Supplemental A

Derivation Details for Asymptotic Bias and Sensitivity Analysis

We organize this supplement into two main parts. In the first part, we provide details for deriving the asymptotic bias one or multiple unobserved mediators (M_U) may present in estimating indirect and direct effects. As described in the manuscript, we focus on the *specific* indirect effect via M_O alone and the direct effect from X to Y. The Law of Iterated Expectations (LIE) is integral to the derivation. We first introduce the true model and then write the true model in terms of the conditional mean. Afterward, we derive several correlations that are useful in later derivation. Next, we apply the LIE to derive the asymptotic bias as a function of parameters in the true model. We also provide the formulas for bias as a function of correlations. To follow, we derive the error variances and constraints of parameters and then discuss the directions of asymptotic bias and how asymptotic bias changes with some parameters. We also present the derivation for omitting multiple unobserved mediators and the combined effect of measurement error and an omitted M_U. All variables are standardized in this derivation.

The second part presents details about extending Frank's (2000) Impact Threshold for Confounder Variable (ITCV) into mediation models. We first provide an overview of how Frank's (2000) ITCV approach works in a linear regression model and then present how to incorporate measurement error into the sensitivity analysis. For more details of ITCV in linear models, see Frank (2000) and Xu et al. (2019). The main text has also discussed how to extend ITCV into mediation models using a joint significance test.

The True Model (zero measurement error)

$$M_{\mathrm{O}} = k \cdot M_{\mathrm{U}} + a_1 \cdot X + \epsilon_{M_{\mathrm{O}}} \tag{A1.1}$$

$$M_{\mathrm{U}} = a_2 \cdot X + \epsilon_{M_{\mathrm{U}}} \tag{A1.2}$$

$$Y = b_1 \cdot M_{\mathrm{O}} + b_2 \cdot M_{\mathrm{U}} + c \cdot X + \epsilon_Y \tag{A1.3}$$

True Model in Terms of the Conditional Mean

$$E(M_{\mathrm{O}}|M_{\mathrm{U}}, X) = k \cdot M_{\mathrm{U}} + a_1 \cdot X \tag{A2.1}$$

$$E(M_{\mathrm{U}}|X) = a_2 \cdot X \tag{A2.2}$$

$$E(Y|M_{\mathrm{O}}, M_{\mathrm{U}}, X) = b_1 \cdot M_{\mathrm{O}} + b_2 \cdot M_{\mathrm{U}} + c \cdot X \tag{A2.3}$$

Correlations

***Correlation Between X and M*U**

From Equation A1.2,

$$cov(X, M_{\mathrm{U}}) = a_2 \cdot var(X) \tag{A3.1}$$

Assuming all variables are standardized, then:

$$\rho_{X,M_{\mathrm{U}}} = a_2 \tag{A3.2}$$

***Correlation Between X and M*O**

From Equation A1.1,

$$cov(X, M_{\mathrm{O}}) = k \cdot cov(X, M_{\mathrm{U}}) + a_1 \cdot var(X) \tag{A3.3}$$

Assuming all variables are standardized, then:

$$\rho_{X,M_{\mathrm{O}}} = ka_2 + a_1 \tag{A3.4}$$

***Correlation Between M*O *and M*U**

From Equation A1.1,

$$cov(M_{\mathrm{O}}, M_{\mathrm{U}}) = k \cdot var(M_{\mathrm{U}}) + a_1 \cdot cov(X, M_{\mathrm{U}}) \tag{A3.5}$$

Assuming all variables are standardized, then:

$$\rho_{M_{\mathrm{O}},M_{\mathrm{U}}} = k + a_1 a_2 \tag{A3.6}$$

***Correlation Between Y and M*U**

From Equation A1.3,

$$cov(M_U, Y) = b_1 \cdot cov(M_U, M_O) + b_2 \cdot var(M_U) + c \cdot cov(M_U, X) \tag{A3.7}$$

Assuming all variables are standardized, then:

$$\rho_{Y,M_U} = b_1 k + b_1 a_1 a_2 + b_2 + c a_2 \tag{A3.8}$$

Asymptotic Bias if Omitting M_U

According to the LIE, the true model excluding M_U can be written as follows:

$$E(M_O|X) = E[E(M_O|M_U, X)|X] = k \cdot E(M_U|X) + a_1 \cdot X \tag{A4.1}$$

$$E(Y|M_O, X) = E[E(Y|M_O, M_U, X)|M_O, X] = b_1 \cdot M_O + b_2 \cdot E(M_U|M_O, X) + c \cdot X \tag{A4.2}$$

Now write:

$$E(M_U|X) = \beta_1 \cdot X \tag{A4.3}$$

$$E(M_U|M_O, X) = \beta_2 \cdot M_O + \beta_3 \cdot X \tag{A4.4}$$

Then, we get:

$$E(M_O|X) = k\beta_1 \cdot X + a_1 \cdot X \tag{A4.5}$$

$$E(Y|M_O, X) = b_1 \cdot M_O + b_2 \cdot \beta_2 \cdot M_O + b_2 \cdot \beta_3 \cdot X + c \cdot X \tag{A4.6}$$

Therefore, the omission of M_U results in the following asymptotically biased estimates:

$$\tilde{a}_1 = a_1 + k\beta_1 \tag{A4.7}$$

$$\tilde{b}_1 = b_1 + b_2\beta_2 \tag{A4.8}$$

$$\tilde{c} = c + b_2\beta_3 \tag{A4.9}$$

Above, $k\beta_1$ represents the bias for $\tilde{a}_1$, and $b_2\beta_2$ represents the bias for $\tilde{b}_1$. Accordingly, the bias for the indirect effect $\tilde{a}_1\tilde{b}_1$ is equal to: $[(a_1 + k\beta_1) \cdot (b_1 + b_2\beta_2) - a_1 b_1]$. Lastly, $b_2\beta_3$ represents the bias for $\tilde{c}$.

We now provide the derivations for β_1, β_2, and β_3. First, based on Equation A4.3, β_1 represents the regression coefficient of X when we regress M_U on X (at the population level). Therefore,

$$\beta_1 = \rho_{X,M_\mathrm{U}} = a_2 \tag{A4.10}$$

Similarly, per Equation A4.4, β_2 represents the regression coefficient of M_O when we regress M_U on M_O and X, meaning:

$$\beta_2 = \frac{\rho_{M_\mathrm{O},M_\mathrm{U}} - \rho_{X,M_\mathrm{U}} \cdot \rho_{X,M_\mathrm{O}}}{1 - \rho_{X,M_\mathrm{O}}^2} \tag{A4.11}$$

Following the same logic, β_3 represents the regression coefficient of X when we regress M_U on M_O and X (at the population level), meaning:

$$\beta_3 = \frac{\rho_{X,M_\mathrm{U}} - \rho_{M_\mathrm{O},M_\mathrm{U}} \cdot \rho_{X,M_\mathrm{O}}}{1 - \rho_{X,M_\mathrm{O}}^2} \tag{A4.12}$$

To note, Equations A3.1–A3.8 derive all the correlation elements relevant to Equation A4.11 and Equation A4.12. In sum, we arrive at following estimates with the omission of M_U:

$$\tilde{a}_1 = a_1 + k \cdot a_2 \tag{A4.13}$$

$$\tilde{b}_1 = b_1 + b_2 \cdot k \cdot \frac{1 - a_2^2}{1 - (k \cdot a_2 + a_1)^2} \tag{A4.14}$$

$$\tilde{c} = c + b_2 \cdot \frac{a_2 - (k + a_1 \cdot a_2) \cdot (k \cdot a_2 + a_1)}{1 - (k \cdot a_2 + a_1)^2} \tag{A4.15}$$

$$\tilde{a}_1 \tilde{b}_1 = a_1 \cdot b_1 + k \cdot a_2 \cdot b_1 + (a_1 + k \cdot a_2) \cdot \frac{b_2 \cdot k \cdot \left(1 - a_2^2\right)}{1 - (k \cdot a_2 + a_1)^2} \tag{A4.16}$$

Bias as a Function of Correlations

Per Equation A1.1, k represents the regression coefficient of M_U when we regress M_O on M_U and X. Similarly, per Equation A1.3, b_2 represents the regression coefficient of M_U when we regress Y on M_O, M_U, and X. Thus, we can write the results in terms of correlations as follows:

$$\tilde{a}_1 = a_1 + k \cdot \beta_1 = a_1 + \rho_{X,M_\mathrm{U}} \cdot \frac{\rho_{M_\mathrm{O},M_\mathrm{U}} - \rho_{X,M_\mathrm{O}} \cdot \rho_{X,M_\mathrm{U}}}{1 - \rho_{X,M_\mathrm{U}}^2} \tag{A5.1}$$

$$\tilde{b}_1 = b_1 + b_2 \cdot \beta_2 = b_1 + \frac{1}{1+2\cdot\rho_{M_O,M_U}\cdot\rho_{X,M_U}\cdot\rho_{X,M_O}-\rho^2_{M_O,M_U}-\rho^2_{X,M_U}-\rho^2_{X,M_O}} \cdot \big(\rho_{Y,M_U} + \rho_{Y,M_O} \cdot \rho_{X,M_U} \cdot \rho_{X,M_O} + \rho_{X,Y} \cdot \rho_{M_O,M_U} \cdot \rho_{X,M_O} - \rho_{Y,M_U} \cdot \rho^2_{X,M_O} - \rho_{Y,M_O} \cdot \rho_{M_O,M_U} - \rho_{X,Y} \cdot \rho_{X,M_U}\big) \cdot \frac{\rho_{M_O,M_U}-\rho_{X,M_U}\cdot\rho_{X,M_O}}{1-\rho^2_{X,M_O}} \tag{A5.2}$$

$$\tilde{c} = c + b_2 \cdot \beta_3 = c + \frac{1}{1+2\cdot\rho_{M_O,M_U}\cdot\rho_{X,M_U}\cdot\rho_{X,M_O}-\rho^2_{M_O,M_U}-\rho^2_{X,M_U}-\rho^2_{X,M_O}} \cdot \big(\rho_{Y,M_U} + \rho_{Y,M_O} \cdot \rho_{X,M_U} \cdot \rho_{X,M_O} + \rho_{X,Y} \cdot \rho_{M_O,M_U} \cdot \rho_{X,M_O} - \rho_{Y,M_U} \cdot \rho^2_{X,M_O} - \rho_{Y,M_O} \cdot \rho_{M_O,M_U} - \rho_{X,Y} \cdot \rho_{X,M_U}\big) \cdot \frac{\rho_{X,M_U}-\rho_{M_O,M_U}\cdot\rho_{X,M_O}}{1-\rho^2_{X,M_O}} \tag{A5.3}$$

To note, the above derivation uses the following two formulas for regression coefficients.

(a) When we regress Y on X1 and X2, the standardized solution equals:

$\begin{bmatrix} 1 & \rho_{X1,X2} \\ \rho_{X2,X1} & 1 \end{bmatrix}^{-1} \begin{pmatrix} \rho_{Y,X1} \\ \rho_{Y,X2} \end{pmatrix}$. From this, we get the coefficient of X1 is equivalent to (in terms of correlation): $\frac{\rho_{Y,X1}-\rho_{Y,X2}\rho_{X1,X2}}{1-\rho^2_{X1,X2}}$.

(b) When we regress Y on X1, X2, and X3, the standardized solution is equal to

$\begin{bmatrix} 1 & \rho_{X1,X2} & \rho_{X1,X3} \\ \rho_{X2,X1} & 1 & \rho_{X2,X3} \\ \rho_{X3,X1} & \rho_{X3,X2} & 1 \end{bmatrix}^{-1} \begin{pmatrix} \rho_{Y,X1} \\ \rho_{Y,X2} \\ \rho_{Y,X3} \end{pmatrix}$. From this, we get the coefficient of X1 is equivalent to (in terms of correlations):

$\frac{\rho_{Y,X1}+\rho_{Y,X2}\cdot\rho_{X1,X3}\cdot\rho_{X2,X3}+\rho_{Y,X3}\cdot\rho_{X1,X2}\cdot\rho_{X2,X3}-\rho_{Y,X1}\cdot\rho^2_{X2,X3}-\rho_{Y,X2}\cdot\rho_{X1,X2}-\rho_{Y,X3}\cdot\rho_{X1,X3}}{1+2\cdot\rho_{X1,X2}\cdot\rho_{X2,X3}\cdot\rho_{X1,X3}-\rho^2_{X1,X2}-\rho^2_{X2,X3}-\rho^2_{X1,X3}}$.

Conditions for Asymptotically Unbiased Estimates When Omitting M_U

Per Equation A4.13, $\tilde{a}_1 = a_1$ if either k or $a_2 = 0$. Of note, when $k = 0$, there is no causal relation between the two mediators, meaning that, in such a scenario, the equation represents a parallel mediation model (recall Figure 1B in the main text). In our illustrative example, this occurs when IMPORT has no effect on PMI. Asymptotically unbiased results can also occur when $a_2 = 0$ or when *X* does not affect M_U. Such a condition mirrors cases in which M_U is a confounder of the M_O –*Y* relation, but not caused by *X*.

Based on Equation A4.14, $\tilde{b}_1 = b_1$, if either (a) $b_2 = 0$ (when M_U does not affect *Y*); (b) $k = 0$ (when M_U does not affect M_O, such as in the case of a parallel two-mediator model); or (c) $a_2^2 = 1$, which means the effect of *X* on M_U equals –1 or 1. If condition (c) is true, then *X* entirely determines M_U,[1] which, in our illustrative example in the main text, would occur if IMPORT was fully determined by COND.

In terms of the direct effect ($\tilde{c}$), Equation A4.15 indicates that $\tilde{c} = c$ if $a_2 = (k + a_1 \cdot a_2) \cdot (k \cdot a_2 + a_1)$ or if $b_2 = 0$. The first condition is equivalent to $\rho_{X,M_U} = \rho_{M_O,M_U} \cdot \rho_{X,M_O}$, where ρ_{X,M_U}, ρ_{M_O,M_U}, and ρ_{X,M_O} represent the correlation between (a) *X* and M_U, (b) M_O and M_U, and (c) *X* and M_O, respectively. Note that $\rho_{X,M_U} = \rho_{M_O,M_U} \cdot \rho_{X,M_O}$ when *X* and M_U are uncorrelated conditional on M_O. That is, when M_U contains no unique information about *X* (in terms of linear relationships), omitting M_U does not have an impact on the estimated direct effect from *X* to *Y*. In our example in the main text, this means the correlation between COND and IMPORT is equal to the correlation between PMI and IMPORT multiplied by the correlation between COND and PMI. Meanwhile, the second condition ($b_2 = 0$) indicates that there is no effect of M_U on *Y*. In sum, if the mediation effect via M_U is nonzero (i.e., $a_2 \neq 0$ and $b_2 \neq 0$), and M_U has a nonzero effect on M_O (i.e., $k \neq 0$), the estimates for a_1, b_1, and c are biased (i.e., $a_1 \neq \tilde{a}_1$, $b_1 \neq \tilde{b}_1$, $c \neq \tilde{c}$). The only exception is $c = \tilde{c}$ when $a_2 = (k + a_1 \cdot a_2) \cdot (k \cdot a_2 + a_1)$.

In sum, to obtain an asymptotically unbiased specific indirect effect via M_O when omitting M_U, M_U must have no effect on M_O. This is equivalent to a parallel two-mediator model

[1] We standardize all variables for derivation purposes. Therefore, the condition, $a_2^2 = 1$, is equivalent to a zero-error term, indicating the M_U depends entirely on the treatment variable.

(i.e., Figure 1B). Under a parallel two-mediator model, we can accurately obtain the specific indirect effect via M_{O} and estimates for a_1 and b_1. However, to obtain an asymptotically unbiased direct effect when omitting M_{U}, either one of the following two conditions must be met: $\rho_{X,M_{\mathrm{U}}} = \rho_{M_{\mathrm{O}},M_{\mathrm{U}}} \cdot \rho_{X,M_{\mathrm{O}}}$ or $b_2 = 0$. Thus, the conditions under which asymptotically unbiased direct and indirect effects occur are different, meaning one cannot simultaneously obtain both accurate direct and indirect effects when M_{U} is omitted for the model.

Derivation of Error Variance for Standardized Variables

Based on Equations A1.1–A1.3, we can derive the following equations:

$$var(M_{\mathrm{O}}) = var(k \cdot M_{\mathrm{U}}) + var(a_1 \cdot X) + 2 \cdot cov(k \cdot M_{\mathrm{U}}, a_1 \cdot X) + var\left(\epsilon_{M_{\mathrm{O}}}\right) \quad \text{(A6.1)}$$

$$var(M_{\mathrm{U}}) = var(a_2 \cdot X) + var\left(\epsilon_{M_{\mathrm{U}}}\right) \quad \text{(A6.2)}$$

$$var(Y) = var(b_1 \cdot M_{\mathrm{O}}) + var(b_2 \cdot M_{\mathrm{U}}) + var(c \cdot X) + 2 \cdot cov(b_1 \cdot M_{\mathrm{O}}, b_2 \cdot M_{\mathrm{U}}) + 2 \cdot cov(b_1 \cdot M_{\mathrm{O}}, c \cdot X) + 2 \cdot cov(b_2 \cdot M_{\mathrm{U}}, c \cdot X) + var(\epsilon_Y) \quad \text{(A6.3)}$$

Subsequently, given our assumption that all variables are standardized, we identify the following constraints on parameters:

$$1 - k^2 - a_1^2 - 2ka_1a_2 > 0 \quad \text{(A6.4)}$$

$$1 - a_2^2 > 0 \quad \text{(A6.5)}$$

$$1 - b_1^2 - b_2^2 - c^2 - 2b_1b_2 \cdot (k + a_1a_2) - 2b_1c \cdot (ka_2 + a_1) - 2a_2b_2c > 0 \quad \text{(A6.6)}$$

Discussion About the Direction of Asymptotic Bias

Regarding a_1**,** based on Equation A4.13, $\tilde{a}_1 - a_1 > 0$ if k and a_2 have the same direction. That is, we *overestimate* a_1 when k and a_2 are either both positive or both negative. For the same reason, if $k > 0$ and $a_2 < 0$ or $k < 0$ and $a_2 > 0$, we *underestimate* a_1.

Regarding b_1, based on Equation A4.14, the sign of $\tilde{b}_1 - b_1$ depends on whether k and b_2 have the same direction. To see this, first, per Equation A6.5, $1 - a_2^2 > 0$. Second, $1 -$

$(k \cdot a_2 + a_1)^2 > 0$ because $(k \cdot a_2 + a_1)^2 = k^2 \cdot a_2^2 + a_1^2 + 2ka_1a_2 < k^2 + a_1^2 + 2ka_1a_2 < 1$, where the last inequality is based on Equation A6.4. Thus, we *overestimate* b_1 when k and b_2 are either both positive or both negative, and if $k > 0$ and $b_2 < 0$ or $k < 0$ and $b_2 > 0$, we *underestimate* b_1.

Lastly, in terms of *c*, based on Equation A4.15, the sign of $\tilde{c} - c$ depends on whether b_2 and $a_2 - (k + a_1 \cdot a_2) \cdot (k \cdot a_2 + a_1)$ have the same direction. We have already shown that $1 - (k \cdot a_2 + a_1)^2 > 0$ when we discuss the bias for estimating b_1. Additionally, from Equations A3.1 – A3.8, $a_2 - (k + a_1 \cdot a_2) = \rho_{X,M_U} - \rho_{M_O,M_U} \cdot \rho_{X,M_O}$.

Changes in Asymptotic Bias With M_U-related Parameters

Given the scope of the paper, we only discuss how asymptotic bias changes with *k.* See Lin (2019) for discussions with regards to a_2 and b_2. To start, based on Equations A4.13 – A4.15, we can derive the following:

$$\frac{\partial \tilde{a}_1}{\partial k} = a_2 \quad \text{(A7.1)}$$

$$\frac{\partial \tilde{b}_1}{\partial k} = \frac{(1-a_2^2)b_2 \cdot [a_2^2k^2 + (1-a_1^2)]}{[1-(a_1+a_2k)^2]^2} \quad \text{(A7.2)}$$

$$\frac{\partial \tilde{c}}{\partial k} = \frac{b_2(1-a_2^2)\{-2a_2k + a_1[(a_1+a_2k)^2-1]\}}{\left(-1+a_1^2+2a_1a_2k+a_2^2k^2\right)^2} \quad \text{(A7.3)}$$

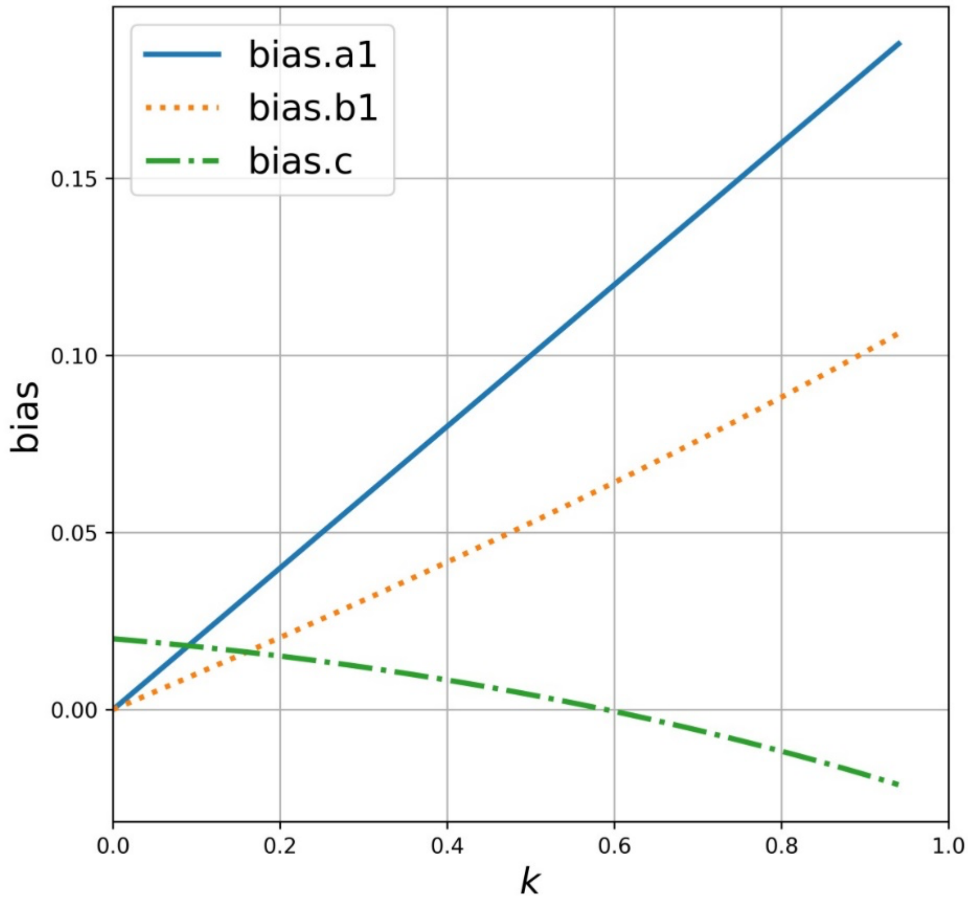
When all parameters are positive and follow the constraints given by Equations 6.4–6.6, we can show that $\frac{\partial \tilde{a}_1}{\partial k}$ and $\frac{\partial \tilde{b}_1}{\partial k}$ are always positive. For $\frac{\partial \tilde{c}}{\partial k}$, we can show that it is always negative by: $(a_1 + a_2k)^2 = a_1^2 + a_2^2k^2 + 2ka_1a_2 < a_1^2 + k^2 + 2ka_1a_2 < 1$, where the last inequality is based on Equation A6.4.

To interpret, when a_2 and b_2 are fixed values, the magnitude of *k* has a *positive* effect on the asymptotic bias of the estimated *indirect* effect via M_O alone ($X \rightarrow M_O \rightarrow Y$) and a *negative* effect on the estimated *direct* effect ($X \rightarrow Y$). In other words, as the effect from M_U to M_O

increases, the asymptotic bias (and relative bias) of $\tilde{a}_1\tilde{b}_1$ increases, while the asymptotic bias of $\tilde{c}$ decreases. To visualize *k*'s effect on asymptotic bias, we plot out the asymptotic bias for the estimates of a_1, b_1, and *c*, assuming varying values of *k* and fixed values of other path coefficients (i.e., $a_1 = a_2 = 0.2$, and $b_2 = 0.1$). As shown in Figure A.1 below, the magnitude of bias for $X \rightarrow M_O$ and $M_O \rightarrow Y$ increases as k increases, resulting in a more serious overestimation of the indirect effect via M_O. In comparison, the positive bias of the direct effect $X \rightarrow Y$ decreases to zero when $k = 0.6$ and becomes further negative as k increases to 1. That is, when k is relatively small, the model tends to overestimate c; however, this inverts as k increases, meaning when k is relatively large, the model tends to underestimate c. Importantly, Equation A4.15 also implies that the asymptotic bias in the estimation of *c* does not depend on its true value. From a practical perspective, this suggests that when the actual direct effect $X \rightarrow Y$ is zero ($c = 0$), omitting M_U can yield a nonzero direct effect from $X \rightarrow Y$, even in exceptionally large samples.

Figure A.1

How Asymptotic Biases Change with Different Levels of k



Alternative Decomposition of the Total Effects in a Serial Two Mediator Model

The discussions above focus on the *specific* indirect effect via M_O *alone* ($X \rightarrow M_O \rightarrow Y$) and the direct effect from X to Y. As discussed in the introduction of the main text, when the goal is to decompose the total effect from X to Y, there are multiple ways to do so when more than one mediator exists in the model. Here, we discuss the decomposition of the total effect in a serial two-mediator model into (a) an indirect effect via M_O and (b) a direct effect not via M_O. Importantly, the indirect effect includes both of the following pathways: $X \rightarrow M_O \rightarrow Y$ and $X \rightarrow M_O \rightarrow M_U \rightarrow Y$. Meanwhile, the direct effect includes both $X \rightarrow M_U \rightarrow Y$, and $X \rightarrow Y$. Using this decomposition, the true indirect effect equals $a_1 b_1 + a_2 k b_1$, and the true direct effect equals $a_2 b_2 + c$. When we omit M_{U}, we use $\tilde{a}_1 \tilde{b}_1$ and $\tilde{c}$ to estimate these two effects. Hence, the following equations present the asymptotic bias:

$$\tilde{a}_1 \tilde{b}_1 - (a_1 b_1 + a_2 k b_1) = b_2 \cdot k \cdot \frac{1-a_2^2}{1-\rho_{X,M_{\mathrm{O}}}{}^2} \cdot \rho_{X,M_{\mathrm{O}}}, \quad \text{(A8.1)}$$

$$\tilde{c} - (a_2 b_2 + c) = b_2 \cdot k \cdot \frac{a_2^2 - 1}{1-\rho_{X,M_{\mathrm{O}}}{}^2} \cdot \rho_{X,M_{\mathrm{O}}}. \quad \text{(A8.2)}$$

Note that sum of these asymptotic biases equal zero. This is intuitive because both $\tilde{a}_1 \tilde{b}_1 + \tilde{c}$, and $(a_1 b_1 + a_2 k b_1) + (a_2 b_2 + c)$ adds up to the total effect from *X* to *Y*. Accordingly, if the indirect effect is positively asymptotically biased, the direct effect must be negatively biased. Based on Equations A8.1–8.2, when b_2, *k*, and $\rho_{X,M_{\mathrm{O}}}$ have the same sign, then the indirect effect via M_{O} is overestimated according to: $b_2 \cdot k \cdot \frac{1-a_2^2}{1-\rho_{X,M_{\mathrm{O}}}{}^2} \cdot \rho_{X,M_{\mathrm{O}}}$. The direct effect not via M_{O} is underestimated by same amount. Noticeably, the larger $b_2 \cdot k$, the larger the positive bias for the indirect effect (when $b_2 \cdot k \cdot \rho_{X,M_{\mathrm{O}}} > 0$). This is intuitive because $b_2 \cdot k$ captures how M_{U} confounds the estimation of the effect $M_{\mathrm{O}} \rightarrow Y$. Furthermore, Equations A8.1 and A8.2 highlight an interesting finding that both effects under this decomposition are asymptotically unbiased when $k = 0$ (parallel mediation model).

Multiple Unobserved Mediators M_{U}

Below, we show the effect on the estimation of direct and indirect effects with the omission of two unobserved mediators: M_{U1} and M_{U2}. First, we assume there is no causal relationship between M_{U1} and M_{U2}. As such, we can write the true model as follows.

$$M_{\mathrm{O}} = k_1 \cdot M_{\mathrm{U1}} + k_2 \cdot M_{\mathrm{U2}} + a_1 \cdot X + \epsilon_{M_{\mathrm{O}}} \quad \text{(A9.1)}$$

$$M_{\mathrm{U1}} = a_{21} \cdot X + \epsilon_{M_{\mathrm{U1}}} \quad \text{(A9.2)}$$

$$M_{\mathrm{U2}} = a_{22} \cdot X + \epsilon_{M_{\mathrm{U2}}} \quad \text{(A9.3)}$$

$$Y = b_1 \cdot M_{\mathrm{O}} + b_{21} \cdot M_{\mathrm{U1}} + b_{22} \cdot M_{\mathrm{U2}} + c \cdot X + \epsilon_Y \quad \text{(A9.4)}$$

Following the same approach, we can then write the true model in terms of conditional mean as:

$$E(M_{\mathrm{O}}|M_{\mathrm{U1}}, M_{\mathrm{U2}}, X) = k_1 \cdot M_{\mathrm{U1}} + k_2 \cdot M_{\mathrm{U2}} + a_1 \cdot X \quad \text{(A10.1)}$$

$$E(M_{\mathrm{U1}}|X) = a_{21} \cdot X \quad \text{(A10.2)}$$

$$E(M_{\mathrm{U2}}|X) = a_{22} \cdot X \quad \text{(A10.3)}$$

$$E(Y|M_{\mathrm{O}}, M_{\mathrm{U1}}, M_{\mathrm{U2}}, X) = b_1 \cdot M_{\mathrm{O}} + b_{21} \cdot M_{\mathrm{U1}} + b_{22} \cdot M_{\mathrm{U2}} + c \cdot X \quad \text{(A10.4)}$$

According to the LIE, we can further write the true model excluding M_{U1} and M_{U2} as:

$$E(M_{\mathrm{O}}|X) = E[E(M_O|M_{\mathrm{U1}}, M_{\mathrm{U2}}, X)|X] = k_1 \cdot E(M_{\mathrm{U1}}|X) + k_2 \cdot E(M_{\mathrm{U2}}|X) + a_1 \cdot X \quad \text{(A10.5)}$$

$$E(Y|M_{\mathrm{O}}, X) = E[E(Y|M_{\mathrm{O}}, M_{\mathrm{U1}}, M_{\mathrm{U2}}, X)|M_{\mathrm{O}}, X]$$

$$= b_1 \cdot M_{\mathrm{O}} + b_{21} \cdot E(M_{\mathrm{U1}}|M_{\mathrm{O}}, X) + b_{22} \cdot E(M_{\mathrm{U2}}|M_{\mathrm{O}}, X) + c \cdot X \quad \text{(A10.6)}$$

Now we write the following equations:

$$E(M_{\mathrm{U1}}|X) = \beta_{11} \cdot X$$

$$E(M_{\mathrm{U2}}|X) = \beta_{12} \cdot X \quad \text{(A11.1)}$$

$$E(M_{\mathrm{U1}}|M_{\mathrm{O}}, X) = \beta_{21} \cdot M_{\mathrm{O}} + \beta_{31} \cdot X \quad \text{(A11.2)}$$

$$E(M_{\mathrm{U2}}|M_{\mathrm{O}}, X) = \beta_{22} \cdot M_{\mathrm{O}} + \beta_{32} \cdot X \quad \text{(A11.3)}$$

Next, plug these Equations A11.1–A11.3 into Equations A10.5 and A10.6, we obtained the following formulas for biased estimates.

$$\tilde{a}_1 = a_1 + k_1\beta_{11} + k_2\beta_{12} \quad \text{(A12.1)}$$

$$\tilde{b}_1 = b_1 + b_{21}\beta_{21} + b_{22}\beta_{22} \quad \text{(A12.2)}$$

$$\tilde{c} = c + b_{21}\beta_{31} + b_{22}\beta_{32} \quad \text{(A12.3)}$$

To calculate the asymptotic bias due to the omission of more than one unobserved mediator, we use the same approach above for omitting a single M_{U}. Specifically, we substitute all β with parameters and/or correlations. We can also extend two M_{U} to N M_{U}, as follows.

$$\tilde{a}_1 = a_1 + \sum_N^i k_i \cdot a_{2i} \tag{A13.1}$$

$$\tilde{b}_1 = b_1 + \sum_N^i b_{2i} \cdot k_i \cdot \frac{1-a_{2i}^2}{1-(k_i \cdot a_{2i}+a_1)^2} \tag{A13.2}$$

$$\tilde{c} = c + \sum_N^i b_{2i} \cdot \frac{a_{2i}-(k_i+a_1 \cdot a_{2i}) \cdot (k_i \cdot a_{2i}+a_1)}{1-(k_i \cdot a_{2i}+a_1)^2} = c + \sum_N^i b_{2i} \cdot \frac{\rho_{X,M_{Ui}}-\rho_{M_O,M_{Ui}} \cdot \rho_{X,M_O}}{1-(k_i \cdot a_{2i}+a_1)^2} \tag{A13.3}$$

In effect, Equations A13.1–A13.3 show that when estimating a_1, b_1 and c, the asymptotic bias caused by all the N unobserved mediators is the sum of the asymptotic bias caused by each unobserved mediator. For example, when there are two omitted mediators ($N = 2$), we get asymptotically unbiased estimates if (a) both unobserved mediators satisfy the conditions we delineated in previous sections for asymptotically unbiased results, or (b) the additional bias from the second omitted mediator cancels out the bias from the first omitted mediator. Both conditions are very difficult to meet in practice, as condition b requires that the two asymptotic biases (i.e., bias from M_{U1} and M_{U2}) have the same magnitude but in opposite directions. When the two biases have the same direction (e.g., both positive), then the presence of a second unobserved mediator increases overall asymptotic bias. This is consistent with what Fritz et al. (2016) discussed for pretreatment confounders and what Clarke (2005) pointed out for multiple omitted variables from a model: the bias can increase, decrease, or remain the same when excluding more than one confounder from the model.

To further illustrate this point, we consider a special scenario: the true underlying model is a parallel mediation model with two omitted mediators (Figure 4B in the main manuscript). In such model, per Equation A8.1, $k_1 = k_2 = 0$. As such, $\tilde{a}_1 = a_1$ and $\tilde{b}_1 = b_1$, meaning the specific indirect effect via M_O is asymptotically unbiased. Meanwhile, $\tilde{c} = c + a_{21}b_{21} + a_{22}b_{22}$, meaning that the indirect effects through the two unobserved mediators are mistakenly attributed to the direct effect from X to Y. In fact, this result applies to N unobserved mediators in a parallel mediation model. When setting $k_1 = k_2 = \ldots = k_N = 0$, the indirect effect via M_O is always asymptotically unbiased, irrespective of the total number of unobserved mediators in the true model. At the same time, the direct effect from *X* to *Y* is asymptotically biased unless the indirect effects via the other omitted mediators cancel each other out.

The scenarios quickly become complex when we assume that more than one causally related mediator is excluded from the model. To offer a straightforward example, we look at the sequential mediation model with more than one omitted mediator. Figure 5C in the main manuscript shows a special case when $N = 2$, and the following equations define such a model:

$$M_{\mathrm{U1}} = a_2 \cdot X + \varepsilon_{M_{\mathrm{U1}}} \tag{A14.1}$$

$$M_{\mathrm{U2}} = k_1 \cdot M_{\mathrm{U1}} + \varepsilon_{M_{\mathrm{U2}}} \tag{A14.2}$$

$$M_{\mathrm{O}} = k_2 \cdot M_{\mathrm{U2}} + \varepsilon_{M_{\mathrm{O}}} \tag{A14.3}$$

$$Y = b_1 \cdot M_{\mathrm{O}} + c \cdot X + \varepsilon_Y \tag{A14.4}$$

Accordingly, applying the LIE, we can derive the following equation:

$$E(M_{\mathrm{O}}|X) = E[E(M_{\mathrm{O}}|M_{\mathrm{U2}}, M_{\mathrm{U1}}, X)|X] = k_1 k_2 a_2 \cdot X \tag{A15.1}$$

$$E(Y|M_{\mathrm{O}}, X) = E[E(Y|M_{\mathrm{O}}, M_{\mathrm{U1}}, M_{\mathrm{U2}}, X)|M_{\mathrm{O}}, X] = b_1 \cdot M_{\mathrm{O}} + c \cdot X. \tag{A15.2}$$

Thus, $\tilde{a}_1 = k_1 k_2 a_2$, while $\tilde{b}_1 = b_1$ and $\tilde{c} = c$. In other words, only $\tilde{a}_1$ is asymptotically biased (note that $a_1 = 0$).

This approach can also be extended for N M_{U}. In such a case, $\tilde{a}_1 = a_2 \prod_i^N k_i$. In other words, when we exclude all the unobserved mediators, we obtain asymptotically unbiased estimates for b_1 and c, while the estimate of a_1 has a bias equivalent to $a_2 \prod_i^N k_i$. In short, the estimated direct effect of X to M_{O} picks up all the intermediated processes via those omitted mediators. Note, however, that unlike the scenarios with independent omitted mediators, it is impossible for unobserved mediators to cancel each other out. In fact, if all k and a_2 take non-zero values, the estimate for b_1 is always asymptotically biased.

We can further generalize the underlying mediating process to serial mediator models and allow omitted mediators to be causally related. Below shows an example when $N = 2$. Compared to Equations A9.1–9.4, the only difference is the model for M_{U2}. In this case, M_{U1} has an effect on M_{U2} (Equation A16.3).

$$M_{\mathrm{O}} = k_1 \cdot M_{\mathrm{U1}} + k_2 \cdot M_{\mathrm{U2}} + a_1 \cdot X + \epsilon_{M_{\mathrm{O}}} \tag{A16.1}$$

$$M_{\mathrm{U1}} = a_{21} \cdot X + \epsilon_{M_{\mathrm{U}}} \tag{A16.2}$$

$$M_{\mathrm{U2}} = a_{22} \cdot X + l \cdot M_{\mathrm{U1}} + \epsilon_{M_{\mathrm{U}}} \tag{A16.3}$$

$$Y = b_1 \cdot M_{\mathrm{O}} + b_{21} \cdot M_{\mathrm{U1}} + b_{22} \cdot M_{\mathrm{U2}} + c \cdot X + \epsilon_Y \tag{A16.4}$$

Following the same approach, we can then write the true model in terms of conditional mean as:

$$E(M_{\mathrm{O}}|M_{\mathrm{U1}}, M_{\mathrm{U2}}, X) = k_1 \cdot M_{\mathrm{U1}} + k_2 \cdot M_{\mathrm{U2}} + a_1 \cdot X \tag{A17.1}$$

$$E(M_{\mathrm{U1}}|X) = a_{21} \cdot X \tag{A17.2}$$

$$E(M_{\mathrm{U2}}|X,\ M_{\mathrm{U1}}) = a_{22} \cdot X + l \cdot M_{\mathrm{U1}} \tag{A17.3}$$

$$E(Y|M_{\mathrm{O}}, M_{\mathrm{U1}}, M_{\mathrm{U2}}, X) = b_1 \cdot M_{\mathrm{O}} + b_{21} \cdot M_{\mathrm{U1}} + b_{22} \cdot M_{\mathrm{U2}} + c \cdot X \tag{A17.4}$$

Equations A10.5–10.6, A11.1–11.3, and A12.1–12.3 still hold as before. At first glance of Equations A12.1–12.3, we may argue that the overall bias is still the sum of the bias due to each unmeasured mediator. However, note that β_{12}, β_{22}, and β_{32} in Equations A11.1–11.3 now take on different values as M_{U2} is not only a function of X but also M_{U1}. Specifically, $\beta_{12} = a_{22} + l \cdot a_{21}$; β_{22} is the regression coefficient of M_{O} when we regress M_{U2} on M_{O} and X; β_{32} is the regression coefficient of X when we regress M_{U2} on M_{O} and X. In fact, the parameter l that measures the effect of M_{U1} on M_{U2} plays a role in each of β_{12}, β_{22}, and β_{32}. Thus, it would be inaccurate to simply say that the overall bias is the sum of the bias due to each unmeasured mediator.

Combined Effects of Measurement Errors in M_{O} and Y and the Unobserved M_{U}

In this section, we consider measurement error in M_{O} and Y. Since we assume all variables are standardized, we focus on standardized coefficients biased by measurement error in both the predictors and the outcome. Based on Fritz et al. (2016) we know the following is true when there is no confounding variable in a single-mediator model, where the reliability of M and Y are r_{MM} and r_{YY}, and $\omega = \frac{r_{MM} - r_{XM}^2}{1 - r_{XM}^2}$, which is the reliability partialling out X. a_{ME}, b_{ME}, and c_{ME} are the estimates biased by measurement error. a_T, b_T, and c_T are the true values.

$$a_{\mathrm{ME}} = a_T \cdot \sqrt{r_{\mathrm{MM}}} \tag{A18.1}$$

$$b_{\mathrm{ME}} = b_T \cdot \omega \cdot \frac{\sqrt{r_{\mathrm{YY}}}}{\sqrt{r_{\mathrm{MM}}}} \tag{A18.2}$$

$$c_{\mathrm{ME}} = \sqrt{r_{\mathrm{YY}}} \cdot [c_T + a_T \cdot b_T \cdot (1 - \omega)] \tag{A18.3}$$

By combining these three equations with Equations A4.13–A4.15, we can solve for the combined effects of omitted M_{U} and measurement error, where the reliability of M_{O} and Y are represented as r_{MM} and r_{YY}, respectively, using the following equations: Importantly, note that ω now represents the reliability partialling out both X and M_{U}.

$$\tilde{a}_{1\mathrm{ME}} = a_1 \cdot \sqrt{r_{\mathrm{MM}}} + k_{\mathrm{ME}} \cdot a_{2\mathrm{ME}} \tag{A19.1}$$

$$\tilde{b}_{1\mathrm{ME}} = b_1 \cdot \omega \cdot \frac{\sqrt{r_{\mathrm{YY}}}}{\sqrt{r_{\mathrm{MM}}}} + b_{2\mathrm{ME}} \cdot k_{\mathrm{ME}} \cdot \frac{1 - a_{2\mathrm{ME}}^2}{1 - (k_{\mathrm{ME}} \cdot a_{2\mathrm{ME}} + a_{1\mathrm{ME}})^2} \tag{A19.2}$$

$$\tilde{c}_{\mathrm{ME}} = \sqrt{r_{YY}} \cdot [c + a_1 \cdot b_1 \cdot (1 - \omega)] + b_{2\mathrm{ME}} \cdot \frac{r_{X,M_{\mathrm{U}}} - r_{M_{\mathrm{O}},M_{\mathrm{U}}} \cdot r_{X,M_{\mathrm{O}}}}{1 - (k_{\mathrm{ME}} \cdot a_{2\mathrm{ME}} + a_{1\mathrm{ME}})^2} \tag{A19.3}$$

where ω is the reliability of M_{O} after partialling out X and M_{U}. Specifically, ω equals $\frac{r_{\mathrm{MM}} + 2 \cdot r_{X,M_{\mathrm{O}}} \cdot r_{M_{\mathrm{O}},M_{\mathrm{U}}} \cdot r_{X,M_{\mathrm{U}}} - r_{\mathrm{MM}} \cdot r_{X,M_{\mathrm{U}}}^2 - r_{X,M_{\mathrm{O}}}^2 - r_{M_{\mathrm{O}},M_{\mathrm{U}}}^2}{1 + 2 \cdot r_{X,M_{\mathrm{O}}} \cdot r_{X,M_{\mathrm{U}}} \cdot r_{M_{\mathrm{O}},M_{\mathrm{U}}} - r_{X,M_{\mathrm{O}}}^2 - r_{X,M_{\mathrm{U}}}^2 - r_{M_{\mathrm{O}},M_{\mathrm{U}}}^2}$. The terms k_{ME}, $a_{2\mathrm{ME}}$, and $b_{2\mathrm{ME}}$ are the path coefficients biased by measurement error, and $r_{X,M_{\mathrm{U}}}$, $r_{M_{\mathrm{O}},M_{\mathrm{U}}}$, and $r_{X,M_{\mathrm{O}}}$ are the observed correlations also biased by measurement error.

Alternatively, we can also combine Equations A19.1–A19.3 with Equations A5.1–A5.3 to solve for the combined effects of omitted M_U and measurement error in terms of correlations, such that:

$$\tilde{a}_{1\mathrm{ME}} = a_1 \cdot \sqrt{r_{\mathrm{MM}}} + r_{X,M_\mathrm{U}} \cdot \frac{r_{M_\mathrm{O},M_\mathrm{U}} - r_{X,M_\mathrm{O}} \cdot r_{X,M_\mathrm{U}}}{1 - r_{X,M_\mathrm{U}}^2} \tag{A20.1}$$

$$\tilde{b}_{1\mathrm{ME}} = b_1 \cdot \omega \cdot \frac{\sqrt{r_{\mathrm{YY}}}}{\sqrt{r_{\mathrm{MM}}}} + \frac{r_{Y,M_\mathrm{U}} + r_{Y,M_\mathrm{O}} \cdot r_{X,M_\mathrm{U}} \cdot r_{X,M_\mathrm{O}} + r_{Y,X} \cdot r_{M_\mathrm{O},M_\mathrm{U}} \cdot r_{X,M_\mathrm{O}} - r_{Y,M_\mathrm{U}} \cdot r_{X,M_\mathrm{O}}^2 - r_{Y,M_\mathrm{O}} \cdot r_{M_\mathrm{O},M_\mathrm{U}} - r_{Y,X} \cdot r_{X,M_\mathrm{U}}}{1 + 2 \cdot r_{M_\mathrm{O},M_\mathrm{U}} \cdot r_{X,M_\mathrm{U}} \cdot r_{X,M_\mathrm{O}} - r_{M_\mathrm{O},M_\mathrm{U}}^2 - r_{X,M_\mathrm{U}}^2 - r_{X,M_\mathrm{O}}^2} \cdot \frac{r_{M_\mathrm{O},M_\mathrm{U}} - r_{X,M_\mathrm{U}} \cdot r_{X,M_\mathrm{O}}}{1 - r_{X,M_\mathrm{O}}^2} \tag{A20.2}$$

$$\tilde{c}_{\mathrm{ME}} = \sqrt{r_{\mathrm{YY}}} \cdot [c + a_1 \cdot b_1 \cdot (1 - \omega)] + \frac{r_{Y,M_\mathrm{U}} + r_{Y,M_\mathrm{O}} \cdot r_{X,M_\mathrm{U}} \cdot r_{X,M_\mathrm{O}} + r_{Y,X} \cdot r_{M_\mathrm{O},M_\mathrm{U}} \cdot r_{X,M_\mathrm{O}} - r_{Y,M_\mathrm{U}} \cdot r_{X,M_\mathrm{O}}^2 - r_{Y,M_\mathrm{O}} \cdot r_{M_\mathrm{O},M_\mathrm{U}} - r_{Y,X} \cdot r_{X,M_\mathrm{U}}}{1 + 2 \cdot r_{M_\mathrm{O},M_\mathrm{U}} \cdot r_{X,M_\mathrm{U}} \cdot r_{X,M_\mathrm{O}} - r_{M_\mathrm{O},M_\mathrm{U}}^2 - r_{X,M_\mathrm{U}}^2 - r_{X,M_\mathrm{O}}^2} \cdot \frac{r_{X,M_\mathrm{U}} - r_{M_\mathrm{O},M_\mathrm{U}} \cdot r_{X,M_\mathrm{O}}}{1 - r_{X,M_\mathrm{O}}^2} \tag{A20.3}$$

Note that, in this case, *r* represents the correlations biased by measurement error.

If the true values of a_2, b_2, and *c* are known, then first calculate the true correlations based on these parameter coefficients using Equations A3.1–3.8. Then, calculate the correlations biased by measurement error and plug these correlations into the formulas provided above to calculate the corresponding $\tilde{a}_{1\mathrm{ME}}$, $\tilde{b}_{1\mathrm{ME}}$, and $\tilde{c}_{\mathrm{ME}}$. We generated Tables 2–4 in the main text using this approach.

The Impact Threshold for a Confounding Variable (ITCV) in Linear Models

The general idea of Frank's (2000) ITCV approach is to evaluate how strong the two confounder-related correlations must be to change the test statistic of the predictor of interest below a critical value for statistical significance. To illustrate, below, we describe how to apply this approach for the following bivariate regression model: $Y = \beta_0 + \beta_1 X + \epsilon$. After fitting such a model with observed data, we get a statistically significant estimated effect $\hat{\beta}_1$ and its corresponding standard error $SE(\hat{\beta}_1)$. We also get the *t* statistic for *X* as $t(\hat{\beta}_1) = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)}$. From here, we calculate a corresponding correlation level between *X* and *Y* using the formula: $r_{Y,X} = \frac{t(\hat{\beta}_1)}{\sqrt{nobs - ncov - 2 + t(\hat{\beta}_1)^2}}$, where *nobs* represents the sample size of the dataset and $ncov$ represents the number of covariates in the sample (in this case $ncov$ equals to 0).

Next, we assume that one confounding variable (*CV*) is added to the model, and thus, the the following formula: $r_{Y,X|CV} = \frac{r_{Y,X} - r_{Y,CV} \cdot r_{X,CV}}{\sqrt{1 - r_{Y,CV}{}^2}\sqrt{1 - r_{X,CV}{}^2}}$ solves for the partial correlation between *X* and *Y* conditional on the *CV*. As such, when $r_{Y,X|CV}$ becomes smaller than a threshold for making an inference ($r^{\#}$), then the inference regarding *X* changes because of the *CV*. For a threshold based on statistical significance, we calculate $r^{\#}$ from $t_{critical}$: $r^{\#} = \frac{t_{critical}}{\sqrt{nobs - ncov - 2 + t_{critical}^2}}$, where $t_{critical}$ is the *t* critical value for a given significance level.

Now, assume $r_{Y,CV} = r_{X,CV}$. Frank (2000) shows that this is a conservative assumption because it favors the challenger of the inference by maximizing the *impact* of the *CV*, where *impact* is defined as the product of $r_{Y,CV}$ and $r_{X,CV}$. We walk through the derivation here briefly. First, $t(\hat{\beta}_1) = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} = \frac{r_{Y,X} - r_{Y,CV} \cdot r_{X,CV}}{\sqrt{\frac{1 - r_{X,CV}^2 - (r_{Y,X}^2 + r_{Y,CV}^2 - 2 r_{Y,X} \cdot r_{Y,CV} \cdot r_{X,CV})}{df}}}$ (see Equation (2) in Frank, 2000 for details for how to get this). Second, we observe that $r_{Y,CV} \cdot r_{X,CV}$ appears in both the numerator and denominator for $t(\hat{\beta}_1)$; thus, we define $r_{Y,CV} \cdot r_{X,CV} = k$ to characterize the overall impact of the confounder *CV*, from which we get $t(\hat{\beta}_1) = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} = \frac{r_{Y,X} - k}{\sqrt{\frac{1 - r_{X,CV}^2 - (r_{Y,X}^2 + r_{Y,CV}^2 - 2 r_{Y,X} \cdot k)}{df}}}$. Because the inference is invalid once $t(\hat{\beta}_1)$ becomes *less than* a critical value, we seek to minimize $t(\hat{\beta}_1)$ with respect to $r_{X,CV}$ and $r_{Y,CV}$. In other words, we favor the challenger of the inference by maximizing the *impact* of the confounder *CV*. Now, for a given value of *k*, $t(\hat{\beta}_1)$ is minimized when the denominator is maximized, which happens when $1 - r_{X,CV}^2 - (r_{Y,X}^2 + r_{Y,CV}^2 - 2r_{Y,X} \cdot k)$ is maximized. Using a LaGrange multiplier based on the constraint that $r_{Y,CV} \cdot r_{X,CV} = k$, the maximum occurs when $r_{X,CV}^2 = r_{Y,CV}^2$ (see Equation (4) in Frank, 2000 for the detailed derivation). In other words, the constrained relative minimum of $t(\hat{\beta}_1)$ occurs when $r_{X,CV}^2 = r_{Y,CV}^2 = k$. For a confounder (vs a suppressor), $r_{X,CV}$ and $r_{Y,CV}$ take the same sign, meaning $r_{X,CV} = r_{Y,CV}$. The same results can be applied to a partial correlation when observed covariates are included in the model (see Xu et al., 2018).

Thus we define $r_{Y,CV} \cdot r_{X,CV} = r_{Y,CV} \cdot r_{Y,CV} = r_{X,CV} \cdot r_{X,CV} = impact$. By setting the conditional correlation $r_{Y,X|CV}$ equal to $r^{\#}$, then $r_{Y,X|CV} = \frac{r_{Y,X} - r_{Y,CV} \cdot r_{X,CV}}{\sqrt{1 - r_{Y,CV}{}^2}\sqrt{1 - r_{X,CV}{}^2}} = \frac{r_{Y,X} - impact}{1 - impact} = r^{\#}$.

Hence, to alter the inference, $impact = \frac{r_{Y,X} - r^{\#}}{1 - |r^{\#}|}$. That is, $impact$ represents the smallest product of $r_{Y,CV}$ and $r_{X,CV}$ needed to change the statistical inference regarding *X* from statistically significant to not significant. This smallest product occurs when $|r_{Y,CV}| = |r_{X,CV}| = \sqrt{impact}$ and $r_{Y,CV} \cdot r_{X,CV} > 0$. Other combinations of $r_{Y,CV}$ and $r_{X,CV}$ can also alter the inference but their product would need to be larger than *impact* to do so. As such, Frank (2000) also calls this smallest *impact* the ITCV, representing Impact Threshold for a Confounding Variable to invalidate an inference.

In essence, the ITCV approach quantifies how strong a confounder must be to change a statistical inference. The larger the impact threshold, the more robust the inference is, as the confounder needs to show stronger correlations with *X* and *Y* to change the inference. We can extend the calculation above to multivariate scenarios. In such cases, *impact* is defined as the product of two partial correlations conditional on all the other covariates in the linear model. For example, for a model with two covariates Z_1 and Z_2: $Y = \beta_0 + \beta_1 X + \beta_2 Z_1 + \beta_3 Z_2 + \epsilon$, the *impact* to alter the inference regarding *X* represents the product of the partial correlation between (a) *Y* and *CV* conditional on both Z_1 and Z_2 ($r_{Y,CV|Z_1,Z_2}$) and (b) *X* and *CV* conditional on both Z_1 and Z_2 ($r_{X,CV|Z_1,Z_2}$). See Frank (2000) and Xu et al. (2019) for more details.

Incorporate Measurement Error into ITCV

For simplicity, we discuss this in a general regression model. The input we need from users includes: $\hat{\beta}_1$; sd(*x*); sd(*y*); the overall R^2; number of covariates in the model (*ncov*), and the total sample size (*nobs*). $\hat{\beta}_1$, se($\hat{\beta}_1$) represent the estimated coefficient and standard error of the predictor of interest *x*. The term *Z* represents a vector of all the covariates *z*. We also ask users to input $R_{(ME)yx}$, which is the zero-order correlation between *y* and *x* biased by measurement error. Note that in case users may not have access to the original dataset, we also provide a backup plan to calculate this value by making certain assumptions. See more explanations below.

From the input, we can calculate the *t* ratio $t_{(ME)yx|Z}$ by dividing the estimated effect $\hat{\beta}_1$ by the standard error se($\hat{\beta}_1$). Note that this *t* value does *not* correct for measurement error. We can also get $t_{critical}$, based on the number of covariates in the model, using $t_{critical}$, $r_{critical} = \frac{t_{critical}}{\sqrt{(nobs-ncov-2)+t^2_{critical}}}$. Then we continue with following steps.

Step 1: we first calculate $r_{(ME)yx|Z}$ based on $t_{(ME)yx|Z}$, using the following formula.

$$r_{(ME)yx|Z} = \frac{t_{(ME)yx|Z}}{\sqrt{(nobs-ncov-2)+t^2_{(ME)yx|Z}}} \qquad \text{(A21.1)}$$

Now we calculate some important intermediate products that we need in the calculation.

$$R^2_{(ME)yZ} = \frac{r^2_{(ME)yx|Z} - R^2}{r^2_{(ME)yx|Z} - 1} \tag{A21.2}$$

$$R^2_{(ME)xZ} = 1 - \frac{sd(y)^2 \cdot (1 - R^2)}{sd(x)^2 \cdot nobs \cdot se(\hat{\beta}_1)^2} \tag{A21.3}$$

$$R_{(ME)yx} = r_{(ME)yx|Z}\sqrt{1 - R^2_{(ME)yZ}}\sqrt{1 - R^2_{(ME)xZ}} + R_{(ME)yZ}R_{(ME)xZ} \tag{A21.4}$$

Note that in A21.4, we try to get the correlation between *y* and *x* (biased by measurement error). In practice, such information is generally obtained from a zero-order/unconditional correlation. However, sometimes users may not have access to such information, especially if they do not have access to the original dataset. In such cases, we use A21.4 to calculate this value. However, in order for A21.4 to work, we need to assume that each element in *Z* is equally weighted in predicting *x* and *y* (e.g., $r_{yZ_1} = r_{yZ_2} = \cdots = r_{yZ_L}$ and $r_{xZ_1} = r_{xZ_2} = \cdots = r_{xZ_L}$).

Step 2: we correct for measurement error in $r_{(ME)yx|Z}$.

$$r_{yx|Z} = \frac{\frac{R_{(ME)yx}}{\sqrt{r_{YY}}\sqrt{r_{XX}}} - \frac{R_{(ME)yZ}R_{(ME)xZ}}{\sqrt{r_{YY}}\ \ \sqrt{r_{XX}}}}{\sqrt{1 - \frac{R^2_{(ME)yZ}}{r_{YY}}}\sqrt{1 - \frac{R^2_{(ME)xZ}}{r_{XX}}}} \tag{A21.5}$$

Note that in Step 2, we assume $r_{xz} \cdot r_{yz} > 0$ for all *z*. That is, all *z* are confounders, with r_{xz} and r_{yz} taking the same sign. If this assumption does not hold, we should group covariates in *Z* by suppressor ($r_{xz} \cdot r_{yz} < 0$) versus confounder ($r_{xz} \cdot r_{yz} > 0$) and employ double adjustment. However, assuming all covariates are conservative in the sense that it will generate an $r_{yx|Z}$ that is as small as possible from $r_{(ME)yx|Z}$ for given reliability levels.

Step 3: we calculate ITCV using $r_{yx|Z}$ (assuming $r_{yx|Z} > r_{critical}$).

$$ITCV = \frac{r_{yx|Z} - r_{critical}}{1 - r_{critical}} \tag{A21.6}$$

Supplemental B

R Code to Conduct Sensitivity Analysis Using ConMed and konfound

This supplement walks through the R code to conduct sensitivity analysis using ConMed and konfound packages. First, we present a table summarizing the main functions in both packages and their corresponding use cases. Second, we describe the installation and use of the main functions in both packages. Lastly, we provide some guidance with regards to what function to use under different scenarios.

Table B.1

Overview of Functions and Use Cases

Package	Function	Use Case
`ConMed`	`conmed_ind`	Evaluate the robustness of inference regarding the *specific indirect* effect via mediator *M* alone. The omitted confounder may impact the inference regarding both the *a* pathway (*X* → *M*) and the b pathway (*M* → *Y*) at the same time.
`ConMed`	`conmed_ind_ME`	Evaluate the robustness of inference regarding the *specific indirect* effect via mediator *M* alone, accounting for potential *measurement error* in treatment *X*, and/or mediator *M*, and/or outcome *Y*.
`konfound`	`pkonfound`	Evaluate the robustness of inference regarding any *direct* effect, such as the direct effect from treatment *X* to outcome *Y* (*X* → *Y*) or the direct effect from mediator *M* to outcome *Y* (*M* → *Y*). Examples include an omitted pretreatment confounder that potentially confounds the relationship from *X* to *M*.

Note. All functions can be used in models with any number of observed covariates.

Installation

For ConMed, you can install the development version from GitHub with:

```
install.packages("devtools")
devtools::install_github("qinyun-lin/ConMed")
```

For konfound, you can install the CRAN version with:

`install.packages("konfound")` or

development version from GitHub with:

`devtools::install_github("jrosen48/konfound")`.

Note that the package ConMed may be incorporated into the konfound package in near future. Please check [here]() for most recent updates.

Use of ConMed and konfound

Use `conmed_ind` for evaluating the robustness of inference regarding the *indirect* effect via mediator *M*. This is especially useful for those concerned about an omitted confounder potentially impacting the inference regarding both the *a* pathway (*X* → *M*) and the *b* pathway (*M* → *Y*) at the same time.

```
library(ConMed)
conmed_ind(est_eff_a, std_err_a, est_eff_b, std_err_b, n_obs,
n_covariates_a, n_covariates_b, alpha, tails)
```

`est_eff_a` specifies the estimated effect of the *a* pathway (*X* → *M*);

`std_err_a` specifies the standard error of the *a* pathway (*X* → *M*);

`est_eff_b` specifies the estimated effect of the *b* pathway (*M* → *Y*);

`std_err_b` specifies the standard error of the *b* pathway (*M* → *Y*);

`n_obs` specifies the sample size;

`n_covariates_a` specifies the number of covariates in the regression model for estimating the *a* pathway (*X* → *M*);

`n_covariates_b` specifies the number of covariates in the regression model for estimating the *b* pathway (*M* → *Y*);

`alpha` specifies probability of rejecting the null hypothesis with the default value set as 0.05;

`tails` specifies an integer regarding whether hypothesis testing is one-tailed (1) or two-tailed (2) with the default value set as 2.

Use `conmed_ind_ME` for evaluating the robustness of inference regarding the *indirect* effect via mediator *M*, accounting for potential measurement error in *X*, *M*, and/or *Y*. This is especially useful for those concerned about (a) an omitted confounder potentially impacting the inference regarding both the *a* pathway (*X* → *M*) and the *b* pathway (*M* → *Y*) at the same time and (b) measurement error in *X*, *M*, and/or *Y*.

```
library(ConMed)
conmed_ind_ME(est_eff_a, std_err_a, est_eff_b, std_err_b, n_obs, rel_M,
rel_Y, sd_X, sd_M, sd_Y, R2_a, R2_b, rXM, rMY)
```

Most of the arguments are the same as the function of `conmed_ind`, with several additional arguments as follows.

`rel_M, rel_Y, rel_X` specify the reliability level of *M*, *Y*, and *X*, respectively;

`sd_X, sd_M, sd_Y` specify the standard deviation for *X*, *M*, and *Y*, respectively;

`R2_a` specifies the R^2 of the regression model estimating the *a* pathway (*X* → *M*);

`R2_b` specifies the R^2 of the regression model estimating the *b* pathway (*M* → *Y*);

`rXM` specifies the zero-order correlation between *X* and *M*;

`rMY` specifies the zero-order correlation between *M* and *Y*.

Use `pkonfound` for evaluating the robustness of inference regarding any *direct* effect, such as the direct effect from treatment *X* to outcome *Y*, or the direct effect from mediator *X* to outcome *Y*. This could be especially useful in two scenarios. First, although the confounder may impact more than one effect estimate, this function is useful for those interested in any specific direct effect, such as the direct effect from treatment *X* to outcome *Y* (*X* → *Y*). Second, it is also useful for those concerned about a confounder potentially only impacting one effect estimate. For example, when someone has a significant mediation effect and they are concerned about a pretreatment confounder that correlates with *M* and *Y* but not *X*, then they may use this function for the *b* pathway (*M* → *Y*).

```
pkonfound(est_eff, std_err, n_obs, n_covariates, alpha, tails, index =
"IT")
```

`est_eff` specifies the estimated effect of the pathway of interest;

`std_err` specifies the standard error of the pathway of interest;

`n_obs` specifies the sample size;

`n_covariates` specifies the number of covariates in the regression model for estimating the pathway of interest;

`alpha` specifies probability of rejecting the null hypothesis with the default value set as 0.05;

`tails` specifies an integer regarding whether hypothesis testing is one-tailed (1) or two-tailed (2) with the default value set as 2;

`index` specifies what approach we want to use. In this case, we set it as "IT" because we are interested in the correlation-based approach.

Here is an example of using `pkonfound` to evaluate the robustness of inference regarding $\hat{a}$ if researchers are concerned about any omitted pretreatment confounder that correlates with *X* and *M* but not *Y*. This function also works well with any observed covariates as the user only needs to specify the estimated coefficient $\hat{a}$ and its standard error when the observed covariate *Z* is included in the regression model. Assume $\hat{a} = 0.181$, $SE(\hat{a}) = 0.087$ and there is one observed covariate *Z* included in the regression model, then we call `pkonfound(est_eff = 0.181, std_err = 0.087, n_obs = 123, n_covariates_a = 1, alpha = 0.05, tails = 2, index = “IT”)`. Below is a screenshot for the result. In such case, ITCV is the smallest product of $r_{X,CV|Z}$ and $r_{M,CV|Z}$. In other words, the product of $r_{X,CV|Z}$ and $r_{M,CV|Z}$ must be larger than 0.011 to change the inference of $\hat{a}$.

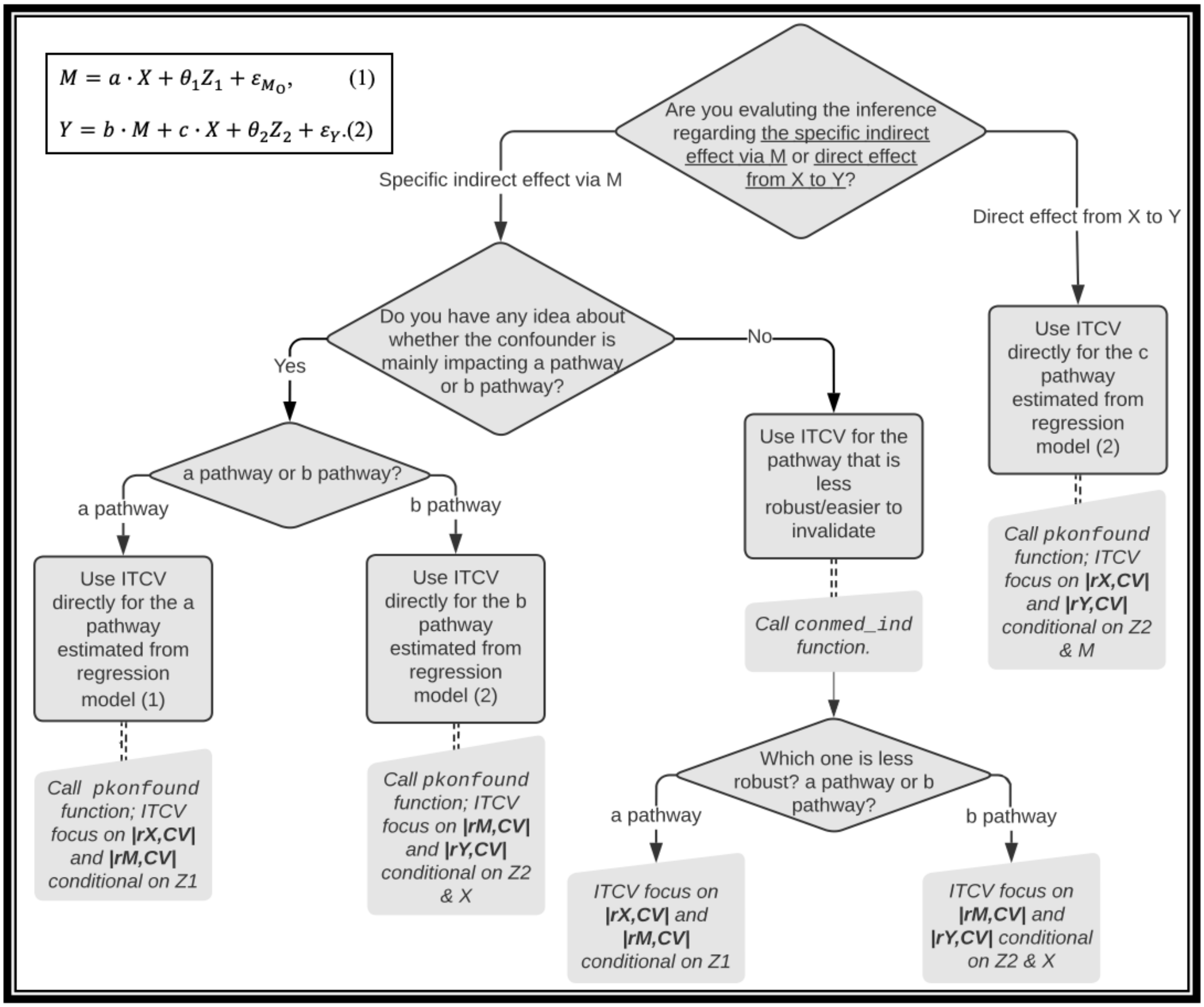
Figure B.1

Example of Using pkonfound for an Omitted Pretreatment Confounder

```
> library(konfound)
> pkonfound(est_eff = 0.181, std_err = 0.087, n_obs = 123, n_covariates = 1, alpha =
 0.05, tails = 2, index = "IT")
Impact Threshold for a Confounding Variable:
The minimum impact to invalidate an inference for a null hypothesis of 0 effect is b
ased on a correlation of  0.103
 with the outcome and at  0.103  with the predictor of interest (conditioning on obs
erved covariates) based on a threshold of
0.179 for statistical significance (alpha = 0.05).

Correspondingly the impact of an omitted variable (as defined in Frank 2000) must be

0.103 X 0.103 = 0.011 to invalidate an inference for a null hypothesis of 0 effect.
See Frank (2000) for a description of the method.

Citation:
Frank, K. (2000). Impact of a confounding variable on the
inference of a regression coefficient. Sociological Methods and Research, 29 (2), 14
7-194
For other forms of output, run ?pkonfound and inspect the to_return argument
For models fit in R, consider use of konfound().
```

Figure B.2

Flow Chart for Application of the Sensitivity Analysis for Different Cases



Note. Z_1 and Z_2 represent observed covariates included in the models.

References

Clarke, K. A. (2005). The Phantom Menace: Omitted Variable Bias in Econometric Research. *Conflict Management and Peace Science*, *22*(4), 341–352. https://doi.org/10.1080/07388940500339183

Frank, K. A. (2000). Impact of a confounding variable on a regression coefficient. *Sociological Methods & Research*, *29*(2), 147–194. https://doi.org/10.1177/0049124100029002001

Fritz, M. S., Kenny, D. A., & MacKinnon, D. P. (2016). The combined effects of measurement error and omitting confounders in the single-mediator model. *Multivariate Behavioral Research*, *51*(5), 681–697. https://doi.org/10.1080/00273171.2016.1224154

Lin, Q. (2019). *Quantifying strength of evidence in education research: accounting for spillover, heterogeneity, and mediation* (Publication No. 27669219) [Doctoral dissertation, Michigan State University]. ProQuest Dissertation and These Global.

Xu, R., Frank, K. A., Maroulis, S. J., & Rosenberg, J. M. (2019). konfound: Command to quantify robustness of causal inferences. *The Stata Journal*, 19(3), 523–550. https://doi.org/10.1177/1536867X19874223