

Causal mediation and sensitivity analysis for mixed-scale data

Statistical Methods in Medical Research

2023, Vol. 32(7) 1249–1266

© The Author(s) 2023



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/09622802231173491

journals.sagepub.com/home/smm

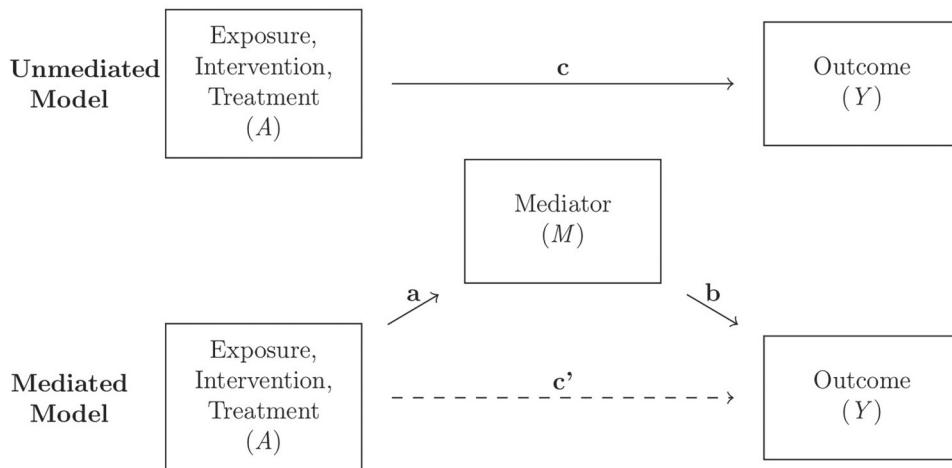


Figure 1. Schematic depiction of a causal structure with a single variable M mediating the effect of the treatment A on the outcome Y . Top: the causal structure ignoring the existence of the mediator. Bottom: the causal structure with the mediator included.

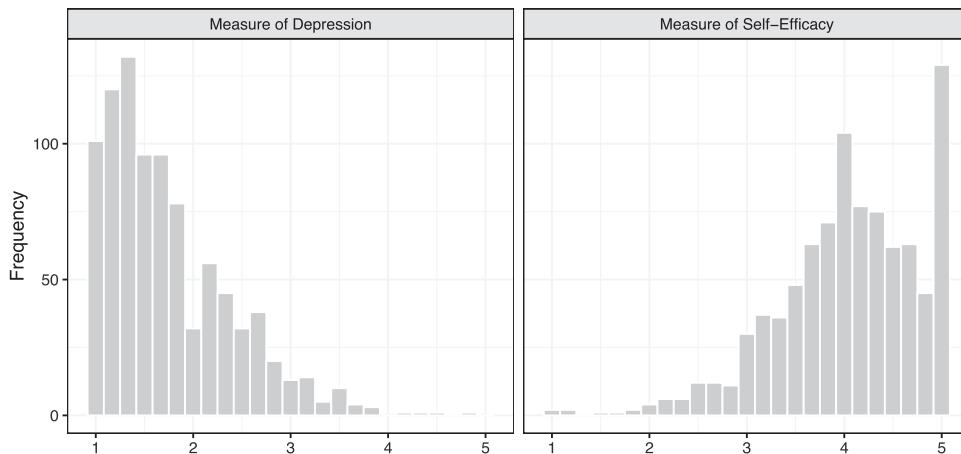


Figure 2. Empirical distribution of (left) measured depression level (right) measured job-search self-efficacy at the end of study.

mixed-scale data. In principle, this framework can be used regardless of the model for the observed data, and we use a zero-one inflated beta regression model to illustrate.

For the sake of reproducibility, we focus on the JOBS II study of Vinokur et al.,³ for which a subset of data is available in the mediation package in R. A description of this dataset is given in Section 2.1. Imai et al.¹ present several analyses of this dataset, essentially operating under the assumption that the mediator (a measure of self-efficacy in finding a job) and the outcome (a measure of depression) are normally distributed. As shown in Figure 2, however, it is apparent that neither the outcome nor the mediator is well-described by a normal distribution; both exhibit skewness and there is a substantial mass at the boundary values of 1 for depression and 5 for self-efficacy.

An additional challenge with mixed-scale models is assessing the sensitivity of inferences to untestable assumptions. As with most estimands in causal inference, it is well-known that the causal mediation effects are not identified on the basis of the observed data distribution alone, and can only be consistently estimated under additional (unfalsifiable) assumptions. A useful benchmark assumption is *sequential ignorability* (SI, Imai et al.¹), which essentially rules out the existence of unmeasured confounders. We found performing sensitivity analysis in the mixed-scale setting to be challenging, as to the best of our knowledge none of the existing proposals for sensitivity analysis can be applied directly. For example, the approaches proposed by Imai et al.¹ are justified by a linear structural equation model (LSEM, Baron and Kenny⁴), which is not applicable in this setting. Similarly, the limited work with non-continuous or categorical data^{5,6} also does not apply directly to the mixed-scale setting. We develop a pair of widely-applicable sensitivity analysis strategies that accomplish the two goals of (i) assessing the extent to which our conclusions are driven by unmeasured confounding and (ii) neither imposing any additional restrictions on, nor adding information about, the distribution of the observed data. Our second

goal is part of a recent trend in causal inference and missing data research of proposing sensitivity analyses that clearly and unambiguously separate the (parametric) assumptions used to model the observed data from the assumptions used to identify the causal effects of interest.^{8,9,7,10}

1.1 Review of existing methods

The traditional approach to mediation analysis uses structural equation modeling (SEM) to quantify mediation; linear structural equation models (LSEMs) are particularly popular.¹¹ However, LSEMs do not generalize easily to non-linear systems.¹² Additionally, Imai et al.¹ make the point that the identification assumptions used in LSEMs are inexorably tied to the choice of parametric model, stating: “[because] the key identification assumption is stated in the context of a particular model, [it is] difficult to separate the limitations of research design from those of the specified statistical model.” Motivated by this argument, Imai et al.¹ proposed a more general approach to mediation analysis using a potential outcomes framework, introduced the nonparametric assumption of sequential ignorability to identify the effects, and showed that the single mediator LSEM is a special case of the potential outcomes framework that is valid as long as the linearity assumption holds.

There is a rich literature addressing the causal mediation problem from the semiparametric perspective. An emphasis in this literature is the development of methods that are both statistically efficient and *multiply robust* in the sense that they produce consistent estimates even if one of several models required for estimation are misspecified.^{13,14} An advantage of these approaches is that one can easily use them with modern machine learning methods via cross-fitting.¹⁵ To the best of our knowledge, however, these methods have not been developed in the context of mixed-scale data. Bayesian nonparametric and semiparametric models based on infinite mixture models have also been proposed,¹⁶ although not for mixed-scale data.

A variety of models have also extended beyond continuous/binary models for the mediator and outcome. These include models for the zero-inflated count, survival, and ordinal data, as well as quantile regression models.^{6,17,18,1}

1.2 Contributions

We make the following contributions in this article. First, we describe how to implement the *g*-formula for computing both mean and quantile causal effects for generic mixed-scale models under the sequential ignorability assumption. Second, we illustrate these concepts using a zero-one inflated beta regression model and argue for its appropriateness on the benchmark JOBS II dataset. Third, we show how to conduct a principled sensitivity analysis to check the sensitivity of our conclusions to the untestable sequential ignorability assumption; the sensitivity parameters we introduce are designed to be unidentified, so that varying them does not affect the distribution of the observed data. We show how to introduce sensitivity parameters that are shifts of the mean on either a linear or logit scale; both of these approaches are very easy to incorporate into our models by post-processing the model fit. These mean-shift assumptions are weaker than the usual sequential ignorability assumption in that they only identify the mean of the potential outcomes rather than their whole distribution, but include the results of sequential ignorability as a special case.

We also present a flexible zero-one inflated beta (ZOIB) model¹⁹ for mediation analysis with boundary-censored data and show how to perform inference with this model. The ZOIB models we use here are conceptually related to other zero-inflated models, such as the zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB), zero-inflated gamma (ZIG), and zero-inflated log-normal (ZILN). See Zuur et al.²⁰ for a review of zero-inflated models for count data and Liu et al.²¹ for semicontinuous zero-inflated models. Like the ZIG and ZILN (and unlike the ZIP and ZINB) models, the ZOIB model is semicontinuous, as the boundary points 0 and 1 occur with probability 0 in a standard beta regression model. These models are cast as a covariate-dependent mixture model, with the ZOIB being a mixture of a beta distribution and a Bernoulli distribution (where the parameters of the beta and Bernoulli distributions themselves are also covariate dependent).

A Bayesian implementation of this model is given at www.github.com/theodds/ZOIBMediation. Our model is implemented in Stan, with both the average and quantile causal mediation effects computed using a Monte Carlo implementation of the *g*-formula. The Bayesian framework provides a straightforward approach to incorporating uncertainty in the sensitivity parameters through the use of informative priors, which can be elicited from subject-matter experts.²² In principle, however, one could also apply Frequentist inference using the nonparametric bootstrap.²³

1.3 Outline

In Section 2 we review the potential outcomes framework for mediation analysis and argue for the use of mixed-scale models on the JOBS II dataset. In Sections 3 and 4 we present our framework for causal mediation analysis, show how

to compute the mean and quantile mediation effects, and develop our zero-one inflated beta regression model. In Section 5, we present two alternative assumptions to sequential ignorability which allow for a sensitivity analysis, and show that these assumptions identify the average causal mediation effects. In Section 6, we illustrate our methodology on synthetic data and real data. We conclude in Section 7 with a discussion. Proofs are in the Appendix. Some additional algorithms, additional sensitivity analyses, and Markov chain Monte Carlo diagnostics are given in the Supplemental Material.

2 Notation and definitions of causal effects

2.1 JOBS II dataset

For the sake of reproducibility, we motivate concepts and illustrate our methods on a subset of the JOBS II dataset³ which is available publicly in the `mediation` package in R. The JOBS II data was used to evaluate the potential benefits of participation in a job-search skills seminar in southeastern Michigan. Subjects were recently unemployed adults during 1991–1993. Participants were pre-screened and classified according to their risk of depression and anxiety. High-risk participants, along with a random sample of low-risk participants, were invited to participate in the study. Prior to the seminar, questionnaires were sent out to the respondents. The questionnaires covered a range of topics about the respondent, including their loss of employment, the quality of work-life at their previous job, their health problems, and their history of substance abuse. The primary baseline covariates, which we denote as X_i , include measures of depression at baseline, education, income, race, marital status, age, sex, previous occupation, and level of economic hardship. The participants were randomly assigned to treatment and control groups. The treatment group, $A_i = 1$ was assigned to attend a seminar that taught participants job search skills and coping strategies for dealing with setbacks in the job hunt. The control group, $A_i = 0$, received a booklet of job search tips. Prior to measuring the outcome, but post-intervention, researchers measured an underlying mechanism in the relationship between the intervention and outcome. This *mediator* was a continuous measure of job search self-efficacy, M_i . In this study, two outcomes were measured: a continuous measure of depression using the Hopkins Symptom Checklist²⁴ and a binary variable for employment at the follow-up time. We will focus on the continuous measure of depression, Y_i .

Even for a benchmark dataset like the JOBS II data, which has been analyzed using LSEMs,¹ there is overwhelming evidence that neither the outcome nor the mediator are normally distributed. The observed values of the outcome and mediator, which are supported on [1, 5] and highly skewed, are displayed in Figure 2, and it is apparent from the figure that the assumption of (say) a normally distributed error is untenable.

2.2 The potential outcomes framework

Using the potential outcomes framework, the causal effect of the job training program can be defined as the difference between two potential outcomes. One potential outcome is realized if the subject participates in the training program and the other is realized if the subject does not.

Associated with the outcome and the mediator are *potential outcomes* that would have been observed had the treatment assignment has been different. A potential outcome is defined as the outcome which would have been observed under an exposure a participant did not actually receive. We let $M_i(a)$ denote the value of the mediator had the treatment has been assigned to the value a ; in terms of the JOBS II study, this is the self-efficacy which would have been realized had the treatment for individual i had been fixed at either receiving the treatment ($a = 1$) or not ($a = 0$). Similarly, we let $Y_i(a, m)$ denote the value of the outcome that would have been realized had the treatment for individual i been fixed at a and the mediator fixed at m ; in terms of the JOBS II study, this is the depression level which would have been observed at a given level of self-efficacy under the two treatments.

We link the potential outcomes to the observed data through the *consistency* assumption that $M_i = M_i(A_i)$ and $Y_i = Y_i\{A_i, M_i(A_i)\}$. The primary challenge in estimation of the mediation effects, that is, the effects of changes of M_i and A_i on Y_i , lies in the fact that we cannot observe the counterfactual outcomes $Y_i\{A_i, M_i(1 - A_i)\}$, as this would require observing what would have happened under both $A_i = 1$ and $A_i = 0$.

Throughout, we assume that the distribution of the potential outcomes, the mediator, and the exposure is distributed according to a distribution f_θ in some parametric family $\{f_\theta : \theta \in \Theta\}$. Abusing notation, we use f_θ as the conditional and marginal density/mass functions as needed, with its meaning inferred from context; for example, $f_\theta(A_i = a | X_i = x) = f_\theta(a | x)$ denotes the probability of $A_i = a$ given $X_i = x$.

2.2.1 Sequential ignorability

As a starting point for identifying the causal effects of interest we will use the *sequential ignorability* assumption of Imai et al.¹ For subject i , let \mathcal{X} be the support of the distribution of X_i and let \mathcal{M} be the support of M_i . The SI assumption imposes

the following restrictions on the model parameterized by an unknown θ :

$$\{Y_i(a', m), M_i(a)\} \perp A_i \mid X_i = x, \theta \quad \text{and} \quad (SI1)$$

$$Y_i(a', m) \perp M_i(a) \mid A_i = a, X_i = x, \theta \quad (SI2)$$

for all $a, a' = 0, 1$ and $x \in \mathcal{X}$, where the expression $[U \perp V \mid W = w]$ means that U is conditionally independent of V given $W = w$. Additionally, we require the overlap condition (SI3) that $\Pr_\theta(A_i = a \mid X_i = x) > 0$ and $f_\theta\{M_i(a) = m \mid A_i = a, X_i = x\} > 0$, for all $m \in \mathcal{M}$. In words, SI1 states that, given the observed confounders, the treatment assignment is independent of the potential outcomes $Y_i(a', m)$ and $M_i(a)$; this will hold whenever the treatment assignment is randomized. On the other hand, SI2 states that the assignment of the mediator does not affect the outcome, given the observed treatment and pre-treatment covariates. Of the two assumptions, SI2 is generally the more problematic; for example, in the JOBS II study, the job-search self-efficacy is not randomized by study design, so that SI2 makes the unfalsifiable assertion that all common causes of $M_i(a)$ and $Y_i(a, m)$ have been measured.

2.2.2 Causal mediation effects

We define the following causal mediation effects^{26,25} using the JOBS II study for context. The *natural indirect effect* (NIE), also called the *causal mediation effect*, is defined for $a = 0, 1$ as

$$\delta_i(a) = Y_i\{a, M_i(1)\} - Y_i\{a, M_i(0)\}$$

For example, in the JOBS II study, $\delta_i(0)$ is the effect of increasing/decreasing a subject's self-efficacy from their baseline level to the level we would have observed had they attended the seminar, holding fixed that the subject did not attend the seminar. The *natural direct effect* (NDE) is defined for $a = 0, 1$ as

$$\zeta_i(a) = Y_i\{1, M_i(a)\} - Y_i\{0, M_i(a)\}$$

For example, in the JOBS II study, $\zeta_i(1)$ is the difference between the two potential depression levels for subject i according to whether they participated in the job training seminar or not, under the assumption that their job search self-efficacy is held constant at the level which would have been observed if they had attended the seminar.

Because we cannot observe $Y_i\{a, M_i(a')\}$ when $a \neq a'$, we cannot directly observe either $\delta_i(a)$ or $\zeta_i(a)$. Nevertheless, under sequential ignorability we can estimate the *average mediation effects*

$$\begin{aligned} \delta(a) &= \mathbb{E}_\theta[Y_i\{a, M_i(1)\} - Y_i\{a, M_i(0)\}] \quad \text{and} \\ \zeta(a) &= \mathbb{E}_\theta[Y_i\{1, M_i(a)\} - Y_i\{0, M_i(a)\}] \end{aligned} \quad (1)$$

The mediation effects $\delta(a)$ and $\zeta(a)$ decompose the *average total effect* $\tau = \mathbb{E}_\theta[Y_i\{1, M_i(1)\} - Y_i\{0, M_i(0)\}]$ in the sense that $\delta(1) + \zeta(0) = \delta(0) + \zeta(1) = \tau$. The total effect is analogous to the usual average causal treatment effect (ATE) of the treatment assignment.

Under sequential ignorability, Imai et al.¹ showed that the distribution of the potential outcomes $Y_i\{a, M_i(a')\}$ for any a, a' is nonparametrically identified as

$$\begin{aligned} f_\theta(Y_i\{a, M_i(a')\} = y \mid X_i = x) \\ = \int_{\mathcal{M}} f(Y_i = y \mid M_i = m, A_i = a, X_i = x) f(M_i = m \mid A_i = a', X_i = x) dm \end{aligned} \quad (2)$$

for all $x \in \mathcal{X}$. The marginal distribution of $Y_i\{a, M_i(a')\}$ is then $\int f_\theta(Y_i\{a, M_i(a')\} = y \mid X_i = x) F_X(dx)$ so that it (along with the average direct and indirect effects) is also identified. While there usually will not be a simple analytical expression for (2), it is nevertheless easy to approximate (2) using Monte Carlo integration; this approach was popularized by Robins²⁷ as a tool to implement the *g*-formula in causal inference.

While the average causal mediation effects are the most commonly studied, one may also be interested in causal effects on other aspects of the distribution of the outcome. Let $Q_q(Z)$ denote the q^{th} quantile of a random variable Z . Then the

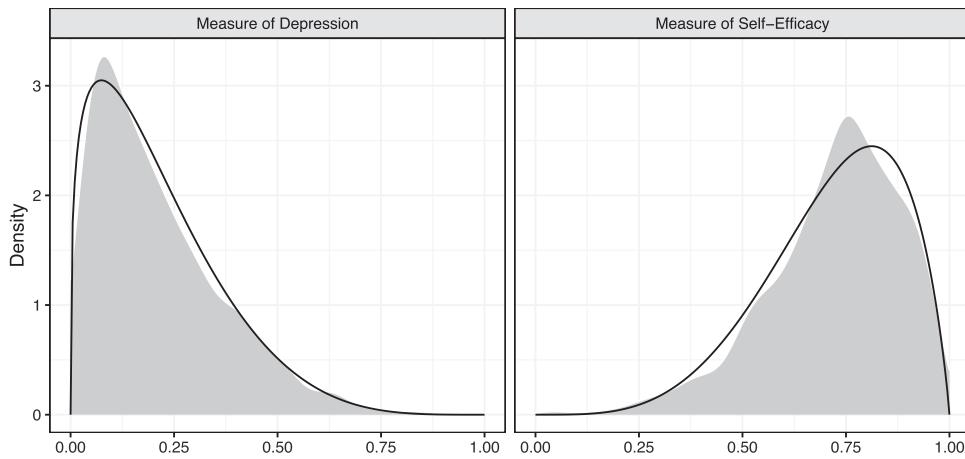


Figure 3. Kernel density estimate (gray) and fitted beta distribution (solid black) of the distribution of Y_i and M_i conditional on $Y_i \notin \{0, 1\}$ and $M_i \notin \{0, 1\}$.

quantile mediation effects at the quantile q are

$$\begin{aligned}\delta_q(a) &= Q_q[Y_i\{a, M_i(1)\}] - Q_q[Y_i\{a, M_i(0)\}] \quad \text{and} \\ \zeta_q(a) &= Q_q[Y_i\{1, M_i(a)\}] - Q_q[Y_i\{0, M_i(a)\}].\end{aligned}\tag{3}$$

Because (2) fully identifies the distribution of $Y_i\{a, M_i(a')\}$, the quantile mediation effects are also identified under SI. We note that, rather than the difference in the quantiles, one might be tempted to define $\delta_q(a) = Q_a[Y_i\{a, M_i(1)\} - Y_i\{a, M_i(0)\}]$ (and similarly for $\zeta_q(a)$), which represents the causal mediation effect as a quantile of the differences. There are two issues which arise from doing this. First, SI is not sufficient to identify $\delta_q(a)$, as SI does not identify the joint distribution of the potential outcomes. Second, we lose the decomposition property $\tau_q = \delta_q(a) + \zeta_q(a')$ so that the mediation effects no longer serve as a decomposition of the total effect. The terms $\delta_q(a)$ and $\zeta_q(a)$ as defined in (3) do have a compelling causal interpretation: rather than representing a causal effect on the individual level, they capture the causal effect on the quantiles from shifting the entire population from untreated to treated.

3 Observed data models for zero-one inflated data

Estimating the causal mediation effects under SI requires only that we estimate the distribution of the observed data. Without loss of generality, we assume that Y_i and M_i can be rescaled to lie in the interval $[0, 1]$; for the JOBS II dataset, this can be done with the transformations $Y_i \leftarrow (Y_i - 1)/4$ and $M_i \leftarrow (M_i - 1)/4$, as the measures of depression and self-efficacy were measured on a scale from 1 to 5.

A flexible distribution for zero-one inflated data on $[0, 1]$ is the *zero-one inflated beta* (ZOIB) distribution, which we denote by $Z \sim \text{ZOIB}(\alpha, \gamma, \mu, \phi)$. If $Z \sim \text{ZOIB}(\alpha, \gamma, \mu, \phi)$ then Z is a mixed discrete-continuous random variable such that

$$\begin{aligned}\Pr(Z = 0) &= \alpha \\ \Pr(Z = 1 \mid Z \neq 0) &= \gamma \quad \text{and} \\ [Z \mid Z \notin \{0, 1\}] &\sim \text{Beta}\{\mu\phi, (1 - \mu)\phi\}\end{aligned}\tag{4}$$

The parameterization of the beta distribution in (4) is chosen so that μ is the mean of the beta distribution, that is, $\mathbb{E}_\theta(Z \mid Z \notin \{0, 1\}) = \mu$. The mean of the $\text{ZOIB}(\alpha, \gamma, \mu, \phi)$ distribution is given by

$$\mathbb{E}_\theta(Z) = (1 - \alpha)\gamma + (1 - \alpha)(1 - \gamma)\mu\tag{5}$$

Figure 3 shows that the beta distribution is effective at modeling the shape of the data for the continuous part of the JOBS II data, while the parameters α and γ allow for an increased chance of observing the boundary values $Z_i = 1$ and $Z_i = 0$.

Our ZOIB model assumes that $[Y_i \mid M_i = m, A_i = a, X_i = x] \sim \text{ZOIB}(\alpha_i^Y, \gamma_i^Y, \mu_i^Y, \phi_i^Y)$ and $[M_i \mid A_i = a, X_i = x] \sim \text{ZOIB}(\alpha_i^M, \gamma_i^M, \mu_i^M, \phi_i^M)$. We model the parameters of these ZOIB distributions with generalized linear models of the form

$$\begin{aligned} \text{logit}(\alpha_i^Y) &= (X_i, M_i)^\top \beta_\alpha^Y(A_i), & \text{logit}(\gamma_i^Y) &= (X_i, M_i)^\top \beta_\gamma^Y(A_i) \\ \text{logit}(\mu_i^Y) &= (X_i, M_i)^\top \beta_\mu^Y(A_i), & \log(\phi_i^Y) &= (X_i, M_i)^\top \beta_\phi^Y(A_i) \end{aligned} \quad (6)$$

The dependence of the β 's on A_i is included to allow for heterogeneous effects of the covariates and mediator; the homogeneous model is included as a special case where only the intercept varies with A_i . Similar models are specified for $(\alpha_i^M, \gamma_i^M, \mu_i^M, \phi_i^M)$. As a default, all of the regression coefficients are given flat Normal($0, \tau^2$) priors, where τ^2 is taken to be large after centering and scaling the covariates X_i (except for the intercept) to have mean 0 and standard deviation 1.

To estimate the mediation effects we also require a model for the distribution F_X of the covariates. As a default, we assume that F_X is discretely supported on the observed values of the X_i 's, that is, $\Pr_\theta(X_i = x_j) = \omega_j$ where (x_1, \dots, x_N) are the observed values of the covariates. We then specify an improper *Bayesian bootstrap*²⁸ prior for $\omega = (\omega_1, \dots, \omega_N)$, that is, $\pi(\omega) = \prod_i \omega_i^{-1}$. After observing the data, the posterior distribution of ω is Dirichlet($1, \dots, 1$), and can be sampled exactly. Specifying a Bayesian bootstrap prior for F_X avoids the notoriously difficult task of estimating F_X via density estimation, and has been shown in other settings to result in improved theoretical properties of Bayesian causal inference methods.²⁹

4 Posterior computation and inference

We divide inference into two steps:

1. Draw a set of approximate samples $\theta_1, \dots, \theta_B$ from the posterior distribution of θ .
2. For each θ_b , compute $\delta(a)$, $\zeta(a)$, and τ , yielding approximate samples from the posterior distribution for these mediation effects.

For the first step, we use the probabilistic programming language *Stan*, which implements an adaptive version of Hamiltonian Monte Carlo (HMC) to sample θ_b 's.³⁰ The sole exception to this sampling scheme is that we sample $\omega \sim \text{Dirichlet}(1, \dots, 1)$ directly from the posterior distribution.

4.1 Average mediation effects

Due to the nonlinearities of the ZOIB model, the mediation effects $\delta(a)$, $\zeta(a)$, and τ are not available in closed form and must be approximated. To compute the mediation effects, we use a Monte Carlo implementation of the *g*-formula. The idea is to note that, because (2) identifies the distribution of $Y_i \{a, M_i^*(a')\}$ for all a, a' , we can simulate K new realizations $Y_i^* \{a, M_i^*(a')\}$ for $i = 1, \dots, N$ from the model, in which case

$$\begin{aligned} \delta(a) &\approx K^{-1} \sum_{i,k} \omega_i [Y_{ik}^* \{a, M_{ik}^*(1)\} - Y_{ik}^* \{a, M_{ik}^*(0)\}] & \text{and} \\ \zeta(a) &\approx K^{-1} \sum_{i,k} \omega_i [Y_i^* \{1, M_{ik}^*(a)\} - Y_i^* \{0, M_{ik}^*(a)\}] \end{aligned} \quad (7)$$

are unbiased estimators of $\delta(a)$ and $\zeta(a)$. The approximations in (7) are less efficient than using the true values $\delta(a)$ and $\zeta(a)$ because they contain Monte Carlo error, but are *conservative* in the sense that using them results in valid inference. The approximations can be improved by using various tricks to eliminate the Monte Carlo error. One improvement is to notice that we can decrease the variance of (7) by replacing $Y_{ik}^* \{a, M_{ik}^*(a')\}$ with the conditional expectation $\mathbb{E}_\theta[Y_{ik}^* \{a, M_{ik}^*(a')\} \mid M_{ik}^*(a'), X_i]$; for the ZOIB model, this is given by $(1 - \alpha_{ik}^Y)\gamma_{ik}^Y + (1 - \alpha_{ik}^Y)(1 - \gamma_{ik}^Y)\mu_{ik}^Y$ where α_{ik}^Y , γ_{ik}^Y and μ_{ik}^Y are given by (6) with A_i evaluated at a and M_i evaluated at $M_{ik}^*(a')$.

The Monte Carlo integration strategy is summarized in Algorithm 1, and it applies to any model. In Algorithm 2, we give the special case of our ZOIB regression models. In these algorithms, $F_M^-(u \mid A_i = a, X_i = x) = \inf \{m : F_M(m \mid A_i = a, X_i = x) \geq u\}$ denotes the *generalized inverse* of the cumulative distribution function of $[M_i \mid A_i = a, X_i = x]$.

4.2 Quantile mediation effects

Equation (2) can also be used to form a Monte Carlo estimate of the quantile mediation effects, although the implementation is somewhat more subtle. If we have a sample of $Y_{ik}^* \{a, M_i^*(a')\}$'s from the marginal density $f_\theta(Y_i \{a, M_i(a')\} = y)$ then we

Algorithm 1 Monte Carlo g -formula estimate of average causal effects.**Input:** $\theta, K, \{X_i\}_{i=1}^N$

- 1: Sample $\omega \sim \text{Dirichlet}(1, \dots, 1)$
- 2: **for** $k = 1, \dots, K$ and $i = 1, \dots, N$ **do**
- 3: Sample $U_{ik} \sim \text{Uniform}(0, 1)$
- 4: Set $M_{ik}^*(a) = F_M^-(U_{ik} \mid A = a, X = X_i, \theta)$ for $a = 0, 1$.
- 5: Set $Y_{ik}^* \{a, M_{ik}^*(a')\} = \mathbb{E}_\theta \{Y_i \mid M_i = M_{ik}^*(a'), A_i = a, X_i\}$ for $a, a' = 0, 1$.
- 6: **end for**
- 7: Approximate $\delta(a), \zeta(a), \tau$ with $\hat{\delta}(a), \hat{\zeta}(a), \hat{\tau}$, respectively, where

$$\begin{aligned}\hat{\delta}(a) &= K^{-1} \sum_{i,k} \omega_i [Y_i^* \{a, M_i^*(1)\} - Y_i^* \{a, M_i^*(0)\}] \\ \hat{\zeta}(a) &= K^{-1} \sum_{i,k} \omega_i [Y_i^* \{1, M_i^*(a)\} - Y_i^* \{0, M_i^*(a)\}] \\ \hat{\tau} &= K^{-1} \sum_{i,k} \omega_i [Y_i^* \{1, M_i^*(1)\} - Y_i^* \{0, M_i^*(0)\}]\end{aligned}$$

- 8: **for** $a = 0, 1$
- 8: **return** $\{\hat{\delta}(0), \hat{\delta}(1), \hat{\zeta}(0), \hat{\zeta}(1), \hat{\tau}\}$

can approximate its q^{th} quantile as $Q_q(\mathbb{F}_{aa'})$, where $\mathbb{F}_{aa'}$ is the empirical distribution of the $Y_{ik}^* \{a, M_{ik}^*(a')\}$'s and $Q_q(F)$ is the q^{th} quantile of F . We can then calculate (3) using the approximation

$$\delta_q(a) \approx Q_q(\mathbb{F}_{a1}) - Q_q(\mathbb{F}_{a0}) \quad \text{and} \quad \zeta_q(a) \approx Q_q(\mathbb{F}_{1a}) - Q_q(\mathbb{F}_{0a}) \quad (8)$$

Note that for this to be valid we must sample the covariates X_{ik}^* 's used to generate $M_{ik}^*(a')$ and $Y_{ik}^* \{a, M_{ik}^*(a')\}$ according to ω_i , rather than averaging over ω as in (7). This results in higher Monte Carlo error in (8) than in (7).

Reducing Monte Carlo error in (8) can also be done, although it requires different strategies; for example, it is no longer valid to replace $Y_{ik}^* \{a, M_{ik}^*(a')\}$ with its mean. One may take K very large, but this can substantially increase computation time. Another trick is to construct the joint distribution of $\{Y_{ik}^* \{a, M_{ik}^*(a')\} : a, a' \in \{0, 1\}\}$ in a way which makes the potential outcomes highly correlated. Interestingly, because (3) depends only on the marginal distributions of the potential outcomes, it is invariant to our choice of joint distribution; hence, this does not actually imply any additional restrictions on the model. To ensure a strong dependence between the $Y_i^* \{a, M_i^*(a')\}$'s, we simulate $M_i^*(a)$ and $Y_i^*(a, m)$ to be *comonotone*,³¹ that is, we simulate $U, V \sim \text{Uniform}(0, 1)$ and apply the probability integral transform to get $M_i^*(a) = F_M^-(U \mid A_i = a, X_i)$ and $Y_i^* \{a, M_i^*(a')\} = F_Y^-(V \mid M_i = M_i^*(a'), A_i = a, X_i)$ (note that the same U and V are used for different values of a and a'). This, combined with taking K to be modestly large (say, $K = 10$) is sufficient to effectively eliminate the Monte Carlo error.

Our general algorithm for approximating the quantile mediation effects is given in Algorithm 3, with the extension to the specific setting of the ZOIB model being derived in the same way Algorithm 2 was derived from Algorithm 1.

4.3 Assessing the Monte Carlo error

Linero²³ introduced a method for computing and (in the case where the effects are approximately normal) correcting for the Monte Carlo error in the types of estimators we have proposed; code for implementing this is available at www.github.com/theodds/AGC. This approach requires $K > 1$ and, in all cases we have considered, the Monte Carlo error is negligible for $K = 2$. Linero²³ also shows that naive estimators that are not designed to eliminate Monte Carlo error can be very inefficient unless K is taken rather large (say, $K \geq 10$).

5 Sensitivity analysis

Because SI is an untestable assumption, it is essential to assess the extent to which the conclusions of an SI-based analysis are sensitive to the existence of unmeasured confounders, that is, SI2. Accordingly, we now present a framework for

Algorithm 2 Monte Carlo g -formula for ZOIB model.**Input:** $\theta, K, \{X_i\}_{i=1}^N$

1: Sample $\omega \sim \text{Dirichlet}(1, \dots, 1)$
 2: **for** $k = 1, \dots, K$ and $i = 1, \dots, N$ **do**
 3: Sample $U \sim \text{Uniform}(0, 1)$
 4: **for** $a = 0, 1$ **do**
 5: Compute $\text{logit } \alpha^M \leftarrow X_i^\top \beta_\alpha^M(a)$, $\text{logit } \gamma^M \leftarrow X_i^\top \beta_\gamma^M(a)$,
 $\text{logit } \mu^M \leftarrow X_i^\top \beta_\mu^M(a)$, $\log \phi^M \leftarrow X_i^\top \beta_\phi^M(a)$.
 6: Set

$$M_{ik}^*(a) \leftarrow \begin{cases} 0 & \text{if } U < \alpha^M, \\ 1 & \text{if } U > 1 - (1 - \alpha^M)\gamma^M, \\ F_{\text{Beta}}^{-1}\{U' \mid \mu^M \phi^M, (1 - \mu^M) \phi^M\} & \text{otherwise} \end{cases}$$

where $U' = (U - \alpha^M)/[(1 - \alpha^M)(1 - \gamma^M)]$.

7: **for** $a' = 0, 1$ **do**
 8: Compute $\text{logit } \alpha^Y \leftarrow (X_i, M_{ik}^*(a))^\top \beta_\alpha^Y(a')$,
 $\text{logit } \gamma^Y \leftarrow (X_i, M_{ik}^*(a))^\top \beta_\gamma^Y(a')$,
 $\text{logit } \mu^Y \leftarrow (X_i, M_{ik}^*(a))^\top \beta_\mu^Y(a')$,
 $\log \phi^Y \leftarrow (X_i, M_{ik}^*(a))^\top \beta_\phi^Y(a')$.
 9: Set $Y_{ik}^*(a', M_{ik}^*(a)) \leftarrow (1 - \alpha^Y)\gamma^Y + (1 - \alpha^Y)(1 - \gamma^Y)\mu$.
 10: **end for**
 11: **end for**
 12: **end for**
 13: Approximate $\delta(a), \zeta(a), \tau$ with $\hat{\delta}(a), \hat{\zeta}(a), \hat{\tau}$, respectively, where

$$\begin{aligned} \hat{\delta}(a) &= K^{-1} \sum_{i,k} \omega_i [Y_{ik}^* \{a, M_{ik}^*(1)\} - Y_{ik}^* \{a, M_{ik}^*(0)\}] \\ \hat{\zeta}(a) &= K^{-1} \sum_{i,k} \omega_i [Y_{ik}^* \{1, M_{ik}^*(a)\} - Y_{ik}^* \{0, M_{ik}^*(a)\}] \\ \hat{\tau} &= K^{-1} \sum_{i,k} \omega_i [Y_{ik}^* \{1, M_{ik}^*(1)\} - Y_{ik}^* \{0, M_{ik}^*(0)\}] \end{aligned}$$

14: **for** $a = 0, 1$
 14: **return** $\{\hat{\delta}(0), \hat{\delta}(1), \hat{\zeta}(0), \hat{\zeta}(1), \hat{\tau}\}$

performing sensitivity analysis using the mixed-scale models we have developed. Without loss of generality, we assume that the data has been scaled so that both $Y_i(a, m)$ and $M_i(a)$ take values in $[0, 1]$. As a guiding principle, we require that any sensitivity parameters be *pure* in the sense that varying them does not alter the fit of the model to the data. This allows us to *independently* assess (i) goodness-of-fit for the observed data model and (ii) the impact of SI2 failing.

5.1 Sensitivity on the logit scale

We propose an approach to sensitivity analysis that allows for dependence between $Y_i(a, m)$ and $M_i(a)$ even after accounting for X_i and A_i . We replace assumption SI2 with the following two assumptions:

SI2A Conditional on X_i , the potential outcomes $M_i(0)$ and $M_i(1)$ are jointly distributed according to a Gaussian copula with correlation $\rho \in [0, 1]$. More precisely, we have $M_i(a) = F_M^{-1}\{\Phi(Z_{ia}) \mid A_i = a, X_i = x\}$ where $F_M^{-1}(u \mid A_i = a, X_i = x)$ denotes the generalized inverse cdf of M_i given $A_i = a$ and $X_i = x$, and (Z_{i0}, Z_{i1}) is jointly standard normal with correlation ρ .

Algorithm 3 Monte Carlo g -formula estimate of quantile causal effects.**Input:** $\theta, K, \{X_i\}_{i=1}^N$

- 1: Sample $\omega \sim \text{Dirichlet}(1, \dots, 1)$
- 2: **for** $k = 1, \dots, K$ and $i = 1, \dots, N$ **do**
- 3: Sample $X_{ik}^* = X_j$ with probability ω_j
- 4: Sample $U_{ik}, V_{ik} \sim \text{Uniform}(0, 1)$
- 5: Set $M_{ik}^*(a) = F_M^-(U_{ik} \mid A = a, X = X_{ik}^*, \theta)$ for $a = 0, 1$.
- 6: Set $Y_{ik}^* \{a, M_{ik}^*(a')\} = F_Y^-(V_{ik} \mid M = M_{ik}^*(a'), A = a, X = X_{ik}^*, \theta)$ for $a, a' = 0, 1$.
- 7: **end for**
- 8: Approximate $\delta_q(a), \zeta_q(a), \tau_q$ with $\hat{\delta}_q(a), \hat{\zeta}_q(a), \hat{\tau}_q$, respectively, where

$$\begin{aligned}\hat{\delta}_q(a) &= Q_q(\mathbb{F}_{a1}) - Q_q(\mathbb{F}_{a0}) \\ \hat{\zeta}_q(a) &= Q_q(\mathbb{F}_{1a}) - Q_q(\mathbb{F}_{0a}) \\ \hat{\tau}_q &= Q_q(\mathbb{F}_{11}) - Q_q(\mathbb{F}_{00})\end{aligned}$$

for $a = 0, 1$, where $\mathbb{F}_{aa'}$ is the empirical distribution of the $Y_{ik}^* \{a, M_{ik}^*(a')\}$'s.

- 9: **return** $\{\hat{\delta}_q(0), \hat{\delta}_q(1), \hat{\zeta}_q(0), \hat{\zeta}_q(1), \hat{\tau}_q\}$

SI2B Conditional on $X_i, M_i(0)$, and $M_i(1)$, the mean of $Y_i(a, m)$ is given by

$$\mathbb{E}_\theta \{Y_i(a, m) \mid M_i(a), M_i(a'), X_i\} = \text{expit}[\text{logit } r_y(m, a, x) + \lambda \{M_i(a) - m\}]$$

where $r_y(m, a, x) = \mathbb{E}_\theta(Y_i \mid M_i = m, A_i = a, X_i = x)$.

SI2B has been chosen specifically so that it reproduces the inferences under SI2 when $\lambda = 0$ while leaving the sensitivity parameters λ and ρ unidentified so that they can be varied freely without changing the fit of the model to the data. The most closely related sensitivity analysis framework which we are aware of is the “hybrid” approach of Albert and Wang,⁵ although this approach differs importantly in that the hybrid approach replaces the term $\lambda \{M_i(a) - m\}$ with a term of the form $\lambda(a - a')$, which is similarly designed to drop out of the distribution of the observed data.

To motivate the choice of SI2B (in particular, the term $M_i(a) - m$), we note that a natural way to induce correlation between the potential mediators and outcomes is to add an additional linear term to the regression model; that is, if we (for the sake of simplicity) replace our ZOIB model with a logistic regression model for the mean, we could incorporate $M_i(a)$ into the linear predictor as $\beta_0^Y + X_i^\top \beta_X^Y + a \beta_A^Y + m \beta_M^Y + \lambda M_i(a)$, with the term $\lambda M_i(a)$ capturing any association due to unmeasured confounding between $Y_i(a, m)$ and $M_i(a)$. The issue with using this expression directly is that λ is confounded with β_M^Y ; adding and subtracting λm , however, gives

$$\beta_0^Y + X_i^\top \beta_X^Y + a \beta_A^Y + m(\beta_m^Y + \lambda) + \lambda \{M_i(a) - m\}$$

Because $M_i(A_i) - M_i(A_i) = 0$, the term $\lambda \{M_i(a) - m\}$ disappears from the distribution of the observed data; consequently, only the term $\beta_m^Y + \lambda$ is identified. Our decision to write our assumption as in SI2B (with λm subtracted off explicitly) only has the effect of reparameterizing the above model in terms of the identified parameter $\beta_m^Y + \lambda$. While this argument only holds exactly for the logistic regression mean model, the intuition is the same for generic models: subtracting λm allows us to parameterize the distribution of the observed data with identifiable parameters.

The following proposition establishes that $\mathbb{E}_\theta[Y_i \{a, M_i(a')\}]$ is identified for all a, a' so that the average causal mediation effects are also identified. A proof is given in Appendix 7.

Proposition 1. Suppose that SI1, SI2A, SI2B, and SI3 hold. Then we have

$$\begin{aligned}\mathbb{E}_\theta[Y_i \{a, M_i(a')\}] &= \iiint \text{expit}\{\tilde{r}_y(m', a, x) + \lambda(m - m')\} \\ &\quad \times \text{Normal}\{(z_0, z_1)^\top \mid (0, 0)^\top, \Sigma\} dz_0 dz_1 F_X(dx)\end{aligned}$$

where $\tilde{r}_y(m, a, x) = \text{logit } r_y(m, a, x)$, $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, and we define $m' = F_M^-(\Phi(z_{a'}) \mid A_i = a', X_i = x)$ and $m = F_M^-(\Phi(z_a) \mid A_i = a, X_i = x)$ in the integral.

As in Section 4, there is no analytical expression for $\mathbb{E}_\theta[Y_i\{a, M_i(a')\}]$, and hence we must resort to Monte Carlo integration. Fortunately, by noting that the approximation

$$\mathbb{E}_\theta[Y_i\{a, M_i(a')\}] \approx K^{-1} \sum_{i,k} \omega_i \text{expit}[\tilde{r}_y\{M_i^\star(a'), a, X_i\} + \lambda\{M_i^\star(a) - M_i^\star(a')\}]$$

is unbiased, it is straight-forward to modify Algorithm 1 to compute $\mathbb{E}_\theta[Y_i\{a, M_i(a')\}]$ under this assumption as well. A procedure for computing the mediation effects under SI2A and SI2B is given in Algorithm S.1 of the Supplemental Material.

As there is no information in the data about the sensitivity parameters ρ and λ , it is essential that we are able to elicit plausible ranges for their values. To gain better intuition about the role of λ , suppose that we had instead posited the logistic regression model $\text{logit } r_y(m, a, x) = \beta_0^Y + x^\top \beta_X^Y + a \beta_A^Y + m \beta_M^Y$. In this case, we can rewrite

$$\text{logit } \mathbb{E}_\theta\{Y_i(a, m) \mid M_i(a), M_i(a'), X_i\} = \beta_0^Y + x^\top \beta_X^Y + a \beta_A^Y + m(\beta_M^Y - \lambda) + M_i(a)\lambda$$

In words, $\lambda = 0$ (which produces the same inferences as SI2) attributes the entirety of the association between M_i and Y_i to a causal effect β_M^Y from shifting the value of m in the potential outcome $Y_i(a, m)$. When $\lambda \neq 0$ this association is instead parsed into an effect of size $(\beta_M^Y - \lambda)$ for shifting m in $Y_i(a, m)$ and a residual association of size λ between $Y_i(a, m)$ and $M_i(a)$, above-and-beyond the causal effect of shifting m due to unmeasured confounding. In the absence of subject-matter expertise about the likely values of λ , one can use weak prior knowledge about the magnitude of the mediation effect to narrow down the range of plausible values. In the JOBS II study, we do this by constructing a pilot estimate of $r_y(m, a, x)$ of the form $\text{logit } \hat{r}_y(m, a, x) = \hat{\beta}_0^Y + x^\top \hat{\beta}_X^Y + a \hat{\beta}_A^Y + m \hat{\beta}_M^Y$. In most cases, we feel that it is reasonable to assume that the effect of unmeasured confounding will not dominate the causal effect associated to m , so that $\lambda \in [-\hat{\beta}_M^Y, \hat{\beta}_M^Y]$ is a conservative collection of plausible values of λ .

The parameter ρ measures dependence in the mediator process $M_i(\cdot)$. Since $|\rho| \leq 1$ and is typically believed to be positive, we recommend repeating the sensitivity analysis for a small number of ρ 's in $[0, 1]$. For the JOBS II data, we consider $\rho \in \{0, 0.5, 0.95\}$, with the results for $\rho \in \{0, 0.5\}$ given in the Supplemental Material.

Remark 1. The use of copulas in SI2A bears a passing resemblance to our use of comonotone random variables to reduce Monte Carlo error in Section 4. In this case, however, the choice of ρ does impact the model, and so we can no longer reduce the Monte Carlo error by making $M_{ik}^\star(a)$ and $M_{ik}^\star(a')$ comonotone.

Remark 2. In the Supplemental Material, we provide a scheme for performing sensitivity analysis on a linear, rather than logit, scale. The linear scale sensitivity analysis has the advantage of not requiring the specification of ρ .

6 Illustrations

6.1 Application to JOBS II

We apply our methodology to the JOBS II dataset using SI as a benchmark. As potential confounders we include numeric variables measuring economic hardship (econ_hard), baseline depression (depress1), and age (age), as well as the categorical variables measuring (sex), race (nonwhite), income bracket (income), occupation (occ), marital status (marital), and education level (educ).

We posit generalized linear models for each component of the ZOIB models with a homogeneous effect for the treatment and mediator. That is, we use linear predictors of the form $X_i^\top \beta^M + \theta^M A_i$ for the mediator and linear predictors of the form $X_i^\top \beta^Y + \theta^Y A_i + \eta M_i$ for the outcome (with separate coefficients for each model component). Note that the homogeneity assumption implies neither that $\delta(0) = \delta(1)$ nor that $\zeta(0) = \zeta(1)$ due to the nonlinearity of the link functions in (6).

The observed data models for M_i and Y_i are fit using Markov chain Monte Carlo (MCMC) in Stan, with a total of 8000 samples collected over four parallel chains and 4000 samples discarded to burn-in. There is no evidence of failure of the chains to mix: all traceplots indicate rapid mixing (see Figures S.4 and S.5 in the Supplemental Material), all values of the Gelman-Rubin diagnostic³² \hat{R} are effectively 1 (minimum of 0.9991, maximum of 1.003), and the minimal effective sample size across all of the monitored parameters is 1807.

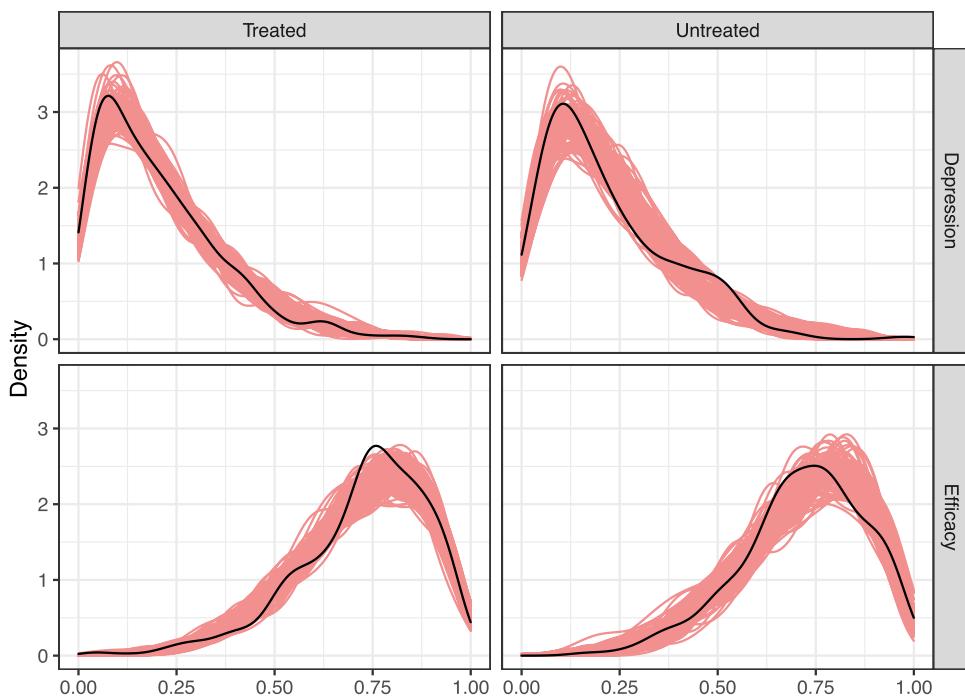


Figure 4. Kernel density estimates of the non-boundary proportion of the original data (black) and 100 replicated datasets (red).

Table 1. Effect estimates for the JOBS II data using the zero-one inflated beta (ZOIB) formulation for the outcome (depression) and mediator (efficacy) assuming sequential ignorability.

Effect	Est.	SD	Lower	Upper	Z-score	P-value
$\delta(0)$	-0.0110	0.0108	-0.0330	0.0098	-1.0131	0.3110
$\delta(1)$	-0.0102	0.0101	-0.0308	0.0089	-1.0144	0.3104
$\zeta(0)$	-0.0282	0.0403	-0.1065	0.0491	-0.7000	0.4839
$\zeta(1)$	-0.0275	0.0400	-0.1058	0.0490	-0.6880	0.4915
τ	-0.0385	0.0416	-0.1202	0.0415	-0.9244	0.3553

We now perform posterior predictive checks to compare the observed data Y_i and M_i to replicate datasets simulated from the fitted model. The goal of these checks is to assess how well the fitted model aligns with the observed data. In Figure 4, we check the fit of the beta distribution to the continuous part of the observed mediator/outcome distributions by comparing a kernel density estimate of the observed-data distribution of Y_i and M_i to 100 replicated datasets sampled from the posterior predictive distribution; the top row shows density estimates for the depression level under each treatment level for the continuous part of the data, while the bottom row shows the same for job-search self-efficacy. The posterior predictive distribution produces datasets that closely match the observed data, suggesting that the beta model for the continuous part of the data is adequate.

In Figure 5, we check the fit of the logistic regression models to the boundary points 0 and 1 by comparing the observed proportion of 0s and 1s for the outcome and mediator to what is observed in replicated datasets. Again, there is close agreement between the model and simulated data.

Since the outcomes in the analysis were scaled by the transformation $y \mapsto \frac{y-1}{4}$, the causal estimates can be brought back to their original scale by multiplying the estimates by 4. The results for the average causal mediation effects on the original scale are given in Table 1.

Although the average causal mediation effect estimates are small (less than a tenth of a point for all effects) and not statistically significant, all treatment effect estimates are negative, which would imply that participation in the job training seminar decreased the depression of subjects both directly through their participation in the seminar and indirectly through the effect of the training seminar on the self-efficacy of participants in finding a job. Of particular interest for us is the decomposition $\tau = \delta(0) + \zeta(1)$, which captures (i) the benefit of increasing self-efficacy generally for those who do not

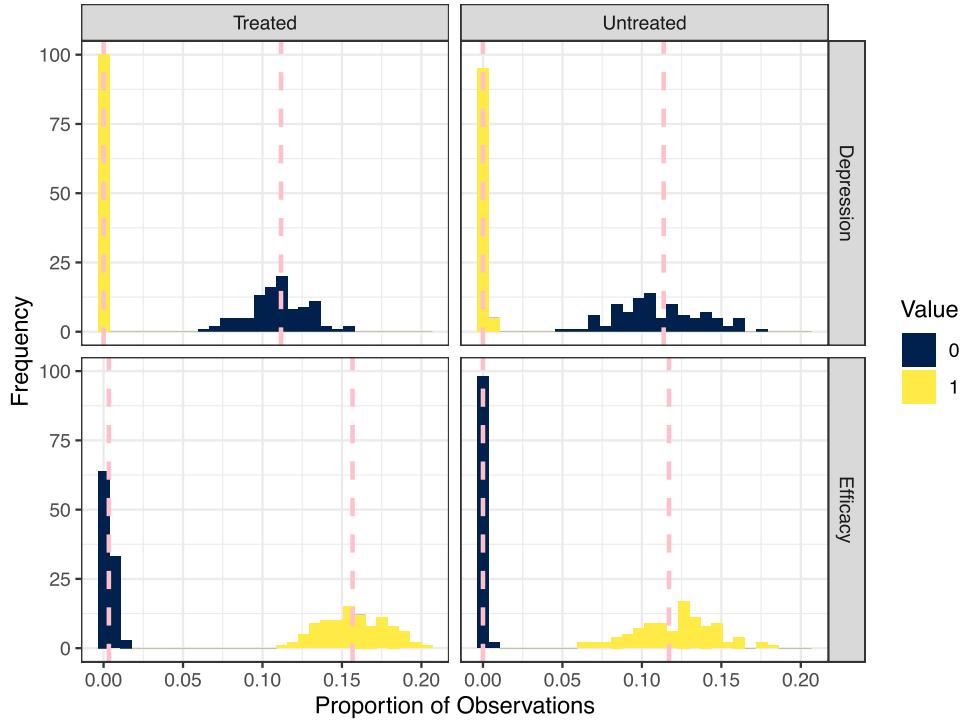


Figure 5. Histograms of the proportions of individuals taking the boundary values 0 (in dark blue) and 1 (in light yellow) across 100 replicated datasets for the outcome (depression, first row) and mediator (efficacy, second row), separately for the treated and untreated groups. The proportions of the observed data taking boundary values are given by the pink vertical dashed lines.

receive treatment and (ii) the additional benefit of the seminar above and beyond its effect on increasing a subject's self-efficacy. However, there is little evidence for either a direct or indirect effect of the treatment on the outcome; in particular, the signs of the mediation effects are uncertain. Similar results are obtained for the quantile mediation effects.

6.2 Sensitivity analysis

While there is no evidence for either a direct or indirect treatment effect under SI, we may be concerned that the effects are being masked by unobserved confounding. We now apply the sensitivity analysis techniques introduced in Section 5 to assess the impact of unmeasured confounding.

Setting $\rho = 0.95$, Figure 6 shows how inferences about the mediation effects change as λ is varied across a range of plausible values. To calibrate λ , we fit a linear model to the conditional mean logit $\mathbb{E}(Y_i | M_i, A_i, X_i)$ using quasi-likelihood and then considered values of λ no more than twice as large in magnitude than the estimated effect of M_i in this model; this corresponds to the belief that most of the association of Y_i with M_i should be attributable to a causal effect of the mediator rather than confounding between the outcome and mediator processes.

From Figure 6, inferences for the direct effects $\zeta(a)$ are robust to unmeasured confounding between $Y_i(a, m)$ and $M_i(a)$, while inferences for the indirect effects $\delta(a)$ are less robust. When λ is negative, there is less uncertainty in $\delta(a)$, although the estimates are also pulled towards zero. For larger values of λ , the estimates of $\delta(a)$ are larger (in magnitude), but also more uncertain. The substantive conclusion remains the same: there is little evidence regarding the sign of either the direct or indirect effects.

We provide a more in-depth sensitivity analysis in the Supplemental Material. In particular, we consider both varying ρ and a *linear* variant of assumption SI2B. While the substantive conclusions don't change from varying ρ , we do see that ρ interacts strongly with λ , with the trends for $\rho = 0, 0.5$ being markedly nonlinear.

6.3 Simulation example

We evaluate Algorithm 2 under a variety of different data generating mechanisms based on the JOBS II data. To devise relevant simulation settings, we first fit our model to the JOBS II data and then modified the estimated coefficients of the fitted beta-regression model. We consider homogeneous effects for both the mediator and outcome, and write ξ^M and ξ^Y

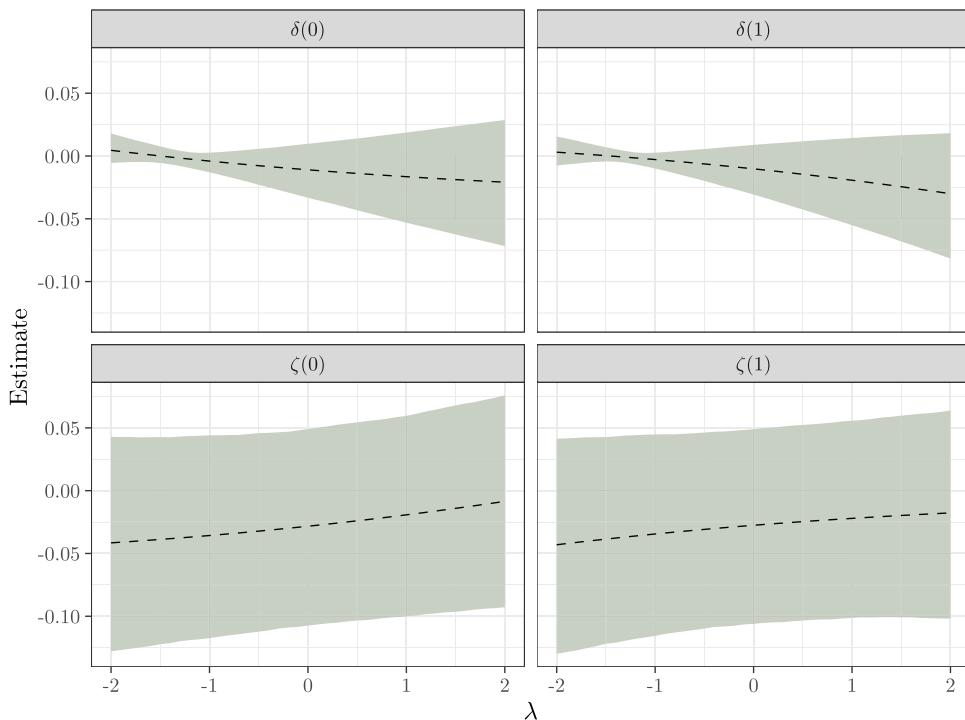


Figure 6. Sensitivity of inferences about $\delta(a)$ and $\zeta(a)$ to changes in the sensitivity parameter λ under assumptions SI2A and SI2B. The dashed line is the posterior mean, and the bands delimit a pointwise 95% credible interval.

for the estimated coefficient for the effect of A_i on M_i and Y_i , respectively. The following features of the data generating mechanism were varied.

Sample Size We consider $N \in \{899, 1798\}$, which is equal to the size of the JOBS II dataset and twice the size of the JOBS II dataset.

Treatment Effect on Mediator We consider $\xi^M \in \{0, \hat{\xi}^M, 10\hat{\xi}^M\}$; the first setting corresponds to no indirect effect of treatment, the second to a realistic indirect effect of the treatment, and the last to a very large indirect effect on the treatment.

Direct Treatment Effect on the Outcome We consider $\xi^Y \in \{0, \hat{\xi}^Y, 10\hat{\xi}^Y\}$, the choices of which are analogous to the ones for the treatment effect on the mediator.

We consider five data generating mechanisms: (1) no mediation, where $(\xi^Y, \xi^M) = (\hat{\xi}^Y, 0)$; (2) complete mediation, where $(\xi^Y, \xi^M) = (0, \hat{\xi}^M)$; (3) strong no mediation, where $(\xi^Y, \xi^M) = (10\hat{\xi}^Y, 0)$; (4) strong complete mediation, where $(\xi^Y, \xi^M) = (0, 10\hat{\xi}^M)$; and (5) no modifications, where $(\xi^Y, \xi^M) = (\hat{\xi}^Y, \hat{\xi}^M)$. For each scenario, we used simulated 200 datasets to compute the bias of the effect estimates, the root-mean-squared error (RMSE), coverage of nominal 95% credible intervals, and the average length of a nominal 95% interval.

Table 2 summarizes the results for each scenario. For readability, all entries of the table are multiplied by 100. Prior to the simulation experiment, we computed the true direct, indirect, and total effects using Monte Carlo integration with 90,799 samples (101 times the size of the original data). For each simulated dataset, we collected a total of 2000 samples across eight parallel chains, with 250 burn-in samples per chain. Our method performs well in terms of bias for both sample sizes and, as expected, we observe lower RMSEs for the larger sample size. The 95% credible intervals are slightly conservative, particularly for the indirect effect; across all scenarios and effects, the smallest coverage probability was 94%. Ultimately, the results show that our approach to computing the mediation effects, while tending to be conservative, appears to work well.

6.4 Robustness to model misspecification

We now consider two simulation experiments designed to answer the following questions: (i) is the ZOIB model robust to model misspecification when the data is semi-continuous but the continuous part is not a beta regression? and (ii) does the

Table 2. JOBS II simulation results for average mediation effects using Algorithm 2.

Scenario	Effect	Truth	n = 899				n = 1798			
			Bias	RMSE	Coverage	Length	Bias	RMSE	Coverage	Length
1	$\delta(0)$	-0.26	0.01	0.31	99.00	1.53	0.03	0.22	99.00	1.07
	$\delta(1)$	-0.26	0.02	0.29	100.00	1.45	0.03	0.21	99.50	1.02
	$\zeta(0)$	-0.98	0.00	1.06	97.50	4.26	0.06	0.72	95.50	3.02
	$\zeta(1)$	-0.97	0.01	1.05	97.00	4.23	0.07	0.72	95.50	3.01
	τ	-1.23	0.02	1.11	96.50	4.41	0.09	0.74	95.50	3.13
	$\delta(0)$	-0.41	0.10	0.33	97.50	1.55	0.02	0.22	98.50	1.08
2	$\delta(1)$	-0.40	0.09	0.32	98.00	1.53	0.01	0.22	98.50	1.06
	$\zeta(0)$	0.22	0.01	1.11	95.00	4.33	0.01	0.73	95.50	3.06
	$\zeta(1)$	0.24	0.00	1.10	95.00	4.30	0.00	0.73	95.00	3.03
	τ	-0.17	0.10	1.18	94.50	4.49	0.02	0.80	95.00	3.17
	$\delta(0)$	-0.11	-0.11	0.30	98.50	1.49	-0.13	0.26	96.50	1.05
3	$\delta(1)$	-0.10	-0.02	0.16	99.50	0.93	-0.04	0.12	99.00	0.65
	$\zeta(0)$	-9.75	0.06	1.03	95.50	3.98	-0.12	0.74	94.50	2.81
	$\zeta(1)$	-9.73	0.14	1.03	94.00	3.96	-0.02	0.73	95.50	2.79
	τ	-9.84	0.03	1.04	94.50	4.08	-0.15	0.75	94.00	2.88
	$\delta(0)$	-1.34	0.05	0.38	96.00	1.64	0.06	0.28	97.50	1.14
4	$\delta(1)$	-1.37	0.09	0.38	94.00	1.64	0.11	0.29	97.00	1.14
	$\zeta(0)$	0.29	0.02	1.11	95.00	4.37	-0.07	0.72	97.00	3.09
	$\zeta(1)$	0.26	0.06	1.08	94.50	4.25	-0.02	0.70	97.50	3.01
	τ	-1.08	0.11	1.13	94.50	4.43	0.04	0.74	96.50	3.14
	$\delta(0)$	-0.33	0.02	0.33	98.50	1.53	0.02	0.22	99.50	1.08
5	$\delta(1)$	-0.34	0.06	0.31	98.50	1.44	0.05	0.21	99.00	1.01
	$\zeta(0)$	-0.99	-0.07	0.95	96.50	4.28	-0.07	0.74	95.00	3.03
	$\zeta(1)$	-1.00	-0.04	0.94	96.50	4.24	-0.03	0.73	95.00	3.01
	τ	-1.33	-0.01	1.05	96.50	4.42	-0.02	0.78	96.50	3.13

Each value in the table is multiplied by 100. The columns correspond to bias, the true values of the effect, the root-mean-squared error (RMSE), the coverage of nominal 95% credible intervals, and the average length of a nominal 95% credible interval. Scenarios 1–5 correspond to the following: (1) $\xi^M = 0$ and $\xi^Y = \hat{\xi}^Y$; (2) complete mediation, where $(\xi^Y, \xi^M) = (0, \hat{\xi}^M)$; (3) $\xi^M = 0$ and $\xi^Y = 10$; (4) complete mediation, where $(\xi^Y, \xi^M) = (0, 10\hat{\xi}^M)$; (5) no modifications, that is, $(\xi^Y, \xi^M) = (\hat{\xi}^Y, \hat{\xi}^M)$, where $\hat{\xi}^M$ and $\hat{\xi}^Y$ are the estimated coefficients for the effect of A_i on M_i and Y_i , respectively.

commonly-used linear structural equation modeling (LSEM) framework perform well with semi-continuous data despite assuming that the underlying distribution is continuous?

To assess the robustness of the ZOIB model to model misspecification, we generate data under a *censored regression* model, where latent variables $M'_i(a)$ and $Y'_i(a, m)$ are modeled using normal linear models such that we observe (say) $M_i(a) = 0$ if $M'_i(a) < 0$ and $M_i(a) = 1$ if $M'_i(a) > 1$. We generated plausible linear models by fitting unconstrained linear models to the JOBS II data.

Results for the censored regression simulation are given in Table 3, with the estimands being the median causal mediation effects. For comparison, results for fitting an unconstrained LSEM are also given. Both LSEM and the ZOIB models have performance that does not reach the nominal coverage level, with the ZOIB performing slightly worse overall in terms of RMSE, interval length, and coverage and slightly better in terms of bias. Interestingly, LSEM seems to perform well when the ground-truth is the censored linear model despite the fact that it does not respect the semicontinuous nature of the data; considering that LSEM is correctly specified except for the fact that it does not capture the boundary behavior correctly, this is not entirely unexpected.

While LSEM performs reasonably well when the underlying data-generating process is a censored linear regression, our next simulation shows that the LSEM can perform poorly when the underlying data-generating process is a ZOIB model. This simulation setting is also based on the JOBS II data, but with stronger treatment effects in the beta part of the regression ($\xi^M = 2$ and $\xi^Y = -0.5$) and with a constant precision. Results are given in Table 4. For this setting, we find that the LSEM is heavily biased and inefficient, with coverage probabilities as low as 16%. The ZOIB model, while not perfect, is more efficient, less biased, and is much closer to the nominal coverage levels.

Table 3. Results for the censored regression simulation. All quantities in the table (except for coverage) are multiplied by 100. CI length denotes the average length of a 95% credible interval while coverage denotes the coverage probability of a nominal 95% credible interval.

Method	Parameter	Bias	RMSE	Standard error	CI length	Coverage
LM	δ_0	0.016	0.194	0.203	0.795	0.960
	δ_1	0.016	0.194	0.203	0.795	0.960
	ζ_0	0.260	0.994	0.935	3.669	0.950
	ζ_1	0.260	0.994	0.935	3.669	0.950
	τ	0.276	1.049	0.952	3.736	0.910
ZOIB	δ_0	-0.028	0.220	0.234	0.919	0.980
	δ_1	-0.024	0.215	0.232	0.909	0.970
	ζ_0	0.179	1.188	1.042	4.091	0.910
	ζ_1	0.183	1.191	1.039	4.077	0.910
	τ	0.155	1.241	1.063	4.172	0.890

ZOIB: zero-one inflated beta; CI: credible interval; RMSE: root-mean-squared error; LM: linear model.

Table 4. Results for the ZOIB simulation comparing the LSEM and ZOIB approaches. All quantities in the table (except for coverage) are multiplied by 100. CI length denotes the average length of a 95% credible interval while coverage denotes the coverage probability of a nominal 95% credible interval.

Method	Parameter	Bias	RMSE	Standard error	CI length	Coverage
LM	δ_0	-0.402	0.683	0.521	2.044	0.910
	δ_1	-1.454	1.556	0.521	2.044	0.170
	ζ_0	1.839	1.951	0.638	2.504	0.160
	ζ_1	0.786	1.022	0.638	2.504	0.790
	τ	0.384	0.518	0.382	1.499	0.820
ZOIB	δ_0	0.088	0.415	0.373	1.485	0.905
	δ_1	0.173	0.242	0.188	0.752	0.905
	ζ_0	-0.051	0.219	0.235	0.938	0.981
	ζ_1	0.034	0.394	0.396	1.575	0.933
	τ	0.122	0.257	0.221	0.887	0.905

ZOIB: zero-one inflated beta; CI: credible interval; LSEM: linear structural equation modeling; RMSE: root-mean-squared error; LM: linear model.

Summarizing, this simulation study shows that the ZOIB model performs reasonably well when the data-generating process is either a censored linear regression or ZOIB, while we found that LSEM performed well only when the data-generating process was a censored linear regression.

7 Discussion

In many practical situations, either the mediator or outcome (or both) will have a mixed-scale distribution, necessitating the use of models beyond the usual linear and generalized linear models. We proposed a zero-one inflated beta distribution after scaling the data to lie in $[0, 1]$. This family of distributions includes many distributional shapes.

The framework proposed here is flexible enough for users to adapt Algorithm 1 and Algorithm 3 to fit essentially arbitrary zero-one inflated models; for example, it is straight-forward to adapt this approach to handle censored regression models.³³ It is also straight-forward to extend our approach to a Bayesian nonparametric setting using Dirichlet process mixture models¹⁶ and nonparametric Bayesian additive regression tree models.³⁴

Just as important as the model we have proposed is our description of how to perform a sensitivity analysis with this type of data. The framework for sensitivity analysis we presented is simple, interpretable, and also extends easily to other model specifications such as the censored linear regression model.

While we have taken a Bayesian approach in this paper, there is nothing in principle which stops us from applying our approach in the Frequentist framework. Taking a Frequentist approach using the bootstrap helps ensure that the estimated standard errors of the parameters remain honest even when the model is misspecified; on the other hand, the estimands may no longer correspond to the de-jure causal effects of interest resulting in a tradeoff that is not worthwhile. Bayesian approaches are also well-suited to settings with hierarchical structures, where software tools such as Stan make it very

easy to consider many different models, and also allows for researchers to incorporate subject-matter expertise via the prior distribution. A potential downside is that, depending on the number of parameters, the Bayesian approach may result in high computational costs. As shown by Linero,²³ the g -computation framework we used here can be used with the nonparametric bootstrap as well; hence, we could have performed maximum likelihood estimation and used the bootstrap to perform uncertainty quantification rather than using Bayesian inference.

In this work, we assumed that all relevant causal effects are defined with respect to the population F_X that the covariates were sampled from. However, it can easily be modified to accommodate a stratified sample from a target population G_X . For example, in Algorithm 3, given a population defined by G_X we simply replace sampling X_{ik}^* from F_X (which is sampled from the Bayesian bootstrap) with sampling X_{ik}^* from G_X . In the case of stratified sampling on a binary covariate V_i , with $\Pr_\theta(V_i = 1) = \varpi$ known, we might model the distribution of $[X_i \mid V_i = 1]$ and $[X_i \mid V_i = 0]$ with independent Bayesian bootstraps G_X^1 and G_X^0 and then sample $X_{ik}^* \sim \varpi G_X^1 + (1 - \varpi) G_X^0$.

Code and a package which replicates our analysis is available online at www.github.com/theodds/ZOIBMediation.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This material is based upon work supported by the National Science Foundation under Grant No. DMS-2144933 and the National Institute of Health under Grant Nos. NIH/NICHD-R01HD093055, NIH/NIMH-R01MH121364, and NIH/NIMH-R01MH121627.

ORCID iD

Antonio R Linero  <https://orcid.org/0000-0002-9531-5667>

Supplemental material

Supplemental material for this article is available online.

References

1. Imai K, Keele L and Tingley D. A general approach to causal mediation analysis. *Psychol Methods* 2010; **15**: 309.
2. Rubin DB. Direct and indirect causal effects via potential outcomes. *Scand J Stat* 2004; **31**: 161–170.
3. Vinokur AD, Price RH and Schul Y. Impact of the JOBS intervention on unemployed workers varying in risk for depression. *Am J Community Psychol* 1995; **23**: 39–74.
4. Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* 1986; **51**: 1173.
5. Albert JM, Wang W. Sensitivity analyses for parametric causal mediation effect estimation. *Biostatistics* 2015; **16**: 339–351.
6. Wang W, Albert JM. Estimation of mediation effects for zero-inflated regression models. *Stat Med* 2012; **31**: 3118–3132.
7. Franks A, D'Amour A and Feller A. Flexible sensitivity analysis for observational studies without observable implications. *J Am Stat Assoc* 2019; **115**: 1730–1746.
8. Linero AR. Bayesian nonparametric analysis of longitudinal studies in the presence of informative missingness. *Biometrika* 2017; **104**: 327–341.
9. Linero AR, Daniels MJ. Bayesian approaches for missing not at random outcome data: the role of identifying restrictions. *Stat Sci* 2018; **33**: 198–213.
10. Scharfstein DO, Nabi R, Kennedy EH, et al. Semiparametric sensitivity analysis: unmeasured confounding in observational studies. *arXiv preprint arXiv:2104.08300*, 2021.
11. Gunzler D, Chen T, Wu P, et al. Introduction to mediation analysis with structural equation modeling. *Shanghai Arch Psychiatry* 2013; **25**: 390.
12. MacKinnon DP, Dwyer JH. Estimating mediated effects in prevention studies. *Eval Rev* 1993; **17**: 144–158.
13. Tchetgen Tchetgen EJ, Shpitser I. Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis. *Ann Stat* 2012; **40**: 1816.
14. Zheng W, van der Laan MJ. Targeted maximum likelihood estimation of natural direct effects. *Int J Biostat* 2012; **8**. <https://doi.org/10.2202/1557-4679.1361>
15. Farbmacher H, Huber M, Lafférs L, et al. Causal mediation analysis with double machine learning. *Econom J* 2022; **25**: 277–300.
16. Kim C, Daniels MJ, Marcus BH, et al. A framework for Bayesian nonparametric inference for causal effects of mediation. *Biometrics* 2017; **73**: 401–409.
17. Cheng J, Cheng NF, Guo Z, et al. Mediation analysis for count and zero-inflated count data. *Stat Methods Med Res* 2018; **27**: 2756–2774.

18. VanderWeele TJ. Causal mediation analysis with survival data. *Epidemiology* 2011; **22**: 582.
19. Ospina R, Ferrari SLP. A general class of zero-or-one inflated beta regression models. *Comput Stat Data Anal* 2012; **56**: 1609–1623.
20. Zuur AF, Ieno EN, Walker NJ, et al. Zero-truncated and zero-inflated models for count data. In *Mixed effects models and extensions in ecology with R*, Springer, 2009, pP.261–293.
21. Liu L, Shih Y-CT, Strawderman RL, et al. Statistical analysis of zero-inflated nonnegative continuous data: a review. *Stat Sci* 2019; **34**: 253–279.
22. Hogan JW, Daniels MJ and Hu L. A Bayesian perspective on assessing sensitivity to assumptions about unobserved data. In: Molenberghs G, Fitzmaurice G, Kenward MG, Tsiatis A, and Verbeke G, (eds), *Handbook of Missing Data Methodology*. CRC Press, 2014.
23. Linero AR. Simulation-based estimators of analytically intractable causal effects. *Biometrics* 2022; **78**: 1001–1017. <https://doi.org/10.1111/biom.13499>
24. Derogatis LR, Lipman RS, Rickels K, et al. The Hopkins symptom checklist (HSCL): a self-report symptom inventory. *Behav Sci* 1974; **19**: 1–15.
25. Pearl J. Direct and indirect effects. In: Breese J and Koller D (eds), *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA. Morgan Kaufmann, 2001, pp. 411–420.
26. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 1992; **3**: 143–155.
27. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survival effect. *Math Model* 1986; **7**: 1393–1512.
28. Rubin DB. The Bayesian bootstrap. *Ann Stat* 1981; **9**: 130–134.
29. Ray K, van der Vaart A. Semiparametric Bayesian causal inference. *Ann Stat* 2020; **48**: 2999–3020.
30. Carpenter B, Gelman A, Hoffman M, et al. Stan: a probabilistic programming language. *J Stat Softw* 2016; **20**: 1–37.
31. Deelstra G, Dhaene J and Vanmaele M. An overview of comonotonicity and its applications in finance and insurance. In: Di Nunno G and Øksendal B (eds), *Advanced Mathematical Methods for Finance*. Berlin, Heidelberg: Springer, 2011, pp. 155–179. https://doi.org/10.1007/978-3-642-18412-3_6
32. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci* 1992; **7**: 457–472.
33. McDonald JF, Moffitt RA. The uses of tobit analysis. *Rev Econ Stat* 1980; **62**: 318–321.
34. Li Y, Linero AR and Murray J. Adaptive conditional distribution estimation with Bayesian decision tree ensembles. *J Am Stat Assoc* 2022. <https://doi.org/10.1080/01621459.2022.2037431>

Appendix. Proof of Proposition I

By iterated expectation and SI2B, we have

$$\begin{aligned}\mathbb{E}_\theta[Y_i\{a, M_i(a')\}] &= \mathbb{E}_\theta\mathbb{E}_\theta[Y_i\{a, M_i(a')\} \mid M_i(a), M_i(a'), X_i] \\ &= \mathbb{E}_\theta \text{expit}[\tilde{r}_y\{M_i(a'), a, X_i\} + \lambda\{M_i(a) - M_i(a')\}]\end{aligned}$$

By SI2A, we have that $M_i(a) = F_M^-\{\Phi(Z_{ia}) \mid A_i = a, X_i = x\}$ and $M_i(a') = F_M^-\{\Phi(Z_{ia'}) \mid A_i = a', X_i = x\}$, where (Z_{i0}, Z_{i1}) is jointly standard normal with correlation matrix Σ . Therefore

$$\begin{aligned}\mathbb{E}_\theta \text{expit}[\tilde{r}_y\{M_i(a'), a, X_i\} + \lambda\{M_i(a) - M_i(a')\}] \\ &= \iiint \text{expit}[\tilde{r}_y(m', a, x) + \lambda(m - m')] \\ &\quad \times \text{Normal}\{(z_0, z_1)^\top \mid (0, 0)^\top, \Sigma\} dz_0 dz_1 F_X(dx)\end{aligned}$$

where $m \equiv F_M^-(z_a \mid A_i = a, X_i = x)$ and $m' = F_M^-(z_{a'} \mid A_i = a', X_i = x)$, completing the proof.