Abstract: Advice regarding the analysis of observational studies of exposure effects usually is against adjustment for factors that occur after the exposure, as they may be caused by the exposure (or mediate the effect of exposure on outcome), so potentially leading to collider stratification bias. However, such factors could also be caused by unmeasured confounding factors, in which case adjusting for them will also remove some of the bias due to confounding. We derive expressions for collider stratification bias when conditioning and confounding bias when not conditioning on the mediator, in the presence of unmeasured confounding (assuming that all associations are linear and there are no interactions). Using simulations, we show that generally neither the conditioned nor the unconditioned estimate is unbiased, and the trade-off between them depends on the magnitude of the effect of the exposure that is mediated relative to the effect of the unmeasured confounders and their relations with the mediator. We illustrate the use of the bias expressions via three examples: neuroticism and mortality (adjusting for the mediator appears the least biased option), glycated hemoglobin levels and systolic blood pressure (adjusting gives smaller bias), and literacy in primary school pupils (not adjusting gives smaller bias). Our formulae and simulations can inform quantitative bias analysis as well as analysis strategies for observational studies in which there is a potential for unmeasured confounding.

Keywords: Bias; Confounding; Mediation; Observational study

Epidemiology 2021;32:00–00

To Adjust or Not to Adjust? When a “Confounder” Is Only Measured After Exposure

Rolf H. H. Groenwold, a,b Tom M. Palmer, c,d and Kate Tilling d

In observational epidemiologic studies, often the interest lies in the causal effect of an exposure on a certain outcome, yet this may be biased by (unmeasured) confounding. 1 To control for confounding, generally, it is advised to condition on (a sufficient set of) confounders, that is, variables that cause both exposure and outcome. 2–5 Textbooks on epidemiology, as well as research articles about confounding adjustment, generally advise against controlling for variables that are measured after exposure has started. 6–12 One reason is that adjustment for postexposure variables may lead to collider stratification bias. Another reason is that such variables may actually be mediators of the causal relationship between exposure and outcome, and conditioning on such variables may introduce bias. 5,11 As indicated by VanderWeele: “…we often refrain from adjusting for covariates that occur temporally subsequent to the exposure.”

Nevertheless, in applied research, postexposure variables are sometimes adjusted for in the analysis of an observational study. For example, in a cohort study of the effect of prevalent vitamin K antagonist use on renal function, adjustment was made for measurements of hemoglobin and glomerular filtration rate at cohort entry (i.e., after initiation of vitamin K antagonist use). 14 In a case–control study of very hot tea consumption and esophageal cancer, adjustment was made for the confounder “oral hygiene” measured at the time of outcome assessment. 15 In another case–control study of migraine and the risk of stroke, adjustment was made for several confounders, again measured at the time of outcome assessment. 16 One may wonder whether these are examples of bad practice or whether there could be instances in which adjustment for a postexposure variable could, in fact, be beneficial, particularly if the postexposure variable may carry information about unmeasured confounders.

Bias formulas for unmeasured confounding have been described, for example, in Arah et al 17 and VanderWeele and Arah. 18 However, these do not consider scenarios where a variable related to an unmeasured confounder may also be caused by the exposure. Here, we consider that scenario where...
an effect of the exposure, or a mediator, is associated with an unmeasured confounder and derive expressions for the bias when conditioning and when not conditioning on the mediator. The use of these formulae is illustrated by three examples: a study of neuroticism and mortality, a study of glycated hemoglobin levels and systolic blood pressure (SBP), and a study of literacy in primary school pupils.

METHODS

Notation and Setup

To derive expressions for the bias due to unmeasured confounding and the bias due to conditioning on a mediator that is (possibly) associated with an unmeasured confounder, we consider the model depicted in Figure 1. X denotes the exposure, Y the outcome of interest, and M is a mediator of the effect of X on Y. The variables $U_1$ and $U_2$ are unmeasured variables. $U_1$ is a confounder of the $X \rightarrow Y$-relation, and $U_1$ also causes $M$. $U_2$ is a confounder of the $M \rightarrow Y$-relation, but not of the $X \rightarrow Y$-relation. Each variable is a linear combination of the variables affecting it (indicated by the directed arrows in Figure 1) plus an error term. We assume there are no interactions between variables in their effects on outcome or mediator. Furthermore, we assume that all variables are normally distributed, all associations are linear and additive, all errors uncorrelated, and there is no interaction between variables. The regression coefficients of the linear models are denoted by $\beta$, for example, $\beta_{ym}$ for the effect of X on M. The variances of the errors are denoted by $\sigma^2$, for example, $\sigma^2_y$ for the variance of the error of Y. We assume an ordinary least squares (OLS) estimator is used to obtain estimates of the effect of the exposure, or a mediator, is associated with an unmeasured confounder and derive expressions for the bias when conditioning and when not conditioning on the mediator.

Expression (1) can be simplified when assuming all error terms are equal to 1 (i.e., $\sigma^2_{y1} = 1$, $\sigma^2_{y2} = 1$). Then

$$\hat{\beta}_{xy} = \beta_{xy} + \beta_{ym} \beta_{my} + \frac{\beta_{x1} \sigma^2_y (\beta_{y1} + \beta_{ym} \beta_{my})}{\beta_{x1}^2 \sigma^2_{y1} + \sigma^2_y}$$

Here, $\beta_{xy}$ is the direct effect of X on Y, $\beta_{ym}$ the direct effect of M on $Y$, etc. The bias in the OLS estimator of the total effect of X on Y is given by

$$\beta_{xy} = \beta_{xy} + \beta_{ym} \beta_{my} + \frac{\beta_{x1} \sigma^2_y (\beta_{y1} + \beta_{ym} \beta_{my})}{\beta_{x1}^2 \sigma^2_{y1} + \sigma^2_y}, \text{or}$$

$$\beta_{xy} = \frac{\beta_{x1} \sigma^2_y \beta_{y1}}{\beta_{x1}^2 \sigma^2_{y1} + \sigma^2_y} + \beta_{ym} \beta_{my} + \frac{\beta_{x1} \sigma^2_y \beta_{ym} \beta_{my}}{\beta_{x1}^2 \sigma^2_{y1} + \sigma^2_y}$$

The two components of this second expression represent the two back-door paths from X to Y: the first term represents the path $X \leftarrow U_1 \rightarrow Y$ and the second term represents the path $X \leftarrow U_1 \rightarrow M \rightarrow Y$. Note that if $U_1$ does not cause X ($\beta_{x1} = 0$), then there is no bias due to unmeasured confounding. If $U_1$ does not cause Y directly (i.e., $\beta_{y1} = 0$), then there is a nonzero bias due to unmeasured confounding that is mediated by M.

Expression (1) can be simplified when assuming all error terms are equal to 1 (i.e., $\sigma^2_{y1} = 1$, $\sigma^2_{y2} = 1$). Then

$$\hat{\beta}_{xy} = \beta_{xy} + \beta_{ym} \beta_{my} + \frac{\beta_{x1} \sigma^2_y (\beta_{y1} + \beta_{ym} \beta_{my})}{\beta_{x1}^2 + 1}$$

If the unmeasured confounder does not cause the mediator, that is, $\beta_{ym} = 0$, this bias expression simplifies to $\beta_{x1} \beta_{y1} \beta_{ym} / (\beta_{x1}^2 + 1)$.

Bias due to conditioning on a mediator

To explore under what scenarios conditioning on M may remove (part of) the confounding effect by $U_1$, we derive an
expression for the OLS estimator of the relation between $X$ and $Y$, when conditioning on $M$. Derivations of the general expressions are presented in the eAppendix; http://links.lww.com/EDE/B760. For clarity, here we assume all error terms are equal to 1 (i.e., $\sigma_{y_{1m}} = 1$, $\sigma_{y_{2m}} = 1$, etc.). In that case, conditioning on $M$ is expected to result in the observed relation between $X$ and $Y$ to be

$$\hat{\beta}_{x_{1m}} = \frac{1}{(\beta_{x_{1m}} + 1)(\beta_{y_{2m}} + 1) + \beta_{x_{1m}}\beta_{y_{2m}}} - (2)$$

When $\beta_{x_{2m}} = 0$, expression (2) reduces to

$$\hat{\beta}_{x_{1m}} = \frac{\beta_{y_{2m}}(\beta_{x_{1m}} + 1) + \beta_{y_{2m}}\beta_{y_{1m}}}{(\beta_{x_{1m}} + 1) + \beta_{y_{2m}}}$$

which reduces even further to

$$\hat{\beta}_{x_{1m}} = \frac{\beta_{y_{2m}}(\beta_{x_{1m}} + 1) + \beta_{y_{2m}}\beta_{y_{1m}}}{(\beta_{y_{2m}} + 1)}$$

if the unmeasured confounder $U_1$ does not cause the mediator $M$, that is, when $\beta_{y_{1m}} = 0$. When $\beta_{x_{2m}} \neq 0$, yet $\beta_{y_{1m}} = 0$, expression (2) reduces to

$$\hat{\beta}_{x_{1m}} = \frac{\beta_{y_{2m}} + \beta_{y_{2m}}\beta_{y_{1m}}}{(\beta_{x_{1m}} + 1) + \beta_{y_{2m}}} - \frac{\beta_{x_{2m}}\beta_{y_{1m}}}{(\beta_{x_{2m}} + 1)}$$

If we assume no unmeasured confounding by $U_1$, for example, because $\beta_{y_{1m}} = 0$, this reduces further to

$$\hat{\beta}_{x_{1m}} = \frac{\beta_{y_{2m}} + \beta_{y_{2m}}\beta_{y_{1m}}}{(\beta_{x_{1m}} + 1) + \beta_{y_{2m}}}$$

Even though there is no unmeasured confounding of the $X-Y$ relation by $U_1$, clearly this expectation of the observed relation between $X$ and $Y$ differs from the total effect of $X$ on $Y$ ($\beta_{x_{1m}} + \beta_{x_{2m}}\beta_{y_{1m}}$), which is partly due to blocking of the direct effect of $X$ on $Y$ via $M$ and partly due to collider stratification bias (opening the path $X \rightarrow M \leftarrow U_2 \rightarrow Y$).

The bias expressions were supported by statistical simulations. Details of these simulations and their results are presented in the eAppendix; http://links.lww.com/EDE/B760.

**Illustration of Bias Formulas**

To illustrate the use of the bias formulae (1) and (2), we first present a numerical example. The bias formulae can be applied by specifying the parameters. First, we could assume all the error terms ($\sigma_{y_{1m}}$, $\sigma_{y_{2m}}$, $\sigma_{x_{1}}$, $\sigma_{x_{2}}$, and $\sigma_{y}$) are equal to 1. Next, based on the literature or own experience the other parameters could be specified. For example, we could assume that a one-unit increase in the unmeasured variable $U_1$ causes an increase in each of $X$, $M$, and $Y$ of $0.3$ ($\beta_{y_{1m}} = \beta_{y_{1m}} = \beta_{y_{1m}} = 0.3$). Likewise, we could assume that a one-unit increase in the unmeasured variable $U_2$ causes an increase in each of $M$ and $Y$ of $0.2$ ($\beta_{y_{2m}} = \beta_{y_{2m}} = \beta_{y_{2m}} = 0.2$). Also, we assume that a one-unit increase in $X$ causes and increase in $M$ of $0.4$ ($\beta_{x_{1m}} = 0.4$) and that a one-unit increase in $M$ causes an increase in $Y$ of $0.7$ ($\beta_{y_{2m}} = 0.7$). Finally, we assume that $X$ has no direct effect on $Y$ (i.e., $\beta_{x_{1m}} = 0$). With these numbers, we can apply the bias formulae and calculate the bias. Without adjustment for $U_1$, the bias in the estimated $X-Y$ relation is $0.14$, whereas with adjustment for $U_1$, the bias is $-0.25$.

Figure 2 illustrates the impact of various model parameters on the bias of the OLS estimator of the $X-Y$ relation due to unmeasured confounding by $U_1$, with and without adjustment for the mediator $M$. These results are based on expressions (1) and (2). Specifically, the values of error terms ($\sigma_{y_{1m}}$, $\sigma_{y_{2m}}$, $\sigma_{x_{1}}$, $\sigma_{x_{2}}$, and $\sigma_{y}$) were set to $1$. The value of the model coefficients ($\beta_{x_{1m}}$, $\beta_{x_{2m}}$, $\beta_{y_{1m}}$, $\beta_{y_{2m}}$, $\beta_{y_{1m}}$, $\beta_{y_{2m}}$, $\beta_{x_{1m}}$, and $\beta_{x_{2m}}$) were each set to $0.00, 0.25, 0.50, 0.75$, or $1.00$, while $\beta_{x_{1m}}$ was set to $0.00$. This means that in total $78,125$ different configurations of the parameter values were investigated. In each panel, the bias is plotted against the variation in one of the parameters. The distribution of the bias (represented as a box-plot) is due to the variation in the other parameter; that is, each is marginal over the values of the other parameters.

The figure shows situations when conditioning on a mediator may actually lead to a larger bias, compared with not conditioning on a mediator, for example, when the indirect effect via $M$ becomes larger (indicated by increasing values of $\beta_{x_{2m}}$ and $\beta_{y_{2m}}$). Particularly when $\beta_{x_{1m}}$ and $\beta_{y_{1m}}$ have larger values, adjustment may be preferred, however those settings do include scenarios in which adjustment can lead to a relatively large negative bias. For the parameters $\beta_{x_{2m}}$ and $\beta_{y_{2m}}$, both approaches tend to result in bias, although adjustment for the mediator $M$ can result in bias that is larger in magnitude, owing to collider stratification bias. Note that in this illustration, all parameters of the bias formulae have a zero or positive value. However, if, for example, we take the same parameter values as mentioned above (i.e., $\beta_{x_{1m}} = 0.1$, $\beta_{x_{1m}} = 0.3$, $\beta_{y_{1m}} = 0.7$, $\beta_{y_{1m}} = 0.2$), but set $\beta_{x_{1m}}$ and $\beta_{y_{1m}}$ to negative values (i.e., $\beta_{x_{1m}} = -0.3$, $\beta_{y_{1m}} = -0.4$, then the bias after adjustment for the mediator would be positive ($0.25$) whereas not adjusting would result in a negative bias ($-0.14$).

**Example 1—Neuroticism and Mortality**

To illustrate the use of the expressions for bias, presented in the previous sections, we consider a study of the relation between neuroticism and mortality risk, in which...
neuroticism and self-rated health were measured at cohort entry. In this case, self-rated health scores could have been affected by preexisting neuroticism, yet at the same time could be related to unmeasured health status (see Figure 3 for a graphical representation). Conditioning on self-rated health scores may control for confounding but may also induce a bias due to conditioning on the mediator. In this situation, the data analyst needs to identify whether the bias introduced by conditioning on the mediator is likely to be greater than the bias introduced by not conditioning on it. This assessment could inform the analysis strategy.

In terms of the model presented in the previous section, \(X\) would represent neuroticism, \(Y\) mortality, \(M\) self-rated health score, and \(U_1\) would represent unmeasured health status. For brevity, we assume there is no confounding of the \(M-Y\) relation by a variable \(U_1\). If we assume that there is no direct effect of self-rated health on mortality, but only indirectly via health status, we set \(\beta_{uv} = 0\) and remove the arrow from self-rated health to mortality from the graph. Furthermore, we assume for ease of interpretation that the effect of \(X\) on \(M\) is equal to the effect of \(U\) on \(M\) and that the effect of \(U\) on \(X\) is equal to the effect of \(U\) on \(Y\). In addition, all error terms (\(\sigma\)) were set to 1.

Figure 4 shows the difference in absolute bias of not conditioning on self-rated health score versus conditioning on self-rated health score for various scenarios. In those scenarios in which the confounding effect of \(U_1\) is relatively small (\(\beta_{ux} = 0.1, \beta_{uy} = 0.1\)), both analytical approaches give small bias. However, when the confounding effect is larger, differences are more pronounced. If the unmeasured confounder is positively associated with \(X\) and \(Y\) (e.g., \(\beta_{ux} = 0.4, \beta_{uy} = 0.4\)), adjustment for \(M\) reduces the bias. In contrast, if the unmeasured confounder is negatively associated with \(X\) and \(Y\) (e.g., \(\beta_{ux} = -0.3, \beta_{uy} = -0.3\)), adjustment for \(M\) leads to an increase in bias. The reason for the latter phenomenon is collider stratification bias: upon adjustment for \(M\), the relation between \(U_1\) and \(X\) becomes stronger, thus increasing the confounding effect by \(U\).

Arguably, a more realistic scenario is that in which neuroticism has a negative effect on self-rated health (e.g., \(\beta_{xm} = 0.5\)), while health status has a positive effect on self-rated health (e.g., \(\beta_{um} = 1\)). Furthermore, the health status could be negatively associated with neuroticism and mortality (e.g., \(\beta_{ux} = -0.3\) and \(\beta_{uy} = -0.3\)). Given these values, without adjustment for the health status the bias in the effect of neuroticism is expected to be 0.08. However, when adjusting for self-rated health, this bias is expected to be −0.03. This suggests that, in this scenario, adjustment for self-rated health gives a smaller bias. Although the actual values of the different parameters will be context-specific, an analysis like this may
guide data analysts in decisions about analysis strategies and sensitivity analyses of adjustment for a mediator that is caused by an unmeasured confounder, in this case self-rated health scores. Further work could include using more realistic values for the coefficients, if these could be obtained for example from external data or expert opinion.

Example 2—Glycated Hemoglobin Levels and Systolic Blood Pressure

Suppose an interested reader finds an article that describes a study of the relation between glycated hemoglobin levels (HbA1C) and SBP. This relation is estimated using linear regression. Age, sex, and body mass index (BMI) are included as covariates in the model to adjust for confounding. In this cohort study, BMI was measured 2 months after HbA1C levels were measured, for example, because routine data from general practitioners were used. The reader wonders whether the researchers actually should have adjusted for BMI, or if this could 'adjust away' part of the effect of HbA1C on SBP. Therefore, she decides to conduct a sensitivity analysis.

Based on the literature and her own experience, she expresses possible ranges of the relations between pre-HbA1c-measurement BMI ($\text{BMI}_0$), HbA1C, post-HbA1c-measurement BMI ($\text{BMI}_1$), and SBP (Figure 5). As in the previous example, all error terms ($\sigma$'s) are assumed to be 1 and $U_j$ is considered absent. For each relation, the possible range is then split in 10 equidistant values. For each of the 100,000 combinations of the values of the five relations, the possible bias due to conditioning on $\text{BMI}_1$ (the mediator that is related to the unmeasured confounder $\text{BMI}_0$) is compared with the possible bias due to omitting the unmeasured confounder $\text{BMI}_0$.

Figure 6 shows the relative bias of conditioning versus not conditioning on $\text{BMI}_1$, which indicates that in 17.3% of the scenarios it would be better not to condition on $\text{BMI}_1$. In the large majority of scenarios (82.7%), however, conditioning on $\text{BMI}_1$ is to be preferred. Based on the assumptions made and the quantification of the bias under two analytical strategies, the reader concludes that she agrees with the choices made by the researchers regarding the analysis of the study of the relation between HbA1C and SBP.

Note that different values of the parameters can give rise to the same difference in absolute bias. For example, when $\beta_{ux}$, $\beta_{yu}$, $\beta_{y0}$, $\beta_{y1}$, and $\beta_{yu}$ have values 0.90, 0.15, 0.17, 1.0, and 0.80, respectively, the difference in absolute bias between adjusting and not adjusting for $\text{BMI}_1$ is approximately zero. Also, when $\beta_{ux}$, $\beta_{yu}$, $\beta_{y0}$, $\beta_{y1}$, and $\beta_{yu}$ have values 1.06, 0.106, 0.172, 1.10, and 0.60, respectively, the difference in absolute bias is approximately zero. However, the magnitude of the bias differs: 0.13 for the first scenario and 0.10 for the second scenario. Furthermore, in both scenarios the bias when not conditioning has a positive sign, whereas the bias when conditioning on the mediator has a negative sign. An alternative approach, like in example 1, would be to consider the (absolute) bias in each of the analytical approaches, instead of their difference; if the bias for both approaches is considered too large, it might be preferable not to conduct a statistical analysis at all but instead put effort in collecting additional data, for example, on $U_i$ (in this example $\text{BMI}_0$).

Example 3—Literacy in Primary School Children

Literacy achievements among primary school pupils may affect their literacy later in life. Suppose researchers are
interested in the effect of literacy in the first year of school on literacy 2 years later (i.e., in the third year in school). Literacy is measured in the first and in the third year by means of the Word Recognition and Phonic Skills Test (WRAPS). Age, sex, and phonological capacity may be confounders of this relation. The latter could be measured using the phonological assessment battery (PHAB). For this example, we use data from the ready to learn randomized controlled trial.\textsuperscript{23} We use the data of 132 children who received the control intervention, with complete information about WRAPS in years 1 and 3, PHAB in years 1 and 2, age, and sex. These data are available via http://reshare.ukdataservice.ac.uk/853287/, where more information can be found, for instance about ethical review of this study.

We estimated that the relation between WRAPS at baseline and WRAPS after 2 years of follow-up was estimated using linear regression. We included age and sex as covariates in the model to adjust for possible confounding. First, we considered the situation in which a measurement of the confounder PHAB in year 1 is available (reference scenario). After adjustment for age, sex, and PHAB in year 1, we estimated a one-unit increase in WRAPS in year 1 to increase WRAPS in year 3 by 0.74 units (95% CI = 0.48 to 1.00). If, however, a measurement of PHAB in year 2 was available, we could adjust for that, because PHAB in year 2 is probably correlated to PHAB in year 1. After adjustment for age, sex, and PHAB in year 2, we estimated a one-unit increase in WRAPS in year 1 to increase WRAPS in year 3 by 0.63 units (95% CI = 0.37 to 0.88).

In this case, the data allow for estimation of the parameters of the bias formulae (1) and (2), where we assume there is no $U_2$ and $X$ represents WRAPS in year 1, $Y$ represents WRAPS in year 3, $U_1$ represents PHAB in year 1, and $M$ represents PHAB in year 2. Based on the estimates of the parameters, conditional on age and sex, we can attribute the following values (rounded to two decimal points) to the parameters:

\[
\sigma_\eta = 51, \sigma_x = 8.5, \sigma_m = 27, \sigma_y = 12, \beta_{xy} = 0.65, \beta_{xm} = -0.34, \beta_{mx} = -0.16, \beta_{mxy} = -0.050, \beta_{xy} = 0.34, \beta_{x} = 0.030.\]

Given these values, without adjustment for the unmeasured confounder ($U_1$/PHAB in year 1) we expect the bias in the effect of WRAPS to be 0.04, which corresponds to the difference in estimates of 0.70 versus 0.74. However, when adjusting for the mediator (M/PHAB in year 2), this bias is expected to be −0.07. This value corresponds to the difference in estimates of 0.70 versus 0.63.

In practice, data would not be available to estimate all parameters. However, estimates may be obtained from analysis of other data sources or retrieved from other source of information. These could then inform a sensitivity analysis.

**DISCUSSION**

In the presence of unmeasured confounding, conditioning on a mediator that is caused by an unmeasured confounder may reduce, but could also amplify, the bias in the OLS estimator of the exposure-outcome relation. In this paper, we provided expressions for the bias in the OLS estimator for this particular situation to provide support in choosing between two potentially biased estimators. Using simulations, we show that generally neither the conditioned or the unconditioned estimate is unbiased, and the trade-off between them depends on the magnitude of the effect of the exposure that is mediated relative to the confounding effect of the unmeasured confounders and their relations with the mediator.

Epidemiologists are increasingly seeing the value of quantitative bias analysis, whereby statistical formulae and external knowledge are used to estimate the potential size and direction of bias (e.g., due to unmeasured confounding) for a given analysis.\textsuperscript{22,24–27} Previous work on the impact of conditioning on a mediator or omitting a confounder already provided expressions for the bias in case of a linear model.\textsuperscript{5,13,28–31} We extended this work by considering the situation where a mediator is caused by an unmeasured confounder. In general, neither of the two approaches considered (i.e., conditioning and not conditioning on the mediator) yields an unbiased estimator. Where others considered similar scenarios, yet focused
on estimation of indirect effects, our interest lay in estimation of the total effect of the exposure on the outcome (and possible bias thereof).

Texts on confounding adjustment generally advise against adjustment for variables that are caused by exposure. Variables measured after initiation of exposure may be caused by (prior) exposure status and therefore are generally not adjusted for. Nevertheless, there are several examples of observational studies where adjustment is made for variables that are measured after exposure started, with the aim to control for confounding. To allow for a fair assessment of the quality of an observational study of causal effects, researchers should report relevant details about the timing of the measurement of confounders. If these are measured after the start of exposure, ideally researchers should explain whether they think the measured value represents a mediator or whether it should be considered to hold information about one or multiple unmeasured founders. Directed acyclic graphs may be helpful to support and explain the arguments.

The expression for the bias when conditioning on the mediator, equation (2), also provides further insight in collider stratification bias. Particularly, if we assume $\beta_{xy}$, $\beta_{ym}$, $\beta_{ym}$, and $\beta_{my}$ are zero, the expression reduces to $\hat{\beta}_{ym} = \frac{-\beta_{om} \beta_{ym} \beta_{my}^{2}}{2 \beta_{om}^{2} + 1}$. In this case, the total effect of $X$ on $Y$ is zero and $\hat{\beta}_{ym}$ is the bias due to conditioning on the collider $M$. Collider stratification bias can arise, for example, due to selective inclusion of subjects into a study or due to missing data. To quantify the potential bias due to collider stratification, bias expressions like the one presented in this paper could be used.

Although the linear model allows us to derive expressions for the bias of the different OLS estimators, an obvious limitation is that the results only apply to a limited number of settings. In particular, we have assumed that all associations are linear, and that there are no interactions. Also, the model considered is a gross simplification of what an observational study looks like in reality. In future work, the presented expressions could be extended to more complex models, including nonlinear models. For example, future work could extend the theorems about collider bias for binary outcome models. Also, extensions could be considered that take possible correlation between measured and unmeasured confounders into account. While we focused on bias in studies with a so-called point-exposure, many exposures are in fact time-varying (neuroticism in the first example is actually more of a time-varying exposure than a point-exposure). Future research could extend our bias formulae to setting of time-varying exposure. We note, however, that our bias expressions already could provide some insight into that situation, by considering the exposure ($X$) and the mediator ($M$) as different measurements of exposure (e.g., $X_1$ and $X_2$). The expressions could then help to quantify the potential impact of (time-varying) confounding and for example support decisions about the frequency with which time-varying confounders are measured.

An alternative to using bias expressions to guide choices regarding data analysis and to inform sensitivity analysis is to use simulated data to assess bias. There are functions in several structural equation modeling packages, which allow the user to generate data from a specified DAG/model. For example, the simulateSEM() function in the dagitty package, the simulateData() function in the lavaan package, and functions in the simMixedDAG package. The user can then fit various proposed models to the data to assess the potential bias in their proposed modeling approach. A simulation approach was also used to assess the likely magnitude of collider bias as an explanation for the obesity paradox. And this web application, http://watzilei.com/shiny/collider/, uses simulated data to investigate collider bias for user-specified models. Additionally, the tipr package implements methods to assess the sensitivity of regression results to unmeasured confounders. The expressions presented in this paper offer a relatively simple first approach as compared to more advanced simulation based sensitivity analysis.

In summary, one of the major drawbacks of observational epidemiology is the possibility of unmeasured confounding. Where possible mediators of the exposure-outcome relation hold information about unmeasured confounders, it is important to estimate the bias incurred by both adjusting and not adjusting for these mediators. The bias expressions presented here are helpful to support choices regarding analysis strategies and to inform sensitivity analyses for observational studies in which there is a potential for unmeasured confounding. The result of such a sensitivity analysis could be that collecting data on currently unmeasured confounders is deemed necessary or the study should be stopped because the risk of bias is considered too large. Alternatively, the results of both analyses (adjusted and not adjusted for the mediator) could be presented and the bias expressions presented here could support discussions about the validity of each of the two analyses. Given that there will usually be uncertainty about the values of the parameters, it is not recommended to present the results of just one of the two analyses without acknowledging that uncertainty. Also, we would like to caution against interpreting our results as evidence that adjustment for postexposure variables is advantageous. Instead, we hope they will be the basis of thoughtful investigations of the validity of observational research.

REFERENCES
1. Hernán MA, Robins JM. Causal Inference: What If. Boca Raton, FL: Chapman & Hall/CRC; 2020.
2. VanderWeele TJ, Shpitser I. A new criterion for confounder selection. Biometrics. 2011;67:1406–1413.
