On the Bias of Complete- and Shifting-Case Meta-Regressions with Missing Covariates

Jacob M. Schauer, Northwestern University

Karina Diaz, Columbia University

Jihyun Lee, The University of Texas at Austin

Therese D. Pigott, Georgia State University

# 1 Introduction

Meta-regression is a useful tool for studying important sources of variation between effects in a meta-analysis (Borenstein, 2009; Tipton et al., 2019a). Analyses of these models in the absence of missing data have been studied thoroughly in the literature (e.g., Berkey et al., 1995; Hedges, 1983b; Hedges et al., 2010; Konstantopoulos, 2011; Viechtbauer, 2007). However, it is common for meta-analytic datasets to be missing data (Pigott, 2001a). In the context of meta-regression, issues with missing data frequently involve missing covariates (Pigott, 2001b; Tipton et al., 2019b).

Precisely how to proceed with a meta-regression with missing covariates remains something of an open question. Statistical guidance suggests that analyses ought to consider the mechanism that causes covariates to be missing (Pigott, 2001b, 2019). However, it appears that doing so is less common in practice for meta-analyses. A recent review found that meta-regressions with missing data tend to take one of two strategies (Tipton et al., 2019b). An analyst may conduct a *complete-case analysis* that excludes any effects for which a relevant covariate is missing (i.e., only analyze complete cases). However, if there are very few such effects, a common approach is to use *shifting units of analysis*, which we refer to in this article as a *shifting-case analysis* (Cooper, 2017). Under a shifting-case analysis, analysts fit a series of meta-regression models on subsets of relevant covariates, so that each model selectively omits certain covariates.

Estimators used in both complete-case and shifting units analyses in some sense ignore effects for which a covariate is missing. Ignoring missing data can potentially lead to biased estimates (see Little & Rubin, 2002; Graham, 2012). Despite authors pointing out such issues in meta-analysis, these methods continue to enjoy widespread use (Pigott, 2019). Existing literature on this discussion has yet to detail precisely how much bias can arise in a complete- or shifting-case analysis, nor is there exhaustive guidance on when these methods produce unbiased estimates. In short, there is an understanding that these methods *can* induce bias, but less is known about *how much* and *under what conditions*.

This article examines the potential bias of complete- and shifting-case analyses. The following section provides a demonstration of these methods on data concerning a meta-analysis of substance abuse interventions (Tanner-Smith et al., 2016). We then introduce a statistical framework for studying bias for incomplete data meta-regressions that incorporates a model for whether or not a data point is observed. Using this framework, we describe conditions under which complete- and shifting-case analyses are unbiased. When these conditions are not met, we derive an approximation for the bias of complete- and shifting-case analyses using standard models for missingness and examine the magnitude of bias.

# 2 Example: Substance Abuse Interventions

Tanner-Smith et al. (2016) conducted a meta-analysis that examined the effects of substance abuse interventions on future substance use among adolescents. The studies included in this meta-analysis involved a variety of different treatment types (e.g., cognitive behavioral therapy, family therapy, and pharmacological therapy) and treatment intensities (measured in hours per week), and were carried out in a variety of contexts, including in-patient and out-patient centers. Tanner-Smith et al. used meta-regression models to study potential moderators of these effects, and their analyses had to contend with a number of effects that were missing covariates. While in practice, models were estimated via the expectation-maximization (EM) algorithm rather than complete- or shifting-case methods, we use a subset of this data in in order to illustrate complete- and shifting-case analyses.

Consider a subset of the Tanner-Smith et al. data comprising 74 effect estimates of substance abuse interventions from 46 studies. These effect estimates involve contrasts between groups in a study that are subjected to different treatment conditions, denoted in the data as *Group 1* and *Group 2*, so that each treatment effect can be thought of as Group 1 minus Group 2. Each effect estimate corresponds to a given contrast within a study. Effect sizes are measures on the scale of bias-corrected standardized mean differences. Often the same group (typically the control group) in a study was used in multiple contrasts, so that effect sizes in this meta-analysis are likely correlated.

Suppose an analysis of interest involves the impact of high- versus low-intensity interventions on treatment effects, where a high-intensity intervention consisted of more than 1.5 hours per week of treatment. Then this analysis might use a pair of binary covariates for each effect: one would indicate whether group 1 received a high-intensity intervention (i.e., if group 1 treatment was high-intensity) and the other would indicate whether group 2 received a high intensity (i.e., if group 2 treatment was high-intensity). The relevant meta-regression model would regress the effect estimates on these two covariates.

In the data, the treatment intensity is missing for some of the effects, and Table ?? summarizes missingness for these covariates. Table ?? shows that only 37 of the 74 (50%) have a reported treatment intensity for both groups (i.e., are both observed), but that 54 (73%) of effects report Group 1’s treatment intensity (i.e., is observed) and 41 (55%) effects report Group 2’s treatment intensity.

A complete-case analysis would include only the 37 effects for which both covariates were observed. Using robust variance estimation to account for dependence between effect sizes, a complete-case analysis would result in the coefficient estimates and standard errors displayed in the first column of Table @ref(tab:cc\_adt\_example). Based on these estimates, when Group 1 receives a high-intensity treatment, we would expect an effect to be larger by (in standard deviation units) than when Group 1 receives a low-intensity treatment, which is statistically significant at the level. Note that the estimated between-effect variance is .

However, the model above is estimated on only half of the data. Concern over using a small proportion of the data, or a relatively few number of effects often leads meta-analysts to opt for a shifting-case analysis. An example of a shifting-case analysis would use the 54 effects for which Group 1’s treatment intensity is observed (i.e., is observed), but only including in the model. Doing so leads to the estimates in second column of Table @ref(tab:cc\_adt\_example). Note that the coefficient estimate for Group 1’s treatment intensity is still positive, but is roughly 60% the magnitude of the estimate in the complete-case model.

Finally, an analogous model in a shifting-case analysis would include the 55 effects for which Group 2’s intensity is observed, and include only that covariate in the model. The third column of Table @ref(tab:cc\_adt\_example) shows that this results in a coefficient estimate for Group 2’s treatment intensity (0.16) that is in the opposite direction of the estimate from the complete case analysis (-0.21).

It should be noted that all of these estimates ought to be interpreted with caution. The complete-case analysis includes only half of the effect sizes, which comprises a missingness rate well beyond what might be considered negligible (Bennett, 2001; Schafer, 1999). The shifting-case analyses include more of the data, but because each shifting-case model omits one of the covariates, these models are not equivalent to the model that includes both covariates (Cooper, 1998). The remainder of this article concerns the impact of omitting effect sizes and/or covariates from meta-regressions on the accuracy of coefficient estimates, and finds that this can induce potentially substantial bias.

# 3 Model and Notation

Suppose a meta-analysis involves effects estimated from studies. For the th effect, let be the estimate of the effect parameter , and let be the estimation error variance of . Denote a vector of covariates that pertain to as . Note that the first element of is a 1, which corresponds to an intercept term in a meta-regression model, and that for corresponds to different covariates. The meta-regression model can be expressed as:

Here, is the vector of regression coefficients. The estimation errors are typically assumed to be normally distributed with mean zero and variance , which is true of some effect sizes, and is an accurate large-sample approximation for others (Cooper et al., 2019). The term represents the random effect such that and .

This is the standard mixed-effects meta-regression model, and it is also consistent with subgroup analysis models (Cooper et al., 2019; Hedges & Vevea, 1998). The vector refers to the parameters of model. Under a fixed-effects model, it is assumed that , in which case , and .

A common assumption in random effects meta-regression is that the random effects are independent and normally distributed with mean zero and variance (Hedges, 1983a; Hedges & Vevea, 1998; Laird & Mosteller, 1990; Viechtbauer, 2005):

In that case, the distribution can be written as

Thus, the joint likelihood for all effects can be wrtiten as:

where is the vector of effect estimates, is the vector of estimation variances, and is the matrix of covariates where each row of is simply the row vector . Note that the functions in both () and () assume that *all* of the covariates are observed. Equation () is referred to as the *complete-data likelihood function* (Gelman, 2014; Little & Rubin, 2002). We note that a meta-regression with no missing data will be accurate if the complete-data model is correctly specified. Thus, to illustrate the properties of incomplete data meta-regression, we assume that the complete-data model is correctly specified.

The vector of regression coefficient estimates for the complete-data model when there is no missing data is typically estimated by

Here, is the diagonal matrix of weights. The covariance matrix of is given by

This model can be expanded to account for dependent effect sizes by assuming that is a vector of effects from the same study, is vector of estimation errors, is a vector of random effects, and has covariance matrix . In this model, is a matrix of covariates for each effect in . While we focus on results for independent effect sizes in this article, we expect that the results hold for dependent effect sizes.

Not all relevant variables may be observed in a meta-analytic dataset. Let be a vector of response indicators that correspond with effect . This article concerns missing covariates, and we assume that and are observed for every effect of interest in a meta-analysis. Thus, each element of corresponds to a covariate . The take a value of either 0 or 1: indicates the corresponding is observed and , indicates a that the corresponding is not observed. Note that is a vector of 0s and 1s of length . For instance, were missing, then .

Denote as the indices of covariates that are observed and be the set of indices for missing covariates. Then, the complete-data model can be written as

Note that the complete-data model depends on entries of , which are unobserved.

When meta-analytic datasets are missing covariates, analyses involve incomplete data. In practice, meta-regressions of incomplete data have largely relied on one of two approaches: complete-case analyses and available-case analyses (Pigott, 2001b, 2019; Tipton et al., 2019b, 2019a). A complete-case analysis (CCA) includes only effects for which all covariates of interest are observed. Available-case meta-regressions typically amount to including a subset of relevant covariates that are completely observed (Pigott, 2019; Tipton et al., 2019b). Analyses of this sort have been referred to as *shifting units of analysis* in the meta-analytic literature (Cooper, 2017). In this article we refer to this analytic approach as a *shifting-case analysis* (SCA).

## 3.1 Complete-Case Estimators

A common approach in meta-regression with missing covariates is to use a complete-case analysis. This approach simply omits rows in the data for which any covariate is missing. Thus, this analysis method only uses effects and covariates for which $R\_i = [1, \ldots, 1] = \mathbbm{1}$.

Let $C = \{i : R\_i = \mathbbm{1}\}$ index all relevant effects such that $R\_i = \mathbbm{1}$, so that is the matrix of covariates such $R\_i = \mathbbm{1}$, is the corresponding subset of effect estimates, and is the corresponding subset of weights. The complete-case analysis estimates coefficients with:

## 3.2 Shifting-Case Estimators

When there are multiple covariates of interest, each of which has some missingness, there may only be a few effects for which all covariates of interest are observed. When that happens, a complete case analysis can be unfeasible. A common solution to this in meta-analysis is to use an available-case analysis (Pigott, 2019). In practice, an *available-case* meta-regression is often equivalent to a shifting-case analysis, referred to in the literature as *shifting units of analysis* (Cooper, 2017; Tipton et al., 2019b).

Shifting-case analyses involve fitting multiple regression models, each including a subset of the covariates of interest. Sometimes this even takes the form of regressing effect estimates on one covariate at a time (Pigott, 2019; Tipton et al., 2019b). A given regression in a shifting-case analysis conditions on a set of missingness patterns where . These patterns are such that for those covariates that are observed *and* included in the analyses, but for covariates that are excluded from the analysis; that is, excluded covariates may be observed or unobserved.

As an example, suppose there are two covariates of interest and . A shifting-case analysis might first regress on observed values of . This regression would include observations for which both and are observed (i.e., ) and observations for which is observed but is missing (i.e., . This would imply that the regression involves effects for which .

Consider a single regression in a shifting case analysis, and let index the covariates included in that model and let be the complement of . Note that this model is used to estimate , which is a subset of the full vector of coefficients . In the example above, where is regressed on only , . Denote as the set of missingness patterns such that all included covariates are observed: $\mathcal{R}\_j = \{R \in \mathcal{R}: R\_S = \mathbbm{1}\}$. Note that contains missingness patterns such that all the included covariates are observed, but any excluded covariates may be either observed or unobserved. Finally, let be the set of effects for which are observed . Then, the shifting-case estimators for are given by:

where contains the columns of that pertain to the covariates that are included in the model, and the rows for which all of those covariates are observed. The matrix contains the rows of for which are observed (and similarly for ).

## 3.3 Missingness Mechanisms

Both the complete- and shifting-case estimators are analyses of incomplete data. Analyses of incomplete data require some assumption about why data are missing, which is referred to as the missingness *mechanism*. The mechanism by which missingness arises is typically modeled through the distribution of . Let denote the parameter (or vector of parameters) that index the distribution of so that the probability mass function of can be written as . Assumptions about the missingness mechanism are therefore equivalent to assumptions about .

Rubin (1976) defined three types of mechanisms in terms of the distribution of . Data could be missing completely at random (MCAR), which means that the probability that a given value is missing is independent of all of the observed or unobserved data:

MCAR implies that probability that a given value is missing is unrelated to anything observed or unobserved, and depends only on the missingness parameter .

Covariates could be missing at random (MAR), which implies the distribution of missingness depends only on observed data:.

MAR differs from MCAR in that missingness might be related to observed values. As an example, if with larger standard errors are less likely to report the racial composition of their samples, then missingness would depend on the (observed) estimation error variances. Data missing according to this mechanism would violate an assumption of MCAR, since missingness is related to an observed value.

Finally, data are said to be missing not at random (MNAR) if the distribution of depends on unobserved data in some way. In the context of the meta-regression data, this would imply that is related to , so that the probability of a covariate not being observed depends on the value of the covariate itself. For instance, data would be MNAR if studies with larger standard errors and a greater proportion of minorities are less likely to report the racial composition of their samples because the likelihood that racial composition is not reported will depend on the composition itsef.

A related concept in missing data is that of *ignorability*, which means that the missingness pattern does not contribute any additional information. When missing data are ignorable, it is not necessary to know (or estimate) in order to conduct inference on (Gelman, 2014; Graham, 2012; Little & Rubin, 2002; van Buuren, 2018). In practice, missing data are ignorable if they are MAR and if and are distinct.

# 4 Conditional Incomplete Data Meta-Regression

Because both complete- and available-case analyses depend on the value of , they can be seen as models that condition on missingness. Models that condition on missingness are not necessarily identical to the complete-data model, which is the model of interest, because the complete-data model does not condition on . Yet, complete- and available-case analyses proceed as if the complete-data and conditional models on missingness are equivalent. Doing so ignores the missingness mechanism and its potential impact on the accuracy of analytic results.

The complete-data model can be related to the conditional models through the distribution of missingness . This approach is referred to as a *selection model* in the missing data literature (Gelman, 2014; Little & Rubin, 2002; van Buuren, 2018). We can write the selection model for meta-regression with missing covariates as:

where indexes the distribution of . Here, refers to the relevant subset of on which the analysis conditions; for a complete-case analysis, $\mathcal{R}\_j = \{\mathbbm{1}\}$.

Equation () describes the conditional model as a function of the complete-data model and a selection model that gives the probability that a given set of covariates are observed. The denominator on the right hand side of () is a normalizing factor that is equivalent to the probability of observing the missingness pattern given the estimation error variance and the observed and unobserved covariates in the vector , and can be written as

Note that when the complete-data model in () is not equivalent to the conditional model in (), the resulting coefficient estimators in a meta-regression can be biased. To see this, we can write:

Here, we see that the expectation of conditional on can be written as the complete-data expectation plus a bias term . If , it follows that conditioning on induces bias in the distribution of used in an analysis. Because the complete-case estimator in () and the shifting-case estimator in () are weighted averages of the , the following sections show that they can be biased if . The precise magnitude of the will depend on the selection model in () and hence on the missingness mechanism.

A standard approach for modeling missingness mechanisms for covariates is to assume follows some log-linear distribution (Agresti, 2013). Various authors have described approaches to modelling for missing covariates in generalized linear models that include logistic and multinomial logistic models (Ibrahim, 1990; Ibrahim et al., 1999; Lipsitz, 1996). Thus, one class of models for missingness would involve the logit probability of observing some missingness patterns :

where , so that would be the intercept term for the logit model for the set of missingness patterns . While log-linear models are not the only applicable or appropriate selection model, we make this assumption at points throughout this article in order to demonstrate conditions under which conditional meta-regressions are inaccurate, and how inaccurate they can be.

## 4.1 Approximate Bias for Log-linear Selection Models

As argued above, the bias of complete-case estimators or shifting-case estimators will depend in some way on the bias induced in by conditioning on . The magnitude and direction of will in turn depend on the missingness mechanism.

It is possible to derive an approximation for when under certain conditions when follows a log-linear distribution. If is the standard fixed- or random effects meta-regression model in equation (), and follows the log-linear model in (), then

where is equivalent to evaluated at and

is the derivative of with respect to evaluated at . The detailed proof is presented in Appendix A.

While the following sections will examine possible values that may take under different selection models, we can gain some insight on bias by examining (). The expression for depends on three main quantities. First, is an increasing of , which is the probability that . This implies that the bias will be greater as the probability of omitting an observation increases. Second, increases in , which means that it will be larger when vary more around the regression line. Finally, depends on . Note that is the derivative of with respect to . If does not depend on , . Thus, depends on the components of the selection model that are functions of and how important those components are.

# 5 Bias in Complete-Case Analyses

Complete-case analyses only include effects for which all relevant covariates are observed. The complete-case coefficient estimator given in equation () conditions on $R\_i = \mathbbm{1}$. As noted above, conditioning on can induce bias, however there are conditions under which the complete case analysis will lead to unbiased coefficient estimates. First, if the covariates are MCAR, so that , then

$$
p(R\_i = \mathbbm{1} | T\_i, X\_i, v\_i, \psi)
= p(R\_i = \mathbbm{1} | \psi)
$$

and hence

$$\begin{equation}
p(T\_i | X\_i, v\_i, R\_i = \mathbbm{1}, \eta, \psi)
= \frac{p(R\_i = \mathbbm{1} | \psi) p(T\_i | x\_i, v\_i, \eta)}{p(R\_i = \mathbbm{1} | \psi)}
= p(T\_i | x\_i, v\_i, \eta)
\label{eq:cc\_mcar}
\end{equation}$$

Thus, the complete-data and conditional models are equivalent. Assuming is MCAR, then in () will be unbiased estimator of . This is consistent with broader results on analyses of MCAR data in primary study context (Little & Rubin, 2002).

A complete-case analysis is also unbiased under slightly less restrictive assumptions. Suppose that , then the complete-data model and the conditional model are equivalent:

$$\begin{equation}
p(T\_i | X\_i, v\_i, R\_i = \mathbbm{1}, \eta, \psi)
= \frac{p(R\_i = \mathbbm{1} | v\_i, \psi) p(T\_i | x\_i, v\_i, \eta)}{p(R\_i = \mathbbm{1} | v\_i, \psi)}
= p(T\_i | x\_i, v\_i, \eta)
\label{eq:cc\_mar}
\end{equation}$$

The assumption that implies that if missingness only depends on the estimation error variances, then a complete case analysis may be unbiased. This is a weaker assumption than MCAR, which requires . Intuitively, if both and are conditionally independent of , then so is the relationship between and . Under this assumption, will be unbiased.

For most effect size indices, variances are functions of the sample sizes within studies . Some effect sizes, such as the -transformed correlation coefficient, have variances that depend entirely on the sample size of a study, while for other effect sizes this is approximately true, such as the standardized mean difference. For such effect sizes, this assumption implies that missingness depends only on the sample size of the study. This may be true, for instance, if smaller studies are less likely to report more fine-grained demographic information regarding their sample out of concern for the privacy of the subjects who participated in the study (and that no other factors affect missingness).

An even weaker assumption can also lead to unbiased estimation with complete cases. When , we can write:

$$\begin{equation}
p(T\_i | X\_i, v\_i, R\_i = \mathbbm{1}, \eta, \psi)
= \frac{p(R\_i = \mathbbm{1} | X\_i, v\_i, \psi) p(T\_i | X\_i, v\_i, \eta)}{p(R\_i = \mathbbm{1} | X\_i, v\_i, \psi)}
= p(T\_i | X\_i, v\_i, \eta)
\label{eq:cc\_mnar}
\end{equation}$$

Thus, if , then the complete-case model will be the same as the complete-data model. Note that can still depend on , including that are not observed, which means that the data are MNAR.

However, when is not independent of or (given ), then analyses can be biased. Let $\mathcal{R}\_1 = \{\mathbbm{1}\}$ so that the complete-case analysis conditions on . Based on equation (), the bias of will depend on the . If we let be the vector of and let be the subset of for which all covariates are observed (i.e., $R\_i = \mathbbm{1}$). Then the bias of the complete-case analysis can be written as

The bias in equation () is a weighted average of individual biases . Hence, the bias will be larger if the are larger.

Precisely, how large the bias in () is will depend on the distribution of and its relationship to effect estimates and their covariates . When follows the log-linear model in (), the approximate bias can be written as

where

is a diagonal matrix where entries refer to the probability that an observation is *not* a complete case,

is a matrix of derivatives, and is a vector of parameters that index the selection model. Note that the bias in () involves which contains the rows of for which $R\_i = \mathbbm{1}$; similarly for .

While () provides a general expression for the approximate bias of , it can be a little difficult to interpret. Loosely, we can see that the bias depends on the probability that observations are omitted due to missingness , as well as some function of the components of the log-linear model . To better intuit this bias, we provide a simple example in the following section.

## 5.1 Example: Complete-Case Analysis with a Single Binary Covariate

Suppose the model of interest includes a single binary covariate covariate , so that the complete data model is

where are the regression coefficients of interest. Note that is the average effect when and is the contrast in mean effects for when versus when .

Because is a scalar, so is ; indicates that is missing, indicates that is observed. A complete-case analysis would include only effects for which is observed (i.e., ). The complete-case estimator for is given by a weighted sum of among the effects for which and :

The complete-case estimator for is given by the difference between the (weighted) mean effect for versus :

Assume that the selection model is log-linear, and that missingness depends on the size of the effect and the value of :

Note that this is an MNAR mechanism, since it depends on . Under this model, depends only on and not , so we can write $H\_1(X\_i) = p(R \neq \mathbbm{1} | T\_i, X\_i, v\_i)\rvert\_{T\_i = X\_i^T \beta}$. As well, and . Given the result in equation (), we can write

Under this selection model, the bias of the complete-case estimator for is:

where is the average estimation error variance among effects for which and .

The expression in () depends on three key quantities, and is an increasing function of each of those quantities. First, the bias increases in , which is essentially the probability that a case is missing any covariates; as covariates are more likely to be missing, the bias will be greater. Second, it is increasing in , the average variation of for which ; the greater the variation, the greater the bias. Finally, the bias depends on , which is the relationship between an being observed and the size of .

To gain better insight into equation (), suppose so that each study has roughly the same estimation error variance. If we assume is on the scale of a standardized mean difference, where is the total sample size used to compute . Various researchers have described conventions for the magnitude of that range from to (Hedges & Pigott, 2001, 2004; Hedges & Schauer, 2019). Thus, we can write from some constant that ranges from 0 to 1.

The parameter is a log-odds ratio, which reflects the odds of a complete case for versus . There are various conventions for the size of an odds ratio that depend on base rates $P[R = \mathbbm{1} | T]$. Conventions used by Cohen (1988) have been interpreted as implying that a “small” odds ratio is about 1.49, a “medium” odds ratio is about 3.45, and a “large” odds ratio is about 9.0. Ferguson (2009) suggests 2.0, 3.0, and 4.0 for small, medium, and large odds ratios, while Chen et al. (2010) provide a range of conventions for different base rates, and their tables are roughly consistent with about 1.5 being a small odds ratio, 2.4 being medium, and 4.1 being large. Haddock et al. (1998) suggests any odds ratio over 3.0 would be considered quite large. Thus, consider a range of odds ratios from about 1.5 to 4.5.

However, the actual size of will depend on the scale of . A difference of is considered quite large for standardized mean differences. A less extreme difference for a standardized mean difference would be no larger than the size of an individual . Conventions for standardized mean differences imply that a “small” effect would be about , a “medium” effect would be , and a “large” effect would be (Cohen, 1988). Thus, meaningful values of might feasibly range from 0.2 to 1.0. These conventions for odds ratios and would imply that relevant values of might range from 0.4 (large with small odds ratio) to over 7.5 (small with large odds ratio).

Based on these conventions, Figure shows the potential (approximate) bias of for this example. Each panel corresponds to a given within-study variance and residual heterogeneity . Panels plot the bias contributed by a single case as a function of the probability of missingness (-axis) and (color). The panels on the bottom few rows and left most columns show that if both is small and is small, then will be less than 0.05. However if is larger and the probability of a complete case is strongly related to (i.e., is large), then the bias can be greater than 0.2 or even 0.5 (Cohen’s ).

It is worth noting that Figure gives the bias for when is positively correlated with being fully observed, and hence . This need not be the case when . When , then the bias of is negative, and would be be a mirror image of those in Figure . Larger, more negative values of would lead to a greater downward bias.

The bias of under selection model () is given by:

where is the mean among effects for which and . As with , the bias of is an increasing function of . If has a strong positive correlation with , then will be larger and so will the bias of .

When all studies have approximately the same estimation error variance so that and , then the bias of is approximately:

The expression in () depends on three quantities. Like , the bias of is an increasing function of and . The bias of also increases as a function of , which can be thought of as a difference in missingness rates between cases where and . This difference will be greater if and are strongly correlated. Taken together, the bias of will be greatest when there are fewer complete cases, missingness is strongly related to the size of effects and the value of the covariate .

To gain insight into the bias of , consider the values of and discussed above. The difference is just a difference in conditional probabilities. For reference, because both and are binary, then would be equal to the correlation between and (assuming equal marginals in a table). Thus, could be as small as 0, but could possibly be as large as 1, though conventions on the size of correlations suggest that would be a “large” value (Cohen, 1988).

Figure shows the potential bias of for this example assuming the values discussed above. Each panel corresponds to a given amount of heterogeneity , and within panels the bias is shown as a function of the difference (-axis) and (color). Figure highlights that the relationship between and () and between and (-axes) can affect the magnitude of the bias. If is strongly correlated with both and the bias can be as large as = 0.3 or 0.4. However, the less depends on or , the lower the bias is.

It is worth noting that the missingness mechanism in this example is MNAR. However, the bias is largely a function of , which is the coefficient in the selection model for the effect size . The data would be MAR under a selection model where in (). In that case, the bias of would be identical to the biases presented above.

For context, suppose the analysis of interest in the substance abuse example involved only the indicator for whether Group 1 received a high-intensity intervention. The missingness rate for Group 1’s intensity is 27%. Using a logit model analogous to @ref(eq:cc\_loglinear), we get an estimate of . Based on the heterogeneity estimated in the example and the mean within-study variance , this would give a bias of about in the intercept .

# 6 Bias in Shifting-Case Analyses

Shifting-case analyses (SCA) are a common approach in meta-regression when there are very few complete cases across multiple covariates. These analyses involve fitting multiple regression models, where each model omits some of the covariates of interest. In this sense, shifting-case analyses can be thought of as a set of regression models. Consider one model from that set, which estimates regression coefficients for some subset of the relevant covariates using the estimator in equation (). Recall that refers to the set of covariates omitted from the model, and that the estimator conditions on a set of missingness patterns . The set of missingness patterns is such that so that all included covariates are observed.

To understand the conditions under which is unbiased, we can write a shifting-case model as:

The model in () is slightly different from the models in the previous sections in that all of the functions depend on the covariates included in a given regression rather than the complete set of relevant covariates . Thus, the function can be thought of as a partial-data model, since it omits relevant covariates. The partial-data model need not be equivalent to the complete-data model because the former conditions only on and not the full set of covariates . These models would only be equivalent if . That is, unless the excluded covariates are completely unrelated to effect size (given the covariates included in the SCA model), then will be biased even if are completely observed.

The model in () suggests a very strict set of conditions for which is unbiased. First, missingness must be independent of effect sizes. This arises if or . This is a similar assumption as that made for unbiased complete-case analyses, and amounts to effect sizes (and potentially covariates) being independent of missingness.

Second, any excluded covariates must be completely irrelevant given the included covariates: . This amounts to for all . A related assumption is that , which would imply that the complete data likelihood involves no interactions between and and that and are orthogonal. Note that conditions on omitted covariates *and* omitted observations must hold in order for to be unbiased.

When the assumptions about omitted variables and cases are not met, will be biased. Just how biased will depend on a number of factors, including the amount of missingness, the missingness mechanism, and the relevance of any excluded covariates. The bias can be expressed as:

where is the matrix of omitted covariates and comprises the coefficients for the omitted covariates. The term is a vector of biases due to missingness and is the subset of for which . Note that the are the biases due solely to missingness as in equation ():

The expression in () shows that a shifting-case analysis suffers from two sources of bias. The first source, captured in the first term in (), is a function of the coefficients for the excluded covariates . This is referred to in the statistical and econometric literature as *omitted variable bias* (e.g., Farrar & Glauber, 1967; Mela & Kopalle, 2002). Omitted variable bias arises even if no are missing, and is related to the issue of multicollinearity in linear models. In fact, if the columns in and are orthogonal, so that the omitted variables are independent of the included variables, then the omitted variable bias will be zero. When the omitted variables are not orthogonal to the included variables, the bias will be nonzero, and it will depend in large part on the contribution of the omitted variables . The estimator will have greater bias if the coefficients for the omitted variables are larger and the omitted covariates are correlated with the included covariates .

The second term in () captures the bias due to ignoring observations missing . This *missingness bias* is a function of , which is itself a vector of biases for each effect, and it can be understood in terms of its individual components . Because the are of the same form for the complete-case versus shifting-case models, the missing data bias for a shifting-case analysis is governed by similar factors as the complete-case analyses, and are quite possibly similar in magnitude. Based on (), will be positive if is positively correlated with whether , and will be greater in magnitude when that correlation is larger.

Taken together, the bias of shifting-case analysis can be greater than the bias of a complete-case analysis. This occurs if the omitted variable bias and the missingness bias are in the same direction (e.g., both are positive). For both biases to be in the same direction, correlation between and the omitted variables must be in the same direction as the correlation between the probability that is observed. If, however, the omitted variable and missingness biases are in opposite directions, this can reduce the bias of a shifting-case estimator. It is worth noting, however, that it will almost always be impossible to confirm the direction of biases, since they depend on potentially unobserved covariates.

## 6.1 Example: Shifting-Cases Analysis with Two Binary Covariates

Suppose and and are binary covariates such that

If there is missingness in both and , then so that indicates both covariates are observed, and indicates only is observed. If missingness is such that for very few effect estimates, then a shifting-case analysis might involve regressing on the observed values of and then on the observed values of .

The first regression would take only rows for which is observed, so that and the excluded could be either 0 or 1. The shifting-case estimators follow from equation ():

Assume that missigness follows the following log-linear model:

Note that this gives the log-odds that an effect is included in the model given and , and that is not involved. Further, because the distribution of depends on , the mechanism is MNAR.

Given the selection model in (), the bias of the coefficient estimators can be written as:

Here are the missingness biases as defined above, and whose approximate values is given in (). To distinguish from the from the complete-case example, we use the notation.

Both the bias of and depend on two terms. The first term in each expression is the omitted variable bias, and the second term in each expression is the missingness bias. Consider the omitted variable biases. When effects are estimated with roughly the same precision, so that , then the omitted variable biases reduce to

The omitted variable biases for each coefficient can be seen as depending on two quantities. Both () and () are increasing in , which is the contribution of to the complete-data model. The omitted variable bias for is also increasing in . The bias for in () is increasing in . Because both and are binary, this difference is roughly equivalent to their Pearson correlation (assuming equal marginals). If , then their correlation is zero, and the omitted variable bias will be zero. But if and are correlated, the bias of will depend on how strongly correlated and are, and how big is.

Figure shows the omitted variable bias of (left plot) and as a function of . Both the bias and are shown on the scale of Cohen’s . In the left plot is the proportion of when . In the right plot, , which is roughly the correlation between and . Note that because can be intuited as (roughly) a Pearson correlation, the values in the figure include 0, 0.1 (i.e., a “small” correlation), 0.3 (medium correlation), and 0.5 (large correlation) (Cohen, 1988).

The figure shows that if so that is independent of given , that both and will be unbiased. However, when is nonzero, both estimators will be biased. If and are highly correlated, or if when with high probability, the bias of both estimators will about as large as a “small” effect (i.e., ) when is larger than 0.2. For the bias will be less than about when or if .

Figure does not take into account any bias induced by missingness. However, because the missingness mechanism in () is the same as the mechanism for the complete-case example (), the missingness bias for is the same as that for in (), which is shown in Figure . Likewise, the missingness bias for is the same as that for in (), which is shown in Figure .

Thus, the total bias of will be the sum of the omitted variable biases shown in Figure and the missingness biases shown in Figure . If both the omitted and missingness biases are on the higher end, the total bias of might be as large as to over 1.0. Likewise, the total bias of will be the sum of the omitted variable biases shown in Figure and the missingness biases shown in Figure , and can be larger than .

As noted above, the missingness bias and omitted variable bias can be in the different directions. For instance, if but , then the omitted variable bias for will be negative, but the missingness bias will be positive. In such cases, the bias of the shifting case estimators could be smaller than the bias of the complete-case estimators. However, because the biases depend on unknown (and potentially unobserved) quantities, it will often be impossible to empirically verify the magnitude or direction of the bias.

In the substance abuse example, consider the regression that omits Group 2’s covariates. Recall that the estimate for on complete cases was . Among the observed data, we find that and . Given these values, the omitted variable bias for the intercept would be , and the omitted variable bias for would be . Note, however, that the total bias further depends on the bias due to missingness. Based on empirical estimates in the previous section, this would give a bias in the estimate of of about .

# 7 Discussion

This article described a selection model approach to studying the bias of two common methods for conducting meta-regressions with missing covariates: complete-case and shifting-case analyses. Under certain assumptions regarding the selection model, we obtained expressions for the approximate bias of coefficient estimators. These expressions were presented in a general form, which was then unpacked by way of examples.

We found that both complete-case and shifting-case analyses will produce biased coefficient estimates unless certain conditions are met. Complete-case estimators are unbiased if the probability that all relevant covariates are observed is (conditionally) independent of the effect size estimate. Shifting-estimators are unbiased if, in addition to effect sizes being independent of missingness, the covariates omitted from a model have no relationship with the effect size. When these conditions are not met, the bias of coefficient estimates can be substantial—as large as to —depending on the missingness mechanism, the missingness rate, an the relevance of any omitted covariates.

An important aspect of these results is that bias will depend on unknown parameters and unobserved data. This means that it will be impossible to empirically verify the magnitude or direction of the bias. Even the estimated biases from the substance abuse data, which were on the order of about may not be entirely accurate, as so much of that data is missing. Further, it will require strong assumptions regarding the missingness mechanism to correct any bias. These assumptions may be buttressed by theory about scientific report, data collection, and data curation.

It is not immediately clear how commonly the conditions required for unbiased complete- and shifting-case estimators arise. Recent empirical work on examining missingness in meta-analytic datasets found that effect sizes can be strongly correlated with missingness, though this is not always the case (Schauer et al., under review). Further, the issues of multicollinearity and confounding in meta-regression, including those discussed by Lipsey (2003), would suggest that omitting variables in a shifting-case analysis are likely to induce bias.

Based on these results, our primary recommendation is that analysts attempt to understand the missingness mechanisms and patterns in their data. This can leverage knowledge about standard reporting and coding practices, as well as exploratory analyses (Schauer et al., under review). If there is very little missingness, or if there is good reason to assume that missingness is uncorrelated with effect size estimates, a complete-case analysis may be a reasonable option. However, we would discourage analysts from continuing to use shifting-case analyses because it would seem unlikely that omitted variable biases are zero in practice.

We would also suggest analysts investigate the feasibility of alternative estimation methods. Ibrahim (1990) describes an EM algorithm for generalized linear models with missing covariates, and Ibrahim et al. (1999) extend that algorithm when covariates are MNAR. In addition, full-information maximum likelihood (FIML) has long been used in linear models (Graham, 2009, 2012), and has shown some promise for meta-regression involving continuous covariates. Finally, multiple imputation (MI) has become something of a standard approach for handling missing data across a number of fields (Little & Rubin, 2002; Rubin, 1987; van Buuren, 2018).

However, employing any of these alternative strategies is not necessarily straightforward for meta-analysts. To our knowledge, the EM algorithm for missing covariates has yet to be implemented in standard meta-analytic software. Although FIML for meta-regression model is available in SEM framework (Cheung, 2019), the approach has not been empirically validated under various conditions. How best to specify quality imputation models for MI analyses is something of an open question for meta-regression, as is the potential inaccuracies incurred by using poor imputation models. Research on and clear implementation of these methods for meta-regression model would seem to be of great use for meta-analysts.

# 8 Approximate Bias for Log-Linear Selection Models

**Proposition:** Suppose is the standard fixed- or random effects meta-regression model in equation (), and suppose follows the log-linear model in (). Then:

where . Therefore, the bias of the conditional expectation is given by:

**Proof:**

In this proof, we drop the subscript for sake of simplicity. Denote

$$\begin{align\*}
H\_j(X\beta, X, v) \equiv H\_j(X, v)
& = P[R =\not\in \mathcal{R}\_j | T, X, v] \rvert\_{T = X\beta} \\
G\_j(X, v)
& = 1 - H\_j(X, v) \\
g\_j(X, v)
& = P[R \in \mathcal{R}\_j | X, v] \text{ as in (\ref{eq:pr\_xv}).}
\end{align\*}$$

Then an approximation for is as follows:

Note that this uses a first order Taylor expansion of the log-linear model at , and thus assumes the are differentiable. The approximation will be more accurate if are small. A more accurate approximation is possible if the are linear in . In that case, only an approximation of the denominator of the log-linear model is required.

# 9 References

Agresti, A. (2013). *Categorical data analysis* (3rd ed). Hoboken, NJ: Wiley.

Bennett, D. A. (2001). How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health*, *25*(5), 464–469.

Berkey, C. S., Hoaglin, D. C., Mosteller, F., & Colditz, G. A. (1995). A random-effects regression model for meta-analysis. *Statistics in Medicine*, *14*(4), 395–411. <https://doi.org/10.1002/sim.4780140406>

Borenstein, M. (2009). *Introduction to meta-analysis*. Chichester, U.K.: John Wiley & Sons. Retrieved from <http://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=427912>

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics - Simulation and Computation*, *39*(4), 860–864. <https://doi.org/10.1080/03610911003650383>

Cheung, M. W.-L. (2019). *Handling missing covariates in mixed-effects meta-analysis with full-information maximum likelihood*. Presented at the Society for Research Synthesis Methods, Chicago, IL. Retrieved from <http://www.srsm.org/uploads/4/6/1/3/46138157/abstract_-_mike_cheung.pdf>

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Hillsdale, N.J: L. Erlbaum Associates.

Cooper, H. M. (1998). *Synthesizing research: A guide for literature reviews* (3rd ed). Thousand Oaks, Calif: Sage Publications.

Cooper, H. M. (2017). *Research synthesis and meta-analysis: a step-by-step approach* (Fifth Edition). Los Angeles: SAGE.

Cooper, H. M., Hedges, L. V., & Valentine, J. C. (Eds.). (2019). *Handbook of research synthesis and meta-analysis* (3rd edition). New York: Russell Sage Foundation.

Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: The problem revisited. *The Review of Economics and Statistics*, *49*(1), 92. <https://doi.org/10.2307/1937887>

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*(5), 532–538. <https://doi.org/10.1037/a0015808>

Gelman, A. (2014). *Bayesian data analysis* (Third edition). Boca Raton: CRC Press.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*(1), 549–576. <https://doi.org/10.1146/annurev.psych.58.110405.085530>

Graham, J. W. (2012). *Missing Data*. New York, NY: Springer New York. Retrieved from <http://link.springer.com/10.1007/978-1-4614-4018-5>

Haddock, C. K., Rindskopf, D., & Shadish, W. R. (1998). Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods*, *3*(3), 339–353. <https://doi.org/10.1037/1082-989X.3.3.339>

Hedges, L. V. (1983a). A random effects model for effect sizes. *Psychological Bulletin*, *93*(2), 388–395. <https://doi.org/10.1037/0033-2909.93.2.388>

Hedges, L. V. (1983b). Combining independent estimators in research synthesis. *British Journal of Mathematical and Statistical Psychology*, *36*(1), 123–131. <https://doi.org/10.1111/j.2044-8317.1983.tb00768.x>

Hedges, L. V., & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods*, *6*(3), 203–217. <https://doi.org/10.1037/1082-989X.6.3.203>

Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods*, *9*(4), 426–445. <https://doi.org/10.1037/1082-989X.9.4.426>

Hedges, L. V., & Schauer, J. M. (2019). Statistical analyses for studying replication: Meta-analytic perspectives. *Psychological Methods*, *24*(5), 557–570. <https://doi.org/10.1037/met0000189>

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65. <https://doi.org/10.1002/jrsm.5>

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, *3*(4), 486–504. <https://doi.org/10.1037/1082-989X.3.4.486>

Ibrahim, J. G. (1990). Incomplete data in generalized linear models. *Journal of the American Statistical Association*, *85*(411), 765–769. <https://doi.org/10.1080/01621459.1990.10474938>

Ibrahim, J. G., Lipsitz, S. R., & Chen, M.-H. (1999). Missing covariates in generalized linear models when the missing data mechanism is non-ignorable. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *61*(1), 173–190. <https://doi.org/10.1111/1467-9868.00170>

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis: Three-level meta-analysis. *Research Synthesis Methods*, *2*(1), 61–76. <https://doi.org/10.1002/jrsm.35>

Laird, N. M., & Mosteller, F. (1990). Some statistical methods for combining experimental results. *International Journal of Technology Assessment in Health Care*, *6*(1), 5–30. <https://doi.org/10.1017/S0266462300008916>

Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *The ANNALS of the American Academy of Political and Social Science*, *587*(1), 69–81. <https://doi.org/10.1177/0002716202250791>

Lipsitz, S. (1996). A conditional model for incomplete covariates in parametric regression models. *Biometrika*, *83*(4), 916–922. <https://doi.org/10.1093/biomet/83.4.916>

Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. Hoboken, NJ, USA: John Wiley & Sons, Inc. Retrieved from <http://doi.wiley.com/10.1002/9781119013563>

Mela, C. F., & Kopalle, P. K. (2002). The impact of collinearity on regression analysis: The asymmetric effect of negative and positive correlations. *Applied Economics*, *34*(6), 667–677. <https://doi.org/10.1080/00036840110058482>

Pigott, T. D. (2001a). A review of methods for missing data. *Educational Research and Evaluation*, *7*(4), 353–383. <https://doi.org/10.1076/edre.7.4.353.8937>

Pigott, T. D. (2001b). Missing predictors in models of effect size. *Evaluation & the Health Professions*, *24*(3), 277–307. <https://doi.org/10.1177/01632780122034920>

Pigott, T. D. (2019). Handling missing data. In Harris Cooper, Larry V. Hedges, & Jeffrey C. Valentine (Eds.), *The Handbook for Research Synthesis and Meta-analysis* (3rd ed.). New York: Russell Sage.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. <https://doi.org/10.1093/biomet/63.3.581>

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical Methods in Medical Research*, *8*(1), 3–15. <https://doi.org/10.1177/096228029900800102>

Schauer, J. M., Dìaz, K., & Pigott, T. D. (under review). Exploratory analyses for missing data in meta-analyses.

Tanner-Smith, E. E., Steinka-Fry, K. T., Kettrey, H. H., & Lipsey, M. W. (2016, December). Adolescent substance use treatment effectiveness: A systematic review and meta-analysis. Office of Justice Programs.

Tipton, E., Pustejovsky, J. E., & Ahmadi, H. (2019a). A history of meta-regression: Technical, conceptual, and practical developments between 1974 and 2018. *Research Synthesis Methods*, *10*(2), 161–179. <https://doi.org/10.1002/jrsm.1338>

Tipton, E., Pustejovsky, J. E., & Ahmadi, H. (2019b). Current practices in meta-regression in psychology, education, and medicine. *Research Synthesis Methods*, *10*(2), 180–194. <https://doi.org/10.1002/jrsm.1339>

van Buuren, S. (2018). *Flexible Imputation of Missing Data, Second Edition* (2nd ed.). Second edition. | Boca Raton, Florida : CRC Press, [2019] |: Chapman and Hall/CRC. Retrieved from <https://www.taylorfrancis.com/books/9780429492259>

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, *30*(3), 261–293. <https://doi.org/10.3102/10769986030003261>

Viechtbauer, W. (2007). Accounting for heterogeneity via random-effects models and moderator analyses in meta-analysis. *Zeitschrift Für Psychologie / Journal of Psychology*, *215*(2), 104–121. <https://doi.org/10.1027/0044-3409.215.2.104>