Published at https://ijate.net/ · · · https://dergipark.org.tr/en/pub/ijate · · · *Research Article*

Point and Interval Estimators of an Indirect Effect for a Binary Outcome

Hyung Rock Lee [iD][1,*], **Jaeyun Sung** [iD][2], **Sunbok Lee** [iD][3]

[1]University of Central Arkansas, Department of Exercise & Sport Science, Conway, AR U.S.A.
[2]Lyon College, Department of Political Science, Batesville, AR U.S.A.
[3]Ewha Womans University, Department of Education, Seoul, South Korea

Abstract: Conventional estimators for indirect effects using a difference in coefficients and product of coefficients produce the same results for continuous outcomes. However, for binary outcomes, the difference in coefficient estimator systematically underestimates the indirect effects because of a scaling problem. One solution is to standardize regression coefficients. The residual from a regression of a predictor on a mediator, which we call the residualized variable in this paper, was used to address the scaling problem. In simulation study 1, different point estimators of indirect effects for binary outcomes are compared in terms of the means of the estimated indirect effects to demonstrate the scaling problem and the effects of its remedies. In simulation study 2, confidence and credible intervals of indirect effects for binary outcomes were compared in terms of powers, coverage rates, and type I error rates. The bias-corrected (BC) bootstrap confidence intervals performed better than did other intervals.

1. INTRODUCTION

Mediation analysis tests hypotheses about the mechanism through which a focal independent variable influences an outcome of interest. In mediation analysis, a third intermediate variable named a mediator accounts for the relationship between the independent variable and the outcome, and the effect of the independent variable on the outcome via the mediator is referred to as an indirect effect (Baron & Kenny, 1986). In the literature, two estimators have been widely used to estimate the indirect effect: the difference in coefficients of two nested regression models (Clogg, Petkova, & Shihadeh, 1992; Freedman & Schatzkin, 1992) and the product of coefficients in a path model (Alwin & Hauser, 1975; Bollen, 1987; Sobel, 1982).

For a given sample, the estimates of the two estimators are exactly the same when the outcome is continuous (MacKinnon, Warsi, & Dwyer, 1995). However, when the outcome is binary, the estimates from the two estimators are not the same (Breen, Karlson, & Holm, 2013; MacKinnon, Lockwood, Brown, Wang, & Hoffman, 2007). For a binary outcome, the difference in coefficients estimator underestimates the indirect effect because the regression coefficients of two nested probit or logit models are estimated in different scales (Allison, 1999; Karlson, Holm, & Breen, 2012; Winship & Mare, 1983). One solution to the scaling problem

---

is to use standardized regression coefficients for the difference in coefficients estimator (MacKinnon & Dwyer, 1993; Winship & Mare, 1983). Breen et al. (2013) proposed another solution in which a residualized variable was used to address the scaling issue in the use of the difference in coefficients estimator for a binary outcome. Traditionally, confidence or credible intervals based on the delta (Sobel, 1982), bootstrap (Bollen & Stine, 1990; MacKinnon, Lockwood, & Williams, 2004), and Bayesian methods (Yuan & MacKinnon, 2009) have been widely used to make statistical inference about indirect effects. Previous studies on the indirect effect for a continuous outcome showed that the normality assumption about the sampling distribution might not be valid in small samples in which the true sampling distribution is asymmetric (Bollen & Stine, 1990; MacKinnon et al., 2004). Given the various methods researchers may choose for testing indirect effects for a binary outcome, the performances of those methods are not fully compared yet.

This study aims to compare various point and interval estimators of the indirect effect for a binary outcome using Monte Carlo simulation studies. The point estimators in our study include the conventional difference in coefficients estimator, the difference in coefficients estimator with standardized regression coefficients, the difference in coefficients estimator with residualized variables, and the product of coefficients estimator. Also, the interval estimates or confidence intervals of the indirect effects obtained using the delta, bootstrap, and Bayesian methods are also of interest. In simulation study 1, the point estimators were compared in terms of the means of estimated indirect effects across replications. In simulation study 2, the confidence intervals based on the delta, bootstrap, and Bayesian methods were compared in terms of powers, type I error rates, and coverage rates. We first present the scaling problem using the difference in coefficients estimator for a binary outcome. Then, we describe two solutions to the scaling problem. The delta, bootstrap, and Bayesian methods are briefly introduced before the method section, in which more details on the Monte Carlo simulation studies are presented.

1.1. The Scaling Problem in Estimating an Indirect Effect for a Binary Outcome

The indirect effects for binary outcomes are frequently of interest in social science. For example, in prevention studies, the outcomes of interest are often binary variables such as heart disease or drug use incidence. Since prevention programs are typically designed to change some mediating constructs that are assumed to be related to the outcomes of interest, the success of prevention programs can be evaluated by testing the indirect effect of prevention programs on binary outcomes via mediating constructs (MacKinnon & Dwyer, 1993; MacKinnon et al., 2007).

Binary outcomes in mediation analysis can be modeled using probit or logit regressions. However, the indirect effect estimated by the difference in probit or logit regression coefficients can be inaccurate because the coefficients of two nested probit or logit regressions are measured in different scales and therefore are not directly comparable (Allison, 1999; Karlson et al., 2012; Winship & Mare, 1983). For a more detailed discussion of the scaling issue, let us consider the following simple mediation model in which the latent response variable y∗ is used to model a binary outcome:

$$y^* = \beta_1 + \beta_{yx}x + e_1 , \tag{1}$$

$$y^* = \beta_2 + \beta_{yx.m}x + \beta_{ym.x}m + e_2 , \tag{2}$$

$$m = \beta_3 + \beta_{mx}x + e_3 , \tag{3}$$

where y^* is a continuous latent response variable, β_{yx} is the total effect of x on y^*, $\beta_{yx.m}$ is the direct effect of x on y^* net of m, $\beta_{ym.x}$ is the direct effect of m on y^* net of x, and e_1, e_2, and

e_3 represent error terms. In the latent response variable formulation, a continuous latent response variable y^* is introduced to represent the propensity of the occurrence of a certain category in a categorical outcome. Then, a categorical outcome is considered an observed categorical indicator of an unobserved continuous latent response variable (Muth´en, 1979, 1984). For a binary outcome y, a continuous latent response variable y^* is related to the binary outcome y via a threshold τ as follows:

$$y = 1 \ \ if \ \ y^* > \tau \ \ or \ \ 0 \ \ if \ \ y^* \leq \tau, \tag{4}$$

where the threshold τ is typically assumed to be zero for an identification purpose. Note that, in Equations 1 and 2, the specific form of the model for a binary outcome, i.e., a probit or logit model, is determined by the distribution of an error term: normally distributed error terms result in probit models, and logistically distributed error terms result in logit models (Winship & Mare, 1983).

The scaling issue in estimating indirect effects using probit or logit models can be illustrated by considering the relationship between the regression coefficients in a latent response variable formulation and those in probit or logistic models (Allison, 1999; Breen et al., 2013; Karlson et al., 2012). To examine the relationship, let us assume that the error distributions in Equations 1 and 2 follow normal distributions. That is, $e_1 = \sigma_1 u$ and $e_2 = \sigma_2 u$, where u is a random variable following a standard normal distribution, and σ_1 and σ_2 are scale factors. Then, the probit model for Equation 1 can be described as follows:

$$g[\Pr(y = 1|x)] = g[\Pr(y^* > 0|x)] \tag{5}$$

$$= g\left[\Phi\left(\frac{E(y^*|x)}{\sqrt{V(y^*|x)}}\right)\right] \tag{6}$$

$$= \frac{E(y^*|x)}{\sqrt{V(y^*|x)}} \tag{7}$$

$$= \frac{\beta_1 + \beta_{yx}x}{\sigma_1} \tag{8}$$

$$= b_1 + b_{yx}x, \tag{9}$$

where g is the probit link function or the inverse of the cumulative distribution function of a standard normal distribution, Φ is the cumulative distribution function of a standard normal distribution, and b_1 and b_{yx} are the regression coefficients in the probit model. Similarly, the probit model for Equation 2 can be expressed as the following equation:

$$g[\Pr(y = 1|x, m)] = \frac{\beta_2 + \beta_{yx.m}x + \beta_{ym.x}m}{\sigma_2} = b_2 + b_{yx.m}x + b_{ym.x}m. \tag{10}$$

Then, by comparing the regression coefficients from the latent response variable formulation in Equations 1 and 2, and those from the probit model in Equations 9 and 10, we obtain the following equations:

$$b_{yx} = \frac{\beta_{yx}}{\sigma_1}, \tag{11}$$

$$b_{yx.m} = \frac{\beta_{yx.m}}{\sigma_2}, \tag{12}$$

$$b_{yx} - b_{yx.m} = \frac{\beta_{yx}}{\sigma_1} - \frac{\beta_{yx.m}}{\sigma_2} \neq \beta_{yx} - \beta_{yx.m}. \tag{13}$$

By using logistically distributed error terms and logit link functions, it can be shown that Equations 11, 12, and 13 are also valid for a logit model. Notice that the probit or logit regression coefficients b_{yx} and $b_{yx.m}$ in Equations 11 and 12 involve different scale factors σ_1 and σ_2, which implies that the coefficients of two nested probit or logit models are not directly comparable because they are measured in different scales. Furthermore, because the model in Equation 2 has an additional variable m, the residual variance of the model in Equation 2 should be smaller than that of the model in Equation 1, i.e., $\sigma_2 \leq \sigma_1$. Therefore, the difference in probit or logit regression coefficients in Equation 13, i.e., $b_{yx} - b_{yx.m}$, would underestimate the true amount of an indirect effect, i.e., $\beta_{yx} - \beta_{yx.m}$ (Breen et al., 2013; MacKinnon et al., 2007).

1.2. The Solutions to the Scaling Problem

One solution to the scaling problem is to make the scale equivalent across nested models by standardizing regression coefficients before estimating indirect effects (MacKinnon, 2008; MacKinnon & Cox, 2012; Winship & Mare, 1983). A residualized variable was recently used to address the scaling problem (Breen et al., 2013; Karlson et al., 2012). Those two approaches are briefly illustrated in this section.

1.2.1. Standardized Coefficients

In order to make the scale equivalent or comparable across two nested probit or logit models, Winship and Mare (1983) suggested to standardize regression coefficients using the variance of a latent response variable y^*. For probit models, the variances of y^* in Equations 1 and 2 can be obtained using the following equations (MacKinnon, 2008):

$$Var[y^*] = b_{yx}^2 Var[x] + 1. \tag{14}$$

$$Var[y^*] = b_{yx.m}^2 Var[x] + b_{ym.x}^2 Var[m] + 2b_{yx.m}b_{ym.x}Cov[x,m] + 1 \tag{15}$$

For logit models, the constant 1 in Equations 14 and 15 needs to be replaced by $\pi^2/3$, which is the variance of the standard logistic distribution. Then, the standardized coefficients can be obtained by dividing probit or logit coefficients in Equations 9 and 10 by the square root of the variances of y^* in Equations 14 and 15, and the indirect effect using the standardized coefficients can be expressed as follows:

$$\bar{b}_{yx} - \bar{b}_{yx.m} = \frac{b_{yx}}{\sqrt{Var[y^*]}} - \frac{b_{yx.m}}{\sqrt{Var[y^*]}}, \tag{16}$$

where $\bar{b}_{yx}$ and $\bar{b}_{yx.m}$ represent the standardized regression coefficients for b_{yx} and $b_{yx.m}$, respectively. The standard error of $\bar{b}_{yx} - \bar{b}_{yx.m}$, which is needed for statistical inferences, can be expressed as the following equation

$$SE[\bar{b}_{yx} - \bar{b}_{yx.m}] = \sqrt{SE[\bar{b}_{yx}] + SE[\bar{b}_{yx.m}] - 2Cov[\bar{b}_{yx}, \bar{b}_{yx.m}]}, \tag{17}$$

where $SE[\bar{b}_{yx}] = SE[b_{yx}]/\sqrt{Var[y^*]}$ and $SE[\bar{b}_{yx.m}] = SE[b_{yx.m}]/\sqrt{Var[y^*]}$. For logit models, $Cov[\bar{b}_{yx}, \bar{b}_{yx.m}]$ can be obtained using the formula described in Freedman and Schatzkin (1992). However, we are unaware of any analytical method for calculating the covariance between coefficients of two nested probit models. In our Monte Carlo simulation study, therefore, the bootstrap method was used to construct the confidence intervals of $\bar{b}_{yx} - \bar{b}_{yx.m}$.

1.2.2. Residualized Variables

Breen et al. (2013) proposed another solution to the scaling problem in estimating the indirect effect for a binary outcome. In their method, the mediator m in Equation 2 is replaced by the residualized variable $\tilde{m}$ as shown in the following equation:

$$y^* = \beta_2 + \beta_{yx.\tilde{m}} x + \beta_{y\tilde{m}.x} \tilde{m} + e_4, \qquad (18)$$

where $\tilde{m}$ is the residual from a regression of m on x, $e_4 = \sigma_4 u$, u is a standard normal distribution for probit models and a standard logistic distribution for logit models, and σ_4 is a scale factor. Since x-residualized $\tilde{m}$ is uncorrelated with x, adding $\tilde{m}$ will not change the coefficient of x, which gives the following equation:

$$\beta_{yx} = \beta_{yx.\tilde{m}} \qquad (19)$$

Also, it can be shown that the model in Equation 18 can be obtained by reparameterizing the model in Equation 2, which implies that the residuals in Equations 2 and 18 should be the same:

$$\sigma_2 = \sigma_4 \qquad (20)$$

Given Equations 19 and 20, we have the following equation:

$$b_{yx.\tilde{m}} - b_{yx.m} = \frac{\beta_{yx.\tilde{m}}}{\sigma_4} - \frac{\beta_{yx.m}}{\sigma_2} = \frac{\beta_{yx} - \beta_{yx.m}}{\sigma_2} = \frac{\beta_{mx}\beta_{ym.x}}{\sigma_2} = b_{mx}b_{ym.x}, \qquad (21)$$

where b_{mx} is used to represent β_{mx} for notational consistency. Note that, unlike in Equation 13 in which β_{yx} and $\beta_{yx.m}$ are measured on difference scales, $\beta_{yx.\tilde{m}}$ and $\beta_{yx.m}$ in Equation 21 are measured on the same scale. Therefore, Equation 21 implies that $b_{yx.\tilde{m}} - b_{yx.m}$ measures the change in the coefficient of x due to the inclusion of m, or an indirect effect, on the same scale (Karlson et al., 2012). Another implication of Equation 21 is that it provides the exact decomposition of the total effect $b_{yx.\tilde{m}} = \beta_{yx}/\sigma_2$ into the direct $b_{yx.m} = \beta_{yx.m}\sigma_2$ and indirect $b_{mx}b_{ym.x} = \beta_{mx}\beta_{ym.x}/\sigma_2$ effects.

1.3. Confidence Intervals for Indirect Effects

Confidence intervals have been widely used as interval estimators for indirect effects because they are more informative than hypothesis tests. Confidence intervals can provide information about the variability and direction of the true effect as well as the binary decision on the statistical significance (Gardner & Altman, 1986; Harlow, Mulaik, & Steiger, 2013). Three types of confidence intervals for indirect effects have been discussed in the literature: confidence intervals based on the delta method (Sobel, 1982), the bootstrap method (Bollen & Stine, 1990; MacKinnon et al., 2004), and the Bayesian method (Yuan & MacKinnon, 2009).

A delta method is a general approach for approximating asymptotic standard errors of the non-linear function of statistics. Once the standard error is obtained using the delta method, confidence intervals can be constructed by assuming that the sampling distribution of the non-linear function of statistics follows a normal distribution. However, the normality assumption on the sampling distribution may not be valid in practice. For example, confidence intervals using the delta method performed poorer than did confidence intervals using the bootstrap method for small samples in which the true sampling distribution of the indirect effect deviates from the normal distribution (Bollen & Stine, 1990; MacKinnon et al., 2004).

Unlike the delta method, the bootstrap method does not assume any specific form of the sampling distribution. In the bootstrap method, the analytical derivation of the sampling distribution in the asymptotic theory is replaced with the sampling distribution's empirical

construction. Bootstrap samples of the same size as the original sample are randomly drawn from the original sample with replacement, and then the statistic of interest is calculated for the bootstrap samples to construct the empirical sampling distribution of the statistic. Because no distributional assumption is required in the bootstrap method, confidence intervals using the bootstrap method can be asymmetric to reflect the asymmetric nature of the true sampling distribution. The asymmetric confidence intervals using the bootstrap method may perform better than the symmetric confidence intervals using the delta method when the actual sampling distribution deviates from a normal distribution.

One of the simplest bootstrap confidence intervals is the percentile bootstrap confidence interval in which the lower and upper bounds of $100(1-\alpha)\%$ confidence intervals are defined as α and $1-\alpha/2$ percentiles of the values of the statistic calculated from the bootstrap samples. Note that the justification for the percentile bootstrap confidence interval requires the existence of a monotone transformation of the statistic such that the transformed statistic on the transformed scale is symmetrical and centered on the observed statistic. However, such transformation rarely exists in practice, and therefore the percentile bootstrap confidence intervals are often incorrect. This limitation led to the development of the bias-corrected (BC) bootstrap confidence intervals in which bias in the sampling distribution of the statistic is adjusted using a correction factor (Davison & Hinkley, 1997). More specifically, let $\hat{\theta}$ and $\hat{\theta}_{(b)}$ be the statistics that are calculated from the original and b-th bootstrap sample respectively, where b = 1, ..., B and B is the total number of bootstrap samples. In the BC bootstrap confidence intervals, the estimated is defined as the z score of the percentile of the observed θˆ. That is, $\hat{z}_0 = \Phi^{-1}(p/B)$, where p is the number of $\hat{\theta}_{(b)}$s that are less than $\hat{\theta}$ and Φ^{-1} is the inverse cumulative distribution function for a standard normal distribution. Then, the upper and lower bounds of $100(1-\alpha)\%$ confidence intervals are defined as $2\hat{z}_0 + z_{1-\alpha/2}$ and $2\hat{z}_0 + z_{\alpha/2}$, respectively (MacKinnon et al., 2004; Carpenter & Bithell, 2000).

Confidence intervals using the delta method and the bootstrap method are based on the frequentist approach in which an unknown parameter is treated as an unknown fixed value. In the frequentist approach, a confidence interval gives an estimated range of values that are likely to include the unknown fixed value of the parameter. On the contrary, the Bayesian approach treats an unknown parameter as a random variable with a probability distribution. In the Bayesian approach, prior information on the parameter of interest is quantified as a prior distribution, and the Bayes theorem is used to update the prior distribution to the posterior distribution by incorporating the observed data. All knowledge and uncertainty about the unknown parameter can be inferred from the posterior distribution. A credible interval in the Bayesian approach is the counterpart of the confidence interval in the frequentist approach, and the 95% credible interval is defined as the range between 0.025 and 0.975 percentiles of the posterior distribution. Yuan and MacKinnon (2009) pointed out that the Bayesian method is appealing for studies with complex mediation models and small samples because the Bayesian method does not impose restrictive normality assumptions on the sampling distribution of estimates.

So far, we have discussed various point and interval estimators of indirect effects for binary outcomes. As we mentioned earlier, this study aims to compare various point and interval estimators using Monte Carlo simulation studies. In the following sections, more detail on the Monte Carlo simulation studies are discussed.

2. Simulation Study

2.1. Simulation Study 1

2.1.1. Simulation Description

Simulation study 1 was designed to demonstrate the difference between various point estimators of indirect effects for binary outcomes in terms of averages of estimated indirect effects across replications. Data sets for a simulation were generated based on Equations 2 and 3. The effect sizes of $\beta_{yx.m}$, $\beta_{ym.x}$, and β_{mx} in Equations 2 and 3 were set equal to one another for simplicity and set at 0.14, 0.39, and 0.59 to represent small, medium, and large effect sizes, respectively (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002). The independent variable x was sampled from a standard normal distribution. Given x, the mediator m was generated based on Equation 3 under the assumption that the error term e_3 follows a standard normal distribution. Then, a continuous latent response variable y^* was generated based on Equation 2 by setting the error term e_2 as a standard normal distribution for a probit model and a standard logistic distribution for a logit model. Note that the scale factor σ_2 was defined as $e_2 = \sigma_2 u$, where u is a standard normal distribution for a probit model and a standard logistic distribution for a logit model. Therefore, the scale factor σ_2 in this simulation study was one because $e_2 = u$ in our study. Sample sizes were set at 50, 100, 200, 500, 1000, and 5000 following MacKinnon et al. (2007). In all, 3 (effect sizes) × 6 (sample sizes) = 18 conditions were simulated for both probit or logit models, and each simulation condition was replicated 1000 times. The R software package (R Core Team, 2014) was used to generate data sets and to estimate the probit or logistic regression coefficients. For each simulation condition, the averages of estimated indirect effects from five different estimators were calculated: two product of coefficients estimators and three difference in coefficients estimators. To be more specific, an example showing how to calculate five different estimates for indirect effects is presented below. The data set in the example was generated based on the previously described procedure.

2.1.2. An Example

Given a generated data set, the probit regressions described in Equations 9 and 10, and the linear regression described in Equation 3 were fitted to give the following regression coefficients:

$$g[\Pr(y = 1|x)] = 0.0508(0.0433) + 0.7536(0.0524)\text{x}, \tag{22}$$

$$g[\Pr(y = 1|x, m)] = 0.0346(0.0461) + 0.5656(0.0586)\text{x} + 0.5927(0.0521)\text{m}, \tag{23}$$

$$m = 0.0391(0.0309) + 0.5249(0.0311)x, \tag{24}$$

where the numbers in parentheses indicate the standard errors for the coefficients. The conventional product of coefficients estimate and the difference in coefficients estimate are $\hat{b}_{mx}\,\hat{b}_{ym.x} = 0.5249 \times 0.5927 = 0.3111$ and $\hat{b}_{yx} - \hat{b}_{yx.m} = 0.7536 - 0.5656 = 0.1880$, respectively. Note that the difference in coefficients estimator underestimates the indirect effect because of the scaling issue.

On the other hand, the variances of y^* described in Equations 14 and 15 can be obtained as the following:

$$Var[y^*] = b_{yx}^2 Var[x] + 1 = (0.7536)^2(0.9839) + 1 = 1.5588 \tag{25}$$

$$Var[y^*] = b_{yx.m}^2 Var[x] + b_{ym.x}^2 Var[m] + 2b_{yx.m}b_{ym.x}Cov[x, m] + 1 \tag{26}$$

$$= (0.5656)2(0.9839) + (0.5927)2(1.2229) + 2(0.5656)(0.5927)(0.5165) + 1 \tag{27}$$

$$= 2.0906. \tag{28}$$

With the variances of y^*, the standardized regression coefficients can be obtained by dividing the regression coefficients by the square root of the variances of y^*: $\overline{\hat{b}}_{yx}$= $0.7536/\sqrt{1.5588}$ = 0.6036, $\overline{\hat{b}}_{yx.m}$= $0.5656/\sqrt{2.0906}$ = 0.3912, and $\overline{\hat{b}}_{ym.x}$= $0.5927/\sqrt{2.0906}$ = 0.4099. Then, the rescaled product of coefficients and difference in coefficients estimates can be obtained using the standardized regression coefficients: $\hat{b}_{mx}\overline{\hat{b}}_{ym.x}$= $0.5249 \times 0.4099 = 0.2152$ and $\overline{\hat{b}}_{yx} - \overline{\hat{b}}_{yx.m}$= $0.6036 - 0.3912 = 0.2124$. Note that two rescaled estimates are very similar, but not the same.

Finally, the difference in coefficients estimate can be obtained using the residualized m, which is the residual of m in Equation 24. In our study, the residualized m is denoted as $\tilde{m}$. Given $\tilde{m}$, another probit regression can be fitted to the data to give the following regression coefficients:

$$g[\Pr(y = 1|x, m)] = 0.0577(0.0461) + 0.8767(0.0596)x + 0.5927(0.0521)\tilde{m}. \quad (29)$$

Then, the difference in coefficients estimate is $\hat{b}_{yx.\tilde{m}} - \hat{b}_{yx.m}$= $0.8767 - 0.5656 = 0.3111$. Note that $\hat{b}_{yx.\tilde{m}} - \hat{b}_{yx.m}$ and $b_{mx} \hat{b}_{ym.x}$ are exactly the same. In this example, we have demonstrated how to calculate five different estimates of indirect effects for binary outcomes: $\hat{b}_{mx}\hat{b}_{ym.x}, \hat{b}_{yx} - \hat{b}_{yx.m}, \hat{b}_{yx}\hat{b}_{yx.m}, \overline{\hat{b}}_{yx} - \overline{\hat{b}}_{yx.m}$, and $\hat{b}_{yx.\tilde{m}} - \hat{b}_{yx.m}$. The results of simulation study 1 are presented below.

2.1.3. Results

The averages of estimated indirect effects from the five different estimators for probit and logit models are presented in Tables 1 and 2, respectively. Note that true parameter values for some estimators were unknown. In this simulation, the effect sizes of $\beta_{yx.m}$, $\beta_{ym.x}$, and β_{mx} were manipulated. Therefore, the true values of $b_{yx.m}$ and $b_{ym.x}$ can be calculated using Equation 10 in which the scale factor σ_2 can be set to one because we have used the standard normal and the standard logistic distributions as error distributions. Also, the true value of b_{mx} was known because β_{mx} was just relabeled as b_{mx} for notational consistency, i.e., $\beta_{mx} = b_{mx}$. However, the true values of other coefficients, byx and byx.m, were unknown. Therefore, following MacKinnon et al. (2007), the averages of estimated indirect effects for samples of 106 were calculated across 1000 replications and were considered as true values. The estimated true values were exactly the same as the known true values up to four decimal points. For example, when $\beta_{yx.m} = \beta_{ym.x} = \beta_{mx}$= 0.14, the true value of $b_{mx}b_{ym.x}$ was $0.14 \times 0.14 = 0.0196$, which was exactly the same as the value obtained for samples of 106.

For probit models, several trends can be identified from the results presented in Table 1. First, the conventional difference in coefficients estimator, $\hat{b}_{yx} - \hat{b}_{yx.m}$, underestimated the indirect effect compared to the conventional product of coefficients estimator, $\hat{b}_{mx}\hat{b}_{ym.x}$. For example, when $\beta_{yx.m} = \beta_{ym.x} = \beta_{mx}$= 0.59 and the sample size is 106, the means of estimated indirect effects from $\hat{b}_{mx}\hat{b}_{ym.x}$ and $\hat{b}_{yx} - \hat{b}_{yx.m}$ were 0.3481 and 0.2180, respectively. The result showed the scaling problem in directly comparing regression coefficients of two nested probit models. As can be seen from Equation 13, the regression coefficients of two nested probit models are measured in different scales, and therefore are not directly comparable. The difference in estimated indirect effects from the two estimators increased as the effect size of coefficients increased, which is consistent with MacKinnon et al. (2007).

Second, the difference in coefficients estimator and the product of coefficients estimator with standardized coefficients, which are $\hat{b}_{mx} \overline{\hat{b}}_{ym.x}$ and $\overline{\hat{b}}_{yx} - \overline{\hat{b}}_{yx.m}$, yielded very similar, but not identical, results across all simulation conditions. This result showed that the use of the standardized regression coefficients can reduce the difference in estimated indirect effects from

the difference in coefficients and the product of coefficients estimators, but the difference still exists. That is, the total effect can not be exactly decomposed into the direct and indirect effects with the estimators using the standardized regression coefficients, which might cause some problems in calculating the proportion of the indirect effect in the total effect or the effect size of the indirect effect.

Lastly, the conventional product of coefficients estimator, $\hat{b}_{mx}\hat{b}_{ym.x}$ and the difference in coefficients estimator with the residualized variable, $\hat{b}_{yx.\tilde{m}}-\hat{b}_{yx.m}$, produced exactly the same estimate for the indirect effect. The result indicates that the total effect can be exactly decomposed into the direct and indirect effects by using the residualized variable (Breen et al., 2013). The results for logit models in Table 2 also show similar patterns, as shown in Table 1.

Table 1. *Averages of Estimated Indirect Effects from Different Estimators (Probit Models).*

		Products		Differences		
n	β	$\hat{b}_{mx}\,\hat{b}_{ym.x}$	$\hat{b}_{mx}\,\bar{\bar{b}}_{ym.x}$	$\hat{b}_{yx}-\hat{b}_{yx.m}$	$\bar{\bar{b}}_{yx}-\bar{\bar{b}}_{yx.m}$	$\hat{b}_{yx.\tilde{m}}-\hat{b}_{yx.m}$
50	0.14	0.0223	0.0203	0.0180	0.0208	0.0223
100	0.14	0.0204	0.0192	0.0171	0.0192	0.0204
200	0.14	0.0196	0.0189	0.0173	0.0189	0.0196
500	0.14	0.0195	0.0189	0.0177	0.0189	0.0195
1000	0.14	0.0196	0.0191	0.0180	0.0191	0.0196
5000	0.14	0.0198	0.0194	0.0182	0.0194	0.0198
10^6	0.14	0.0196	0.0192	0.0181	0.0192	0.0196
50	0.39	0.1675	0.1341	0.1185	0.1287	0.1675
100	0.39	0.1593	0.1278	0.1151	0.1271	0.1593
200	0.39	0.1596	0.1279	0.1169	0.1300	0.1596
500	0.39	0.1546	0.1260	0.1143	0.1276	0.1546
1000	0.39	0.1523	0.1263	0.1148	0.1266	0.1523
5000	0.39	0.1526	0.1266	0.1149	0.1267	0.1526
10^6	0.39	0.1521	0.1265	0.1150	0.1265	0.1521
50	0.59	0.4075	0.2383	0.2076	0.2422	0.4075
100	0.59	0.3714	0.2362	0.2189	0.2379	0.3714
200	0.59	0.3598	0.2346	0.2178	0.2349	0.3598
500	0.59	0.3518	0.2333	0.2184	0.2335	0.3518
1000	0.59	0.3515	0.2339	0.2185	0.2338	0.3515
5000	0.59	0.3492	0.2336	0.2183	0.2336	0.3492
10^6	0.59	0.3481	0.2332	0.2180	0.2332	0.3481

Notes. The number in each cell represents the averages of estimated indirect effects across 3000 replications for a given condition. The effect sizes of coefficients were set to be equal, i.e., $\beta = \beta_{mx} = \beta_{ym.x} = \beta_{yx.m}$. The bar over a coefficient indicates that the coefficient is standardized using the variance of a latent response variable y^*. The hat over a coefficient indicates that it is a estimated value. $\tilde{m}$ represents the x-residualized m variable, i.e., the residual of m when m is regressed on x.

Table 2. *Averages of Estimated Indirect Effects from Different Estimators (Logit Models).*

		Products		Differences		
n	β	$\hat{b}_{mx}\,\hat{b}_{ym.x}$	$\hat{b}_{mx}\,\bar{\bar{b}}_{ym.x}$	$\hat{b}_{yx}-\hat{b}_{yx.m}$	$\bar{\bar{b}}_{yx}-\bar{\bar{b}}_{yx.m}$	$\hat{b}_{yx.\tilde{m}}-\hat{b}_{yx.m}$
50	0.14	0.0200	0.0102	0.0154	0.0096	0.0200
100	0.14	0.0213	0.0113	0.0185	0.0108	0.0213
200	0.14	0.0204	0.0110	0.0187	0.0107	0.0204
500	0.14	0.0202	0.0110	0.0190	0.0108	0.0202
1000	0.14	0.0192	0.0105	0.0183	0.0103	0.0192
5000	0.14	0.0196	0.0107	0.0188	0.0105	0.0196
10^6	0.14	0.0196	0.0107	0.0188	0.0106	0.0196
50	0.39	0.1675	0.1341	0.1185	0.1287	0.1675
100	0.39	0.1593	0.1278	0.1151	0.1271	0.1593
200	0.39	0.1596	0.1279	0.1169	0.1300	0.1596
500	0.39	0.1546	0.1260	0.1143	0.1276	0.1546
1000	0.39	0.1523	0.1263	0.1148	0.1266	0.1523
5000	0.39	0.1526	0.1266	0.1149	0.1267	0.1526
10^6	0.39	0.1521	0.1265	0.1150	0.1265	0.1521
50	0.59	0.4075	0.2383	0.2076	0.2422	0.4075
100	0.59	0.3714	0.2362	0.2189	0.2379	0.3714
200	0.59	0.3598	0.2346	0.2178	0.2349	0.3598
500	0.59	0.3518	0.2333	0.2184	0.2335	0.3518
1000	0.59	0.3515	0.2339	0.2185	0.2338	0.3515
5000	0.59	0.3492	0.2336	0.2183	0.2336	0.3492
10^6	0.59	0.3481	0.2332	0.2180	0.2332	0.3481

Notes. The number in each cell represents the averages of estimated indirect effects across 3000 replications for a given condition. The effect sizes of coefficients were set to be equal, i.e., $\beta = \beta_{mx} = \beta_{ym.x} = \beta_{yx.m}$. The bar over a coefficient indicates that the coefficient is standardized using the variance of a latent response variable y^*. The hat over a coefficient indicates that it is a estimated value. $\tilde{m}$ represents the x-residualized m variable, i.e., the residual of m when m is regressed on x.

2.2. Simulation Study 2

In simulation study 2, confidence and credible intervals of the product of coefficients estimator, $\hat{b}_{mx}\,\hat{b}_{ym.x}$, were constructed using the delta, bootstrap, and Bayesian methods, and their performance were compared in terms of powers, type I error rates, and coverage rates. As can be seen from the simulation study 1, the values of $\hat{b}_{yx.\tilde{m}} - \hat{b}_{yx.m}$ were exactly the same as the values of $\hat{b}_{mx}\hat{b}_{ym.x}$. Also, testing $H_0\colon \hat{b}_{mx}\,\hat{b}_{ym.x} = 0$ and $H_0\colon \hat{b}_{mx}\,\bar{\bar{b}}_{ym.x} = 0$ are equivalent.

Mplus (L. K. Muth´en & Muth´en, 2010) was used to construct confidence and credible intervals. Note that the bootstrap confidence interval implemented in Mplus is the BC bootstrap confidence interval. In this simulation, date sets were generated using only the probit model because Mplus limits the data generation to the Probit model. In order to model binary outcomes, the weighted least square estimation should be used in Mplus. However, the weighted least square estimation allows only the probit link. Also, the Bayesian estimation in Mplus only allows the probit link. Therefore, data sets in this simulation were generated using the probit model by setting the error term e2 in Equation 2 as a standard normal distsribution, and then the probit model was used to estimate relevant coefficients to calculate estimates for indirect effects.

Data sets were generated using the latent response variable as described in simulation study 1. However, unlike in the simulation study 1, only the effect sizes of β_{mx} and $\beta_{ym.x}$ were set equal to one another, and the effect sizes of β_{mx} and $\beta_{yx.m}$ were set at 0.14, 0.39, and 0.59. Sample sizes were set at 50, 100, 200, 500, and 1000. In all, 32 (effect sizes) × 5 (sample sizes) = 45 conditions were simulated to calculate powers and coverage rates. For type I error rates, β_{mx} and $\beta_{ym.x}$ are set equal to zeros, and $\beta_{yx.m}$ were set at 0.14, 0.39, and 0.59. Therefore, 3 (effect sizes) × 5 (sample sizes) = 15 conditions were simulated for type I error rates. Each simulation condition was replicated 3000 times.

Given data sets, confidence or credible intervals were constructed using Mplus. In constructing confidence intervals using the delta and bootstrap methods, the weighted least square estimation was used by setting the ESTIMATOR = WLSMV option, and the indirect effects were defined using the MODEL CONSTRAINT command. In constructing credible intervals, the Bayesian estimation was used by setting the ESTIMATOR = BAYES option. Because the option for bootstrap confidence intervals was not compatible with the built-in Monte Carlo facility in Mplus, author-written R code was used to automatically run Mplus and extract relevant estimates from generated output files. The coverage rate was evaluated using the criteria suggested by Bradley (1978); the confidence interval is considered to be liberally, moderately, or strictly robust if the coverage rate falls within the range [.925, .975], [.940, .960], or [.945, .955], respectively.

2.2.1. Results

Confidence and credible intervals for the indirect effect estimator, $\hat{b}_{mx} \, \hat{b}_{ym.x}$, were constructed using the delta, bootstrap, and Bayesian methods, and their powers, coverage rates, and type I error rates are presented in Tables 3, 4, and 5. Several trends can be identified from the tables. First, the powers of the BC bootstrap confidence intervals were higher than the powers of other confidence or credible intervals in almost every simulation condition. The differences in powers among methods are prominent when the sample sizes and effect sizes are small. For example, the powers of the confidence and credible intervals using the delta, bootstrap, and Bayesian methods were 0.048, 0.217, and 0.125 respectively when the sample size is 200, and $\beta_{mx} = \beta_{ym.x} = \beta_{yx.m} = 0.14$. The only exceptions were the conditions in which sample sizes were 50, and $\beta_{mx} = \beta_{ym.x} = 0.59$. In those conditions, the powers of the credible intervals using the Bayesian method were little bit higher than the powers of the confidence intervals using the bootstrap method.

Second, the BC bootstrap confidence intervals performed better than did other intervals in terms of coverage rates. In our study, confidence and credible intervals were constructed with a 95% confidence level. Therefore, the nominal coverage rate of confidence and credible intervals is .95. In Tables 3, 4, and 5, The values marked with *, **, and *** indicate that the coverage rates are liberally [.925, .975], moderately [.94, .96], and strictly [.945, .955] robust based on the criteria suggested by Bradley (1978). As can be seen from the tables, the BC bootstrap confidence intervals were more robust than other intervals. The coverage rates of the intervals seemed to become close to the nominal level of .95 as the sample sizes and effect sizes increase. However, the pattern is less clear for sample sizes.

Third, as shown in Table 3, the type I error rates of the delta, bootstrap, and Bayesian methods were very close to zero in all simulation conditions. Because confidence and credible intervals were constructed with a 95% confidence level in this study, the nominal type I error rate is .05. Therefore, our results indicate that the tests of indirect effects using confidence and credible intervals are very conservative.

Table 3. *Powers, Coverage Rates, and Type I Error Rates of Confidence and Credible Intervals for* $\hat{b}_{mx} \; \hat{b}_{ym.x}$

n	βmx	βym.x	Delta			Bootstrap Pow			Bayesian Pow		
n	$\hat{\beta}_{mx}$ $\beta_{ym.x}$	$\beta_{yx.m}$	Pow	Cov	Typ	Pow	Cov	Typ	Pow	Cov	Typ
50	.14(0) .14 (0)	.14	.002	.965*	.000	.024	.976	.002	.020	.990	.002
100	.14(0) .14 (0)	.14	.008	.929*	.000	.072	.936*	.005	.027	.986	.003
200	.14(0) .14 (0)	.14	.048	.908	.000	.217	.938*	.004	.125	.946***	.001
500	.14(0) .14 (0)	.14	.353	.929*	.000	.640	.960**	.003	.623	.939*	.002
1000	.14(0) .14 (0)	.14	.844	.936*	.001	.930	.958**	.002	.873	.954**	.001
50	.14(0) .14 (0)	.39	.002	.972*	.000	.021	.980	.002	.014	.989	.002
100	.14(0) .14 (0)	.39	.008	.929*	.000	.071	.929*	.006	.024	.982	.002
200	.14(0) .14 (0)	.39	.044	.907	.000	.190	.927*	.004	.114	.946***	.002
500	.14(0) .14 (0)	.39	.345	.922	.000	.627	.956**	.002	.597	.939*	.001
1000	.14(0) .14 (0)	.39	.831	.937*	.000	.920	.954**	.001	.862	.954***	.001
50	.14(0) .14 (0)	.59	.001	.964*	.000	.021	.979	.004	.014	.993	.001
100	.14(0) .14 (0)	.59	.007	.940**	.000	.069	.931*	.005	.022	.980	.002
200	.14(0) .14 (0)	.59	.042	.913	.000	.192	.928*	.004	.112	.940**	.001
500	.14(0) .14 (0)	.59	.314	.930*	.000	.601	.958**	.003	.580	.940**	.002
1000	.14(0) .14 (0)	.59	.805	.942**	.000	.921	.960**	.003	.849	.964*	.001

Notes. Pow=powers, Cov=coverage rates, and Typ=type I error rates. Each condition was replicated 3000 times. Type I error rates were calculated by setting $\beta_{mx} = \beta_{ym.x} = 0$ as indicated by zeros within the parentheses. Bootstrap indicates the BC bootstrap confidenc intervals. The values marked with *, **, and *** indicate that the coverage rates are liberally [.925, .975], moderately [.94, .96], and strictly [.945, .955] robust (Bradley,1978).

Table 4. *Powers and Coverage Rates of Confidence and Credible Intervals for* $\hat{b}_{mx} \; \hat{b}_{ym.x}$

n	βmx	βym.x	Delta Pow		Bootstrap Pow	Cov	Bayesian Pow	
n	$\hat{\beta}_{mx}$ $\beta_{ym.x}$	$\beta_{yx.m}$	Pow	Cov	Pow	Cov	Pow	Cov
50	.39 .39	.14	.158					
50	.39 .39	.14	.158	.922	.401	.955***	.398	.942**
100	.39 .39	.14	.645	.941**	.831	.962*	.750	.946***
200	.39 .39	.14	.979	.936*	.990	.953***	.990	.938*
500	.39 .39	.14	1.000	.951***	1.000	.953***	1.000	.924*
1000	.39 .39	.14	1.000	.942**	1.000	.948***	1.000	.955***
50	.39 .39	.39	.129	.925*	.383	.959**	.375	.949***
100	.39 .39	.39	.614	.938*	.810	.960**	.723	.935*
200	.39 .39	.39	.968	.936*	.980	.951***	.983	.934*
500	.39 .39	.39	1.000	.942**	1.000	.944**	1.000	.926*
1000	.39 .39	.39	1.000	.946***	1.000	.950***	1.000	.954***
50	.39 .39	.59	.119	.925*	.368	.962*	.364	.944**
100	.39 .39	.59	.553	.938*	.779	.955***	.691	.933*
200	.39 .39	.59	.961	.940**	.977	.954***	.982	.935*
500	.39 .39	.59	1.000	.947***	1.000	.947***	1.000	.930*
1000	.39 .39	.59	1.000	.953***	1.000	.955***	1.000	.960**

Notes. Pow=powers and Cov=coverage rates. Each condition was replicated 3000 times.
Bootstrap indicates the BC bootstrap confidence intervals. The values marked with *, **, and *** indicate that the coverage rates are liberally [.925, .975], moderately [.94, .96], and strictly [.945, .955] robust (Bradley, 1978).

Table 5. *Powers and Coverage Rates of Confidence and Credible Intervals for $\hat{b}_{mx}\ \hat{b}_{ym.x}$*

n	β_{mx}	$\beta_{ym.x}$		Delta		Bootstrap		Bayesian	
n	β_{mx}	$\beta_{ym.x}$	$\beta_{yx.m}$	Pow		Pow	Cov	Pow	
				Pow	Cov	Pow	Cov	Pow	Cov
50	.39	.39	.14	.158					
50	.59	.59	.14	.606	.950	.811	.954***	.842	.943**
100	.59	.59	.14	.979	.947***	.987	.950***	.750	.946***
200	.59	.59	.14	1.000	.952***	1.000	.947***	.990	.938*
500	.59	.59	.14	1.000	.953***	1.000	.951***	1.000	.924*
1000	.59	.59	.14	1.000	.950***	1.000	.946***	1.000	.955***
50	.59	.59	.39	.555	.943**	.769	.954***	.808	.936*
100	.59	.59	.39	.974	.952***	.984	.953***	.981	.935*
200	.59	.59	.39	1.000	.948***	1.000	.953***	1.000	.932*
500	.59	.59	.39	1.000	.947***	1.000	.946***	1.000	.926*
1000	.59	.59	.39	1.000	.954***	1.000	.954***	1.000	.955***
50	.59	.59	.59	.723	.948***	.754	.948***	.791	.931*
100	.59	.59	.59	.954	.934*	.971	.944**	.961	.923
200	.59	.59	.59	1.000	.944**	1.000	.944**	1.000	.917
500	.59	.59	.59	1.000	.953***	1.000	.954***	1.000	.932*
1000	.59	.59	.59	1.000	.954***	1.000	.954***	1.000	.959**

Notes. Pow=powers and Cov=coverage rates. Each condition was replicated 3000 times.
Bootstrap indicates the BC bootstrap confidence intervals. The values marked with *, **, and *** indicate that the coverage rates are liberally [.925, .975], moderately [.94, .96], and strictly [.945, .955] robust (Bradley, 1978).

3. DISCUSSION

The indirect effect has been estimated in two ways: the difference in coefficients or the product of coefficients. Unlike for continuous outcomes, the difference in coefficients estimator for binary outcomes systematically underestimates the indirect effect because the estimator compares regression coefficients that are measured in different scales. To address the scaling issue, it was proposed to use standardized regression coefficients (Winship & Mare, 1983; MacKinnon, 2008) or residualized variables (Breen et al., 2013). The simulation study 1 was designed to contrast those estimators of indirect effects for binary outcomes in terms of the averages of estimated indirect effects. On the other hand, confidence or credible intervals have been widely used to test indirect effects. In the simulation study 2, confidence or credible intervals of the product of coefficients estimator for binary outcomes were constructed using the delta, bootstrap, and Bayesian methods, and their performance were compared in terms of powers, coverage rates, and type I error rates.

In the simulation study 1, five different point estimators were compared in terms of the averages of estimated indirect effects. The results in Tables 1 and 2 showed that the conventional difference in coefficients estimator ($\hat{b}_{yx} - \hat{b}_{yx.m}$) systematically underestimated the indirect effects compared to the conventional product of coefficients estimator ($\hat{b}_{mx}\ \hat{b}_{ym.x}$). The discrepancy between the two estimators can be reduced by using the standardized regression coefficients. That is, estimated indirect effects from $\hat{b}_{mx}\ \bar{\bar{b}}_{ym.x}$ and $\bar{\bar{b}}_{yx} - \bar{\bar{b}}_{yx.m}$ were similar but not identical. The estimated indirect effects from the conventional product of coefficients estimator ($\hat{b}_{mx}\ \hat{b}_{ym.x}$) and the difference in coefficients estimator using the residualized variable ($\hat{b}_{yx.\tilde{m}} - \hat{b}_{yx.m}$) were exactly the same, which indicates the exact decomposition of the total effect into the direct and indirect effects. In all, the conventional difference in
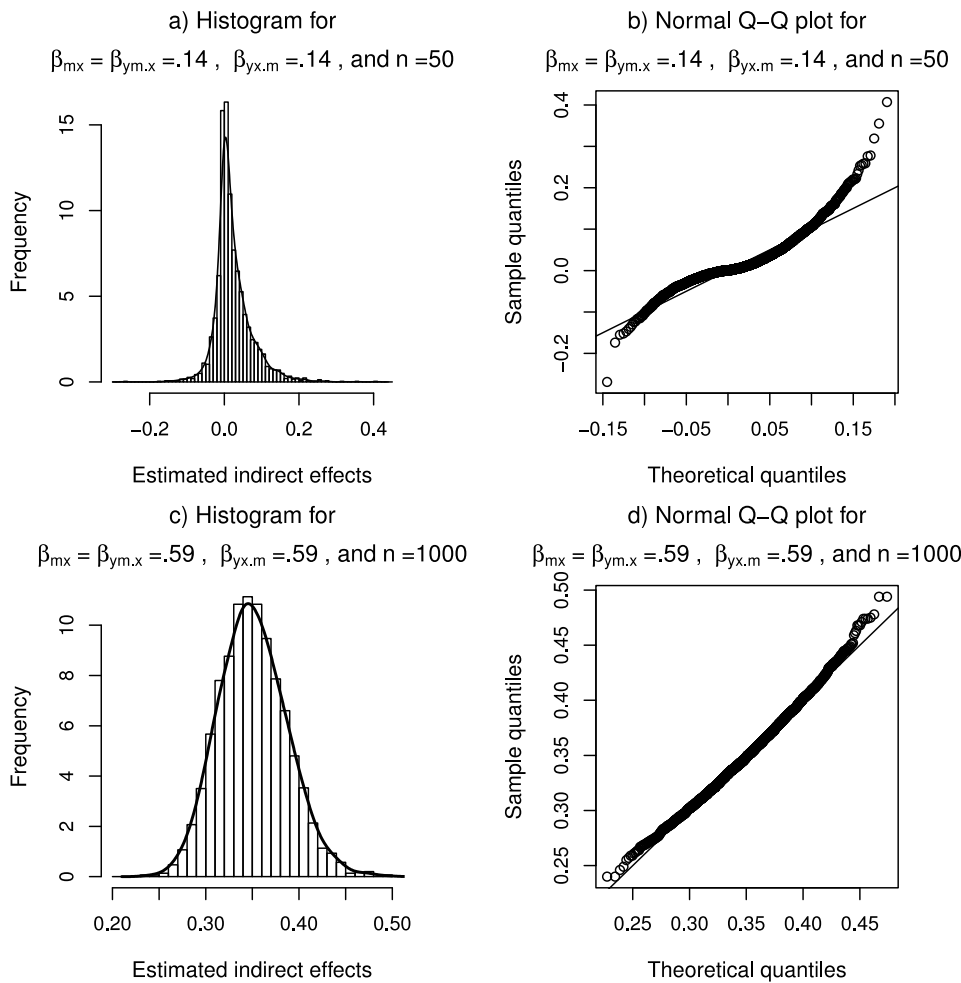
coefficients estimator $(\hat{b}_{yx} - \hat{b}_{yx.m})$ should not be used for binary outcomes. Also, the counterparts of $\hat{b}_{yx.\tilde{m}} - \hat{b}_{yx.m}$ and $\bar{\bar{b}}_{yx} - \bar{\bar{b}}_{yx.m}$ are the $\hat{b}_{mx}\hat{b}_{ym.x}$ and $\hat{b}_{mx}\bar{\bar{b}}_{ym.x}$, respectively.

Here, it may be interesting to discuss the effect size measure of indirect effects for a binary outcome. At first, it may seem that the exact decomposition is necessary to interpret the effect size measures that are defined as the ratio of the indirect effect to the total effect, $b_{mx}b_{ym.x}/(b_{yx.m} + b_{mx}b_{ym.x})$ as proportion. However, as Preacher and Kelley (2011) pointed out, $b_{mx}b_{ym.x}/(b_{yx.m} + b_{mx}b_{ym.x})$ is not a proportion, and can exceed one and even be negative in some cases. On the other hand, in our previous example, the values of $b_{mx}b_{ym.x}/(b_{yx.m} + b_{mx}b_{ym.x})$ that were calculated using the original and standardized regression coefficients were exactly the same. The value of $b_{mx}b_{ym.x}/(b_{yx.m} + b_{mx}b_{ym.x})$ was $(0.5249 \times 0.5927)/(0.5656 + 0.5249 \times 0.5927) = 0.3548$ for the original coefficients, and the value of $b_{mx} \bar{\bar{b}}_{ym.x}/(\bar{\bar{b}}_{yx.m} + b_{mx} \bar{\bar{b}}_{ym.x})$ was $(0.5249 \times 0.5927 \sqrt{2.0906})/(0.5656 \sqrt{2.0906} + 0.5249 \times 0.5927/\sqrt{2.0906}) = 0.3548$ for the standardized coefficients. It is apparent that the two values should be the same because the numerator and denominator in the effect size for original coefficients are divided by the same scaling factor. Moreover, Breen et al. (2013) also suggested to use $b_{mx}b_{ym.x}/(b_{yx.m} + b_{mx}b_{ym.x})$ as the effect size measures for the indirect effect with a residualized variable, which gives exactly the same effect size of 0.3548.

Confidence and credible intervals have been widely used as interval estimators for indirect effects. In the literature, different methods for constructing interval estimators have been compared for indirect effects with continuous outcomes (MacKinnon et al., 2004; Yuan & MacKinnon, 2009). In this study, confidence and credible intervals using the delta, bootstrap, and Bayesian methods were compared for indirect effects with binary outcomes. The results in Tables 3, 4, and 5 showed that the BC bootstrap confidence intervals performed better than did other intervals in terms of powers, coverage rates, and type I error rates, especially when the sample sizes and effect sizes are small. This result is expected because the sampling distributions of estimators tend to deviate from the normal distribution in small samples (MacKinnon et al., 2004; Bollen & Stine, 1990; Yuan & MacKinnon, 2009). In Figure 1, histograms and normal Q-Q plots of estimated indirect effects for the worst- and best- case scenarios in our study are presented to demonstrate how much the sampling distribution of indirect effects could deviate from the normal distribution depending on the sample sizes and effect sizes. For the worst case scenario, the histogram and normal Q-Q plot for $\beta_{mx} = \beta_{ym.x} = .14$, $\beta_{yx.m} = .14$, and n = 50 are presented in Figures 1a and 1b, which show clear deviation from the normal distribution. On the contrary, the histogram and normal Q-Q plot for the best case scenario, where $\beta_{mx} = \beta_{ym.x} = .59$, $\beta_{yx.m} = .59$, and n = 1000, were very close to those for the true normal distribution. Therefore, the poor performance of the confidence intervals using the delta method seems to be reasonable because the assumption of the normal sampling distribution in the delta method is not valid in small samples.

On the other hand, the comparison between the BC bootstrap confidence intervals and the Bayesian credible intervals is interesting. Both methods do not require any specific form of the sampling distribution. The sampling distribution in the bootstrap method is empirically constructed, and the posterior distribution in the Bayesian method is updated from the prior distribution. With the flexibility in the form of the sampling distribution, the two methods can capture the possible asymmetric nature of the true sampling distribution in small samples. Therefore, the better performance of the two methods over the delta method can be understood as the result of the flexible assumption about the sampling distribution.

Figure 1. *Histograms and Normal Q-Q plots.*

a) Histogram for
$\beta_{mx} = \beta_{ym.x} = .14$, $\beta_{yx.m} = .14$, and n = 50

b) Normal Q–Q plot for
$\beta_{mx} = \beta_{ym.x} = .14$, $\beta_{yx.m} = .14$, and n = 50

c) Histogram for
$\beta_{mx} = \beta_{ym.x} = .59$, $\beta_{yx.m} = .59$, and n = 1000

d) Normal Q–Q plot for
$\beta_{mx} = \beta_{ym.x} = .59$, $\beta_{yx.m} = .59$, and n = 1000

Note. The values of estimated indirect effects for histograms and normal Q-Q plots come from the bootstrap method. The estimated indirect effects were exactly the same for both bootstrap and delta methods. The estimated indirect effects from the Bayesian estimation were little bit different from those from the bootstrap and delta methods, but produced very similar histogram and normal Q-Q plot. In Figures (a) and (c), solid lines indicate density plots for the corresponding histograms.

Interestingly, the BC confidence intervals showed better performance than did the Bayesian credible intervals in most simulation conditions. One of the possible explanations may be the use of the default non-informative prior in Mplus. In general, inferences in the Bayesian method are made based on the posterior distribution, which is proportional to the product of the prior and likelihood distributions. The non-informative prior, which is the default prior in Mplus, is typically used when there is no prior knowledge on the parameter of interest. In such a case, the likelihood distribution is the only dominant factor for estimating the posterior distribution. Note that, even in the Bayesian method, we still need an assumption about the form of the likelihood distribution. Therefore, the use of non-informative prior may make the estimation procedure less flexible in capturing the asymmetric nature of the true sampling distribution because the estimation procedures heavily rely on the pre-specified form of the likelihood distribution. In our study, the average widths of intervals using the delta, bootstrap, and Bayesian methods were 0.193, 0.276, and 0.250 respectively when $\beta_{mx} = \beta_{ym.x} = .14$, $\beta_{yx.m} = .14$, and n = 50, and were 0.144, 0.144, and 0.149 respectively when

$\beta_{mx} = \beta_{ym.x} = .59$, $\beta_{yx.m} = .59$, and n = 1000. In small samples, the BC confidence intervals were most wide, whereas all intervals were quite similar in their average widths in large

samples. This might indicate that the BC confidence intervals are most flexible in capturing possible asymmetric nature of the true sampling distribution.

This study compared various point and interval estimators of the indirect effect for a binary outcome. The conventional difference in coefficients estimator should be avoided in estimating the indirect effect for a binary outcome because of the scaling problem. For interval estimations, the BC bootstrap confidence intervals seem to perform better than the intervals based on the delta and Bayesian methods. In this study, only non-informative prior was used in the Bayesian method. It would be interesting to compare the bootstrap method with the Bayesian methods with different prior distributions. Also, this study did not consider the case where the moderator is binary, which would be another interesting study.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the author(s).

Authorship contribution statement

Hyung Rock Lee: Investigation, Resources, Visualization, Software, Formal Analysis, and Writing - original draft. **Jaeyun Sung**: Methodology, Visualization, Supervision, and Validation. **Sunbok Lee**: Methodology, Software, Formal Analysis, Supervision, and Validation.

ORCID

Hyung Rock Lee   https://orcid.org/0000-0002-7415-9466
Jaeyun Sung   https://orcid.org/0000-0001-7461-3123
Sunbok Lee   https://orcid.org/0000-0002-0924-7056

5. REFERENCES

Allison, P. D. (1999). Comparing logit and probit coefficients across groups. *Sociological Methods & Research, 28,* 186-208.

Alwin, D. F., & Hauser, R. M. (1975). The decomposition of effects in path analysis. *American Sociological Review, 40*, 37-47.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.

Bollen, K. A. (1987). Total, direct, and indirect effects in structural equation models. *Sociological Methodology, 17*, 37-69.

Bollen, K. A., & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology, 20*, 15-140.

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology, 31,* 144-152.

Breen, R., Karlson, K. B., & Holm, A. (2013). Total, direct, and indirect effects in logit and probit models. *Sociological Methods & Research, 42*, 164-191.

Carpenter, J., & Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? a practical guide for medical statisticians. *Statistics in Medicine, 19*, 1141-1164.

Clogg, C. C., Petkova, E., & Shihadeh, E. S. (1992). Statistical methods for analyzing collapsibility in regression models. *Journal of Educational and Behavioral Statistics, 17*, 51-74.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge University Press.

Freedman, L. S., & Schatzkin, A. (1992). Sample size for studying intermediate endpoints within intervention trials or observational studies. *American Journal of Epidemiology, 136*, 1148-1159.

Gardner, M. J., & Altman, D. G. (1986). Confidence intervals rather than p values: estimation rather than hypothesis testing. *British Medical Journal (Clinical Research), 292*, 746-750.

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (2013). *What if there were no significance tests?* Psychology Press.

Karlson, K. B., Holm, A., & Breen, R. (2012). Comparing regression coefficients between same-sample nested models using logit and probit a new method. *Sociological Methodology, 42,* 286-313.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Routledge.

MacKinnon, D. P., & Cox, M. C. (2012). Commentary on mediation analysis and categorical variables: The final frontier by dawn iacobucci. *Journal of Consumer Psychology: the official journal of the Society for Consumer Psychology, 22 ,* 600-602.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review, 17*, 144-158.

MacKinnon, D. P., Lockwood, C. M., Brown, C. H., Wang, W., & Hoffman, J. M. (2007). The intermediate endpoint effect in logistic and probit regression. *Clinical Trials, 4*, 499-513.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*, 83-104.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39*, 99-128.

MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research, 30*, 41-62.

Muthen, B. (1979). A structural probit model with latent variables. *Journal of the American Statistical Association, 74*, 807-811.

Muthen, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*, 115-132.

Muthen, L. K., & Muthen, B. O. (2010). *Mplus: Statistical analysis with latent variables: User's guide.* Muthen & Muthen.

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: quantitative strategies for communicating indirect effects. *Psychological Methods, 16*, 93-115.

R Core Team. (2014). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org/

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology, 13*, 290-312.

Winship, C., & Mare, R. D. (1983). Structural equations and path analysis for discrete data. *American Journal of Sociology, 89*, 54-110.

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods, 14*, 301-322.