –.289
Disparity reduction ($\delta_c(1)$)	–.290	–.401	–.402	–.350	–.342
(95% CI)	–.481 to –.107	–.312 to –.037	–.376 to –.112	–.599 to –.107	–.593 to –.125
% reduction	31.8	41.6	41.7	36.3	35.6

Note: Estimators are as follows: 2, product-of-coefficients estimator; 3, ratio of mediator probability weighting estimator; 4, inverse odds ratio weighting estimator; 5, single-mediator imputation estimator; 6, multiple-mediator imputation estimator. Baseline covariates are mean-centered. CI = confidence interval.

completion exist between Black women and White men. Given the condition, the following methods are available: product-of-coefficients, RMPW, IORW, single-mediator imputation, and multiple-mediator imputation (2, 3, 4, 5, and 6 in Table 1).

Which method should be used among these multiple options? In our case, the sample size is 1,978, and the ratio is 0.319. Given the sample size and ratio, the simulation study suggests the product-of-coefficients and the multiple-mediator imputation methods should work well. If investigators are willing to assume no other nonlinear terms except for the group-mediator interaction, the product-of-coefficients method should be considered. If other nonlinear terms are modeled, the imputation method should be considered. The RMPW and IORW methods are also available options, but caution is required as the confidence interval for disparity reduction obtained from nonparametric bootstraps may be narrower than expected for ratios smaller than 0.5.

Summary of Findings From the Working Example

Table 3 shows estimates for disparity reduction and remaining obtained from different estimation methods. We begin by noting that the initial disparity for Black women compared with White men is $\tau(1, 0) = -0.965$, with the 95 percent confidence interval bounded away from zero, which means Black women's CVH is significantly worse (unhealthier) than White men at the confidence level of 95 percent among those who have the average level of age and genetic vulnerability. The initial disparity is slightly smaller for the regression-based method ($\tau(1, 0) = -0.913$). Once the disparity is observed, social scientists would also want to know how to reduce the disparity, for example, by increasing Black women's college completion rate to the level of White men.

The estimand $\delta_c(1)$ ranges from −0.290 (using estimator 2) to −0.402 (using estimator 4) and the confidence intervals (using all three estimators) do not cover zero. Given the assumptions, this means the CVH disparity between Black women and White men would be significantly reduced (by 31.8 percent to 41.7 percent) if we intervened to increase Black women's college completion rate to the level of White men. Note that, compared with the confidence interval of the disparity reduction for the product-of-coefficients method (−0.481, −0.107), the CIs from the weighting methods are narrower (i.e., −0.312, −0.037 for estimator 3; −0.376, −0.112 for estimator 4). This result is consistent with the simulation result with ratios of 0.3 or smaller.

In this example, the statistical uncertainty reflected in the 95 percent confidence interval is greater than the variability in the estimate across different methods. Moreover, the same conclusion is derived from different estimation methods. Yet it is important to note that a different conclusion could be derived depending on estimation methods, particularly when the sample size is small or when the ratio is even smaller than 0.319.

DISCUSSION

Estimation of disparity reduction and remaining is challenging because of the added burden of modeling intermediate confounders. Therefore, it is crucial to use an estimation method that reduces the modeling burden while maintaining good performance. Using both simulation and real-data examples, this article investigated the performance of six methods for estimating disparity reduction and remaining that use different strategies to reduce modeling burdens. We found that the methods imposing a restrictive modeling assumption perform best as long as the assumption is satisfied. For instance, with a continuous mediator, the regression-based estimators provide a precise estimate with the 95 percent coverage rate reaching the nominal level.

The other estimators use the observed distribution of variables. Of these, the weighting (RMPW and IORW) and single-mediator imputation estimators rely on modeling a mediator. The weighting methods for binary mediators perform poorly when the group-mediator association is smaller than half the size of the mediator-outcome association. A low coverage rate of the weighting estimators obtained from nonparametric bootstraps with ratios less than 0.5 is particularly worrisome as it could inflate the type 1 error rate. The single-mediator imputation estimator provides a high coverage rate when the ratio is less than 1 (continuous mediator) or 0.5 (binary mediator), which could inflate the type 2 error rate. In contrast, the multiple-mediator imputation estimator relies on modeling intermediate confounders, and thus the performance does not depend on the ratio. However, the performance of this estimator could be affected by the ratio for X if only a single intermediate confounder exists. This study highlights the need to carefully consider each method's modeling strategies and choose the most appropriate method for a given research project. Applied researchers may choose a method on the basis of availability or familiarity. However, it is crucial to select an optimal method on the basis of the data conditions (variable type of the mediator, number of intermediate confounders, sample size, ratio of the $R - M$ association to the $M - Y$ association) and whether the method's modeling assumptions are met. In the motivating example, we demonstrated how to choose an optimal method after examining data conditions and assessing the plausibility of modeling assumptions. In choosing an optimal method, we gave primary consideration to coverage rather than the bias or variance of the estimates. As shown in the example, the statistical uncertainty reflected in the 95 percent confidence intervals is greater than the range of estimates obtained from different estimation methods, which reflects the bias and variance associated with different methods. In contrast, a low or high coverage rate inflates type 1 and 2 errors, potentially leading to a wrong conclusion.

There are several limitations to our study that could drive future research. First, this study only addresses one way of defining disparity reduction and remaining. A different definition of disparity reduction and remaining exists (Jackson 2021; Jackson and VanderWeele 2018; Lundberg 2022), and the performance of estimation methods for different definitions is unknown. Therefore, the simulation study could be extended to an alternative definition of disparity reduction and remaining. Second, we used the ratio metric as an important condition in our simulation study, but the metric is useful only for continuous and binary mediators. Should categorical mediators with more than two discrete values be used, the metric must be redefined, and the performance of methods should be reexamined. Finally, the current study only addresses issues of estimating disparity reduction and remaining when the identification assumptions, such as no omitted confounding, are met. However, the assumptions are strong, and thus they may not be met in many empirical settings. Therefore, it is crucial to examine whether identification assumptions are met, as the bias due to violations of identification assumptions could be more extensive than that due to modeling assumptions.

Acknowledgments

We are thankful to the editor and the three anonymous reviewers for their valuable feedback in enhancing the quality of our article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by a grant from the American Educational Research Association, which receives funds for its AERA Grants Program from the National Science Foundation under award NSF-DRL #1749275. Opinions reflect those of the authors and do not necessarily reflect those the American Educational Research Association or the National Science Foundation.

ORCID iDs

Soojin Park https://orcid.org/0000-0003-0288-5589
Chioun Lee https://orcid.org/0000-0002-6886-8397

Data Accessibility Statement

Code to replicate all analyses can be found in the personal GitHub repository of the first author at https://github.com/soojinpark33/Optimal-Method-for-CDA.

Supplemental Material

Supplemental material for this article is available online.

References

Bauer, Greta R., and Ayden I. Scheim. 2019. "Methods for Analytic Intercategorical Intersectionality in Quantitative Research: Discrimination as a Mediator of Health Inequalities." *Social Science & Medicine* 226:236–45.

Blinder, Alan S. 1973. "Wage Discrimination: Reduced Form and Structural Estimates." *Journal of Human Resources* 8(4):436–55.

Cohen, Jacob. 2013. *Statistical Power Analysis for the Behavioral Sciences*. London: Taylor & Francis.

Didelez, Vanessa, Philip Dawid, and Sara Geneletti. 2012. "Direct and Indirect Effects of Sequential Treatments." arXiv. Retrieved June 16, 2023. https://arxiv.org/abs/1206.6840.

Efron, Bradley. 1982. *The Jackknife, the Bootstrap, and Other Resampling Plans*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Freedman, Laurence S., and Arthur Schatzkin. 1992. "Sample Size for Studying Intermediate Endpoints within Intervention Trials or Observational Studies." *American Journal of Epidemiology* 136(9): 1148–59.

Fryer, Roland G. 2011. “Racial Inequality in the 21st Century: The Declining Significance of Discrimination.” Pp. 855–971 in *Handbook of Labor Economics*, Vol. 4, edited by O. Ashenfelter and D. Card. New York: Elsevier.

Hong, Guanglei, Jonah Deutsch, and Heather D. Hill. 2015. “Ratio-of-Mediator-Probability Weighting for Causal Mediation Analysis in the Presence of Treatment-by-Mediator Interaction.” *Journal of Educational and Behavioral Statistics* 40(3):307–40.

Hou, Feng. 2014. “A General Approach to Effect Decomposition.” *Social Science Quarterly* 95(3): 894–904.

Jackson, John W. 2021. “Meaningful Causal Decompositions in Health Equity Research: Definition, Identification, and Estimation through a Weighting Framework.” *Epidemiology* 32(2):282–90.

Jackson, John W., and Tyler J. VanderWeele. 2018. “Decomposition Analysis to Identify Intervention Targets for Reducing Disparities.” *Epidemiology* 29(6):825–35.

Jackson, John W., and Tyler J. VanderWeele. 2019. “Intersectional Decomposition Analysis with Differential Exposure, Effects, and Construct.” *Social Science & Medicine* 226:254–59.

Kelcey, Ben, Nianbo Dong, Jessaca Spybrook, and Zuchao Shen. 2017. “Experimental Power for Indirect Effects in Group-Randomized Studies with Group-Level Mediators.” *Multivariate Behavioral Research* 52(6):699–719.

Kitagawa, Evelyn M. 1955. “Components of a Difference between Two Rates.” *Journal of the American Statistical Association* 50(272):1168–94.

Lee, Chioun, Soojin Park, and Jennifer M. Boylan. 2021. “Cardiovascular Health at the Intersection of Race and Gender: Identifying Life-Course Processes to Reduce Health Disparities.” *Journals of Gerontology: Series B* 76(6):1127–39.

Lundberg, Ian. 2022. “The Gap-Closing Estimand: A Causal Approach to Study Interventions That Close Disparities across Social Categories.” *Sociological Methods & Research*. doi:10.1177/00491241211055769.

MacKinnon, David P., Chondra M. Lockwood, Jeanne M. Hoffman, Stephen G. West, and Virgil Sheets. 2002. A Comparison of Methods to Test Mediation and Other Intervening Variable Effects. *Psychological Methods* 7(1):83–104.

Nguyen, Trang Quynh, Ian Schmid, and Elizabeth A. Stuart. 2021. “Clarifying Causal Mediation Analysis for the Applied Researcher: Defining Effects Based on What We Want to Learn.” *Psychological Methods* 26(2):255–71.

Oaxaca, Ronald. 1973. “Male-Female Wage Differentials in Urban Labor Markets.” *International Economic Review* 14(3):693–709. https://doi.org/10.2307/2525981

Olkin, Ingram, and Jeremy D. Finn. 1995. “Correlations Redux.” *Psychological Bulletin* 118(1):155–64.

Park, Soojin, Xu Qin, and Chioun Lee. 2022. “Estimation and Sensitivity Analysis for Causal Decomposition in Health Disparity Research.” *Sociological Methods & Research*. doi:10.1177/00491241211067516.

Pearl, Judea. 2001. “Direct and Indirect Effects.” Pp. 411–20 in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*. New York: Association for Computing Machinery.

Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge, UK: Cambridge University Press.

Robins, James. 1986. “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* 7(9):1393–1512.

Robins, James M. 2003. “Semantics of Causal DAG Models and the Identification of Direct and Indirect Effects.” Pp. 70–82 in *Highly Structured Stochastic Systems*, edited by P. J. Green, N. L. Hjort, and S. Richardson. Oxford, UK: Oxford University Press.

Schmidt, William H., Siwen Guo, and Richard T. Houang. 2021. “The Role of Opportunity to Learn in Ethnic Inequality in Mathematics.” *Journal of Curriculum Studies* 53(5):579–600.

Sudharsanan, Nikkil, and Maarten J. Bijlsma. 2021. "Educational Note: Causal Decomposition of Population Health Differences Using Monte Carlo Integration and the *g*-Formula." *International Journal of Epidemiology* 50(6):2098–2107.
Suglia, Shakira F., Karestan C. Koenen, Renée Boynton-Jarrett, Paul S. Chan, Cari J. Clark, Andrea Danese, Myles S. Faith, et al. 2018. "Childhood and Adolescent Adversity and Cardiometabolic Outcomes: A Scientific Statement from the American Heart Association." *Circulation* 137(5):e15–28.
Tchetgen Tchetgen, Eric J. 2013. "Inverse Odds Ratio-Weighted Estimation for Causal Mediation Analysis." *Statistics in Medicine* 32(26):4567–80.
VanderWeele, Tyler J. 2010. "Bias Formulas for Sensitivity Analysis for Direct and Indirect Effects." *Epidemiology* 21(4):540–51.
VanderWeele, Tyler J. 2013. "A Three-Way Decomposition of a Total Effect into Direct, Indirect, and Interactive Effects." *Epidemiology* 24(2):224–32.
VanderWeele, Tyler J., and Whitney R. Robinson. 2014. "On the Causal Interpretation of Race in Regressions Adjusting for Confounding and Mediating Variables." *Epidemiology* 25(4):473–84.
VanderWeele, Tyler, and Stijn Vansteelandt. 2014. "Mediation Analysis with Multiple Mediators." *Epidemiologic Methods* 2(1):95–115.
Western, Bruce, and Becky Pettit. 2005. "Black-White Wage Inequality, Employment Rates, and Incarceration." *American Journal of Sociology* 111(2):553–78.
Winkleby, Marilyn A., Darius E. Jatulis, Erica Frank, and Stephen P. Fortmann. 1992. "Socioeconomic Status and Health: How Education, Income, and Occupation Contribute to Risk Factors for Cardiovascular Disease." *American Journal of Public Health* 82(6):816–20.

Author Biographies

Soojin Park is an assistant professor in the School of Education at the University of California, Riverside. Her research focuses on developing and validating quantitative methods for causal inference, which she uses to investigate the factors contributing to racial and gendered disparities in educational and health outcomes.

Suyeon Kang is a PhD candidate in applied statistics at the University of California, Riverside. Her research interests include causal inference, multivariate analysis, mixture models, statistical computing, and robustification.

Chioun Lee is an assistant professor in the Department of Sociology at the University of California, Riverside. Her research focuses on the social determinants of health, with a particular focus on how gender and race/ethnicity interact with these determinants over the life-course. As a quantitative sociologist, she uses rigorous statistical methods to examine these complex relationships.