2021

Regression-based mediation analysis: a formula for the bias due to an unobserved precursor variable

Steven B. Kim

Joonghak Lee

Journal of the Korean Statistical Society (2021) 50:1058–1076
https://doi.org/10.1007/s42952-021-00105-9
Online ISSN 2005-2863
Print ISSN 1226-3192

RESEARCH ARTICLE

Regression-based mediation analysis: a formula for the bias due to an unobserved precursor variable

Steven B. Kim[1] · **Joonghak Lee**[2]

Received: 26 June 2020 / Accepted: 13 January 2021 / Published online: 30 January 2021

Abstract
Researchers want to know whether the change in an explanatory variable X affects the change in a response variable Y (i.e., X causes Y). In practice, there can be two causal paths from X to Y, the path through a mediating variable M (indirect effect) and the path not through M (direct effect). The parameter estimation and hypothesis testing can be performed by a regression-based mediation model. It is already known that randomization of X is not enough for unbiased estimation, and the bias due to an unobserved variable has been discussed in literature but often overlooked. In this article, we first review the challenge under a simple mediation model, then we provide a formula for the exact bias due to an unobserved precursor variable W, the variable which potentially causes the changes in X, M, and/or Y. We present simulation studies to demonstrate the impact of an unobserved precursor variable on hypothesis testing for indirect effect and direct effect. The simulation results show that the inflation of type I error is serious particularly in a large sample study. To numerically demonstrate the formula of the exact bias, a popular data set published in a journal of statistics education is revisited, and we quantify why the conclusion of data analysis can be different before and after accounting for the precursor variable. The result shall remind the importance of a precursor variable in mediation analysis.

Keywords Mediation analysis · Regression · Bias · Precursor variable

✉ Joonghak Lee
joonghak.lee@pgr.reading.ac.uk

Steven B. Kim
stkim@csumb.edu

1 Department of Mathematics and Statistics, California State University, Monterey Bay, 100 Campus Center, Seaside, CA 93955, USA

2 International Business and Strategy, Henley Business School, University of Reading, Whiteknights, Reading RG6 6UD, UK

1 Introduction

In educational and behavioral research or related areas, researchers often want to answer a scientific question whether the change in X affects the change in Y (i.e., a causal relationship from X to Y). In this case, X is referred to as the explanatory variable (or independent variable), and Y is referred as the response variable (dependent variable or outcome variable). In this article, X is a Bernoulli random variable (one or zero), a discrete random variable, or a continuous random variable, and Y is a continuous random variable. The causal relationship is often denoted by $X \to Y$, and the direction of arrow matters. If a researcher controls the change in X, it is called an experimental study. Otherwise, it is called an observational study. Caveats and challenges of causal inference through an observational study have been widely discussed in various disciplinary areas (Glass et al. 2013; Kang 2014; Rohrer 2018; Adams et al. 2019).

In mediation analysis (Baron and Kenny 1986; Hayes 2013), suppose there are two causal paths from X to Y. The first path is $X \to M \to Y$, and M is referred to as the mediating variable (mediator or intermediate variable). The second path is $X \to Y$ not through M. This mediation model is graphically illustrated in Fig. 1, and it is the simplest mediation model presented by Hayes (2013) which is highly cited (more than 22,000 as of now) by many researchers. The first path is often referred to as the indirect effect, and the second path is often referred to as the direct effect. Hayes (2013) presented more complex mediation models than the one shown in Fig. 1.

Since Hayes (2013) developed the PROCESS macro in the statistical software SPSS, many researchers in social science have applied to prove complex causal paths, and the mediation models are still popular (Caniëls 2019; Garcia et al. 2018; Seli et al. 2017; Zhang et al. 2019; Emery et al. 2019; Zhu et al. 2019; Villaluz and Hechanova 2018; Manuti and Giancaspro 2019). Some of these studies were observational, and some were experimental. For instance, Zhang et al. (2019) asked subjects to report their values of two explanatory variables (workplace ostracism and leader-member exchange), and Seli et al. (2017) randomly assigned subjects to either an experimental group (manipulating motivation) or a control group (no motivation). The term "mediation" has an implicit direction of relationship, and many researchers used the mediation model (or a more complex mediation model) based on data collected in an observational study (i.e., no randomization of X). Even though many researchers have conducted observational studies, the direction of a causal relationship has been justified by a common sense or an acceptable theory.

The purpose of this article is not to discourage the use of a mediation model. The purpose is to remind researchers the impact of an unobserved "precursor variable" (denoted by W in this article) on the probability of concluding the presence of an indirect effect and/or a direct effect. We use the term "precursor variable" to refer that W

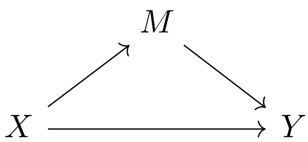

Fig. 1 The simplest mediation model presented by Hayes (2013)

precedes X, M, and Y in a causal relationship as shown in Fig. 2. The necessity of randomization (i.e., controlling X) has been accepted in scientific communities since it was first advocated by Ronald A. Fisher (Fisher 1925; Hall 2007). However, it has been shown that the randomization does not completely remove bias in a certain mediation analysis. VanderWeele (2010) and Imai et al. (2010a, 2010b) presented bias formulas under regression models, and Hong et al. (2018) discussed various methods of sensitivity analysis (Hong et al. 2015, 2018; VanderWeele 2015).

In this article, we review the basic regression-based mediation models (Figs. 1, 2; Hayes 2013), thoroughly present formulas in terms of the regression parameters to quantify the bias in the estimation of indirect effect and direct effect in the absence of W, and use simulations to demonstrate its consequence (inflated Type I error rate). We assume that readers of this article have background knowledge of multiple linear regression models and basic theorems in mathematical statistics including: $Cov(X,X) = V(X)$ and

$$Cov\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_i Y_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_i Cov(X_i, Y_i)$$

where an uppercase letter denotes a random variable and a lowercase letter denotes a constant real number. In Sect. 2, we consider a simple case when the causal relationship from X to Y does not involve a mediator M, and a formula will demonstrate that randomization of X is enough to remove bias in this simple case. In Sect. 3, we consider a case when the causal relationship involves a mediator M, and another formula will clearly demonstrate that randomization of X is not enough to remove bias in this more complicated case. In Sect. 4, we present simulation results which demonstrate seriously inflated Type I error rates even in an experimental study. In Sect. 5, a numerical example is provided based on the data collected in the United States and previously analyzed by Guber (1999).

2 Causal relationship without a mediator

Let X denote an explanatory variable of interest (observed), Y denote a response variable of interest (observed), and W denote an omitted (unobserved) precursor variable which may affect X and/or Y (see Fig. 3). Suppose a researcher assumes the simple linear model

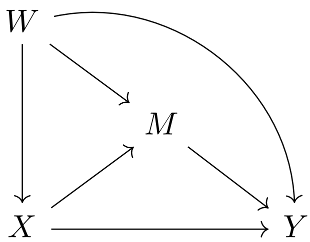

Fig. 2 The simplest mediation model presented by Hayes (2013) with a precursor variable W

$$Y = b_0 + b_1 X + \epsilon\,,$$

and suppose the true relationship among the three random variables, W, X, and Y, is given by the two linear models

$$\begin{aligned} X &= \gamma_0 + \gamma_1 W + \epsilon_1^*\,, \\ Y &= \beta_0 + \beta_1 X + \beta_2 W + \epsilon_2^*\,. \end{aligned}$$

Figure 3 graphically illustrates this situation. If we let $\sigma_W^2 = V(W)$ and $\sigma_i^2 = V(\epsilon_i^*)$ for $i = 1, 2$, then the variances $V(X)$ and $V(Y)$ can be expressed in terms of σ_W^2, σ_1^2, σ_2^2, and the regression parameters.

Here our goal is to express b_1 (the quantity estimated by the researcher) in terms of β_1, β_2, γ_1, σ_W, and σ_1. Under the simple linear model assumed by the researcher,

$$Cov(X, Y) = Cov(X, b_0 + b_1 X + \epsilon) = b_1 V(X)\,,$$

so the estimand b_1 can be expressed as

$$b_1 = \frac{Cov(X, Y)}{V(X)}\,.$$

Under the true relationship,

$$\begin{aligned} Cov(X, Y) &= Cov(X, \beta_0 + \beta_1 X + \beta_2 W + \epsilon_2^*) \\ &= \beta_1 V(X) + \beta_2 Cov(X, W) \\ &= \beta_1 V(X) + \beta_2 Cov(\gamma_0 + \gamma_1 W + \epsilon_1^*, W) \\ &= \beta_1 V(X) + \beta_2 \gamma_1 V(W)\,. \end{aligned}$$

Therefore, the researcher eventually estimates a complex quantity

$$b_1 = \frac{\beta_1 V(X) + \beta_2 \gamma_1 V(W)}{V(X)} = \beta_1 + \beta_2 \gamma_1 \left(\frac{V(W)}{V(X)} \right),$$

where $V(W) = \sigma_W^2$ and $V(X) = \gamma_1^2 \sigma_W^2 + \sigma_1^2$. The researcher can accomplish $b_1 = \beta_1$ by randomization of X (i.e., $\gamma_1 = 0$). When in an observational study (i.e,. $\gamma_1 \neq 0$), the researcher can estimate $b_1 = \beta_1$ when $\gamma_1 \neq 0$, but this is out of researcher's control.

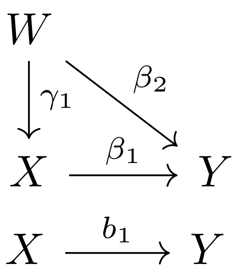

Fig. 3 The true relationship among W, X, and Y (top) and the assumed relationship without W (bottom)

3 Causal relationship with a mediator

Consider a more complex case when M is a mediating variable (observed) in the causal path from X to Y. Suppose a researcher assumes the mediation model (Hayes 2013) with the following two linear models:

$$\begin{aligned} M &= a_0 + a_1 X + \epsilon_1 \,, \\ Y &= b_0 + b_1 X + b_2 M + \epsilon_2 \,. \end{aligned} \tag{1}$$

Suppose the true relationship among the four random variables, W, X, M, and Y, is given by the three linear models

$$\begin{aligned} X &= \gamma_0 + \gamma_1 W + \epsilon_1^* \,, \\ M &= \alpha_0 + \alpha_1 X + \alpha_2 W + \epsilon_2^* \,, \\ Y &= \beta_0 + \beta_1 X + \beta_2 M + \beta_3 W + \epsilon_3^* \,. \end{aligned} \tag{2}$$

Figure 4 graphically illustrates this scenario. Let $\sigma_W^2 = V(W)$ and $\sigma_i^2 = V(\epsilon_i^*)$.

Under the assumed model of Fig. 4 (bottom), $a_1 b_2$ quantifies the indirect effect of X on Y (through M), and b_1 quantifies the direct effect (not through M, but possibly through other mediators). Our goal is to express $a_1 b_2$ and b_1 in terms of the model parameters in the true relationship among W, X, M, and Y which are denoted by the Greek letters in Fig. 4 (top).

3.1 The impact of an unobserved precursor on indirect effect

For the indirect effect $a_1 b_2$, we first note that a_1 and b_2 quantify the relationships among X, M, and Y as follow:

$$\begin{aligned} a_1 &= \frac{Cov(X,M)}{V(X)} \,, \\ b_2 &= \frac{V(X)Cov(M,Y) - Cov(X,Y)Cov(X,M)}{V(X)V(M) - [Cov(X,M)]^2} \,. \end{aligned}$$

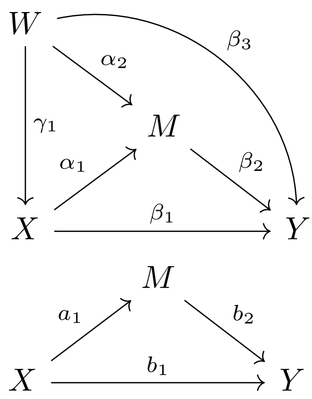

Fig. 4 The true relationship among W, X, M, and Y (top) and the assumed relationship without W (bottom)

Note that a_1 and b_2 can be expressed as

$$a_1 = \alpha_1 + \alpha_2\gamma_1\left(\frac{\sigma_W^2}{\gamma_1^2\sigma_W^2 + \sigma_1^2}\right),$$
$$b_2 = \beta_2 + \alpha_2\beta_3\left(\frac{\sigma_1^2\sigma_W^2}{\alpha_2^2\sigma_1^2\sigma_W^2 + (\gamma_1^2\gamma_W^2 + \sigma_1^2)\sigma_2^2}\right).$$

To this end, the researcher's estimand a_1b_2 for the indirect effect becomes a very complex quantity

$$a_1b_2 = \left(\frac{(\alpha_1 + \alpha_2\gamma_1)\sigma_W^2}{\gamma_1^2\sigma_W^2 + \sigma_1^2}\right)\left(\frac{(\beta_2 + \alpha_2\beta_3)\sigma_1^2\sigma_W^2}{\alpha_2^2\sigma_1^2\sigma_W^2 + (\gamma_1^2\gamma_W^2 + \sigma_1^2)\sigma_2^2}\right). \tag{3}$$

Appendix 1 provides a detail explanation of the derivation of Eq. (3).

There are two cases when $a_1b_2 = \alpha_1\beta_2$. The first case is when the precursor W does not affect the mediator M (i.e., $\alpha_2 = 0$). The second case is when W does not affect both explanatory variable X and response variable Y (i.e., $\gamma_1 = \beta_3 = 0$). In either case, randomization of X (i.e., $\gamma_1 = 0$) is not sufficient to avoid the bias. In particular, when $\alpha_1\beta_2 = 0$ is true with $\alpha_1 \neq 0$ and $\beta_2 = 0$, researchers will estimate $a_1b_2 \neq 0$ which leads to an inflated Type I error rate.

3.2 The impact on direct effect

For the direct effect b_1, it can be similarly shown that

$$b_1 = \frac{V(M)Cov(X,Y) - Cov(M,Y)Cov(X,M)}{V(X)V(M) - [Cov(X,M)]^2}.$$

In the true causal relationship, which starts from the precursor W, it can be shown that

$$b_1 = \beta_1 + \beta_3\left(\frac{\gamma_1\sigma_2^2\sigma_W^2 - \alpha_1\alpha_2\sigma_1^2\sigma_W^2}{\alpha_2^2\sigma_1^2\sigma_W^2 + (\gamma_1^2\sigma_W^2 + \sigma_1^2)\sigma_2^2}\right). \tag{4}$$

Appendix 2 provides a detail explanation of the derivation of Eq. (4).

There are four cases when $b_1 = \beta_1$. The first case is when W does not affect Y (i.e., $\beta_3 = 0$). The second case is when $\gamma_1 = \alpha_1 = 0$, and the third case is when $\gamma_1 = \alpha_2 = 0$. The fourth case is $\gamma_1\sigma_2^2\sigma_W^2 = \alpha_1\alpha_2\sigma_1^2\sigma_W^2$ which is nearly uninterpretable. Again, in any of the four cases, the randomization of X (i.e., $\gamma_1 = 0$) does not guarantee $b_1 = \beta_1$. In particular, when $\beta_1 = 0$ is true, researchers will estimate $b_1 \neq 0$ which leads to an inflated type I error rate. Furthermore, if $\gamma_1 = 0$ and all β_1, α_1, α_2, and β_3 have the same sign, depending on their magnitudes (and magnitudes of σ_W and σ_2), b_1 and β_1 may result in opposite signs which leads to an awkward conclusion.

4 Simulation study

4.1 Simulation designs

To demonstrate the danger of mediation analysis due to an omitted precursor W even with manipulation of X (i.e., $\gamma_1 = 0$), a simulation study was conducted. For all simulation scenarios, we fixed $\beta_1 = 0$, $\beta_2 = 0$, $\sigma_W = 5$, $\sigma_1 = 1$, $\sigma_2 = 5$, and $\sigma_3 = 0.5$ and varied $\alpha_1 = 0, 0.5, 1$, $\alpha_2 = -1, -0.5, 0$, and $\beta_3 = -0.5, 0, 0.5$ to create twenty-seven scenarios such that $\beta_1 = 0$ (no direct effect) and $\alpha_1\beta_2 = 0$ (no indirect effect). For each scenario, we considered sample sizes $n = 50, 100, 500, 1000$, and we estimated (1) the probability of concluding the presence of indirect effect (i.e., $a_1b_2 \neq 0$) and (2) the probability of concluding the presence of direct effect (i.e., $b_1 \neq 0$) based on bias-adjusted 95% confidence intervals (CIs) with 2000 bootstrap samples. The nonparametric method is known to be robust for the mediation analysis, so it is recommended for mediation analysis in practice (Hayes 2009; Preacher and Hayes 2008; Hair et al. 2014). Each scenario was replicated 1000 times for estimating the probability of concluding $a_1b_2 \neq 0$ and the probability of concluding $b_1 \neq 0$. In other words, out of 1000 replications per scenario, the proportion of times that a 95% CI for a_1b_2 excludes zero and the proportion of times that a 95% CI for b_1 excludes zero are calculated.

4.2 Simulation results

The simulation results are summarized in Table 1. In seven scenarios (10, 12, 15, 19, 21, 22, and 24), the probability of concluding $b_1 \neq 0$ and the probability of concluding $a_1b_2 \neq 0$ were substantially greater than 0.05 when $n = 50$, and the probabilities increased as n increased. The bias-adjusted method accurately estimated b_1 and a_1b_2 according to equations (4) and (3) (see Table 2), so there was a higher chance of eliminating $b_1 = 0$ and $a_1b_2 = 0$ as n increased (i.e., shorter CIs around the true values of b_1 and a_1b_2). For the other twenty scenarios, where $b_1 = 0$ and $a_1b_2 = 0$ according to Eqs. (4) and (3), the probabilities of concluding $b_1 \neq 0$ and $a_1b_2 \neq 0$ were close to 0.05. In some of the twenty scenarios, the probabilities slightly exceeded 0.05 when $n = 50$, and these results imply that the bootstrap method may require a larger sample size than $n = 50$ in some cases in order to properly estimate the uncertainty in the interval estimation.

5 Example

Guber (1999) discussed the impact of omitting a (confounding) variable in the association between public school expenditures and academic performance (measured by the average SAT score). In the data, an individual is a state (not a student), and the data consist of all 50 states in the United States.

Table 1 The probability of concluding the presence of direct effect ($b_1 \neq 0$) and the presence of indirect effect ($a_1 b_2 \neq 0$) in the absence of W based on biased-adjusted 95% CIs with 2000 bootstrap samples (1000 replications per scenario)

	Parameters ($\beta_1 = \beta_2 = 0$)					Probability of concluding $b_1 \neq 0$				Probability of concluding $a_1 b_2 \neq 0$			
	α_1	α_2	β_3	b_1	$a_1 b_2$	$n = 50$	$n = 100$	$n = 500$	$n = 1000$	$n = 50$	$n = 100$	$n = 500$	$n = 1000$
1	0.0	− 1.0	− 0.5	0.000	0.000	0.077	0.064	0.045	0.049	0.074	0.059	0.062	0.047
2	0.0	− 1.0	0.0	0.000	0.000	0.074	0.050	0.048	0.055	0.014	0.004	0.005	0.012
3	0.0	− 1.0	0.5	0.000	0.000	0.054	0.069	0.047	0.050	0.072	0.055	0.053	0.057
4	0.0	− 0.5	− 0.5	0.000	0.000	0.073	0.065	0.052	0.059	0.077	0.091	0.040	0.059
5	0.0	− 0.5	0.0	0.000	0.000	0.074	0.058	0.046	0.042	0.003	0.007	0.006	0.005
6	0.0	− 0.5	0.5	0.000	0.000	0.076	0.062	0.057	0.064	0.074	0.068	0.066	0.057
7	0.0	0.0	− 0.5	0.000	0.000	0.058	0.047	0.049	0.051	0.008	0.008	0.002	0.004
8	0.0	0.0	0.0	0.000	0.000	0.052	0.051	0.054	0.052	0.014	0.007	0.010	0.005
9	0.0	0.0	0.5	0.000	0.000	0.066	0.065	0.047	0.040	0.009	0.005	0.005	0.007
10	0.5	− 1.0	− 0.5	0.125	− 0.125	0.089	0.100	0.327	0.589	0.103	0.118	0.350	0.617
11	0.5	− 1.0	0.0	0.000	0.000	0.065	0.075	0.056	0.054	0.015	0.010	0.038	0.051
12	0.5	− 1.0	0.5	− 0.125	0.125	0.097	0.119	0.321	0.608	0.108	0.135	0.376	0.636
13	0.5	− 0.5	− 0.5	0.100	− 0.100	0.073	0.086	0.165	0.276	0.138	0.195	0.530	0.802
14	0.5	− 0.5	0.0	0.000	0.000	0.063	0.050	0.047	0.055	0.016	0.012	0.052	0.066
15	0.5	− 0.5	0.5	− 0.100	0.100	0.083	0.079	0.177	0.275	0.142	0.200	0.524	0.813
16	0.5	0.0	− 0.5	0.000	0.000	0.069	0.066	0.048	0.048	0.014	0.011	0.049	0.063
17	0.5	0.0	0.0	0.000	0.000	0.061	0.054	0.052	0.054	0.008	0.018	0.056	0.070
18	0.5	0.0	0.5	0.000	0.000	0.061	0.052	0.046	0.051	0.012	0.017	0.048	0.071
19	1.0	− 1.0	− 0.5	0.250	− 0.250	0.177	0.295	0.873	0.993	0.202	0.317	0.876	0.993
20	1.0	− 1.0	0.0	0.000	0.000	0.058	0.054	0.049	0.047	0.014	0.036	0.060	0.064
21	1.0	− 1.0	0.5	− 0.250	0.250	0.176	0.252	0.859	0.998	0.223	0.337	0.879	0.992
22	1.0	− 0.5	− 0.5	0.200	− 0.200	0.103	0.140	0.476	0.777	0.282	0.487	0.979	1.000

Table 1 (continued)

	Parameters ($\beta_1 = \beta_2 = 0$)					Probability of concluding $b_1 \neq 0$				Probability of concluding $a_1 b_2 \neq 0$			
	α_1	α_2	β_3	b_1	$a_1 b_2$	$n = 50$	$n = 100$	$n = 500$	$n = 1000$	$n = 50$	$n = 100$	$n = 500$	$n = 1000$
23	1.0	− 0.5	0.0	0.000	0.000	0.069	0.061	0.047	0.053	0.019	0.044	0.078	0.057
24	1.0	− 0.5	0.5	− 0.200	0.200	0.104	0.141	0.487	0.769	0.282	0.501	0.988	1.000
25	1.0	0.0	− 0.5	0.000	0.000	0.060	0.064	0.051	0.058	0.028	0.051	0.070	0.073
26	1.0	0.0	0.0	0.000	0.000	0.064	0.047	0.046	0.058	0.038	0.046	0.054	0.066
27	1.0	0.0	0.5	0.000	0.000	0.058	0.049	0.053	0.049	0.032	0.042	0.084	0.075

In the bold scenarios (10, 12, 15, 19, 21, 22, and 24), the probability of concluding the presence of direct effect and indirect effect increases as the sample size increases

Table 2 The mean of the sampling distribution of bootstrap estimates for the direct effect (b_1) and for the indirect effect ($a_1 b_2$) based on 2000 bootstrap samples (1000 replications per scenario)

	Parameters ($\beta_1 = \beta_2 = 0$)					$E(\hat{b}_1)$				$E(\hat{a}_1 \hat{b}_2)$			
	α_1	α_2	β_3	b_1	$a_1 b_2$	$n = 50$	$n = 100$	$n = 500$	$n = 1000$	$n = 50$	$n = 100$	$n = 500$	$n = 1000$
1	0.0	− 1.0	− 0.5	0.000	0.000	0.005	0.000	− 0.005	0.002	− 0.015	− 0.001	0.002	0.001
2	0.0	− 1.0	0.0	0.000	0.000	0.000	0.000	− 0.000	0.000	− 0.001	− 0.003	− 0.000	− 0.000
3	0.0	− 1.0	0.5	0.000	0.000	0.002	0.002	− 0.001	0.002	− 0.002	− 0.001	− 0.003	0.004
4	0.0	− 0.5	− 0.5	0.000	0.000	− 0.004	0.000	0.001	0.000	− 0.004	− 0.002	0.004	− 0.002
5	0.0	− 0.5	0.0	0.000	0.000	− 0.000	0.000	− 0.000	− 0.000	0.002	0.000	0.000	− 0.000
6	0.0	− 0.5	0.5	0.000	0.000	− 0.001	0.000	− 0.001	0.001	− 0.005	− 0.001	− 0.004	0.002
7	0.0	0.0	− 0.5	0.000	0.000	0.000	− 0.001	− 0.000	− 0.000	0.023	0.001	− 0.005	0.001
8	0.0	0.0	0.0	0.000	0.000	− 0.001	0.000	0.000	− 0.000	− 0.004	0.000	− 0.002	− 0.000
9	0.0	0.0	0.5	0.000	0.000	− 0.004	− 0.000	− 0.000	− 0.000	− 0.017	0.005	− 0.008	− 0.002
10	0.5	− 1.0	− 0.5	0.125	− 0.125	0.129	0.119	0.121	0.124	− 0.129	− 0.114	− 0.125	− 0.126
11	0.5	− 1.0	0.0	0.000	0.000	− 0.000	− 0.000	− 0.000	− 0.000	0.002	− 0.000	− 0.000	0.000
12	0.5	− 1.0	0.5	− 0.125	0.125	− 0.118	− 0.133	− 0.126	− 0.126	0.128	0.136	0.121	0.127
13	0.5	− 0.5	− 0.5	0.100	− 0.100	0.097	0.102	0.100	0.099	− 0.104	− 0.098	− 0.098	− 0.100
14	0.5	− 0.5	0.0	0.000	0.000	− 0.000	0.000	0.000	− 0.000	− 0.005	− 0.001	0.001	− 0.000
15	0.5	− 0.5	0.5	− 0.100	0.100	− 0.103	− 0.100	− 0.100	− 0.100	0.106	0.091	0.098	0.099
16	0.5	0.0	− 0.5	0.000	0.000	0.001	− 0.000	0.000	− 0.000	− 0.002	− 0.000	0.004	− 0.004
17	0.5	0.0	0.0	0.000	0.000	− 0.001	0.000	− 0.000	− 0.000	− 0.002	− 0.001	− 0.001	0.001
18	0.5	0.0	0.5	0.000	0.000	− 0.001	− 0.001	− 0.000	− 0.000	0.001	0.007	− 0.004	0.002
19	1.0	− 1.0	− 0.5	0.250	− 0.250	0.249	0.253	0.249	0.251	− 0.244	− 0.251	− 0.255	− 0.250
20	1.0	− 1.0	0.0	0.000	0.000	0.001	− 0.000	− 0.000	− 0.000	− 0.000	0.002	0.001	− 0.000
21	1.0	− 1.0	0.5	− 0.250	0.250	− 0.249	− 0.252	− 0.246	− 0.252	0.247	0.245	0.248	0.254
22	1.0	− 0.5	− 0.5	0.200	− 0.200	0.199	0.202	0.201	0.201	− 0.192	− 0.201	− 0.198	− 0.199

Table 2 (continued)

	Parameters ($\beta_1 = \beta_2 = 0$)					$E(\hat{b}_1)$				$E(\hat{a}_1\hat{b}_2)$			
	α_1	α_2	β_3	b_1	a_1b_2	$n = 50$	$n = 100$	$n = 500$	$n = 1000$	$n = 50$	$n = 100$	$n = 500$	$n = 1000$
23	1.0	− 0.5	0.0	0.000	0.000	0.000	− 0.000	− 0.000	0.000	− 0.003	0.002	0.001	0.000
24	1.0	− 0.5	0.5	− 0.200	0.200	− 0.199	− 0.201	− 0.205	− 0.201	0.191	0.199	0.201	0.199
25	1.0	0.0	− 0.5	0.000	0.000	− 0.001	− 0.004	− 0.000	− 0.000	0.017	− 0.009	− 0.006	0.001
26	1.0	0.0	0.0	0.000	0.000	− 0.000	0.000	0.000	− 0.000	− 0.001	− 0.002	0.001	− 0.000
27	1.0	0.0	0.5	0.000	0.000	− 0.001	− 0.000	− 0.000	− 0.000	0.001	− 0.002	− 0.004	0.001

In this section, to demonstrate the role of a potential precursor variable, we turn our focus on the association between the state average annual salary of teachers in public schools (denoted X; in thousands of US dollars) and the state average SAT score (denoted Y). If we consider the simple linear regression $Y = c_0 + c_1 X + \epsilon$, the estimated slope is $\hat{c}_1 = -5.5396$. This result suggests a higher average annual salary is associated with a lower average SAT performance ($p = 0.001$).

An important (potential mediating) variable may be the percentage (%) of students taking the SAT in each state (denoted by M). Suppose we model $M = a_0 + a_1 X + \epsilon_1$ and $Y = b_0 + b_1 X + b_2 M + \epsilon_2$ as shown in Eq. (1). The estimated regression parameters are $\hat{a}_1 = 2.7783$, $\hat{b}_1 = 2.1804$, and $\hat{b}_2 = -2.7787$. Conditioning on the potential mediating variable (% SAT takers), it appears that a higher average annual salary is associated with a higher average SAT performance ($p = 0.039$). The opposite signs between $\hat{c}_1 = -5.54$ and $\hat{b}_1 = 2.18$ are due to the fact $\hat{c}_1 = \hat{b}_1 + \hat{a}_1 \hat{b}_2$, where $\hat{a}_1 > 0$ and $\hat{b}_2 < 0$. More SAT takers tend to lower the state average SAT score (Guber 1999).

Given the adjusted estimate $\hat{b}_1 = 2.18$ with the small p value ($p = 0.039$), can we conclude that the state average annual salary and the state average SAT performance are positively associated? Let us consider a potential precursor variable, the state expenditure per students in public schools (denoted W; in thousands of US dollars), as W may affect X, M, and Y. Using the three regression models presented in Eq. (2), we can estimate the regression parameters as shown at the bottom of Fig. 5 with $\hat{\sigma}_W^2 = 1.8201$, $\hat{\sigma}_1^2 = 8.4214$, and $\hat{\sigma}_2^2 = 425.7959$. Note that $\hat{\beta}_1 = -0.31$ with $p = 0.853$ suggest no strong evidence for the positive association. The opposite signs between $\hat{b}_1 = 2.18$ and $\hat{\beta}_1 = -0.31$ can be explained by Eq. (4) with the estimated regression parameters,

$$\hat{b}_1 = \hat{\beta}_1 + \hat{\beta}_3 \left(\frac{\hat{\gamma}_1 \hat{\sigma}_2^2 \hat{\sigma}_W^2 - \hat{\alpha}_1 \hat{\alpha}_2 \hat{\sigma}_1^2 \hat{\sigma}_W^2}{\hat{\alpha}_2^2 \hat{\sigma}_1^2 \hat{\sigma}_W^2 + (\hat{\gamma}_1^2 \hat{\sigma}_W^2 + \hat{\sigma}_1^2) \hat{\sigma}_2^2} \right).$$

The relatively big positive estimates $\hat{\beta}_3 = 13.33$, $\hat{\gamma}_1 = 3.79$, and $\hat{\sigma}_2^2 = 425.8$ could alter from the negative $\hat{\beta}_1 = -0.31$ (or nearly zero) to positive $\hat{b}_1 = 2.18$ (with the small p-value) by omitting the precursor variable (state expenditure) which might affect the whole mechanism among the state average annual salary of teachers, the % students taking SAT, and the state SAT performance. The estimated indirect association $\hat{\alpha}_1 \hat{\beta}_2 = -5.32$ further suggests that a higher state average salary of teachers does not help improving the average SAT performance.

6 Discussion

The primary focus of the simulation study was an inflated Type I error rate (i.e., a higher chance of falsely claiming direct and/or indirect effect) in a mediation analysis due to an unobserved precursor variable. Though the example in Sect. 5 was an observational study, the message was similar. Even for an experimental study, the simulation results alert researchers about the inflated Type I error rate

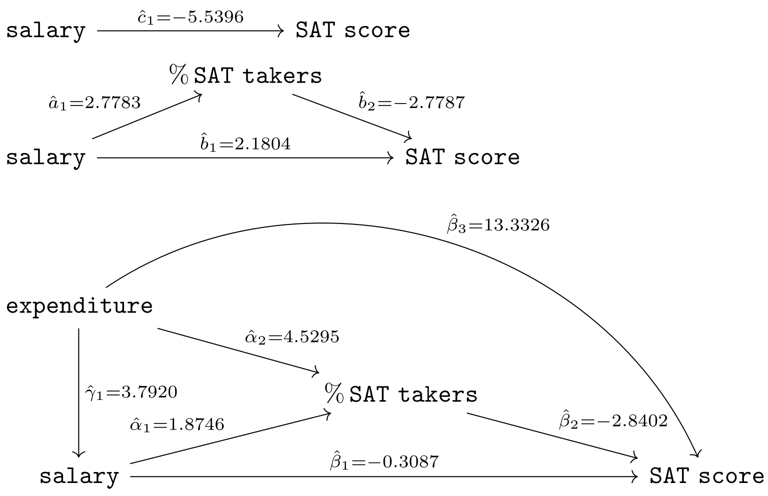

Fig. 5 Changing the direction and magnitude of association between the state average salary (explanatory) and the state average SAT score (response) by including % SAT takers (mediator) and by including state expenditure (precursor) in the analysis. (This is a state-level association, not student-level.)

(worse when n is larger), and the expressions (Eqs. (3) and (4)) explain why the bias exists even after randomization of X (i.e., forcing $\gamma_1 = 0$) in the presence of a mediator (Fig. 4). The uncomfortable fact is that an experimental design, which can change $V(X) = \gamma_1^2\sigma_W^2 + \sigma_1^2 = \sigma_1^2$, cannot fix the bias in particular null cases. For instance, if $\alpha_1 \neq 0$ and $\beta_2 = 0$ (i.e., zero indirect effect), according to Eq. (3),

$$a_1 b_2 = \alpha_1\alpha_2\beta_3\left(\frac{\sigma_1^2\sigma_W^2}{\alpha_2^2\sigma_1^2\sigma_W^2 + \sigma_1^2\sigma_2^2}\right) = \alpha_1\alpha_2\beta_3\left(\frac{\sigma_W^2}{\alpha_2^2\sigma_W^2 + \sigma_2^2}\right)$$

which is independent of $V(X) = \sigma_1^2$. For another instance, if $\beta_1 = 0$ (i.e., zero direct effect), according to equation (4),

$$b_1 = \beta_3\left(\frac{\gamma_1\sigma_2^2\sigma_W^2 - \alpha_1\alpha_2\sigma_1^2\sigma_W^2}{\alpha_2^2\sigma_1^2\sigma_W^2 + (\gamma_1^2\sigma_W^2 + \sigma_1^2)\sigma_2^2}\right) = -\beta_3\left(\frac{\alpha_1\alpha_2\sigma_W^2}{\alpha_2^2\sigma_W^2 + \sigma_2^2}\right)$$

which is independent of $V(X) = \sigma_1^2$ again.

In this article, we focused on one precursor variable. In practice, there can be two or more precursor variables. The take-home messages are clear for reducing bias in the estimation of $\alpha_1\beta_2$ (indirect effect) and β_1 (direct effect). First, randomize X (i.e., $\gamma_1 = 0$) if possible. Second, during data collection, researchers are suggested to record variables which are potentially related to M and Y and adjust them in the regression analysis. A large sample size can tolerate a mild degree of over-fitting due to many W's adjusted in the model. The adjustment will help researchers estimate $\alpha_1\beta_2$ and β_1 with a small bias and perform hypothesis testing with a reduced inflation of Type I error rate.

Appendix 1: Bias in the estimation of indirect effect

Under the assumed model $M = a_0 + a_1X + \epsilon_1$, the covariance between X and M is given by

$$Cov(X, M) = Cov(X, a_1X) = a_1V(X)$$

so,

$$a_1 = \frac{Cov(X, M)}{V(X)}. \tag{5}$$

The assumed model $Y = b_0 + b_1X + b_2M + \epsilon_2$ can be written as $Y - b_1X = b_0 + b_2M + \epsilon_2$, so

$$b_2 = \frac{Cov(M, Y - b_1X)}{V(M)} = \frac{Cov(Y, M) - b_1Cov(X, M)}{V(M)}. \tag{6}$$

Similarly, it can be written as $Y - b_2M = b_0 + b_1X + \epsilon_2$, so

$$b_1 = \frac{Cov(X, Y - b_2M)}{V(X)} = \frac{Cov(X, Y) - b_2Cov(X, M)}{V(X)}.$$

Therefore, Eq. (6) can be written as

$$\begin{aligned} b_2 &= \frac{Cov(Y, M) - \left(\frac{Cov(X,Y) - b_2Cov(X,M)}{V(X)}\right)Cov(X, M)}{V(M)} \\ &= \frac{V(X)Cov(Y, M) - Cov(X, Y)Cov(X, M) + b_2[Cov(X, M)]^2}{V(X)V(M)}. \end{aligned}$$

By solving for b_2,

$$b_2 = \frac{V(X)Cov(Y, M) - Cov(X, Y)Cov(X, M)}{V(X)V(M) - [Cov(X, M)]^2}. \tag{7}$$

Now consider the three true models:

$$\begin{aligned} X &= \gamma_0 + \gamma_1W + \epsilon_1^*, \\ M &= \alpha_0 + \alpha_1X + \alpha_2W + \epsilon_2^*, \\ Y &= \beta_0 + \beta_1X + \beta_2M + \beta_3W + \epsilon_3^*. \end{aligned}$$

From the true relationships among W, X, M, and Y, we have

$$V(X) = V(\gamma_1W + \epsilon_1^*) = \gamma_1^2\sigma_W^2 + \sigma_1^2 \tag{8}$$

and

$$\begin{aligned} Cov(X, M) &= Cov(X, \alpha_1 X + \alpha_2 W) \\ &= \alpha_1 V(X) + \alpha_2 Cov(X, W) \\ &= \alpha_1(\gamma_1^2 \sigma_W^2 + \sigma_1^2) + \alpha_2 \gamma_1 \sigma_W^2 \,. \end{aligned} \tag{9}$$

Therefore, Eq. (5) can be expressed as the true model parameters as

$$a_1 = \alpha_1 + \frac{\alpha_2 \gamma_1 \sigma_W^2}{\gamma_1^2 \sigma_W^2 + \sigma_1^2} \,.$$

To express b_2 in Eq. (7) in terms of the true model parameters, we need to rewrite $V(M)$, $Cov(Y, M)$, and $Cov(X, Y)$ as follows. For $V(M)$, we first express

$$\begin{aligned} M &= \alpha_0 + \alpha_1(\gamma_0 + \gamma_1 W + \epsilon_1^*) + \alpha_2 W + \epsilon_2^* \\ &= (\alpha_0 + \alpha_1 \gamma_0) + (\alpha_1 \gamma_1 + \alpha_2) W + \alpha_1 \epsilon_1^* + \epsilon_2^* \,, \end{aligned}$$

so

$$V(M) = (\alpha_1 \gamma_1 + \alpha_2) \sigma_W^2 + \alpha_1^2 \sigma_1^2 + \sigma_2^2 \,. \tag{10}$$

For $Cov(Y, M)$, note that

$$\begin{aligned} Cov(Y, M) &= Cov(M, Y) \\ &= Cov(M, \beta_1 X + \beta_2 M + \beta_3 W) \\ &= \beta_1 Cov(M, X) + \beta_2 V(M) + \beta_3 Cov(M, W) \,, \end{aligned}$$

where we previously wrote

$$\begin{aligned} Cov(X, M) &= \alpha_1 \sigma_1^2 + (\alpha_1 \gamma_1 + \alpha_2) \gamma_1 \sigma_W^2 \,, \\ V(M) &= (\alpha_1 \gamma_1 + \alpha_2) \sigma_W^2 + \alpha_1^2 \sigma_1^2 + \sigma_2^2 \,. \end{aligned}$$

Further note that $Cov(X, W) = Cov(\gamma_1 W, W) = \gamma_1 \sigma_W^2$, and

$$\begin{aligned} Cov(M, W) &= Cov(\alpha_1 X + \alpha_2 W, W) \\ &= \alpha_1 Cov(X, W) + \alpha_2 \sigma_W^2 \\ &= \alpha_1 \gamma_1 \sigma_W^2 + \alpha_2 \sigma_W^2 \\ &= (\alpha_1 \gamma_1 + \alpha_2) \sigma_W^2 \,. \end{aligned}$$

Therefore,

$$\begin{aligned} Cov(Y, M) = \beta_1 &[\alpha_1 \sigma_1^2 + (\alpha_1 \gamma_1 + \alpha_2) \gamma_1 \sigma_W^2] \\ &+ \beta_2 [(\alpha_1 \gamma_1 + \alpha_2) \sigma_W^2 + \alpha_1^2 \sigma_1^2 + \sigma_2^2] \\ &+ \beta_3 [(\alpha_1 \gamma_1 + \alpha_2) \sigma_W^2] \,. \end{aligned} \tag{11}$$

For $Cov(X, Y)$, we can replace our previous results as

$$\begin{aligned}Cov(X,Y) &= Cov(X, \beta_1 X + \beta_2 M + \beta_3 W)\\ &= \beta_1 V(X) + \beta_2 Cov(X,M) + \beta_3 Cov(X,W)\\ &= \beta_1(\gamma_1^2\sigma_W^2 + \sigma_1^2) + \beta_2[\alpha_1\sigma_1^2 + (\alpha_1\gamma_1 + \alpha_2)\gamma_1\sigma_W^2]\\ &+ \beta_3\gamma_1\sigma_W^2 \,. \end{aligned} \tag{12}$$

After some algebraic work, it can be shown that the denominator of b_2 in Eq. (7) can be simplified as

$$V(X)V(M) - [Cov(X,M)]^2 = \alpha_2^2\sigma_1^2\sigma_W^2 + (\gamma_1^2\sigma_W^2 + \sigma_1^2)\sigma_2^2 \,,$$

and the numerator of b_2 in Eq. (7) can be expressed as

$$\begin{aligned}V(X)Cov(Y,M) - Cov(X,Y)Cov(X,M) = \beta_2[\alpha_2\sigma_1^2\sigma_W^2 + (\gamma_1^2\sigma_W^2 + \sigma_1^2)\sigma_2^2]\\ + \beta_3\alpha_2\sigma_1^2\sigma_W^2 \,.\end{aligned}$$

To this end, we can express b_2 in Eq. (7) as

$$b_2 = \frac{\beta_2[\alpha_2\sigma_1^2\sigma_W^2 + (\gamma_1^2\sigma_W^2 + \sigma_1^2)\sigma_2^2] + \beta_3\alpha_2\sigma_1^2\sigma_W^2}{\alpha_2^2\sigma_1^2\sigma_W^2 + (\gamma_1^2\sigma_W^2 + \sigma_1^2)\sigma_2^2}$$

which simplifies as

$$b_2 = \beta_2 + \beta_3\left(\frac{\alpha_2\sigma_1^2\sigma_W^2}{\alpha_2^2\sigma_1^2\sigma_W^2 + (\gamma_1^2\sigma_W^2 + \sigma_1^2)\sigma_2^2}\right).$$

Therefore, due to an unobserved precursor variable W, researchers would estimate a_1b_2 which is equal to

$$a_1b_2 = \left[\alpha_1 + \alpha_2\gamma_1\frac{\sigma_W^2}{\gamma_1^2\sigma_W^2 + \sigma_1^2}\right]\left[\beta_2 + \alpha_2\beta_3\left(\frac{\sigma_1^2\sigma_W^2}{\alpha_2^2\sigma_1^2\sigma_W^2 + (\gamma_1^2\sigma_W^2 + \sigma_1^2)\sigma_2^2}\right)\right].$$

which is not equal to $\alpha_1\beta_2$ in general.

Appendix 2: Bias in the estimation of direct effect

From Eqs. (6) and (7),

$$
\begin{aligned}
b_1 &= \frac{Cov(X,Y) - b_2 Cov(X,M)}{V(X)} \\
&= \frac{Cov(X,Y) - \left(\frac{Cov(Y,M) - b_1 Cov(X,M)}{V(M)}\right) Cov(X,M)}{V(X)} \\
&= \frac{V(M)Cov(X,Y) - Cov(Y,M)Cov(X,M) + b_1 [Cov(X,M)]^2}{V(X)V(M)},
\end{aligned}
$$

which can be expressed as

$$
b_1 = \frac{V(M)Cov(X,Y) - Cov(Y,M)Cov(X,M)}{V(X)V(M) - [Cov(X,M)]^2}.
$$

Recall Eqs. (10), (8), (12), (11), and (9) for each term of b_1. After some algebraic work, the numerator of b_1 can be simplified as

$$
\begin{aligned}
&V(M)Cov(X,Y) - Cov(Y,M)Cov(X,M) = \beta_1[\alpha_2^2\sigma_1^2\sigma_W^2 + (\gamma_1^2\sigma_W^2 + \sigma_1^2)\sigma_2^2] \\
&+ \beta_3(\gamma_1\sigma_2^2\sigma_W^2 - \alpha_1\alpha_2\sigma_1^2\sigma_W^2),
\end{aligned}
$$

and the denominator of b_1 can be simplified as

$$
V(X)V(M) - [Cov(X,M)]^2 = \alpha_2^2\sigma_1^2\sigma_W^2 + (\gamma_1^2\sigma_W^2 + \sigma_1^2)\sigma_2^2,
$$

To this end, we can express b_1 as

$$
b_1 = \frac{\beta_1(\gamma_1^2\sigma_2^2\sigma_W^2 + \alpha_2^2\sigma_1^2\sigma_W^2 + \sigma_1^2\sigma_2^2) + \beta_3(\gamma_1\sigma_2^2\sigma_W^2 - \alpha_1\alpha_2\sigma_1^2\sigma_W^2)}{\alpha_2^2\sigma_1^2\sigma_W^2 + (\gamma_1^2\sigma_W^2 + \sigma_1^2)\sigma_2^2}
$$

which simplifies as

$$
b_1 = \beta_1 + \beta_3\left(\frac{\gamma_1\sigma_2^2\sigma_W^2 - \alpha_1\alpha_2\sigma_1^2\sigma_W^2}{\alpha_2^2\sigma_1^2\sigma_W^2 + (\gamma_1^2\sigma_W^2 + \sigma_1^2)\sigma_2^2}\right),
$$

and it is not equal to β_1 in general.

References

Adams, R. C., Challenger, A., Bratton, L., Boivin, J., Bott, L., Powell, G., et al. (2019). Claims of causality in health news: A randomised trial. *BMC Medicine*, *17*(1), 91. https://doi.org/10.1186/s12916-019-1324-7.

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*(6), 1173–1182.

Caniëls, M. C. (2019). Proactivity and supervisor support in creative process engagement. *European Management Journal*, *37*(2), 188–197.

Emery, C., Booth, J. E., Michaelides, G., & Swaab, A. J. (2019). The importance of being psychologically empowered: Buffering the negative effects of employee perceptions of Leader–Member Exchange differentiation. *Journal of Occupational and Organizational Psychology*, *92*(3), 566–592.

Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.

Garcia, P. R. J. M., Restubog, S. L. D., Ocampo, A. C., Wang, L., & Tang, R. L. (2018). Role modeling as a socialization mechanism in the transmission of career adaptability across generations. *Journal of Vocational Behavior*, *111*, 39–48.

Glass, T. A., Goodman, S. N., Hernán, M. A., & Samet, J. M. (2013). Causal inference in public health. *Annual Review of Public Health*, *34*, 61–75.

Guber, D. L. (1999). Getting what you pay for: The debate over equity in public school expenditures. *Journal of Statistics Education,7*(2).

Hair J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2014). A primer on partial least squares structural equation modeling (PLS SEM).

Hall, N. S. (2007). R. A. Fisher and his advocacy of randomization. *Journal of the History of Biology*, *40*, 295–325.

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, *76*(4), 408–420.

Hayes, A. F. (2013). *Methodology in the social sciences. Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Guilford Press.

Hong, G., Deutsch, J., & Hill, H. D. (2015). Ratio-of-mediator-probability weighting for causal mediation analysis in the presence of treatment-by-mediator interaction. *Journal of Educational and Behavioral Statistics*, *40*(3), 307–340.

Hong, G., Qin, X., & Yang, F. (2018). Weighting-based sensitivity analysis in causal mediation studies. *Journal of Educational and Behavioral Statistics*, *43*(1), 32–56.

Imai, K., Keele, L., & Tingley, D. (2010a). A general approach to causal mediation analysis. *Psychological Methods*, *15*(4), 309–334.

Imai, K., Keele, L., & Yamamoto, T. (2010b). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, *25*, 51–71.

Kang, J. (2014). Overview and practice of causal inference in observational studies. *Biometrics & Biostatistics International Journal*, *1*(1), 00002.

Manuti, A., & Giancaspro, M. (2019). People make the difference: An explorative study on the relationship between organizational practices, employees' resources, and organizational behavior enhancing the psychology of sustainability and sustainable development. *Sustainability*, *11*(5), 1499.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879–891.

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, *1*(1), 27–42.

Seli, P., Schacter, D. L., Risko, E. F., & Smilek, D. (2017). Increasing participant motivation reduces rates of intentional and unintentional mind wandering. *Psychological Research*, *83*(5), 1057–1069.

VanderWeele, T. J. (2010). Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology*, *21*(4), 540–551.

VanderWeele, T. J. (2015). *Explanation in causal inference: Methods for mediation and interaction*. New York: Oxford University Press.

Villaluz, V., & Hechanova, M. (2018). Ownership and leadership in building an innovation culture. *Leadership & Organization Development Journal*, *40*(2), 138–150.

Zhang, L., Fan, C., Deng, Y., Lam, C. F., Hu, E., & Wang, L. (2019). Exploring the interpersonal determinants of job embeddedness and voluntary turnover: A conservation of resources perspective. *Human Resource Management Journal*, *29*(3), 413–432.
Zhu, H., Wong, N., & Huang, M. (2019). Does relationship matter? How social distance influences perceptions of responsibility on anthropomorphized environmental objects and conservation intentions. *Journal of Business Research*, *95*, 62–70.