Original Article

Check for updates

Bias and Sensitivity Analyses for Linear Front-Door Models

Felix Thoemmes [1] (ORCID) , Yongnam Kim [2]

[1] *Cornell University, Ithaca, NY, USA.* [2] *Seoul National University, Seoul, South Korea.*

Abstract

The front-door model allows unbiased estimation of a total effect in the presence of unobserved confounding. This guarantee of unbiasedness hinges on a set of assumptions that can be violated in practice. We derive formulas that quantify the amount of bias for specific violations, and contrast them with bias that would be realized from a naive estimator of the effect. Some violations result in simple, monotonic increases in bias, while others lead to more complex bias, consisting of confounding bias, collider bias, and bias amplification. In some instances, these sources of bias can (partially) cancel each other out. We present ways to conduct sensitivity analyses for all violations, and provide code that performs sensitivity analyses for the linear front-door model. We finish with an applied example of the effect of math self-efficacy on educational achievement.

Keywords

causal inference, front-door, bias, measurement error, sensitivity analysis

The front-door criterion (Pearl, 1995) is a method to recover an unbiased causal effect in the presence of unobserved confounding. The simplest possible example of the method is a tri-variate model with a putative cause, an intermediary variable (mechanism), and an outcome variable. The cause and the outcome are allowed to have correlated errors, meaning that unobserved confounding can be present. To make it more concrete, consider that the putative cause is self-efficacy and the effect is academic achievement. Self-efficacy was not randomized and no other control variables were available for

adjustment. Under a set of assumptions that we will explain in more detail later, we can still obtain an unbiased causal effect estimate. In particular, we need to have knowledge about the mechanism of the causal effect. For example, we may assume that self-efficacy will affect academic achievement only through changing students' time spent studying. Once we fully know this mechanism (together with additional assumptions about the unconfoundedness of certain paths clarified later), the front-door model allows the estimation of the causal effect of a potentially confounded cause.

The front-door model is not well known to psychological scientists, despite the fact that it is a simple variation of the hugely popular mediation model (Baron & Kenny, 1986). For the benefit of the reader who may not be acquainted with this model, we first outline the basic idea of the front-door model. In our examination of the model, we will limit ourselves to purely linear models, following work by others (Ding & Miratrix, 2015; Pearl, 2015; Thoemmes, 2015). The use of linear models simplifies results, although they do not necessarily generalize to broader classes of models. However, insights from linear models can serve as a tool to aid understanding (Pearl, 2013). For additional work on bias in front-door models, see Glynn and Kashin (2018), or Bellemare et al. (2019).

Throughout our analytic results we use the term "estimator" to describe the procedure or formula that is used to estimate a causal estimand (defined as a causal quantity of interest). We do not discuss efficiency or how to derive standard errors of the front-door estimator. Literature on mediation analysis (Hayes & Scharkow, 2013; MacKinnon et al., 2012; Tofighi & MacKinnon, 2011) can be fruitfully consulted for the derivation of the standard errors of the front-door model. Suggested ways to estimate standard errors for product terms in mediation or front-door models are the delta method (Sobel, 1982), numerical integration (MacKinnon et al., 2007), or bootstrapping (Shrout & Bolger, 2002).

The remainder of this paper is organized as follows. First, we analytically derive bias of the front-door estimator and compare it with a simple regression estimator, which we refer to as the naive estimator. We also explain how measurement error would affect the estimators. Then we show how sensitivity analysis of the front-door model can be implemented. We then illustrate the methods using a real data analysis and conclude with a summary of findings and implications of the front-door models.

Analysis of Front-Door Models

Setup

In the basic front-door model in Figure 1, X transmits a causal effect to Y, via M. The variables X and Y share a common cause, shown as a bi-directed arrow (left). The bi-directed arrow can be replaced with an unobserved (latent) variable that has direct effects going into both variables (right). Assuming $\delta_X \delta_Y = \delta_{XY}$, the two models in

Figure 1 are equivalent. Based on the right graph, we construct an underlying data-generating model expressed in the following linear equations:

$$U = e_U$$
$$X = \delta_X U + e_X$$
$$M = \alpha X + e_M$$
$$Y = \delta_Y U + \beta M + e_Y$$
$$e_U \sim N(0, \sigma_U) \qquad (1)$$
$$e_X \sim N(0, \sigma_X)$$
$$e_M \sim N(0, \sigma_M)$$
$$e_Y \sim N(0, \sigma_Y)$$
$$e_U \amalg e_X \amalg e_M \amalg e_Y$$

Figure 1

Front-Door Model With (Right) and Without (Left) Latent Variable Shown

Note. Latent variables are shown with a dashed ellipse. Observed variables are shown with a solid square.

Disturbance terms, denoted with *e*, are explicitly added in the equation, although they are not shown in the graph. We assume that all disturbance terms are independent of each other, as denoted by the last line of the formulas. For ease of exposition but without loss of generality, we assume that all variables are standardized with zero mean and unit variance. Note that in order to achieve marginal unit variance for each variable, the residual (unexplained) variances of each variable need to be adjusted based on the magnitude of the effects of other variables (Kenny, 1979). While it is not necessary to assume normality of the error terms, we do so because in later data-generating models for our simulation, we will use standardized and normally distributed variables.

Given the model shown in Equation (1), a regular regression model predicting *Y* from *X* alone fails to recover the true causal effect, but the front-door adjustment can yield unbiased estimates, as shown here with a brief example. Assume that the effects δ_X and δ_Y of the unobserved variable *U* on both *X* and *Y* are .625 and .4, respectively. Equivalently, we may say that the coefficient δ_{XY} of the bi-directed arrow is .25 (= .625 × .4). Further assume that the constituent paths α and β of the total effect from *X* to *Y* through

M are both .3. Then, the total causal effect of X on Y, which is the product of the two paths, $.3 \times .3 = .09$, cannot be recovered by the simple regression model of Y on X. The standardized regression coefficient b_{YX} is equivalent to the correlation coefficient between X and Y, and it can be easily computed by path-tracing rules (Wright, 1934). In order to compute the implied correlation of a structural model comprised of standardized variables, one needs to multiply all path coefficients of every trace, and sum these products of traces of every *open* path between two variables. An open path is defined as a path that does not contain a colliding variable, or short, *collider*, which is a variable that has two arrowheads going into it on a single path (e.g., in path $A \rightarrow C \leftarrow B$, C is a collider; see Elwert and Winship, 2014). In our example, there are two open paths between X and Y in Figure 1, $X \rightarrow M \rightarrow Y$ and $X \leftarrow U \rightarrow Y$ (or simply $X \leftrightarrow Y$). Note that both the mediator M and confounder U are not colliders here because there are no two arrowheads that are colliding into them. Further note that the status of a variable as a collider is always path-dependent, which means that a variable can be a collider on one path, but may not be a collider on a different path. The first path yields a product of the path coefficients of $.3 \times .3 = .09$, and the second path, $.625 \times .4 = .25$. Hence the implied correlation between X and Y is $.09 + .25 = .34$. The latter part of this simple expression is often called confounding bias and this is why a regular regression analysis fails to estimate the unbiased effect of X on Y.

However, despite the unobserved confounding via U, the front-door model allows for the estimation of an unbiased effect by estimating the component paths of the total effect without bias. First, the path from X to M is unconfounded, as there are no back-door paths between X and M. The only path from X to M is a direct, front-door path, whose magnitude is simply given by the standardized regression coefficient b_{MX} obtained by regressing M onto X. This is identical to the path coefficient $\alpha = .3$. Second, the path from M to Y is confounded due to the open back-door path $M \leftarrow X \leftrightarrow Y$. This path can be trivially closed by adjusting on X. From a linear regression model predicting Y from M, and adjusting for X, we can derive the standardized regression coefficient $b_{YM|X}$ using the recursive formula (see Appendix, especially Equation (A4) for the derivation),

$$b_{YM|X} \quad = \quad \frac{\rho_{MY} - \rho_{MX}\rho_{YX}}{1 - \rho_{XM}^2}$$

$$= \quad \frac{(\alpha\delta_{XY} + \beta) - \{\alpha(\alpha\beta + \delta_{XY})\}}{1 - \alpha^2}$$

$$= \quad \frac{\alpha\delta_{XY} + \beta - \alpha^2\beta - \alpha\delta_{XY}}{1 - \alpha^2} \qquad (2)$$

$$= \quad \frac{\beta(1 - \alpha^2)}{1 - \alpha^2}$$

$$= \quad \beta \, .$$

Again, note that $\delta_{XY} = \delta_X\delta_Y$ and all variables are standardized. Thus, the second component effect is given by $\beta = .3$. The final step of the front-door adjustment is then to realize that the total effect of X on Y can be derived from the product of the two unbiased component paths (Pearl, 2009). Both standardized regression coefficients, b_{MX} and $b_{YM|X}$, are estimated to be .3, therefore, the product of the two regression coefficients is .09, which is an unbiased estimate of the true total causal effect from X on Y. Next, we extend this work to consider situations in which underlying assumptions are violated.

Assumptions and Sources of Bias

The front-door model rests on a set of assumptions that should be argued for by the researcher. These assumptions are qualitative statements as to why a certain path in a causal graph is indeed believed to be absent. We review these assumptions and then derive expected biases due to violations.

- *Full mediation.* The first assumption is the absence of a direct path from the cause X to the outcome Y. This assumption states that any effect from X needs to be fully mediated by the mechanism M. The researcher needs to argue that any effect of X on Y goes only and fully through some mechanism M. This mechanism could be through a single variable, or multiple variables, but it is a requirement that the totality of these variables accounts for the total effect. If the assumed mechanism is incomplete, meaning that there are additional front-door paths from X to Y that are ignored, then the front-door estimator of the total effect of X on Y will be biased.
- *No confounding between X and M.* The second assumption relates to the unconfoundedness of the relationship between the cause X and the mechanism M. The researcher must assume that there is no open back-door path, and hence no confounding through

common causes, between X and M. Without this assumption, it is impossible to obtain an unbiased estimate of the component path, and thus of the total effect of X on Y.

- *No confounding between M and Y.* The third assumption relates to the conditional unconfoundedness between the mechanism M and the outcome Y, adjusting for X. There can be no additional back-door paths that do not traverse X. Unmeasured confounders of this component path will also make it impossible to estimate the total effect of X on Y without bias.

- *No measurement error.* An additional potential source of bias is measurement error in variables. Measuring only a noisy proxy variable, instead of the true underlying variable, is known to induce bias in causal effects (Greenland & Robins, 1985; Kuroki & Pearl, 2014). In the case of the standardized linear front-door model, measurement error in any of the three variables, X, M, or Y, can induce bias.

Derivation of Bias

We will derive bias for violations of all assumptions, first considering the most general case, and then reducing the complexity of violations. Discussions are restricted to purely linear models, with completely standardized, continuous variables (formulas for unstandardized variables can be found in the Appendix).
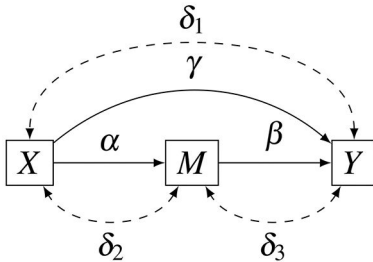
Bias Formulas

We consider a situation in which the violations that we described appear simultaneously, as shown in Figure 2. The component paths from X to M, and from M to Y, are both confounded, and a direct path between X and Y is present. The only violation that we do not consider yet is measurement error, which we will discuss separately in a later section. Note that in this model, we simply use numerated subscripts on the confounding terms, labeled δ_1, δ_2, and δ_3. These bi-directed arrows could have been presented with a latent variable and two arrows emanating from this latent variable. In our simulations, we always include the latent variable. In the model in Figure 2, the true total effect of X on Y is the sum of the two causal paths $\alpha\beta$ and γ. A naive estimator, which simply regresses Y on X, will be biased due to the presence of confounding. The naive estimator is

$$
\begin{aligned}
b_{YX} &= Cov(X,Y)/Var(X) = \rho_{XY} \\
&= (\alpha\beta + \gamma) + \delta_1 + \delta_2\beta.
\end{aligned}
\tag{3}
$$

Figure 2

Front-Door Model With a Direct Path and Confounding on Both Component Paths

The derivation of ρ_{XY} can be easily obtained from Equation (A4) in the Appendix. Bias is added due to the presence of the confounding path between X and Y ($X \leftrightarrow Y$), but also due to the confounder between X and M, and its product with β ($X \leftrightarrow M \rightarrow Y$). By subtracting the true total effect of X on Y (i.e., $\alpha\beta + \gamma$), we can write the bias of the naive estimator as

$$Bias^N = \delta_1 + \delta_2\beta, \tag{4}$$

where the superscript N denotes the naive estimator. Note that δ_3 does not appear in Equation (4) because the assumption of no confounding between M and Y is not required for the estimation of the total effect of X on Y. The non-causal path via δ_3 remains blocked unless M is conditioned on because M is a collider in the path via δ_3 (Elwert & Winship, 2014).

   The bias of the front-door estimator can be derived by using path-tracing rules to compute all implied correlations between two variables, and then the recursive formula to estimate the coefficients from the regression equations that are used for the two component paths. A detailed derivation of this is provided in the Appendix. The first component path from the simple regression of M on X is

$$b_{MX} = \alpha + \delta_2, \tag{5}$$

indicating that δ_2 induces bias in this component path.

   The second component path from M to Y is estimated by the regression of Y on M, adjusting for X, and resolves to

$$b_{YM|X} = \beta + \frac{\delta_3 - \delta_1\delta_2}{1 - (\alpha + \delta_2)^2}. \tag{6}$$

The amount of bias in the second component path can be further parsed into two components. The first bias component of the numerator is δ_3, which is the unobserved confounding via the path $M \leftrightarrow Y$. The second bias component of the numerator is the product of δ_1 and δ_2, but is subtracted. This is due to conditioning on the collider X in the path $M \leftrightarrow X \leftrightarrow Y$. These two bias components are divided by $1 - (\alpha + \delta_2)^2$. Note that $(\alpha + \delta_2)^2$ is the amount of explained variance in M, after adjusting on X, and the square of the correlation between X and M. This term will always be positive, and always increase any existing bias that appears in the numerator. This phenomenon has been studied in the presence of (near-)instrumental variables, and is referred to as *bias amplification* (Myers et al., 2011; Pearl, 2011; Steiner & Kim, 2016).

The front-door estimate is the product of the two biased component paths in Equation (5) and (6):

$$
\begin{aligned}
b_{MX} \times b_{YM|X} &= (\alpha + \delta_2)(\beta + \frac{\delta_3 - \delta_1\delta_2}{1 - (\alpha + \delta_2)^2}) \\
&= \alpha\beta + \delta_2\beta + \frac{(\alpha + \delta_2)(\delta_3 - \delta_1\delta_2)}{1 - (\alpha + \delta_2)^2}.
\end{aligned}
\tag{7}
$$

By subtracting the true total effect $(\alpha\beta + \gamma)$, the bias of the front-door estimator, superscripted by F, is

$$
Bias^F = \delta_2\beta + \frac{(\alpha + \delta_2)(\delta_3 - \delta_1\delta_2)}{1 - (\alpha + \delta_2)^2} - \gamma.
\tag{8}
$$

In addition to the bias component $\delta_2\beta$, the bias of the front-door estimator includes additional bias sources. First, the amount of the direct effect of X on Y (γ) appears in the bias formula and any non-zero value for this parameter influences bias in the front-door estimator in a monotonic fashion. The parameter γ appears in the bias formula as a *subtraction*, therefore, the direct effect of X on Y can offset the other bias components, and thus may decrease the overall bias. Depending on the sign of γ (and other parameters), it may also increase overall bias.

Second, the bias amplification term, $1 - (\alpha + \delta_2)^2$ appears again in the denominator. This is expected because the front-door estimator is the product of Equation (5) and (6). We again note that the expression $\alpha + \delta_2$ is the correlation between X and M. The squared correlation is the amount of variance in M, explained by X, and subtracting this from one is the residual variance in Equation (5). Thus, the higher the correlation between X and M, the stronger the bias amplification will be.

Third, the impact of bias amplification rests on the numerator because only non-zero numerators will be amplified. The numerator consists of two components, $(\alpha + \delta_2)$ and

$(\delta_3 - \delta_1\delta_2)$. If either of those were zero, then the whole fraction becomes zero. The first component is the correlation between X and M. The second component, $\delta_3 - \delta_1\delta_2$, concerns the unobserved common causes. If the unobserved common cause between M and Y (δ_3) takes on the exact same value as the product of the other common causes between X and Y, and X and M $(\delta_1\delta_2)$, then this bias component vanishes. The product term $\delta_1\delta_2$ essentially concerns all back-door paths for the cause X and its effects on either M or Y.

Comparisons of Biases

Direction of Bias Change — Both the naive estimator and the front-door estimator can yield biases when their assumptions are violated. Each violation impacts each bias differently, and this may even result in bias in the opposite direction. In Table 1, we summarize signs of the relationship between specific violations and the sign of change in the resulting bias for the naive and front-door estimators when holding all other parameters constant. It shows that each violation can change each bias for the naive and front-door estimators in different directions. The first entry in Table 1 is γ, representing the violation of the full mediation assumption. Any value of γ has no effect on the bias of the naive estimator ($Bias^N$), hence we denote it by "0". On the other hand, the relationship between γ and the bias of the front-door estimator ($Bias^F$) is negative, meaning that with increases in γ, $Bias^F$ changes in a negative direction, which we denote with "–". Thus, a *positive* violation of the full mediation assumption ($\gamma > 0$) results in *negative* bias to the front-door estimator.

Table 1

Sign of the Relationship Between Specific Violations for Naive and Front-Door Estimators and Resulting Bias Holding All Other Parameters Constant

Violations	Bias of naive estimator $\delta_1 + \delta_2\beta$	Bias of front-door estimator $\delta_2\beta + \frac{(\alpha + \delta_2)(\delta_3 - \delta_1\delta_2)}{1 - (\alpha + \delta_2)^2} - \gamma$
Full mediation (γ)	0	–
No confounding between X & Y (δ_1)	+	$\mathrm{sgn}(-\delta_2(\alpha + \delta_2))$
No confounding between X & M (δ_2)	$\mathrm{sgn}(\beta)$	complex
No confounding between M & Y (δ_3)	0	$\mathrm{sgn}(\alpha + \delta_2)$

Note. $\delta_1 = 0$ is not required for the front-door model.

The next table entry shows how the violation of no confounding between X and Y (δ_1) impacts both the naive and the front-door estimators. A positive value of δ_1 would induce a positive bias for the naive estimator (denoted by "+"). Considering that the

front-door adjustment is a way to circumvent the typical confounding problem between X and Y like δ_1, it might be surprising that δ_1 also changes $Bias^F$. This, however, only happens when other sources of confounding are also present. The direction of $Bias^F$ as a function of δ_1 is given by the sign of $(-\delta_2(\alpha + \delta_2))$. Note that due to the negative sign of the term, δ_1 can induce a negative bias even though all parameters are positive. Only if $\delta_2 = 0$ (or $\alpha + \delta_2 = 0$), will the parameter δ_1 not change $Bias^F$. Thus, any confounding between X and Y affects the front-door estimator only if the confounding between X and M (δ_2) is also present.

The next table entry shows the impact of the violation of no confounding between X and M (δ_2). $Bias^N$ is affected by this violation, and the sign of the bias is determined by the effect of M on Y (β). This is because δ_2 is part of confounding between X and Y, which will be combined with the path $M \rightarrow Y$. However, the impact of δ_2 on $Bias^F$ is complicated and its sign is not easily formulated even in linear models because δ_2 appears in multiple components in the bias formula of the front-door estimator in Equation (8).

Finally, the last entry in Table 1 shows that the violation of no confounding between M and Y does not affect $Bias^N$ while it monotonically changes $Bias^F$ in the direction of the sign of $(\alpha + \delta_2)$. If one wants to predict in which direction $Bias^F$ changes by varying the value of δ_3, it is sufficient to check the sign of the correlation between X and M.
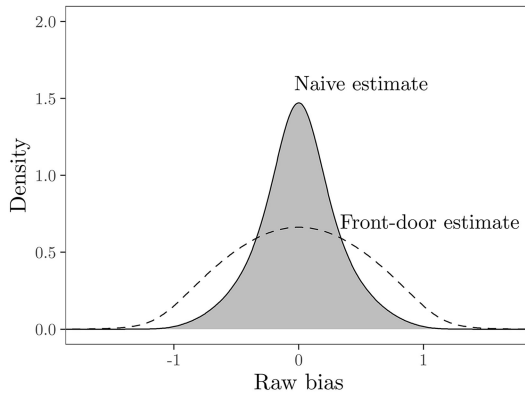
Magnitude of Absolute Bias — Which of the two estimators incurs larger absolute biases? There are many constellations of parameters in which the naive or the front-door estimator would be preferable. We first consider a very broad comparison and examine absolute bias for both estimators over a range of parameter values for the model in Figure 2. We compute bias for a grid of values ranging from $-.9$ to $.9$ (evenly spaced over 9 values) for all nine parameters in the model. Note that for each δ we specify two parameters since the bi-directed arrows are simply a shorthand for " $\leftarrow U \rightarrow$ " where U is an unknown common cause (see Figure 1). For every combination of parameter values, we observe whether the resulting correlation matrix has only positive variances and discard combinations of parameters that violate this condition. This is identical to discarding correlation matrices that are not positive definite. Finally, we compare bias of the two estimators across all parameter values. This broad comparison answers the question whether the bias in the naive estimator is on *average* larger or smaller than the bias in the front-door estimator.

Figure 3 shows the bias for both estimators. For presentation purposes, we display the discrete bias estimates using a smooth kernel density estimate. The naive estimator (solid line) has bias that is concentrated around smaller values. The bias of the front-door estimator (dashed line) can become large, even though this is generally rare. In our simulation, the absolute bias of the naive estimator is larger than or equal to the absolute bias of the front-door estimator ($|Bias^N| \geq |Bias^F|$) in only 27% of cases, indicating that the naive estimator is often preferred in terms of absolute bias. An explanation of this

pattern is that the front-door estimator is subject to bias amplification, and has more potential sources of bias. However, as we shall see in the next section, this pattern is not always true, and can reverse under some instances.

Figure 3

Density of Raw Bias of Both Estimators Averaged Over All Parameter Values

Constraining Violations

From our general bias formula, we can derive what would happen if some violations would be constrained to be zero, meaning that the violation would not appear. There is a total of four constraints that we can consider, namely the three bias terms, δ_1, δ_2, δ_3, and the direct effect γ. We will discuss some, but not all of the constraints in some detail, and then present a summary of all 16 possible combinations of constraints.

First, we can omit the direct path γ by setting it to zero ($\gamma = 0$), that is, we assume that the full mediation assumption is met. The naive estimator would still only be biased due to the presence of $\delta_1 + \delta_2\beta$, and the front-door estimator would not incur bias due to γ. The two bias formulas are,
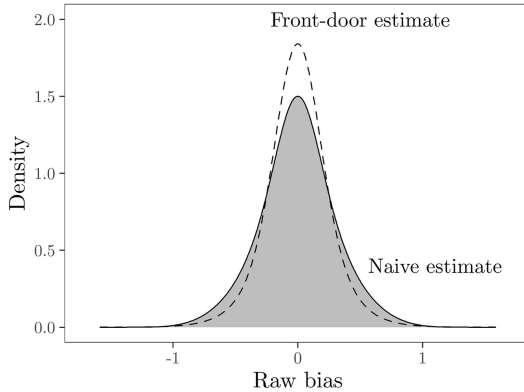
$$
\begin{aligned}
Bias^N_{\gamma = 0} &= \delta_1 + \delta_2\beta; \\
Bias^F_{\gamma = 0} &= \delta_2\beta + \frac{(\alpha + \delta_2)(\delta_3 - \delta_1\delta_2)}{1 - (\alpha + \delta_2)^2},
\end{aligned}
\tag{9}
$$

where we denote the constraining condition as a subscript. The comparison of bias under this constraint is presented in Figure 4. The naive estimator has the same amount of bias as in Equation (4), but the front-door estimator has bias that is more concentrated around zero. The front-door estimates now have less or equal bias than the naive estimates in

67% of all conditions, which is a reversal of the previous finding. This highlights the importance of the full mediation assumption for the front-door adjustment.

Figure 4

Density of Raw Bias of Both Estimator Averaged Over All Parameter Values When $\gamma = 0$

We can constrain the violations further. One can assume that in addition to the full mediation assumption, there is also no confounding between X and M ($\gamma = 0$ & $\delta_2 = 0$). This simplifies both bias formulas. Equation (9) reduces to

$$
\begin{aligned}
Bias^N_{\gamma = 0, \delta_2 = 0} &= \delta_1; \\[2mm]
Bias^F_{\gamma = 0, \delta_2 = 0} &= \frac{\alpha\delta_3}{1 - \alpha^2}.
\end{aligned}
\tag{10}
$$

Under this additional assumption of no confounding between X and M, the front-door estimator has less or equal absolute bias than the naive estimator in 70% of all instances.

If we were to constrain both the direct path and the confounding between M and Y ($\gamma = 0$ & $\delta_3 = 0$), the two bias formulas reduce to

$$
\begin{aligned}
Bias^N_{\gamma = 0, \delta_3 = 0} &= \delta_1 + \delta_2\beta; \\[2mm]
Bias^F_{\gamma = 0, \delta_3 = 0} &= \delta_2\beta + \frac{-\delta_1\delta_2^2 - \alpha\delta_1\delta_2}{1 - (\alpha + \delta_2)^2}.
\end{aligned}
\tag{11}
$$

Here, the front-door estimator is less biased than the naive estimator in 90% of all conditions. We repeat this bias comparison for every possible combination of violations of assumptions. The results are presented in Table 2.

Table 2

Probability That the Absolute Bias of Front-Door Estimators is Less Than (or Equal) to the Absolute Bias of Naive Estimators in the Grid of Parameter Values That We Considered

Condition				Probability of	
$\gamma = 0$	$\delta_1 = 0$	$\delta_2 = 0$	$\delta_3 = 0$	$|Bias^N| \geq |Bias^F|$	$|Bias^N| > |Bias^F|$
Naive estimator better					
				.27	.25
	✓			.14	.09
		✓		.25	.24
			✓	.28	.24
✓	✓			.47	.17
	✓	✓		.06	.00
	✓		✓	.20	.05
✓	✓	✓		.39	.00
		✓	✓	.27	.23
	✓	✓	✓	.15	.00
Front-door estimator better					
✓				.67	.58
✓		✓		.70	.60
✓			✓	.90	.63
✓	✓		✓	1.00	.00
✓		✓	✓	1.00	.74
Both estimators are equal					
✓	✓	✓	✓	1.00	.00

Note. The grid of parameter values we considered ranges from $-.9$ to $.9$ (evenly spaced over 9 values). The check-mark symbol (✓) indicates that the corresponding condition is constrained in computing the probability.

The information contained in the table is complex, as it spans many different constellations of violations of assumptions. To organize the results, we have included sub-headings in the table that sort constellations of assumption violations by which estimator is better, which simply means that it will produce *on average* less bias. In the first segment of the table are all conditions listed in which the naive estimator performs better. What is noteworthy, yet expected, is that the majority of these conditions fulfill the assumption that δ_1 is zero. This assumption states that there is no confounding between X and Y. In the second segment of the table are all conditions in which the front-door estimator is
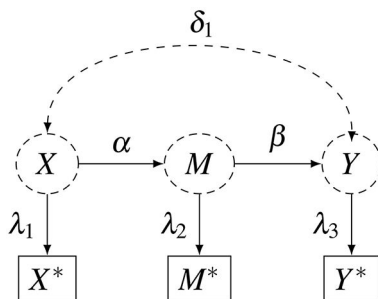
generally preferred. Noteworthy here is that in all instances the assumption that γ is zero is fulfilled. As a reminder, this assumption states that there is full mediation between X and Y, and no direct path. We conclude form this table that the absence of confounding between X and Y is the most critical condition for the naive estimator, and the absence of a direct path between X and Y is the most critical assumption for the front-door estimator. When neither or both of these conditions are met, the naive estimator tends to perform better. Finally, the impact of either δ_2 or δ_3 being zero is comparatively less important. This information can be useful for applied researchers who may have to weight the plausibility of the assumptions. Trivially, and only reported for completeness is the last segment of the table that states that if all assumptions are met, and there is no confounding at all, then both estimators perform equally. R code to replicate our results can be found in Thoemmes (2021).

Measurement Error

We consider presence of independent measurement error in the front-door model, as shown in Figure 5. We conceptualize measurement error as having an imperfect proxy of a variable, with causal relationships among the unobserved latent variables of such proxies. More complex forms of measurement error are described in Hernán and Cole (2009). In Figure 5, observed variables are shown in squares, and have an asterisk superscript, and latent variables are shown in dashed ellipses. Here, we evaluate the impact of measurement error in the absence of the confounding biases, but a fully complex model can be found in the Appendix.

Figure 5

Front Door Model With Measurement Error

The bias due to measurement error is multiplicative, and requires knowledge of the value of λ. The naive estimator from the regression of Y^* on X^* is $b_{Y^*X^*} = \lambda_1\alpha\beta\lambda_3 + \lambda_1\delta_1\lambda_3$, indicating that the measurement error is multiplicative and will attenuate the estimate (since absolute values of standardized loadings are always equal or lower than 1). We

can rearrange this formula to $\lambda_1\lambda_3(\alpha\beta + \delta_1)$ to highlight the fact that the naive estimator is identical to the case without measurement error (i.e., $\lambda_1 = \lambda_3 = 1$), with the only exception that all terms are multiplied by the product of the loadings. At the same time, we can see trivially that measurement error in M is completely inconsequential to the bias in the naive estimator. This is expected, as M is not considered in the estimation of the naive effect.

In the front-door adjustment, the first component path is estimated to be $b_{M^*X^*} = \lambda_1\alpha\lambda_2$, hence it is also incurring attenuation based on the reliabilities of X^* and M^*. The second component path is more complicated, and resolves to:

$$b_{Y^*M^*|X^*} = \frac{\lambda_2\beta\lambda_3(1 - \lambda_1^2\alpha^2) + (1 - \lambda_1^2)(\lambda_2\alpha\delta_1\lambda_3)}{1 - (\lambda_1\alpha\lambda_2)^2}. \tag{12}$$

This expression yields insights into the impact of measurement error. First, the impact of the biasing back-door path $M^* \leftarrow M \leftarrow X \rightarrow Y \rightarrow Y^*$ is only partially removed by conditioning on X^*. The amount of residual bias varies as a function of $1 - \lambda_1^2$. When λ_1 equals 1, meaning no measurement error at all on X^*, the multiplicative terms become zero, and there is no bias due to the confounding path. Conversely, if λ_1 were to drop to zero (meaning that the proxy is pure noise), there would be no reduction in bias at all.

The formula for bias reduces to the following expression if X is measured without error (i.e., $X^* = X$, implying $\lambda_1 = 1$):

$$b_{Y^*M^*|X} = \frac{\lambda_2\beta\lambda_3(1 - \alpha^2)}{1 - (\alpha\lambda_2)^2}. \tag{13}$$

The path between M and Y, labeled as β, is attenuated due to measurement error in M^* and Y^*. If λ_2 were to increase to 1 (meaning no measurement error in M^*), then the attenuation would be only due to λ_3, and thus the measurement error in Y^*. The multiplicative term $1 - \lambda_1^2\alpha^2$ would cancel out of the numerator and denominator in such a situation.

The formula would further reduce to the following expression if both X^* and M^* were measured without error (i.e., $X^* = X$ and $M^* = M$, implying $\lambda_1 = 1$ and $\lambda_2 = 1$, respectively):

$$b_{Y^*M|X} = \frac{\beta\lambda_3(1 - \alpha^2)}{1 - \alpha^2} = \beta\lambda_3. \tag{14}$$

In summary, measurement error complicates the derivation of bias, and tends to attenuate relationships between variables. Measurement error in X mostly leads to residual confounding bias due to back-door paths that are only closed due to conditioning on

noisy proxy variables. Measurement error in all variables tends to attenuate the estimate of both component paths. In addition, measurement error in Y, as opposed to in X or M, does not create bias in estimation of unstandardized regression weights, only in standardized weights by increasing the total variance of Y [1].

Sensitivity Analysis

Using the front-door estimator requires commitment to its untestable assumptions. That means that an applied researcher must—on theoretical grounds—argue that the assumptions hold. A critic could argue that some assumptions might be violated, and in response a sensitivity analysis could be produced. VanderWeele (2010) and VanderWeele and Arah (2011) argue that every time untestable causal assumptions are invoked, a sensitivity analysis is advisable. This analysis can be performed to assess violations of all of the previously mentioned assumptions.

The logic of sensitivity analysis is to consider what would happen to an estimate of an effect, if an assumption was violated. Often it is of interest to argue how large of a violation would have to occur in order for a conclusion to change. This is sometimes conceptualized as an effect changing from a significant (positive or negative) value, to a value that cannot be significantly distinguished from zero (Rosenbaum, 2005). Alternatively, a researcher may ask at what level of violation the observed effect becomes identical to some pre-specified threshold of smallest effect of interest, or at what point it would change to a statistically significant effect of the opposite sign.

Phantom Variable Approach to Sensitivity Analysis

Sensitivity analysis in linear models often relies on analytic solutions to derive closed-form expressions on how an estimator would change if an unobserved confounder was present. The impact of the confounder is typically expressed as a semi-partial R^2, but standardized path coefficients can also be used. The advantage of closed-form solutions is that thresholds of interest (e.g., when an effect ceases to be statistically significant) can immediately be found.

We present an alternative approach, that might be less convenient than closed-form solutions, but highly accessible to practitioners. Our approach to sensitivity analyses uses phantom variables (Rindskopf, 1984) in a linear structural equation model (SEM). A phantom variable is a constructed variable whose variances, and covariances are fixed by the researcher. The term *phantom variable* is used because no data on this variable is observed. In order for any SEM with a phantom variable to be estimable, all parameters that are associated with any phantom variable must be fixed with assumed values. On a

---

1) We thank an anonymous reviewer for pointing out this fact.

conceptual level, a researcher would specify their SEM as usual, and then add phantom variables that specifically violate some of the assumptions. Phantom variables have been used to conduct sensitivity analyses in SEM (Harring et al., 2017), and our approach builds on this work.

In a sensitivity analysis for the front-door estimator, a researcher would define the model and then add as many potential violations as desired, in the form of phantom variables. If it would be desired to introduce a violation of the assumption that there is no unmeasured confounder between X and M, a phantom variable is introduced that has an effect on both these variables. If a violation of the full mediation assumption is of interest, a phantom variable between X and Y (omitted mechanism) is introduced. This process can be repeated for other violations of assumptions, and for joint violations of several assumptions. In the case of single violations, it is often sufficient to vary the degree of the violation (expressed through the magnitude of the path coefficients), and then examine at which values an observed effect ceases to be significant, or alternatively drops below a certain threshold, or becomes significant in the opposite direction. In the case of two violations, a grid of path coefficients for both violations can be constructed. As the number of violations increases, it will often be necessary to explore a large multi-dimensional grid. Then, one might marginalize results over some sensitivity parameter values, or condition on some values of interest. For example, consider a researcher who wants to assess sensitivity to all violations simultaneously. Varying sensitivity parameter values over a large range will yield an extremely large array. To consider the impact of a single violation in this array, the researcher can decide to average over all other violations, and then make statements about the impact of one violation, averaged over all other violations. Alternatively, the researcher could fix all other violations at particular values, and then evaluate the impact of one violation at a time. Computations can be performed in any SEM software, including R using the *lavaan* (Rosseel, 2012), or *blavaan* package (Merkle & Rosseel, 2018). To assist applied researchers, we provide examples of sensitivity analyses and their reproducible code in R can be found in Thoemmes (2021).

Illustration

To illustrate the use of the front-door model and associated sensitivity analyses we analyze data from the Korean Educational Longitudinal Study (KELS) provided by the Korean Educational Development Institute. The KELS contains data from students in Korea collected over several years, and in our example we use the KELS 2009 and 2010 datasets. One possible analysis that we explore here is the relationship between self-reported math efficacy (the belief in one's confidence and proficiency in math) and a standardized math assessment score, here the Korean CSAT (College Scholastic Ability Test) math score. Self-efficacy in math and standardized math scores are most likely confounded through a variety of factors that include person characteristics, school

environments, or family upbringing and parental guidance. We consider as a potential mechanism for the front-door criterion the amount of hours spent studying on math, assessed through student self-report of studying at home, through a tutor, and through an educational television show in Korea (Educational Broadcasting System). For the front-door model to be valid, we need to assume that there are no further mechanisms besides hours studied (M), and further that the relationship between self-efficacy (X) and hours studied (M), and likewise the relationship between hours studied (M) and the SAT math score (Y) are unconfounded. Potential violations of these assumptions can be probed via sensitivity analysis.

Results

For our analysis, we used listwise deletion, resulting in a final sample size of 935. Self-efficacy is scored on a scale with a range in the data from 4 to 16, a mean of 9.8 and standard deviation (SD) of 2.6. Hours studied per week ranged from 3 to 28, with a mean of 10.1, and an SD of 4.4. Finally, SAT scores ranged from 56 to 147, with a mean of 98.9 and SD of 17.4. We present the results of an unadjusted (naive) model, along with the front-door model in Table 3. All standard errors of the front-door model are based on 1,000 bootstrap samples.

Table 3

Illustrative Results of the Causal Effect of Self-Efficacy on Math Score Analyzing the KELS Dataset

Estimate	Unstandardized estimate (*SE*)	Standardized estimate (*SE*)	*p*
Naive estimate	2.00 (.21)	.297 (.03)	< .001
Front-door estimate	.507 (.09)	.075 (.014)	< .001
Component path (X on M)	.592 (.053)	.351 (.031)	< .001
Component path (M on Y)	.856 (.140)	.214 (.035)	< .001

The naive estimate is significant, indicating a two point increase in math scores for each point increase in self-efficacy, or .297 on a standardized metric. The front-door estimate was substantially smaller (however, still significant), with an unstandardized estimate of .507 and .075 on a standardized metric. Intuitively this is plausible, because we would assume that there is potential for confounders that are both positively related to self-efficacy and standardized math scores, such as positive home environment, or intrinsic person characteristics.
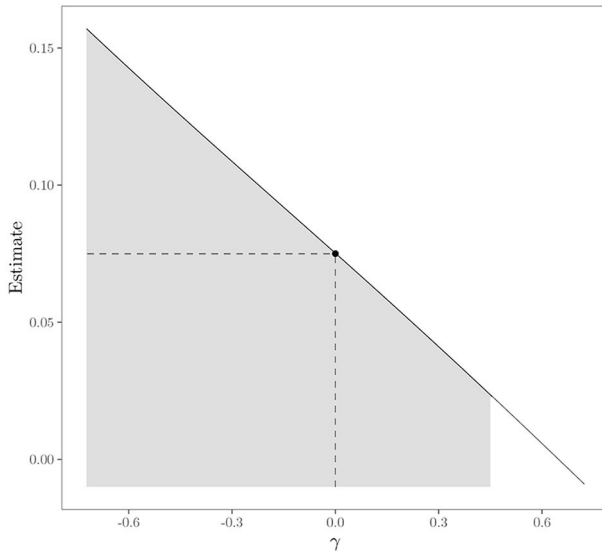
The assumptions that had to be made for the analysis above to be valid are relatively strong. We had to assume no other unmeasured mediators, but it could easily be argued that the hours of studying math is not the only mechanism between self-efficacy and standardized test scores. Possibly, self-efficacy in math also causes less anxiety during

test-taking and thus increases scores on the standardized test. To address this point, one may conduct a sensitivity analysis. We used γ to describe this effect in earlier models. In a sensitivity analysis, γ could take on many different values. For example, if the researcher believes that this unobserved confounding path γ would be .2 (on a standardized metric), we would add a phantom variable on a causal pathway between X and M. To achieve a direct path γ of magnitude .2, we would set the path from X to the intermediary phantom to .447, and the path from the intermediary phantom to the outcome Y to .447 as well. These values are chosen because the direct path γ can be expressed as a product of the two component paths, which in this case would result in .447 × .447 ≈ .2. Typically, we would vary γ over a wide range of values, and in our sensitivity analysis we chose a range of −.85 to .85 on a standardized metric. Figure 6 shows the standardized front-door estimate on the y-axis, and the amount of the violation in form of the γ path on the x-axis. In our particular sample, the observed effect was .075 on the standardized metric, and this effect is plotted as the dashed line, and it is right at γ being zero, which assumes no violation of the assumption. The shaded area under the curve designates for which values of γ the estimate remains significant. As we can see, for any unobserved mechanism with a negative effect, our observed effect would get even larger, and remain significant. For unmeasured mediators with a positive effect, our effect would cease to be significant only if this unmeasured mechanism would exceed 0.455 on a standardized metric. This effect would thus have to be nearly as strong as the one we observed. Arguably, it seems unlikely that an unobserved effect would have an effect of this magnitude, and thus this analysis would bolster our faith in our observed estimate.

The phantom variable approach allows for more complex scenarios. Consider that instead of an unobserved mechanism, we are worried that component paths could be biased by unmeasured confounders. Confounding may occur between self-efficacy (X) and number of hours studying (M), and hours studying (M) and the math score (Y). Intrinsic person characteristics, like conscientiousness, could possibly confound these relationships. To address two possible violations at the same time, two phantom variables can be included, and a sensitivity analysis can be performed over a whole grid of values of δ_2 and δ_3. The resulting contour plot in Figure 7 shows results of this sensitivity analysis. At the intersection where both δ_2 and δ_3 are at 0 is the actual observed front-door estimate (.075). The contour lines are labeled with the value of the estimate that we would observe for different values of the confounding paths, whereas the axes of the plot show the amount of confounding on a standardized metric. The shaded area designates estimates that would reach significance. The plot thus shows regions in which confounding of the component paths would change the conclusion (in terms of statistical significance), or where the conclusions would remain robust. We see that any negative values of confounding between X and M, or M and Y only make the estimate larger. Likewise, the estimate is somewhat robust against positive confounding between X and

Figure 6

Sensitivity of the Standardized Front-Door Estimate to the Direct Effect

Note. The shaded area under the diagonal designates significant estimates.
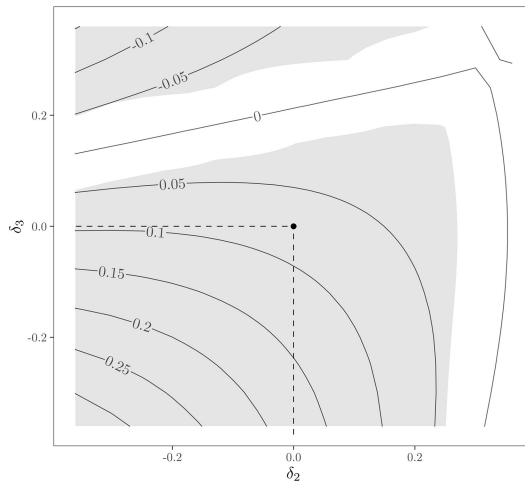
M, and only ceases to be significant for $\delta_2 > .2$. The observed effect is slightly less robust to positive confounding between M and Y and turns to be non-significant for approximately $\delta_2 > .1$.

Discussion

The aim of this paper was to derive the expected bias of the front-door estimator in linear models when necessary assumptions are violated, contrast it with a naive estimator, and to provide guidance on dealing with assumption violations through sensitivity analysis. In the presence of multiple violations, in which confounding and collider biases are introduced, the bias derivation is complex, and depends on a variety of parameters. This complex constellation can yield bias that does not monotonically increase with increases in paths that are inducing bias to the model. Offsetting or cancellation but also amplification of bias can occur. While it is not expected that bias will generally cancel out, it highlights the fact that simply considering the amount of potential confounding is not sufficient, but that also the direction of confounding should be considered when anticipating the amount of bias in a front-door model.

Figure 7

Sensitivity of the Front-Door Estimate to Unobserved Confounding Between X and M, and M and Y

Note. The shaded area designates significant estimates. Note that significant values in the top left of the graph are opposite in sign.

The bias of the front-door estimator was generally larger than that of the naive estimator when all assumptions of the front-door model were violated. However, when the full mediation assumption was met, the front-door estimator often had smaller bias than the naive estimator. This suggests that the front-door model can be preferable over adjusted regression models if the full mediation assumption holds, which resembles results by Glynn and Kashin (2018).

As a limitation of our work, we only consider the linear, tri-variate front-door model with violations of all assumptions, but more complex models that include more than a single cause, mechanism, or outcome, with limited dependent variables, and non-linear relationships, could be considered in future work.

Applied researchers will seldom know the exact magnitude of biasing paths, and therefore it will often be advisable to conduct sensitivity analyses. We offered some suggestions on how to conduct sensitivity analyses for the front-door estimator, and we provide explanation and computer code to perform these analyses using phantom variables. The creation of phantom variables allows convenient and precise modeling of unobserved confounding, and allows applied researchers to gauge the influence of violations of assumptions. We hope that this paper will further the adoption of the front-door adjustment in applied research.

PsychOpen GOLD

Supplementary Materials

The supplementary materials provided are the R code scripts to replicate the results of the current study, sensitivity analyses examples, and the results themselves (see Thoemmes, 2021).

Index of Supplementary Materials

Thoemmes, F. (2021). *front-door bias* [R code, sensitivity analyses examples, study results]. OSF. https://osf.io/ujcqr

References

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. https://doi.org/10.1037/0022-3514.51.6.1173

Bellemare, M. F., Bloem, J. R., & Wexler, N. (2019, June). *The paper of how: Estimating treatment effects using the front-door criterion* (Working Paper). http://marcfbellemare.com/wordpress/wp-content/uploads/2020/06/BellemareBloemWexlerFDCJune2020.pdf

Ding, P., & Miratrix, L. W. (2015). To adjust or not to adjust? Sensitivity analysis of M-bias and butterfly-bias. *Journal of Causal Inference, 3*(1), 41–57. https://doi.org/10.1515/jci-2013-0021

Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology, 40*, 31–53. https://doi.org/10.1146/annurev-soc-071913-043455

Glynn, A. N., & Kashin, K. (2018). Front-door versus back-door adjustment with unmeasured confounding: Bias formulas for front-door and hybrid adjustments with application to a job training program. *Journal of the American Statistical Association, 113*(523), 1040–1049. https://doi.org/10.1080/01621459.2017.1398657

Greenland, S., & Robins, J. M. (1985). Confounding and misclassification. *American Journal of Epidemiology, 122*(3), 495–506. https://doi.org/10.1093/oxfordjournals.aje.a114131

Harring, J. R., McNeish, D. M., & Hancock, G. R. (2017). Using phantom variables in structural equation modeling to assess model sensitivity to external misspecification. *Psychological Methods, 22*(4), 616–631. https://doi.org/10.1037/met0000103

Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science, 24*(10), 1918–1927. https://doi.org/10.1177/0956797613480187

Hernán, M. A., & Cole, S. R. (2009). Invited commentary: Causal diagrams and measurement bias *American Journal of Epidemiology, 170*(8), 959–962. https://doi.org/10.1093/aje/kwp293

Kenny, D. A. (1979). *Correlation and causality*. Wiley.

Kuroki, M., & Pearl, J. (2014). Measurement bias and effect restoration in causal inference. *Biometrika, 101*(2), 423–437. https://doi.org/10.1093/biomet/ast066

MacKinnon, D. P., Cheong, J., & Pirlott, A. G. (2012). Statistical mediation analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 313–331). American Psychological Association.

MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods, 39*(3), 384–389. https://doi.org/10.3758/bf03193007

Merkle, E. C., & Rosseel, Y. (2018). blavaan: Bayesian structural equation models via parameter expansion. *Journal of Statistical Software, 85*(4), 1–30. https://doi.org/10.18637/jss.v085.i04

Myers, J. A., Rassen, J. A., Gagne, J. J., Huybrechts, K. F., Schneeweiss, S., Rothman, K. J., Joffe, M. M., & Glynn, R. J. (2011). Effects of adjusting for instrumental variables on bias and precision of effect estimates. *American Journal of Epidemiology, 174*(11), 1213–1222. https://doi.org/10.1093/aje/kwr364

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika, 82*(4), 669–688. https://doi.org/10.1093/biomet/82.4.669

Pearl, J. (2009). *Causality*. Cambridge University Press.

Pearl, J. (2011). Invited commentary: Understanding bias amplification. *American Journal of Epidemiology, 174*(11), 1223–1227. https://doi.org/10.1093/aje/kwr352

Pearl, J. (2013). Linear models: A useful "microscope" for causal analysis. *Journal of Causal Inference, 1*(1), 155–170. https://doi.org/10.1515/jci-2013 0003.

Pearl, J. (2015). Comment on Ding and Miratrix: "To adjust or not to adjust?". *Journal of Causal Inference, 3*(1), 59–60. https://doi.org/10.1515/jci-2015-0004

Rindskopf, D. (1984). Using phantom and imaginary latent variables to parameterize constraints in linear structural models. *Psychometrika, 49*(1), 37–47. https://doi.org/10.1007/BF02294204

Rosenbaum, P. R. (2005). Sensitivity analysis in observational studies. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science*. John Wiley & Sons.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://www.jstatsoft.org/v48/i02/
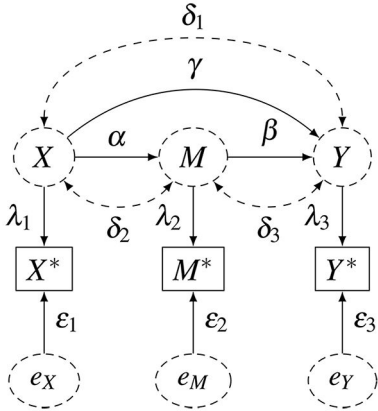
PsychOpen GOLD

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*(4), 422–445. https://doi.org/10.1037/1082-989X.7.4.422

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology, 13*, 290–312. https://doi.org/10.2307/270723

Steiner, P. M., & Kim, Y. (2016). The mechanics of omitted variable bias: Bias amplification and cancellation of offsetting biases. *Journal of Causal Inference, 4*(2), Article 20160009. https://doi.org/10.1515/jci-2016-0009

Thoemmes, F. (2015). M-bias, butterfly bias, and butterfly bias with correlated causes–A comment on Ding and Miratrix (2015). *Journal of Causal Inference, 3*(2), 253–258. https://doi.org/10.1515/jci-2015-0012

Tofighi, D., & MacKinnon, D. P. (2011). Rmediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods, 43*(3), 692–700. https://doi.org/10.3758/s13428-011-0076-x

VanderWeele, T. J. (2010). Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology, 21*(4), 540–551. https://doi.org/10.1097/EDE.0b013e3181df191c

VanderWeele, T. J., & Arah, O. A. (2011). Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology, 22*(1), 42–52. https://doi.org/10.1097/EDE.0b013e3181f74493

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics, 5*(3), 161–215. https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-5/issue-3/The-Method-of-Path-Coefficients/10.1214/aoms/1177732676.full

Appendix: Derivation of Front-Door Formulas

Here we present general unstandardized formulas for the two component effects (i.e., the effect of X on M and the effect of M on Y) of the front-door formula. We consider a linear data-generating model as in Figure A1 where both the confounding and full mediation assumptions are violated, and further all variables are unreliably measured due to independent measurement error. For simplicity but without loss of generality, we assume all exogenous variables, represented by dashed ellipses, such as U_1, U_2, and U_3, which respectively produce the confounding δ_1, δ_2, and δ_3, and e_X, e_M, and e_Y have unit variances.

Figure A1

General Front-Door Model With Measurement Error

The Component Effect of X on M

The population regression coefficient for X^* of the regression of M^* on X^* is expressed by $Cov(X^*, M^*)/Var(X^*)$. As the covariance is expressed as $Cov(X^*, M^*) = \lambda_1\lambda_2\{Var(X)\alpha + \delta_2\}$ (by using path-tracing rules), the regression coefficient for X^*, aiming at the first component effect $(X \rightarrow M)$, is given by

$$b_{M^*X^*} = \frac{\lambda_1\lambda_2\{Var(X)\alpha + \delta_2\}}{Var(X^*)}. \tag{A1}$$

If all variables are standardized, implying that $Var(X) = Var(X^*) = 1$, then Equation (A1) reduces to

$$b_{M^*X^*} = \lambda_1\lambda_2(\alpha + \delta_2). \tag{A2}$$

In addition, if all variables are perfectly measured without error (i.e., $X^* = X$ and $M^* = M$), implying that $\lambda_1 = \lambda_2 = 1$, it further reduces to

$$b_{M^*X^*} = b_{MX} = \alpha + \delta_2, \tag{A3}$$

which is then identical to Equation (5).

The Component Effect of M on Y

First, by applying the path-tracing rules to Figure A1, we can obtain the following population bivariate correlations between the measured variables X^*, M^*, and Y^*:

$$\rho_{Y^*M^*} = \frac{Cov(Y^*, M^*)}{SD(Y^*)SD(M^*)} = \frac{\lambda_2\lambda_3\{Var(M)\beta + \delta_3 + Var(X)\gamma\alpha + \delta_1\alpha + \delta_2\gamma\}}{SD(Y^*)SD(M^*)};$$

$$\rho_{Y^*X^*} = \frac{Cov(Y^*, X^*)}{SD(Y^*)SD(X^*)} = \frac{\lambda_1\lambda_3\{Var(X)\alpha\beta + \delta_2\beta + Var(X)\gamma + \delta_1\}}{SD(Y^*)SD(X^*)}; \qquad \text{(A4)}$$

$$\rho_{X^*M^*} = \frac{Cov(X^*, M^*)}{SD(X^*)SD(M^*)} = \frac{\lambda_1\lambda_2\{Var(X)\alpha + \delta_2\}}{SD(X^*)SD(M^*)}.$$

The unstandardized population partial regression coefficient for M^* of the regression of Y^* on M^* and X^* is expressed as:

$$b_{Y^*M^*|X^*} = \frac{\rho_{Y^*M^*} - \rho_{Y^*X^*}\rho_{X^*M^*}}{1 - \rho_{X^*M^*}^2} \times \frac{SD(Y^*)}{SD(M^*)}. \qquad \text{(A5)}$$

Plugging the correlations from Equation (A4) into Equation (A5), we can obtain the regression coefficient for M^*. First, the denominator part is expressed as:

$$1 - \rho_{X^*M^*}^2 = \frac{Var(X^*)Var(M^*) - \lambda_1^2\lambda_2^2\{Var(X)^2\alpha^2 + \delta_2^2 + 2\alpha\delta_2 Var(X)\}}{Var(X^*)Var(M^*)}. \qquad \text{(A6)}$$

Second, the numerator part is expressed as:

$$\rho_{Y^*M^*} - \rho_{Y^*X^*}\rho_{X^*M^*} = \frac{\lambda_2\lambda_3 Var(X^*)\{Var(M)\beta + \delta_3 + Var(X)\gamma\alpha + \delta_1\alpha + \delta_2\gamma\}}{SD(Y^*)SD(M^*)Var(X^*)}$$

$$- \frac{\lambda_1^2\lambda_2\lambda_3\{Var(X)^2\alpha^2\beta + 2Var(X)\delta_2\alpha\beta + \delta_2^2\beta + Var(X)^2\alpha\gamma + Var(X)\delta_2\gamma + Var(X)\delta_1\alpha + \delta_1\delta_2\}}{SD(Y^*)SD(M^*)Var(X^*)}. \qquad \text{(A7)}$$

Combining the two with the multiplication factor $SD(Y^*)/SD(M^*)$, and slightly rearranging them, we finally have

$$b_{Y^*M^*|X^*} = \beta\left[\frac{\lambda_2\lambda_3\{Var(X^*)Var(M) - \lambda_1^2 K\}}{Var(X^*)Var(M^*) - \lambda_1^2\lambda_2^2 K}\right] + \frac{\lambda_2\lambda_3\{Var(X^*)\delta_3 - \lambda_1^2\delta_1\delta_2\}}{Var(X^*)Var(M^*) - \lambda_1^2\lambda_2^2 K}$$

$$+ \frac{\lambda_2\lambda_3\epsilon_1^2\{Var(X)\alpha\gamma + \delta_1\alpha + \delta_2\gamma\}}{Var(X^*)Var(M^*) - \lambda_1^2\lambda_2^2 K}, \qquad \text{(A8)}$$

where $K = Var(X)^2\alpha^2 + \delta_2^2 + 2\alpha\delta_2 Var(X) = Cov(X, M)^2$. If all variables are standardized, implying that $Var(X) = Var(X^*) = Var(M) = Var(M^*) = 1$, then Equation (A8) reduces to

$$b_{Y^*M^*|X^*} = \beta\left[\frac{\lambda_2\lambda_3\{1 - \lambda_1^2 K\}}{1 - \lambda_1^2\lambda_2^2 K}\right] + \frac{\lambda_2\lambda_3\{\delta_3 - \lambda_1^2\delta_1\delta_2\}}{1 - \lambda_1^2\lambda_2^2 K}$$

$$+ \frac{\lambda_2\lambda_3\epsilon_1^2\{\alpha\gamma + \delta_1\alpha + \delta_2\gamma\}}{1 - \lambda_1^2\lambda_2^2 K}, \qquad \text{(A9)}$$

where $K = \alpha^2 + \delta_2^2 + 2\alpha\delta_2 = Cov(X, M)^2 = Cor(X, M)^2$. In addition, if all variables are perfectly measured without error, implying that $\lambda_1 = \lambda_2 = \lambda_3 = 1$ and $\epsilon_1 = 0$, then Equation (A9) further reduces to

$$b_{Y^*M^*|X^*} = b_{YM|X} = \beta + \frac{\delta_3 - \delta_1\delta_2}{1 - K}, \tag{A10}$$

where $K = (\alpha + \delta_2)^2$. This is identical to Equation (6).

Equation (12) in the section on measurement error can be derived from Equation (A9) by setting $\delta_2 = \delta_3 = 0$ and $\gamma = 0$. That is,

$$\begin{aligned}
b_{Y^*M^*|X^*} &= \beta\left[\frac{\lambda_2\lambda_3(1 - \lambda_1^2\alpha^2)}{1 - \lambda_1^2\lambda_2^2\alpha^2}\right] + \frac{\lambda_2\lambda_3\epsilon_1^2\delta_1\alpha}{1 - \lambda_1^2\lambda_2^2\alpha^2} \\
&= \frac{\lambda_2\beta\lambda_3(1 - \lambda_1^2\alpha^2) + (1 - \lambda_1^2)(\lambda_2\alpha\delta_1\lambda_3)}{1 - (\lambda_1\alpha\lambda_2)^2}.
\end{aligned} \tag{A11}$$

Note that $\epsilon_1^2 = 1 - \lambda_1^2$ because $Var(X^*) = Var(\lambda_1 X + \epsilon_1 e_X) = Var(X) = 1$ and $Var(e_X) = 1$ when variables are standardized.