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

1 Introduction

Meta-regression is a useful tool for studying important sources of variation between effects in a meta-analysis.[1,2] Analyses of these models in the absence of missing data have been studied thoroughly in the literature.[3–7] However, it is common for meta-analytic datasets to be missing data.[8] In the context of meta-regression, issues with missing data frequently involve missing covariates.[9,10]

Precisely how to proceed with a meta-regression when missing covariates remains something of an open question. Statistical guidance suggests that analyses ought to consider the mechanism that causes covariates to be missing.[9,11] 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.[10] An analyst may conduct a *complete-case analysis* (CCA) that excludes any effects for which a relevant covariate is missing (i.e., only analyze complete cases). This is often referred to as "listwise deletion" in data analyses. 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* (SCA).[12] In a SCA, analysts fit a series of meta-regression models on subsets of relevant covariates, so that each model selectively omits certain covariates. This is equivalent to "pairwise deletion" in data analyses.

Both CCA and SCA ignore effects for which a covariate is missing. Ignoring missing data can potentially lead to biased estimates of parameters of interest.[13,14] Despite authors pointing out such issues in meta-analysis, these methods continue to enjoy widespread use.[11] Existing meta-analysis 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.[15] We then introduce a statistical framework for studying bias for incomplete data meta-regressions that incorporates a model for whether or not a covariate is observed. Using this framework, we describe conditions under which CCA and SCA are unbiased. When these conditions are not met, we derive an approximation for the bias of CCA and SCA using standard models for missingness and examine the magnitude of bias. We find that bias is highly dependent on the precise mechanism by which data are missing, and is less reliant on more traditional missingness mechanism classifications (e.g., missing at random vs. not at random).

2 Example: Substance Abuse Interventions

Tanner-Smith et al. [15] 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 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. Typically, researchers avoided no-treatment or placebo conditions in studies over ethical concerns surrounding the failure to treat adolescents with substance abuse disorders. Thus, contrasts within studies (i.e., effect estimates) tended to focus on a specific treatment of interest to the researcher versus some alternate treatment. Effect estimates are reported on the scale of bias-corrected standardized mean differences.

Suppose the 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., $X_1 = 1$ if group 1 treatment was high-intensity) and the other would indicate whether group 2 received a high intensity (i.e., $X_2 = 1$ if group 2 treatment was high-intensity). The relevant meta-regression model would regress the effect estimates on these two covari-

ates.

In the data, the treatment intensity is missing for some of the effects, and Table 1 summarizes missingness for these covariates. Table 1 shows that only 37 of the 74 (50%) have a reported treatment intensity for both groups (i.e., X_1 and X_2 are both observed), but that 54 (73%) of effects report Group 1's treatment intensity (i.e., X_1 is observed) and 41 (55%) effects report Group 2's treatment intensity.

[Table 1 about here.]

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 CCA would result in the coefficient estimates and standard errors displayed in the first column of Table 2. Based on these estimates, when Group 1 receives a high-intensity treatment, we would expect an effect to be larger by $d = 0.44$ (in standard deviation units) than when Group 1 receives a low-intensity treatment, which is statistically significant at the $\alpha = 0.10$ level. Note that the estimated between-effect variance is $\hat{\tau}^2 = 0.08$.

[Table 2 about here.]

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 SCA would use the 54 effects for which Group 1's treatment intensity is observed (i.e., X_1 is observed), but only including X_1 in the model. Doing so leads to the estimates in second column of Table 2. 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 SCA would include the 41 effects for which Group 2's intensity is observed, and include only that covariate in the model. The third column of Table 2 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 CCA (-0.21).

It should be noted that all of these estimates and comparisons between them 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.[16,17] 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.[18] It could even be argued that the parameters in the model with both covariates are not comparable to parameters in models with only one covariate; coefficients in a model with multiple covariates must be interpreted in relation to other variables in the model. The remainder of this article quantifies the bias induced by omitting effect sizes and/or covariates from meta-regressions.

3 Model and Notation

Suppose a meta-analysis involves k effects estimated from collection of studies. For the ith effect, let T_i be the estimate of the effect parameter θ_i, and let v_i be the estimation error variance of T_i. Denote a vector of covariates that pertain to effect estimate T_i as $X_i = [1, X_{i1}, \ldots, X_{ip}]$. Note that the first element of X_i is a 1, which corresponds to an intercept term in a meta-regression model, and that X_{ij} for $j = 1, \ldots p$ corresponds to different covariates. The meta-regression model can be expressed as:

$$T_i | X_i, v_i, \eta = X_i \beta + u_i + e_i \tag{1}$$

Here, $\beta \in \mathbb{R}^{p+1}$ is the vector of regression coefficients. The estimation errors e_i are typically assumed to be normally distributed with mean zero and variance $V[e_i] = v_i$. This assumption is true of some effect size indices and is a very accurate large-sample approximation for others.[19] The term u_i represents the random effect such that $u_i \perp e_i$ and $V[u_i] = \tau^2$. This model is equivalent to the standard mixed-effects meta-regression model, and it is also consistent with subgroup analysis models.[19,20] The vector $\eta = [\beta, \tau^2]$ refers to the parameters of model. Under a fixed-effects model, it is assumed that $\tau^2 = 0$, in which case $\eta = \beta$, and $u_i \equiv 0$.

A common assumption in random effects meta-regression is that the random effects u_i are independent and normally distributed with mean zero and variance τ^2:[20–23]

$$u_i \sim N(0, \tau^2).$$

This could correspond to a scenario of k independent effect estimates presumably from k different studies. In that case, the distribution $p(T|X, v, \eta)$ can be written as

$$p(T_i | X_i, v_i, \eta) = \frac{1}{\sqrt{2\pi(\tau^2 + v_i)}} e^{-\frac{(T_i - X_i\beta)^2}{2(\tau^2 + v_i)}} \tag{2}$$

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

$$p(\mathbf{T} | \mathbf{X}, \mathbf{v}, \eta) = (2\pi)^{-k/2} \left[\prod_{i=1}^{k} (\tau^2 + v_i) \right] e^{-\sum_{i=1}^{k} \frac{(T_i - X_i\beta)^2}{2(\tau^2 + v_i)}} \tag{3}$$

where $\mathbf{T} \in \mathbb{R}^k$ is the vector of effect estimates, $\mathbf{v} \in \mathbb{R}^k$ is the vector of estimation variances, and $\mathbf{X} \in \mathbb{R}^{k \times (p+1)}$ is the matrix of covariates where each row of $\mathbf{X}$ is simply the row vector X_i. Note that the functions in both (2) and (3) assume that *all* of the p covariates are observed. Equation (3) is referred to as

the *complete-data likelihood function*.[13,24] 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

$$\hat{\beta} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{T} \tag{4}$$

Here, $\mathbf{W} = \mathrm{diag}[w_i]$ is the diagonal matrix of weights such that $w_i = 1/(v_i + \tau^2)$. The covariance matrix of $\hat{\beta}$ is given by

$$V[\hat{\beta}] = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \tag{5}$$

Note that the weights involve the true variance component τ^2. In practice, τ^2 must be estimated by $\hat{\tau}^2$, and the resulting weights used in analyses can be written $\hat{w}_i = 1/(v_i + \hat{\tau}^2)$. For the sake of simplicity, we use w_i to derive results in this article, and so results do not depend on variance component estimators. Presumably, use of $\hat{w}_i$ would induce additional variation into analyses.

The substance abuse data contains multiple effect estimates per study that are likely correlated. This differs from the model above. However, we can expand this model to account for dependent effect sizes by assuming that $T_i \in \mathbb{R}^{k_i}$ is a vector of k_i effects from the same study, e_i is vector of estimation errors, u_i is a vector of random effects, and $e_i + u_i$ has covariance matrix Σ_i. In this model, X_i is a matrix of covariates for each effect in T_i. The resulting formulas for the complete-data likelihood function and coefficient estimators will be more complex (including a variance-covariance weight matrix), but they will have a similar form as the independent effect size model.

Not all relevant variables may be observed in a meta-analytic dataset. Let R_i be a vector of response indicators that correspond with effect i. This article concerns missing covariates, and we assume that T_i and v_i are observed for every effect of interest in a meta-analysis. Thus, each element R_{ij} of R_i corresponds to a covariate X_{ij}. The R_{ij} take a value of either 0 or 1: $R_{ij} = 1$ indicates the corresponding X_{ij} is observed and $R_{ij} = 0$, indicates a that the corresponding X_{ij} is not observed. Note that $R_i \in \mathcal{R} \equiv \{0,1\}^p$ is a vector of 0s and 1s of length p. For instance, X_{i2} were missing, this would be indicated by $R_{i2} = 0$.

Denote $O = \{(i,j) : R_{ij} = 1\}$ as the indices of covariates that are observed and $M = \{(i,j) : R_{ij} = 0\}$ be the set of indices for missing covariates. Then, the complete-data model can be written as

$$p(\mathbf{T}|\mathbf{X}, \mathbf{v}, \eta) = p(\mathbf{T}|\mathbf{X}_O, \mathbf{X}_M, \mathbf{v}, \eta). \tag{6}$$

Note that the complete-data model depends on entries of $\mathbf{X}_M$, which are unobserved. It is worth pointing out that the *complete-data model*, which refers to the model with no missing data, is distinct from the *complete-case analysis*, which is an estimation procedure that conditions only on observed data.

3.1   Complete-Case Estimators

A common approach in meta-regression with missing covariates is to use a complete-case analysis.[10,11] 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 $R_i = [1, \ldots, 1] = \mathbb{1}$.

Let $C = \{i : R_i = \mathbb{1}\}$ index all relevant effects i such that $R_i = \mathbb{1}$, so that $\mathbf{X}_C$ is the matrix of covariates such $R_i = \mathbb{1}$, $\mathbf{T}_C$ is the corresponding subset of effect estimates, and $\mathbf{W}_C$ is the corresponding subset of weights. The CCA estimates the coefficients β with:

$$\hat{\beta}_C = (\mathbf{X}_C^T \mathbf{W}_C \mathbf{X}_C)^{-1} \mathbf{X}_C^T \mathbf{W}_C \mathbf{T}_C \tag{7}$$

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.[11] 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*.[10,12]

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.[10,11] In the substance abuse data example, we focused on two covariates of interest X_{i1} and X_{i2}. The SCA first regressed T_i on observed values of X_{i1}. This regression included observations for which both X_{i1} and X_{i2} are observed (i.e., $R_i = [1, 1]$) and observations for which X_{i1} is observed but X_{i2} is missing (i.e., $R_i = [1, 0]$). We then regressed T_i on X_{i2}, which included effects for which $R_i \in \{[1, 1], [0, 1]\}$. In sum, the SCA demonstrated in the previous section involved two regressions, each of which conditioned on different sets of missingness patterns.

To formalize SCA estimators, consider a single regression in an SCA, and let S index the component of X_i (i.e., the intercept term and relevant covariates) included in that model $S = \{j : j = 0$ or X_{ij} in analysis$\}$. Let E be the complement of S so that E indexes the covariates excluded from the regression. Then, the regression is used to estimate and make inferences about coefficients β_S. In the fol-

lowing section, we discuss β_S and its relationship to β, but here assume that the target of inference for an SCA is β and hence β_S comprises a subset of the components of β. For instance, in the first substance abuse SCA regression, T_i was regressed on only X_{i1}, so that $\beta_S = [\beta_0, \beta_1]$.

Denote $\mathcal{R}_j$ as the set of missingness patterns such that all included covariates are observed: $\mathcal{R}_j = \{R \in \mathcal{R} : R_S = \mathbb{1}\}$. Note that $\mathcal{R}_j$ contains missingness patterns such that all the included covariates are observed, but any excluded covariates may be either observed or unobserved. For instance, in the first substance abuse SCA regression of T_i on X_{i1}, the analysis included effects such that $R_i \in \mathcal{R}_1 = \{[1,1], [1,0]\}$. Finally, let U denote the indices i of effects for which X_{iS} are observed; note that U depends on S, so we may write $U(S) = \{i : R_i \in \mathcal{R}_j\}$. Then, the shifting-case estimators for β_S are given by:

$$\hat{\beta}_S = (\mathbf{X}_{US}^T \mathbf{W}_U \mathbf{X}_{US})^{-1} \mathbf{X}_{US}^T \mathbf{W}_U \mathbf{T}_U \tag{8}$$

where $\mathbf{X}_{US}$ contains the columns (S) of $\mathbf{X}$ that pertain to the covariates that are included in the SCA regression, and the rows (U) for which all of those covariates are observed. The matrix $\mathbf{W}_U$ is a square matrix containing the relevant rows and columns of $\mathbf{W}$ for which X_{iS} are observed, while $\mathbf{T}_U$ contains the effect sizes in $\mathbf{T}$ for which X_{iS} is observed.

3.3   Omitted Variables

A common concern in meta-regression is that models may not be able to account for all relevant covariates, either due to sample size constraints or because some covariates were not observed.[25] Such concerns pertain to meta-regressions both with and without missing data. It seems logical that if variables omitted from a model would have been of concern in the absence of missing data, that same concern holds for models estimated with CCA or SCA. Moreover, in SCA, the nature of omitted variables requires some additional considerations, since observed covariates are sequentially omitted from each meta-regression in an SCA.

The implication of omitting observed variables in SCA can be understood via the parameter β_S. It has been noted that there are various conditions under which components of β_S are unequal to their counterparts in β.[26] For instance, it can be the case that $\beta_S = [\tilde{\beta}_0, \tilde{\beta}_1] \neq [\beta_0, \beta_1] \subset \beta$. The difference between β_S and components of β is often referred to as *omitted variable bias* in the statistical and econometric literature.[27,28] This conception inherently assumes that β in the full model is of interest to the analyst. Alternatively one may assume that β_S is of interest, rather than β, so that the components of β_S comprise parameters distinct from β. In this approach, β_S characterizes the relationship between X_{iS} and T_i in a more restricted model that does not account for X_{iE}.

We refer to the difference between β_S and β as omitted variable bias, in keeping with the literature. This is consistent with SCA as a missing data analytic strategy, wherein the target of inference is β, and it reflects the findings of Lipsey (2003), who points out that interpretation of meta-regression coefficients when covariates are omitted can lead to misleading interpretations about the correlates of effective interventions. To that end, subsequent sections present findings on bias induced by omitting observed covariates in an SCA. However, if the intent of the analysis is to examine restricted models, the omitted variable bias presented in this article may be less applicable, though Lipsey's caveats for interpreting such models still apply.

3.4 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 R. Let ψ denote the parameter (or vector of parameters) that index the distribution of R so that the probability mass function of R can be written as $p(R|T, X, v, \psi)$. Assumptions about the missingness mechanism are therefore equivalent to assumptions about $p(R|T, X, v, \psi)$.

Rubin[29] defined three types of mechanisms in terms of the distribution of R. 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:

$$p(\mathbf{R}|\mathbf{T}, \mathbf{X}_O, \mathbf{X}_M, \mathbf{v}, \psi) = p(\mathbf{R}|\psi)$$

MCAR implies that probability that a given value is missing depends only on the missingness parameter ψ.

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

$$p(\mathbf{R}|\mathbf{T}, \mathbf{X}_O, \mathbf{X}_M, \mathbf{v}, \psi) = p(\mathbf{R}|\mathbf{T}, \mathbf{X}_O, \mathbf{v}, \psi)$$

MAR differs from MCAR in that missingness might be related to observed values. As an example, if studies 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 R depends on unobserved data in some way. In the context of the meta-regression data, this would imply that R is related to $\mathbf{X}_M$, 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) ψ in order to conduct inference on η.[13,14,24,30] In practice, missing data are ignorable if they are MAR and if ψ and η are distinct.

4   Conditional Incomplete Data Meta-Regression

Because both complete- and available-case analyses depend on the value of R_i, 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 R_i. Yet, CCA and SCA 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 R_i. This approach is referred to as a *selection model* in the missing data literature.[13,24,30] We can write the selection model for meta-regression with missing covariates as:

$$p(T_i|X_i, v_i, R_i \in \mathcal{R}_j, \eta, \psi) = \frac{p(R_i \in \mathcal{R}_j|T_i, X_i, v_i, \psi)p(T_i|X_i, v_i, \eta)}{p(R_i \in \mathcal{R}_j|X_i, v_i, \psi, \eta)} \tag{9}$$

where ψ indexes the distribution of $R|T, X, v$. Here, $\mathcal{R}_j$ refers to the relevant subset of $\mathcal{R}$ on which the analysis conditions; for a CCA, $\mathcal{R}_j = \{\mathbb{1}\}$.

Equation (9) describes the conditional model as a function of the complete-data model $p(T_i|X_i, v_i, \eta)$ and a selection model $p(R_i \in \mathcal{R}_j|T_i, X_i, v_i, \psi)$ that gives the probability that a given set of covariates are observed. The denominator on the right hand side of (9) is a normalizing factor that is equivalent to the probability of observing the missingness pattern $\mathcal{R}_j$ given the estimation error variance

v_i and the observed and unobserved covariates in the vector X_i, and can be written as

$$p(R_i \in \mathcal{R}_j | X_i, v_i, \psi, \eta) = \int p(R_i \in \mathcal{R}_j | T_i, X_i, v_i, \psi) p(T_i | X_i, v_i, \eta) dT_i \tag{10}$$

Note that when the complete-data model in (2) is not equivalent to the conditional model in (9), the resulting coefficient estimators in a meta-regression can be biased. To see this, we can write:

$$E[T_i | X_i, v_i, R_i \in \mathcal{R}_j] = E[T_i | X_i, v_i] + \delta_{ij} = X_i \beta + \delta_{ij} \tag{11}$$

Here, we see that the expectation of T_i given X_i and R_i can be written as the complete-data expectation $X_i \beta$ (i.e., the regression model) plus a bias term δ_{ij}. The bias term δ_{ij} refers to the bias induced in regression model due to conditioning on missingness pattern $\mathcal{R}_j$, which can affect individual components of η. If $\delta_{ij} \neq 0$, it follows that conditioning on R_i induces bias in the distribution of T_i used in an analysis. Because the CCA estimator (7) and SCA estimator in (8) are weighted averages of the T_i, they can be biased if $\delta_{ij} \neq 0$. The precise magnitude of the δ_{ij} will depend on the selection model in (9) and hence on the missingness mechanism. It is worth noting that the subsequent sections show that bias depends on the precise selection model rather than the class of mechanism (MCAR or MAR).

A standard approach for modeling missingness mechanisms for covariates is to assume R_i follows some log-linear distribution.[31] Various authors have described approaches to modelling R for missing covariates in generalized linear models that include logistic and multinomial logistic models.[32–34] Thus, one class of models for missingness would involve the logit probability of observing some missingness patterns $R_i \in \mathcal{R}_j \subset \mathcal{R}$:

$$\text{logit}[p(R_i \in \mathcal{R}_j | T_i, X_i, v_i)] = \sum_{m=0}^{m_j} \psi_{mj} f_{mj}(T_i, X_i, v_i) \tag{12}$$

Here, $f_{mj}(T_i, X_i, v_i)$ are assumed to be differentiable basis functions of the data and m_j is the number of terms in the selection model. In theory, m_j could be arbitrarily large, but the model is only estimable if $m_j < k$. Finally, we assume $f_{0j}(T_i, X_i, v_i) = 1$, so that ψ_{0j} would be the intercept term for the logit model for the set of missingness patterns $\mathcal{R}_j$.

In general, it will be impossible to know whether a selection model is correctly specified, but the formulation in (12) offers a few important advantages. First, it is fairly general: the only assumption made of the basis functions f_{mj} is that they are differentiable, which means model (12) allows for nonlinear or interaction terms. Second, it expresses the relationships between the probability of the event $\{R_i \in \mathcal{R}_j\}$ and observed variables on the scale of the log odds ratio, a well-understood scale in meta-analysis. Third, it allows for closed-form expressions for the approximate bias of coefficient estimates by virtue of the the

logit link function. Thus, it comprises a large class of models for selection that can be more clearly interpreted.

4.1 Approximate Bias for Log-linear Selection Models

As argued above, the bias of complete-case estimators $\hat{\beta}_C$ or shifting-case estimators $\hat{\beta}_S$ will depend in some way on the bias δ_{ij} induced in T_i by conditioning on $R_i \in \mathcal{R}_j$. The magnitude and direction of δ_{ij} will in turn depend on the selection model.

It is possible to derive an approximation for δ_{ij} under certain conditions. If $p(T_i|X_i, v_i)$ is the standard fixed- or random effects meta-regression model in equation (2), and $p(R_i \in \mathcal{R}_j|T_i, X_i, v_i)$ follows the log-linear model in (12), and the f_{mj} are differentiable with respect to T_i, then

$$\delta_{ij} \approx H_j(X_i\beta, X_i, v_i)(\tau^2 + v_i) \sum_{m=0}^{m_j} \psi_{mj} f'_{mj}(X_i\beta, X_i, v_i) \tag{13}$$

where $H_j(X_i\beta, X_i, v_i)$ is equivalent to $p(R_i \notin \mathcal{R}_j|T_i, X_i, v_i)$ evaluated at $T_i = X_i\beta$ and

$$f'_{mj}(X_i\beta, X_i, v_i) = \left.\frac{\partial f_{mj}}{\partial T_i}\right|_{T_i = X_i\beta}$$

is the derivative of f_{mj} with respect to T_i evaluated at $T_i = X_i\beta$. A more detailed proof is presented in Appendix A.

While the following sections will examine possible values that δ_{ij} may take under different selection models, we can gain some insight on bias by examining (13). The expression for δ_{ij} depends on three main quantities. First, δ_{ij} is an increasing of $H_j(X_i\beta, X_i, v_i)$, which is the probability that $R_i \notin \mathcal{R}_j$. This implies that the bias will be greater as the probability of omitting an observation increases. Second, δ_{ij} increases in the sum of variance components $\tau^2 + v_i$, which means that the bias will be larger when T_i vary more around the regression line. Finally, δ_{ij} depends on $\psi_{mj} f'_{mj}(X_i\beta, X_i, v_i)$. Since f'_{mj} is the derivative of f_{mj} with respect to T, when f_{mj} does not depend on T, then $f'_{mj} = 0$, and hence $\psi_{mj} f'_{mj} = 0$. Thus, δ_{ij} depends on the components of the selection model that are functions of T_i and how strongly those components are related to the probability of observing X_i via the parameter ψ.

5   Bias in Complete-Case Analyses

Complete-case analyses only include effects for which all relevant covariates are observed. The complete-case coefficient estimator $\hat{\beta}_C$ given in equation (7) conditions on $R_i = \mathbb{1}$. As noted above, conditioning

on R_i can induce bias, however there are conditions under which the CCA will lead to unbiased coefficient estimates. These conditions largely amount to whether or not R_i is independent of the effect size estimate T_i, the outcome of meta-regression model. When the distribution of R_i depends on T_i, then complete-case estimators will be biased.

The general condition under which CCA estimators are unbiased is that $R_i \perp T_i$, which occurs for different types of selection models. First, if the covariates are MCAR, then $R_i \perp (T_i, X_i, v_i)$. Alternatively, if the selection model depends only on v_i, but not X_i or T_i, then $R_i \perp (T_i, X_i)|v_i$; this would constitute a MAR mechanism. Finally, if the selection model depends only on v_i and X_i, but not T_i, then $R_i \perp T_i|(X_i, v_i)$, which would correspond to an MNAR mechanism. Under each of these assumptions, it can be shown that the model that conditions on complete cases $R_i = \mathbb{1}$ is identical to the complete-data model, and hence CCA estimators will be unbiased:

$$p(T_i|X_i, v_i, R_i = \mathbb{1}, \eta, \psi) = \frac{p(R_i = \mathbb{1}|T_i, X_i, v_i, \psi)p(T_i|X_i, v_i, \eta)}{p(R_i = \mathbb{1}|X_i, v_i, \eta, \psi)} = p(T_i|X_i, v_i, \eta) \tag{14}$$

This result is consistent with prior work regarding linear regression models with missing covariates.[35,36]

An important aspect of this result is that whether or not a CCA produces unbiased coefficient estimates depends more on the role of T_i in the selection model rather than traditional mechanism classifications of MCAR, MAR, or MNAR. Though various selection models satisfy the conditions of MAR, and similarly with MNAR, the key factor for bias in CCA estimators is the relationship between R_i and T_i. Should $R_i \not\perp T_i$, then CCA estimators can be biased, regardless of whether the mechanism is MAR or MNAR. Similarly, if $R_i \perp T_i$, CCA estimators can be unbiased, regardless of MAR or MNAR.

When R_i is not independent of T_i (given X_i or v_i), then CCA can be biased. Let $\mathcal{R}_1 = \{\mathbb{1}\}$ so that the CCA conditions on $R_i \in \mathcal{R}_1$. Based on equation (11), the bias of $\hat{\beta}_C$ will depend on the δ_{i1}. If we let $\Delta = [\delta_{11}, \ldots, \delta_{k1}]$ be the vector of δ_{i1} and let Δ_C be the subset of Δ for which all covariates are observed (i.e., $R_i = \mathbb{1}$). Then the bias of the complete-case analysis can be written as

$$\text{Bias}[\hat{\beta}_C] = (\mathbf{X}_C^T \mathbf{W}_C \mathbf{X}_C)^{-1} \mathbf{X}_C^T \mathbf{W}_C \Delta_C \tag{15}$$

The bias in equation (15) is a weighted average of individual biases δ_{i1}. Hence, the bias will be larger if the δ_{i1} are larger (and in the same direction).

Precisely, how large the bias in (15) is will depend on the distribution of R_i and its relationship to effect estimates T_i and their covariates X_i. When R_i follows the log-linear model in (12), the approximate

bias can be written as

$$\text{Bias}[\hat{\beta}_C] \approx (\mathbf{X}_C^T \mathbf{W}_C \mathbf{X}_C)^{-1} \mathbf{X}_C^T \mathbf{W}_C \mathbf{H}_{1C} \mathbf{f}_{1C} \psi_1 \tag{16}$$

where

$$\mathbf{H}_1 = \text{diag}[H_1(X_i \beta, X_i, v_i)]$$

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

$$\mathbf{f}_1 = \left[f'_{01}(X_i^T \beta, X_i, v_i), \ldots, f'_{m_1 1}(X_i^T \beta, X_i, v_i) \right]$$

is a $k \times m_1$ matrix of derivatives, and $\psi_1 = [\psi_{01}, \ldots, \psi_{m_1 1}]^T$ is a vector of parameters that index the selection model. Note that the bias in (16) involves $\mathbf{H}_{1C}$ which contains the rows of $\mathbf{H}_1$ for which $R_i = \mathbb{1}$; similarly for $\mathbf{f}_{1C}$.

While (16) provides a general expression for the approximate bias of $\hat{\beta}_C$, it can be a little difficult to interpret. Loosely, we can see that the bias depends on the probability that covariates are missing, reflected in $\mathbf{H}_{1C}$, as well as some function of the components of the log-linear selection model $\mathbf{f}_{1C} \psi_1$. 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 $X_{i1} \equiv X_i \in \{0, 1\}$, so that the complete data model is

$$T_i = \beta_0 + \beta_1 X_i + u_i + e_i \tag{17}$$

where β_0 and β_1 are the regression coefficients of interest. Note that β_0 is the average effect when $X_i = 0$ and β_1 is the contrast in mean effects for when $X_i = 1$ versus when $X_i = 0$.

Because X_i is a scalar, so is R_i; $R_i = 0$ indicates that X_i is missing, $R_i = 1$ indicates that X_i is observed. A CCA would include only effects i for which X_i is observed (i.e., $R_i = 1$). The complete-case estimator for β_0 is given by a weighted sum of T_i among the effects for which $X_i = 0$ and $R_i = 1$:

$$\hat{\beta}_{0C} = \frac{\sum_{i:X_i=0, R_i=1} w_i T_i}{\sum_{i:X_i=0, R_i=1} w_i} \tag{18}$$

The complete-case estimator for β_1 is given by the difference between the (weighted) mean effect for $X_i = 1$ versus $X_i = 0$:

$$\hat{\beta}_{1C} = \frac{\sum_{i:X_i=1, R_i=1} w_i T_i}{\sum_{i:X_i=1, R_i=1} w_i} - \hat{\beta}_{0C} \tag{19}$$

13

Assume that the selection model is log-linear, and that for the sake of simplicity the probability of observing X_i depends on the size of the effect T_i and the value of X_i:

$$\text{logit}[p(R_i = 1|T_i, X_i, v_i)] = \psi_0 + \psi_1 T_i + \psi_2 X_i \tag{20}$$

Note that this is an MNAR mechanism, since the probability X_i is observed depends on X_i itself; a MAR mechanism would involve $\psi_2 = 0$ in equation (20). Because (20) depends on T_i, $\delta_{ij} \neq 0$ for this selection model regardless of MAR or MNAR (i.e., regardless of whether $\psi_2 = 0$ or not), the CCA estimators may be biased.

Under this model, $H_1(X_i\beta, X_i, v_i)$ depends only on X_i and not v_i, so we can write $H_1(X_i) = p(R \neq \mathbb{1}|T_i, X_i, v_i)|_{T_i = X_i^T \beta}$. As well, $f_{11}(T_i, X_i, v_i) = T_i$ and $f_{21}(T_i, X_i, v_i) = X_i$. Given the result in equation (13), we can write

$$\delta_{i1} \approx H_1(X_i)(v_i + \tau^2)\psi_1 \tag{21}$$

Given the selection model in (20), the bias of the complete-case estimator for the intercept, $\hat{\beta}_{0C}$, is:

$$\text{Bias}[\hat{\beta}_{0C}] \approx H_1(0)(\bar{v}_0 + \tau^2)\psi_1 \tag{22}$$

where $\bar{v}_0$ is the average estimation error variance v_i among effects for which $X_i = 0$ and $R_i = 1$. The expression in (22) depends on three key quantities, and is an increasing function of each of those quantities. First, the bias increases in $H_1(0)$, which is an approximation of the probability that X_i is missing among studies for which $X_i = 0$. While under model (20), this probability is a function of T_i and X_i, we can intuit $H_1(0)$ loosely as a missingness rate in X_i among effects for which $X_i = 0$. Second, the bias in (22) is increasing in $\bar{v}_0 + \tau^2$, the average variation of T_i for which $X_i = 0$; the greater the variation, the greater the bias. Because the v_i are typically decreasing in sample size, if studies have smaller samples, the bias will be greater. Finally, the bias depends on ψ_1, which characterizes the relationship between an X_i being observed (i.e., R_i) and T_i. When ψ_1 is positive, larger effect estimates T_i are more likely to have observed X_i and the bias will be positive; if ψ_1 is negative, so that larger effect sizes are more likely to be missing the covariate X_i, then the bias will be negative.

To gain better insight into equation (22), suppose $v_i \approx v = \bar{v}_0$ so that each study has roughly the same estimation error variance. If we assume T_i is on the scale of a standardized mean difference, $v_i \approx 4/n_i$ where n_i is the total sample size used to compute T_i. Various researchers have described conventions for the magnitude of τ^2 that range from $\tau^2 = v/4$ to $\tau^2 = v$.[37–39] Thus, we can write $\tau^2 + v = 4(1 + r)/n$ from some constant r that ranges from 0 to 1.

Further, the parameter ψ_1 is a log-odds ratio, which reflects the odds of a complete case for T_i versus $T_i - 1$. There are various conventions for the size of an odds ratio that depend on base rates $P[R = \mathbb{1}|T]$ that could be interpreted as ranging from 1.5 to as large as 9.0, though various researchers have noted that odds ratios greater than 3.0 or 4.0 could be considered large.[40–43] Thus, we consider a range of odds ratios from about 1.5 to 4.5. However, the actual size of ψ_1 will depend on the scale of a change in effect size $D_T = |T_i - \tilde{T}_i|$. Since it corresponds to a difference, D_T should be no larger than an individual $|T_i|$. Based on conventions in the social and medical sciences (some arbitrary, some empirical), meaningful values of D_T might feasibly range from 0.2 to 1.0.[40,44] These conventions for odds ratios and D_T would imply that relevant values of $|\psi_1|$ might range from 0.4 (large D_T with small odds ratio) to over 7.5 (small D_T with large odds ratio).

Based on these conventions, Figure 1 shows the potential (approximate) bias of $\hat{\beta}_{0C}$ for this example. Each panel corresponds to a given within-study variance $v = 4/n$ and residual heterogeneity τ^2. Panels plot the bias contributed by a single case δ_i as a function of the probability of missingness $H_1(0)$ (x-axis) and ψ_1 (color). The panels on the bottom few rows and left most columns show that if both ψ_1 is small and $\tau^2 + v$ is small, then δ_i will be less than 0.05. However if $\tau^2 + v_i$ is larger and the probability of a complete case is strongly related to T_i (i.e., ψ_1 is large), then the bias can be greater than $d = 0.2$ or even 0.5.

It is worth noting that Figure 1 gives the bias for when T_i is positively correlated with R_i, and hence $\psi_1 > 0$. When $\psi_1 < 0$, then the bias of $\hat{\beta}_{0C}$ is negative, and would be be a mirror image of those in Figure 1. Larger, more negative values of ψ_1 would lead to a greater downward bias.

[Figure 1 about here.]

The bias of the slope coefficient, $\hat{\beta}_{1C}$, under selection model (20) is given by:

$$\text{Bias}[\hat{\beta}_{1C}] \approx \left[H_1(1)(\bar{v}_1 + \tau^2) - H_1(0)(\bar{v}_0 + \tau^2) \right] \psi_1 \tag{23}$$

where $\bar{v}_1$ is the mean v_i among effects for which $X_i = 1$ and $R_i = 1$. As with $\hat{\beta}_{0C}$, the bias of $\hat{\beta}_{1C}$ is an increasing function of ψ_1. If T_i has a strong positive correlation with R_i, then ψ_1 will be larger and so will the bias of $\hat{\beta}_{1C}$.

When all studies have approximately the same estimation error variance so that $v_i \approx v$ and $\bar{v}_0 \approx \bar{v}_1$, then the bias of $\hat{\beta}_{1C}$ is approximately:

$$\text{Bias}[\hat{\beta}_{1C}] \approx [H_1(1) - H_1(0)] (v + \tau^2)\psi_1 \tag{24}$$

15

The expression in (24) is similar to (22), and both expressions depend on similar quantities. Like $\hat{\beta}_{0C}$, the bias of $\hat{\beta}_{1C}$ is an increasing function of $\tau^2 + v$ and ψ_1. The bias of $\hat{\beta}_{1C}$ also increases as a function of $H_1(1) - H_1(0)$, which can be thought of as a difference in missingness rates between cases where $X_i = 1$ and $X_i = 0$. Note, however, that this does not imply that MAR data necessarily leads to an unbiased slope estimate. Recall that H_1 is an approximation of the probability X_i is missing given X_i and T_i in (20): $P[R_i \neq \mathbb{1}|T_i, X_i]$. Even if X_i were MAR (i.e., assuming $\psi_1 \neq 0$ and $\psi_2 = 0$), unbiased slope estimate would further require that the slope be zero: $\beta_1 = 0$. This is because when $\beta_1 \neq 0$, we would expect different rates of missingness among studies for which $X_i = 1$ than $X_i = 0$ because of the relationship between R_i and T_i, as well as the relationship between T_i and X_i. Viewed this way, the bias of $\hat{\beta}_{1C}$ will be greatest when there are fewer complete cases, missingness is strongly related the value of the covariate X_i or to the size of effects (assuming that effects are correlated with X_i).

To gain insight into the magnitude of bias in (24), consider the values of $\psi_1 \in [0.4, 7.5]$ and $\tau^2 + v = 4(1 + r)/n$ discussed above. Note that the difference $H_1(1) - H_1(0) = p(R = 0|X = 1, \eta) - p(R = 0|X = 0, \eta)$ is a difference in conditional probabilities. For reference, because both R_i and X_i are binary, then $p(R = 0|X = 1) - p(R = 0|X = 0)$ would be equal to the correlation between R_i and X_i (assuming equal marginals in a 2×2 table). Thus, $|p(R = 0|X = 1) - p(R = 0|X = 0)|$ could be as small as 0, but could possibly be as large as 1; arbitrary conventions on the size of correlations suggest that $|p(R = 0|X = 1) - p(R = 0|X = 0)| = 0.5$ would be a "large" value.[40]

Figure 2 shows the potential bias of $\hat{\beta}_{1C}$ for this example assuming the values of $\tau^2 + v$, ψ_1, and $H_1(1) - H_1(0)$ discussed above. Each panel corresponds to a given amount of heterogeneity $\tau^2 + v$, and within panels the bias is shown as a function of the difference $H_1(1) - H_1(0)$ (x-axis) and ψ_1 (color). Figure 2 highlights that the relationship between R_i and T_i (ψ_1) and between R_i and X_i (x-axes) can affect the magnitude of the bias. If R_i is strongly correlated with both X_i and T_i the bias can be as large as $d = 0.3$ or 0.4. However, the less R_i depends on T_i or X_i, the lower the bias is.

[Figure 2 about here.]

Recall that the mechanism in these computations is assumed to be MNAR, since ψ_2 in (20) is nonzero. A MAR mechanism would require $\psi_2 = 0$. In that case, the bias for the CCA intercept estimator $\hat{\beta}_{0C}$ is identical to that given in (22). However, the bias in the slope will be slightly different when $\psi_2 = 0$. This is because, as noted noted in (24), the bias in the slope depends (loosely) on the correlation between R and X. Given the form of $H_1(X)$ in this example, it is possible for the bias of $\hat{\beta}_{1C}$ to be greater when $\psi_2 \neq 0$ (MNAR) than when $\psi_2 = 0$ (MAR), which can occur if the correlation between R and T, and R and X are in the same direction (i.e., ψ_1 and ψ_2 are in the same direction). However, when

16

$\psi_2 \neq 0$ (MNAR), the bias of $\hat{\beta}_{1C}$ can also decrease in the magnitude relative to when $\psi_2 = 0$ if ψ_1 and ψ_2 are in the opposite directions.

A key implication of this example is that under the relatively simple selection model in (20), CCA intercept estimators can have substantial bias. This bias does not change even if $\psi_2 = 0$ and the data are MAR. Thus, inferences for the group of studies for which $X_i = 0$ will be biased. Moreover, because inference for the group of studies for which $X_i = 1$ will depend on the intercept estimate, those inferences will also be biased even if the slope estimator $\hat{\beta}_{1C}$ is unbiased.

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, SCA can be thought of as a set of regression models. Consider one model from that set, which estimates regression coefficients for some subset S of the relevant covariates using the estimator $\hat{\beta}_S$ in equation (8). Recall that E refers to the set of covariates omitted from the model, and that the estimator $\hat{\beta}_S$ conditions on a set of missingness patterns $R_i \in \mathcal{R}_j$. The set of missingness patterns $\mathcal{R}_j$ is such that $R_{iS} = 1$ so that all included covariates are observed.

To understand the conditions under which $\hat{\beta}_S$ is unbiased, we can write a shifting-case model as:

$$p(T_i|X_{iS}, v_i, R_i \in \mathcal{R}_j, \eta, \psi) = \frac{p(R_i \in \mathcal{R}_j|T_i, X_{iS}, v_i, \psi)p(T_i|X_{iS}, v_i, \eta)}{p(R_i \in \mathcal{R}_j|X_{iS}, v_i, \eta, \psi)} \tag{25}$$

The model in (25) 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 X_{iS} rather than the complete set of relevant covariates X_i. Thus, the function $p(T_i|X_{iS}, v_i)$ can be thought of as a partial-data model, since it omits some of the relevant covariates. The partial-data model $p(T_i|X_{iS}, v_i)$ need not be equivalent to the complete-data model $p(T_i|X_i, v_i)$ because the former conditions only on X_{iS} and not the full set of covariates X_i. These models would only be equivalent if $T_i \perp X_{iE}|X_{iS}, v_i$. That is, unless the excluded covariates are completely unrelated to effect size (given the covariates included in the SCA model), then $\hat{\beta}_S$ will be biased even if X_{iS} are completely observed.

The model in (25) suggests a very strict set of conditions for which $\hat{\beta}_S$ is unbiased which concern the missingness mechanism and the relevance of excluded covariates in a given shifting-case regression. First, missingness must be independent of effect sizes. This arises if $R_i \perp T_i|X_{iS}, v_i$ or $R_i \perp (T_i, X_{iS})|v_i$,

which is a similar assumption as that made for unbiased CCA. In effect, this assumption implies that missingness is independent of effect sizes T_i (and potentially covariates), but could be correlated with estimation error variances v_i.

Second, any excluded covariates must be completely irrelevant to effect sizes given the included covariates: $T_i \perp X_{iE}|X_{iS}, v_i$. This assumption is equivalent to assuming that $\beta_j = 0$ for all $j \in E$, so that any omitted variables in a given shifting-case regression are assumed to have a coefficient of zero. A related assumption is that $(T_i, X_{iS}) \perp X_{iE}|v_i$, which would imply that the complete-data likelihood involves no interactions between X_{iS} and X_{iE} and that X_{iS} and X_{iE} are orthogonal. Given the nature of many meta-analyses wherein included studies and effects are ostensibly "found objects," correlation among multiple covariates is a common issue in meta-regression.[25] Note that conditions on omitted covariates *and* omitted observations must hold in order for $\hat{\beta}_S$ to be unbiased.

When the assumptions about omitted variables and effect sizes are not met, $\hat{\beta}_S$ will be biased. The magnitude of the bias 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:

$$\text{Bias}[\hat{\beta}_S] = (\mathbf{X}_{US}^T\mathbf{W}_U\mathbf{X}_{US})^{-1}\mathbf{X}_{US}^T\mathbf{W}_U\mathbf{X}_{UE}\beta_E + (\mathbf{X}_{US}^T\mathbf{W}_U\mathbf{X}_{US})^{-1}\mathbf{X}_{US}^T\mathbf{W}_U\Delta_{jU}, \qquad (26)$$

where $\mathbf{X}_{UE}$ is the matrix of omitted covariates and β_E comprises the coefficients for the omitted covariates. The term Δ_j is a vector of biases due to missingness $\Delta_j = [\delta_{1j}, \ldots, \delta_{kj}]$ and Δ_{jU} is the subset of Δ_j for which $R_i \in \mathcal{R}_j$. Note that the δ_{ij} are the biases due solely to missingness as in equation (11).

The expression in (26) shows that a shifting-case analysis suffers from two sources of bias. The first source, captured in the first term in (26), is a function of the coefficients for the excluded covariates β_E, which we refer to as as omitted variable bias. Discussion in a previous section argued that the term *omitted variable bias* assumes that β is the target of inference in an SCA, which may or may not be the case. If β is the target of inference, omitted variable bias arises even if no X_{iS} are missing, and is related to the issue of multicollinearity in linear models. In fact, if the columns in $\mathbf{X}_{US}$ and $\mathbf{X}_{UE}$ 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 in the complete-data model $\mathbf{X}_{UE}\beta_E$. The estimator $\hat{\beta}_S$ will have greater bias if the coefficients for the omitted variables β_E are larger and the omitted covariates $\mathbf{X}_{UE}$ are correlated with the included covariates $\mathbf{X}_{US}$.

The second term in (26) captures the bias due to ignoring missing observations in X_{iS}. This *missingness bias* is a function of Δ_{jU}, which is itself a vector of biases for each effect, and it can be understood

in terms of its individual components δ_{ij}. Because the δ_{ij} are of the same form for the complete-case and shifting-case models, the missing data bias for a SCA is governed by similar factors as the CCA, and are quite possibly similar in magnitude. Based on (13), δ_{ij} will be positive if T_i is strongly correlated with whether $R_i \in \mathcal{R}_j$, and δ_{ij} will be greater in magnitude when that correlation is larger.

Taken together, shifting-case estimators can be even more biased than complete-case estimators. This occurs if the omitted variable and the missingness biases are in the same direction (e.g., both are positive). For both biases to be in the same direction, correlation between T_i and the omitted variables X_{iE} must be in the same direction as the correlation between T_i the probability that X_{iS} 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 $X_i = [1, X_{i1}, X_{i2}]$ and X_{i1} and X_{i2} are binary covariates such that

$$T_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + u_i + e_i \tag{27}$$

If there is missingness in both X_{i1} and X_{i2}, then $R_i \in \{0,1\}^2$ so that $R_i = [1,1]$ indicates both covariates are observed, and $R_i = [1,0]$ indicates only X_{i1} is observed. If missingness is such that $R_i = [1,1]$ for very few effect estimates, then a SCA might involve regressing T_i on the observed values of X_{i1} and then on the observed values of X_{i2}.

The first regression would take only rows for which X_{i1} is observed, so that $R \in \mathcal{R}_1 = \{[1,1],[1,0]\}$ and the excluded X_{i2} could be either 0 or 1. The shifting-case estimators follow from equation (8):

$$\hat{\beta}_{0S} = \frac{\sum_{X_1=0} w_i T_i}{\sum_{X_1=0} w_i}, \qquad \hat{\beta}_{1S} = \frac{\sum_{X_1=1} w_i T_i}{\sum_{X_1=1} w_i} - \hat{\beta}_{0S}$$

Assume that missingness follows the following log-linear model:

$$\mathrm{logit}[p(R_i \in \mathcal{R}_1 | T_i, X_{i1}, v_i)] = \psi_0 + \psi_1 T_i + \psi_2 X_{i1} \tag{28}$$

Note that this gives the log-odds that an effect is included in the model given T_i and X_{i1}, and that X_{i2} is not involved. Further, because the distribution of R_i depends on X_{i1}, the mechanism is MNAR.

Given the selection model in (28), the bias of the coefficient estimators can be written as:

$$\text{Bias}[\hat{\beta}_{0S}] = \beta_2 \frac{\sum_{X_1=0,X_2=1} w_i}{\sum_{X_i=0} w_i} + \frac{\sum_{X_{i1}=0} w_i \tilde{\delta}_{i1}}{\sum_{X_{i1}=0} w_i} \tag{29}$$

$$\text{Bias}[\hat{\beta}_{1S}] = \beta_2 \left(\frac{\sum_{X_{i1}=1,X_{i2}=1} w_i}{\sum_{X_{i1}=1} w_i} - \frac{\sum_{X_{i1}=0,X_{i2}=1} w_i}{\sum_{X_{i1}=0} w_i} \right) + \left(\frac{\sum_{X_{i1}=1} w_i \tilde{\delta}_{i1}}{\sum_{X_{i1}=1} w_i} - \frac{\sum_{X_{i1}=0} w_i \tilde{\delta}_{i1}}{\sum_{X_{i1}=0} w_i} \right) \tag{30}$$

Here $\tilde{\delta}_{i1}$ are the missingness biases as defined above, and whose approximate values is given in (13). To distinguish from the δ_{i1} from the complete-case example, we use the $\tilde{\delta}$ notation.

Both the bias of $\hat{\beta}_{0S}$ and $\hat{\beta}_{1S}$ 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 $w_i \approx w$, then the omitted variable biases reduce to

$$\text{Omitted Var. Bias}[\hat{\beta}_{0S}] = \beta_2 p(X_2 = 1|X_1 = 0) \tag{31}$$

$$\text{Omitted Var. Bias}[\hat{\beta}_{1S}] = \beta_2 \left[p(X_2 = 1|X_1 = 1) - p(X_2 = 1|X_1 = 0) \right] \tag{32}$$

The omitted variable biases for each coefficient can be seen as depending on two quantities. Both (31) and (32) are increasing in β_2, which is the contribution of X_{i2} to the complete-data model. The omitted variable bias for $\hat{\beta}_{0S}$ is also increasing in $p(X_2 = 1|X_1 = 0)$. The bias for $\hat{\beta}_{1S}$ in (32) is increasing in $p(X_2 = 1|X_1 = 1) - p(X_2 = 1|X_1 = 0)$. Because both X_{i1} and X_{i2} are binary, this difference is roughly equivalent to their Pearson correlation (assuming equal marginals). If $X_{i1} \perp X_{i2}$, then their correlation is zero, and the omitted variable bias will be zero. But if X_{i1} and X_{i2} are correlated, the bias of $\hat{\beta}_1$ will depend on how strongly correlated X_{i1} and X_{i2} are, and how big β_2 is.

Figure 3 shows the omitted variable bias of $\hat{\beta}_0$ (left plot) and $\hat{\beta}_1$ (right plot) as a function of β_2. Both the bias and β_2 are shown on the scale of Cohen's d. In the left plot $\pi_{01} = p(X_{i2} = 1|X_{i1} = 0)$ is the proportion of $X_{i2} = 1$ when $X_{i1} = 0$. In the right plot, $\rho_{12} = p(X_{i2} = 1|X_{i1} = 1) - p(X_{i2} = 1|X_{i1} = 0)$, which is roughly the correlation between X_{i1} and X_{i2}. Note that because ρ_{12} 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).[40]

The figure shows that if $\beta_2 = 0$ so that X_{i2} is independent of T_i given X_{i1}, that both $\hat{\beta}_{0S}$ and $\hat{\beta}_{1S}$ will be unbiased. However, when β_2 is nonzero, both estimators will be biased. If X_{i1} and X_{i2} are highly correlated, or if $X_{i2} = 1$ when $X_{i1} = 0$ with high probability, the bias of both estimators will about as large as a "small" effect (i.e., $d = 0.2$) when β_2 is larger than 0.2. For $\hat{\beta}_{1S}$ the bias will be less than about

$d = 0.05$ when $|\beta_2| \le 0.1$ or if $\rho_{12} < 0.5$.

[Figure 3 about here.]

Figure 3 does not take into account any bias induced by missingness. However, because the missingness mechanism in (28) is the same as the mechanism for the complete-case example (20), the missingness bias for $\hat{\beta}_{0S}$ is the same as that for $\hat{\beta}_{0C}$ in (22), which is shown in Figure 1. Likewise, the missingness bias for $\hat{\beta}_{1S}$ is the same as that for $\hat{\beta}_{1C}$ in (23), which is shown in Figure 2.

Thus, the total bias of $\hat{\beta}_{0S}$ will be the sum of the omitted variable biases shown in Figure 3 and the missingness biases shown in Figures 1 and 2. If both the omitted and missingness biases are on the higher end, the total bias of $\hat{\beta}_0$ might be as large as $d = 0.6$ to over 1.0. Likewise, the total bias of $\hat{\beta}_{1S}$ will be the sum of the omitted variable biases shown in Figure 2 and the missingness biases shown in Figure 3, and can be larger than $d = 0.6$.

As noted above, the missingness bias and omitted variable bias can be in the different directions. For instance, if $\beta_2 < 0$ but $\tilde{\delta}_{ij} > 0$, then the omitted variable bias for $\hat{\beta}_{0S}$ 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.

7   Implications for Empirical Example

The theoretical results above suggest that there are conditions under which the coefficient estimates from the CCA and SCA of the substance abuse data in Table 2 are substantially biased. However, it will be difficult, if not impossible, to determine just how biased those estimates are, even given the simplified examples in the previous sections. First, the missingness mechanism is not known for the substance abuse data. Even if we assume that the mechanism follows a log-linear model like that in (20) or (28), the resulting formulas for the bias depend on quantities, such as ψ and η that are not known, and cannot be estimated in the presence of missing data without further assumptions.

However, one approach to examining bias in the estimates presented in Table 2 would involve stochastically imputing the missing X_{ij} in the data. In the same vein as multiple imputation (MI), each set of imputed values constitutes a "complete" dataset from which we can compute the parameters relevant to bias.[13,45] Given an imputed dataset, we can compute (a) the difference in the resulting $\hat{\beta}^{(i)}$ for the ith imputed dataset and $\hat{\beta}_S$, (b) the quantities that govern bias in the formulas above, including ψ, $H(X)$,

and τ^2. This allows us not only to assess the bias, but also to examine which aspects of the missing data are driving it.

As with MI, the accuracy of the resulting estimated quantities depends on the validity of assumptions regarding missingness and the accuracy of the imputation model. Thus, we would urge interpretation of the following results as *potential* biases in the CCA and SCA estimators presented earlier in this article, rather than a precise estimate of the bias. We generated $m = 1,000$ imputations using the `mice` software in the `R` programming language.[46] Estimates of η were computed using `metafor`, specifying a Paule-Mandel estimator for the variance component τ^2.[47] To estimate the log-linear model selection parameters ψ in (28), as well as $H(X)$, we used a logistic regression with the missingness indicator R_{ij} and T_i and X_{ij} as the predictors.

Here, we focus on results for β_0 and β_1. Consider the regression of T_i on X_{i1} reported in Table 2. We can view this as a single regression in a SCA that includes only observations for which X_{i1} is observed. As noted above, the resulting estimators of the intercept β_0 and slope β_1 will exhibit bias due to missingness given in (22) and (23) and bias due to omitting variables as in (31) and (32). Recall that the bias due to missingness in an SCA under this model will be similar to the bias derived for a CCA.

Figure 4 plots the omitted variable bias, missingness bias, and total bias for both $\hat{\beta}_{0S}$ and $\hat{\beta}_{1S}$. Results are reported on the scale of Cohen's d. Omitted variable biases for $\hat{\beta}_{0S}$ range from -0.05 to 0.02 with a mean of -0.01; omitted variable bias for $\hat{\beta}_{1S}$ ranges from -0.31 to 0.11 with a mean of -0.05. Similarly the bias due to missingness could feasibly range from 0.04 to 0.08 with a mean of 0.06 for $\hat{\beta}_{0S}$, while the missingness bias of $\hat{\beta}_{1S}$ might range from 0 to 0.12 with a mean of 0.06. Note that while the omitted variable bias and missingness bias are in opposite directions in this example, this need not be the case in general; both biases could feasibly be in the same direction for other data. In sum, this amounts to a total bias of -0.01 to 0.10 for $\hat{\beta}_{0S}$ and from -0.31 to 0.23 for $\hat{\beta}_{1S}$.

[Figure 4 about here.]

8   Discussion

This article described a selection model approach to study 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 CCA and SCA will produce biased coefficient estimates unless certain con-

ditions are met. While discussion regarding potential bias of these analyses have largely focused on traditional mechanism taxonomy of MCAR, MAR, and MNAR, we found that bias depends more on the precise model for missingness rather than these broader classifications. Certain mechanisms that are MAR or MNAR can lead to unbiased estimates with CCA and SCA, while other MAR or MNAR mechanisms can induce substantial bias. Complete-case estimators are unbiased if the probability that all relevant covariates are observed is (conditionally) independent of the effect size estimate. Shifting-case 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 $d = 0.4$ to $d = 0.8$—depending on the missingness mechanism (i.e., parameters in the selection model), the missingness rate, an the relevance of any omitted covariates.

Results for both CCA and SCA suggest that bias due to missingness will tend to increase in magnitude as a function of the total variation in the data. This means that if studies have small sample sizes (i.e., v_i are large) or there is substantial residual between-effect heterogeneity τ^2, the bias of a CCA or SCA will be greater. Because meta-regression is used to explain between-effect variation τ^2, models capable of explaining much of that variation will have lower bias in CCA and SCA estimates. However, even very modest amounts of residual variation can still result in substantial bias.

An important aspect of these findings 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 $d = \pm 0.1$ 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 reporting, data collection, and data curation.

In addition, 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.[48] Further, the issues of multicollinearity and confounding in meta-regression, including those discussed by Lipsey 25, would suggest that omitting variables in a SCA 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.[48] If there is very little missingness, or if there is a good reason to assume that missingness is uncorrelated with effect size estimates, a CCA may be a reasonable option. However, we would discourage analysts from continuing to use SCA 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[32] describes an EM algorithm for generalized linear models with missing covariates, and Ibrahim, lipsitz, and Chen[33] extend that algorithm when covariates are MNAR. In addition, full-information maximum likelihood (FIML) has long been used in linear models,[14,49] and has shown some promise for meta-regression involving continuous covariates. Finally, multiple imputation has become something of a standard approach for handling missing data across a number of fields.[13,30,45]

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,[50] 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.

9 Highlights

Missing covariates are a common problem when conducting meta-regressions. A common practice for meta-regression analyses has been to ignore effects for which covariates are missing. However, a vast statistical literature suggests that analyses that ignore missing data can only provide accurate estimates of relevant quantitites under certain conditions. In this article, we examine conditions under which ignoring missing covariates in a meta-regression can still lead to unbiased estimation of regression coefficients. We also investigate the possible magnitude and sources of bias when those conditions do not hold. Our findings highlight that substantial bias can be induced by ignoring missing data in a meta-regression.

10 Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. Note that the key findings of this article concern statistical properties of various incomplete data estimators, while the data were used for demonstration purposes only.

A   Approximate Bias for Log-Linear Selection Models

Proposition: Suppose $p(T_i|X_i, v_i)$ is the standard fixed- or random effects meta-regression model in equation (2), and suppose $p(R_i \notin \mathcal{R}_j|T_i, X_i, v_i) = H_j(T_i, X_i, v_i)$ follows the log-linear model in (12). Then:

$$E[T_i|X_i, v_i, R_i \in \mathcal{R}_j] \approx X_i\beta + H_j(X_i\beta, X_i, v_i)(\tau^2 + v_i)\sum_{m=0}^{m_j} \psi_{mj} f'_{mj}(X_i\beta, X_i, v_i) \tag{33}$$

where $f'_{mj}(X_i\beta, X_i, v_i) = \frac{\partial}{\partial T_i} f_{mj}(T_i, X_i, v_i)|_{T_i = X_i^T\beta}$. Therefore, the bias of the conditional expectation is given by:

$$\delta_{ij} \approx H_j(X_i\beta, X_i, v_i)(\tau^2 + v_i)\sum_{m=0}^{m_j} \psi_{mj} f'_{mj}(X_i\beta, X_i, v_i) \tag{34}$$

Proof:

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

$$H_j(X\beta, X, v) \equiv H_j(X, v) = P[R = \notin \mathcal{R}_j|T, X, v]|_{T=X\beta}$$

$$G_j(X, v) = 1 - H_j(X, v)$$

$$g_j(X, v) = P[R \in \mathcal{R}_j|X, v]$$

Then an approximation for $E[T|X, v, R \in \mathcal{R}_j]$ is as follows:

omitting subscripts, Taylor series for exponents in $P[R \in \mathcal{R}_j|T, X, v]$ at $T = X\beta$

$$E[T|X, v, R \in \mathcal{R}_j] = \int \frac{T\exp\left\{-\frac{(T-X\beta)^2}{2(\tau^2+v)} + \sum_i \psi_{ij} f_{ij}(T, X, v)\right\}}{g_j(X, v)\sqrt{2\pi(\tau^2+v)}\left(1 + e^{\sum_i \psi_{ij} f_{ij}(T, X, v)}\right)} dT$$

$$= \int \frac{T}{g_j(X, v)\sqrt{2\pi(\tau^2+v)}} \exp\left\{-\frac{(T-X\beta)^2}{2(\tau^2+v)} + \sum_i \psi_{ij} f_{ij}(X\beta, X, v)\right.$$

$$+ \sum_i \psi_{ij} f'_{ij}(X\beta, X, v)(T - X\beta) - \log\left(1 + e^{\sum_i \psi_{ij} f_{ij}(X\beta, X, v)}\right)$$

$$\left. - G_j(X, v)(\sum_i \psi_{ij} f'_{ij}(X\beta, X, v))(T - X\beta) + O(T^2)\right\} dT$$

$$\approx \int \frac{T}{g_j(X, v)\sqrt{2\pi(\tau^2+v)}} \exp\left\{-\frac{1}{2(\tau^2+v)}\left(T^2 - 2TX\beta - 2T(\tau^2+v)\sum_i \psi_{ij} f'_{ij}(X\beta, X, v)\right.\right.$$

$$\left.\left. + 2T(\tau^2+v)G_j(X, v)(\sum_{i\in\mathcal{D}} \psi_{ij} f'_{ij}(X\beta, X, v)) + \dots\right)\right\} dT$$

$$= X\beta + H_j(X, v)(\tau^2+v)\sum_i \psi_{ij} f'_{ij}(X\beta, X, v) \qquad\blacksquare$$

Note that this uses a first order Taylor expansion of the log-linear model at $T = X\beta$, and thus assumes the f_{mj} are differentiable. The approximation will be more accurate if $\tau^2 + v_i$ are small. A more accurate approximation is possible if the f_{mj} are linear in T_i. In that case, only an approximation of the denominator of the log-linear model is required.

B References

1. Borenstein M. *Introduction to Meta-Analysis.* Chichester, U.K.: John Wiley & Sons; 2009. http:// public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=427912. Accessed May 18, 2020.

2. Tipton E, Pustejovsky JE, Ahmadi H. A history of meta-regression: Technical, conceptual, and practical developments between 1974 and 2018. *Res Syn Meth.* 2019;10(2):161-179. doi:10.1002/jrsm.1338

3. Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random-effects regression model for meta-analysis. *Statist Med.* 1995;14(4):395-411. doi:10.1002/sim.4780140406

4. Hedges LV. Combining independent estimators in research synthesis. *British Journal of Mathematical and Statistical Psychology.* 1983;36(1):123-131. doi:10.1111/j.2044-8317.1983.tb00768.x

5. Hedges LV, Tipton E, Johnson MC. Robust variance estimation in meta-regression with dependent effect size estimates. *Res Synth Method.* 2010;1(1):39-65. doi:10.1002/jrsm.5

6. Konstantopoulos S. Fixed effects and variance components estimation in three-level meta-analysis: Three-level meta-analysis. *Res Syn Meth.* 2011;2(1):61-76. doi:10.1002/jrsm.35

7. Viechtbauer W. Accounting for heterogeneity via random-effects models and moderator analyses in meta-analysis. *Zeitschrift für Psychologie / Journal of Psychology.* 2007;215(2):104-121. doi:10.1027/0044-3409.215.2.104

8. Pigott TD. A review of methods for missing data. *Educational Research and Evaluation.* 2001;7(4):353-383. doi:10.1076/edre.7.4.353.8937

9. Pigott TD. Missing predictors in models of effect size. *Eval Health Prof.* 2001;24(3):277-307. doi:10.1177/01632780122034920

10. Tipton E, Pustejovsky JE, Ahmadi H. Current practices in meta-regression in psychology, education, and medicine. *Res Syn Meth.* 2019;10(2):180-194. doi:10.1002/jrsm.1339

11. Pigott TD. 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; 2019.

12. Cooper HM. *Research Synthesis and Meta-Analysis: A Step-by-Step Approach.* Fifth Edition. Los Angeles: SAGE; 2017.

13. Little RJA, Rubin DB. *Statistical Analysis with Missing Data.* Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2002. http://doi.wiley.com/10.1002/9781119013563. Accessed May 5, 2020.

14. Graham JW. *Missing Data.* New York, NY: Springer New York; 2012. http://link.springer.com/10.1007/978-1-4614-4018-5. Accessed May 5, 2020.

15. Tanner-Smith EE, Steinka-Fry KT, Kettrey HH, Lipsey MW. Adolescent substance use treatment effectiveness: A systematic review and meta-analysis. December 2016.

16. Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res.* 1999;8(1):3-15. doi:10.1177/096228029900800102

17. Bennett DA. How can I deal with missing data in my study? *Aust N Z J Public Health.* 2001;25(5):464-469.

18. Cooper HM. *Synthesizing Research: A Guide for Literature Reviews.* 3rd ed. Thousand Oaks, Calif: Sage Publications; 1998.

19. Cooper HM, Hedges LV, Valentine JC, eds. *Handbook of Research Synthesis and Meta-Analysis.* 3rd edition. New York: Russell Sage Foundation; 2019.

20. Hedges LV, Vevea JL. Fixed- and random-effects models in meta-analysis. *Psychological Methods.* 1998;3(4):486-504. doi:10.1037/1082-989X.3.4.486

21. Hedges LV. A random effects model for effect sizes. *Psychological Bulletin.* 1983;93(2):388-395. doi:10.1037/0033-2909.93.2.388

22. Laird NM, Mosteller F. Some statistical methods for combining experimental results. *Int J Technol Assess Health Care.* 1990;6(1):5-30. doi:10.1017/S0266462300008916

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

24. Gelman A. *Bayesian Data Analysis.* Third edition. Boca Raton: CRC Press; 2014.

25. Lipsey MW. Those confounded moderators in meta-analysis: Good, bad, and ugly. *The ANNALS of the American Academy of Political and Social Science.* 2003;587(1):69-81. doi:10.1177/0002716202250791

26. Farrar DE, Glauber RR. Multicollinearity in regression analysis: The problem revisited. *The Review of Economics and Statistics.* 1967;49(1):92. doi:10.2307/1937887

27. Mela CF, Kopalle PK. The impact of collinearity on regression analysis: The asymmetric effect of neg-

ative and positive correlations. *Applied Economics*. 2002;34(6):667-677. doi:10.1080/00036840110058482

28. Angrist JD, Pischke J-S. *Mostly Harmless Econometrics*. Princeton, N.J.: Princeton University Press; 2009.

29. Rubin DB. Inference and missing data. *Biometrika*. 1976;63(3):581-592. doi:10.1093/biomet/63.3.581

30. van Buuren S. *Flexible Imputation of Missing Data, Second Edition*. 2nd ed. Second edition. | Boca Raton, Florida : CRC Press, [2019] |: Chapman and Hall/CRC; 2018. https://www.taylorfrancis.com/books/9780429492259. Accessed May 5, 2020.

31. Agresti A. *Categorical Data Analysis*. 3rd ed. Hoboken, NJ: Wiley; 2013.

32. Ibrahim JG. Incomplete data in generalized linear models. *Journal of the American Statistical Association*. 1990;85(411):765-769. doi:10.1080/01621459.1990.10474938

33. Ibrahim JG, Lipsitz SR, Chen M-H. Missing covariates in generalized linear models when the missing data mechanism is non-ignorable. *J Royal Statistical Soc B*. 1999;61(1):173-190. doi:10.1111/1467-9868.00170

34. Lipsitz S. A conditional model for incomplete covariates in parametric regression models. *Biometrika*. 1996;83(4):916-922. doi:10.1093/biomet/83.4.916

35. Glynn RJ, Laird NM. Regression estimates and missing data: Complete case analysis. *Cambridge, MA: Harvard School of Public Health, Department of Biostatistics*. 1986.

36. Little RJA. Regression with missing X's: A review. *Journal of the American Statistical Association*. 1992;87(420):1227-1237. doi:10.1080/01621459.1992.10476282

37. Hedges LV, Pigott TD. The power of statistical tests in meta-analysis. *Psychological Methods*. 2001;6(3):203-217. doi:10.1037/1082-989X.6.3.203

38. Hedges LV, Pigott TD. The power of statistical tests for moderators in meta-analysis. *Psychological Methods*. 2004;9(4):426-445. doi:10.1037/1082-989X.9.4.426

39. Hedges LV, Schauer JM. Statistical analyses for studying replication: Meta-analytic perspectives. *Psychological Methods*. 2019;24(5):557-570. doi:10.1037/met0000189

40. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, N.J: L. Erlbaum Associates; 1988.

41. Ferguson CJ. An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice.* 2009;40(5):532-538. doi:10.1037/a0015808

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

43. Haddock CK, Rindskopf D, Shadish WR. Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods.* 1998;3(3):339-353. doi:10.1037/1082-989X.3.3.339

44. Hill CJ, Bloom HS, Black AR, Lipsey MW. Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives.* 2008;2(3):172-177. doi:10.1111/j.1750-8606.2008.00061.x

45. Rubin DB. *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley; 1987.

46. van Buuren S van, Groothuis-Oudshoorn K. **mice** : Multivariate Imputation by Chained Equations in R. *J Stat Soft.* 2011;45(3). doi:10.18637/jss.v045.i03

47. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Soft.* 2010;36(3). doi:10.18637/jss.v036.i03

48. Schauer JM, Dìaz K, Pigott TD, Lee J. Exploratory analyses for missing data in meta-analyses. *Alcohol and Alcoholism.* 2021:1-12. doi:10.1093/alcalc/agaa144

49. Graham JW. Missing data analysis: Making it work in the real world. *Annu Rev Psychol.* 2009;60(1):549-576. doi:10.1146/annurev.psych.58.110405.085530

50. Cheung MW-L. Handling missing covariates in mixed-effects meta-analysis with full-information maximum likelihood. Society for Research Synthesis Methods; 2019; Chicago, IL. http://www.srsm.org/uploads/4/6/1/3/46138157/abstract_-_mike_cheung.pdf. Accessed September 8, 2020.

Table 1: *This table displays the total number and percentage of effect sizes that are missing covariates regarding whether Group 1 or Group 2 received high-intensity interventions in the substance abuse intervention meta-analysis.*

Group 1 Hi-Intensity	Group 2 Hi-Intensity	Count	Percent
Observed	Observed	37	0.50
Observed	Missing	17	0.23
Missing	Observed	4	0.05
Missing	Missing	16	0.22

Table 2: *This table displays the meta-regression results for the model regressing effect sizes on high-intensity indicator variables when using complete- and shifting-case analyses.*

Term	Complete-Case	Shifting-Case Group 1	Shifting-Case Group 2
Intercept	0.11 (SE = 0.06, p = 0.11)	0.14 (SE = 0.06, p = 0.02)	0.15 (SE = 0.06, p = 0.03)
Group 1 Hi-Int.	0.44 (SE = 0.16, p = 0.06)	0.27 (SE = 0.15, p = 0.07)	–
Group 2 Hi-Int.	-0.21 (SE = 0.26, p = 0.46)	–	0.16 (SE = 0.26, p = 0.54)
Variance Comp. τ^2	0.08	0.06	0.09
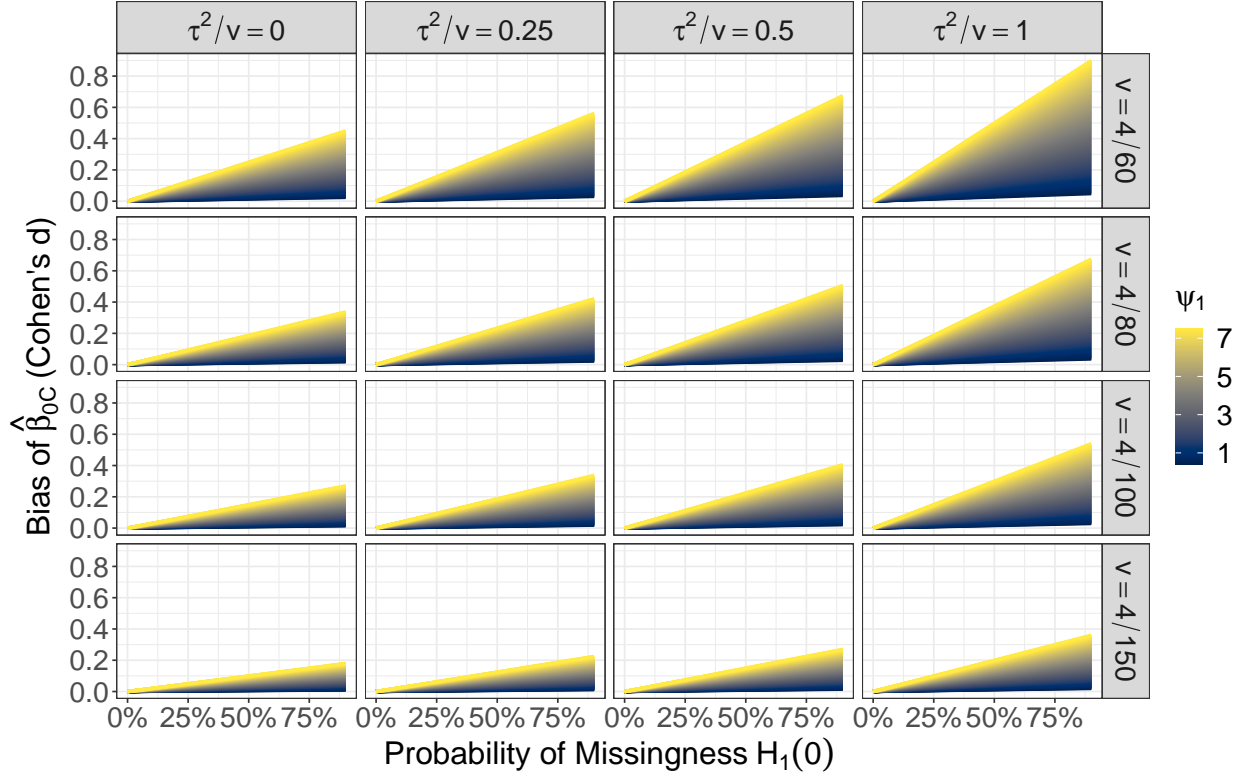
Figure 1: This figure plots the bias of the intercept estimate $\hat{\beta}_{0C}$ (y-axis) of the example. Bias is shown as a function of the average sampling variance v, residual heterogeneity τ^2, the probability of missingness when $X_1 = 0$, $H_1(0)$ (x-axis), and the correlation between missingness and the effect size as measured by ψ_1 (color). Note that ψ_1 is a log-odds ratio for effect sizes on the scale of Cohen's d.
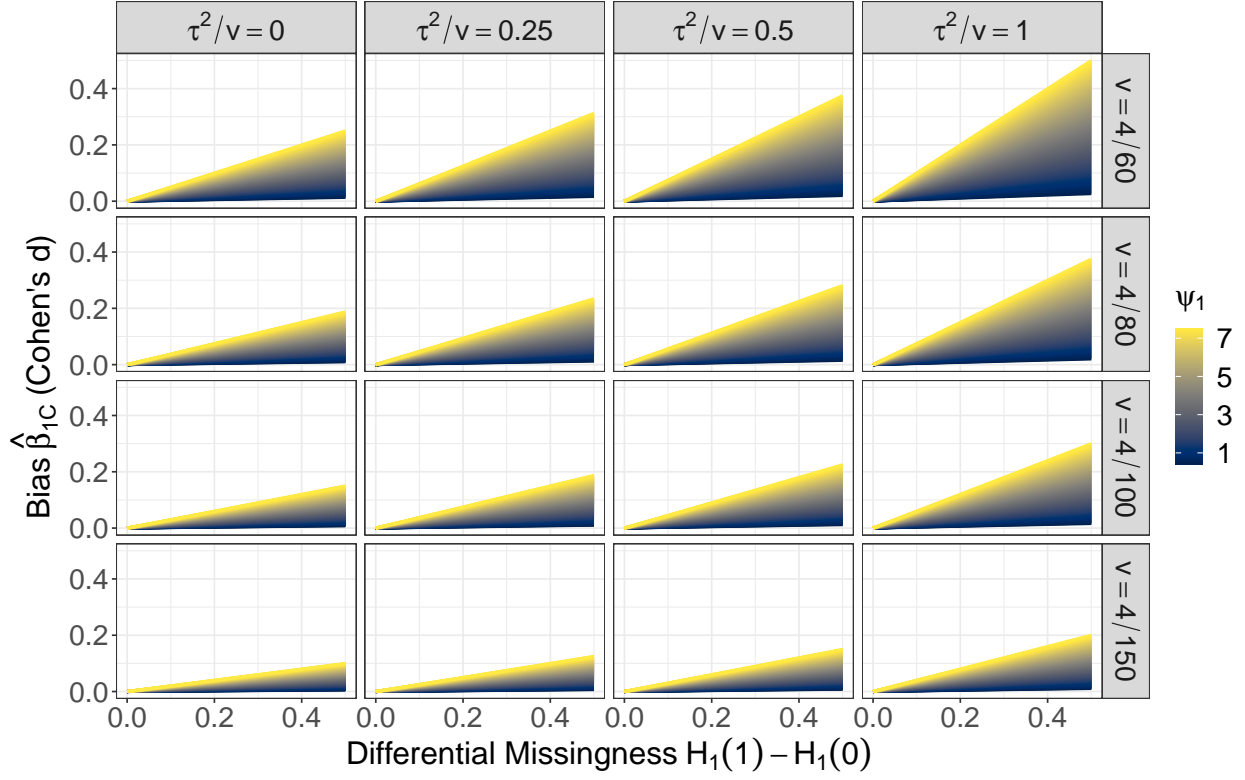
Figure 2: This figure plots the bias of $\hat{\beta}_{1C}$ (y-axis). Each panel corresponds to a given value of residual heterogeneity τ^2 and estimation error variance v. Within panels, the bias of $\hat{\beta}_{1C}$ is plotted as function of differential missingness rates ($p(R = 0|X = 1) - p(R = 0|X = 0)$), which is analogous to the correlation between the value of X and whether it is observed. Bias is also shown as a function of ψ_1 which is the relationship between the probability of observing X and the effect size T. Bias is shown on the scale of Cohen's d and ψ_1 is on the scale of a log-odds ratio.
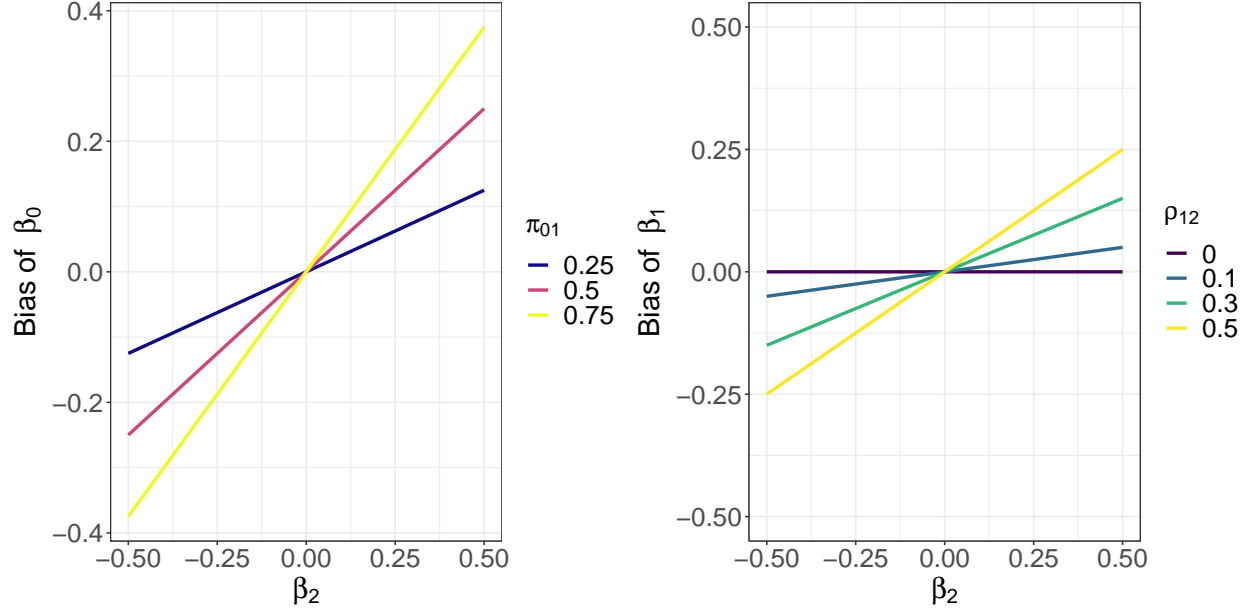
Figure 3: This figure shows the omitted variable bias of $\hat{\beta}_{0S}$ and $\hat{\beta}_{1S}$ for the model in a model (27) as a function of the omitted variable coefficient β_2. The bias (y-axis) and β_2 (x-axis) are on the scale of Cohen's d. The bias displayed is solely due to omitting X_{i2} from a model (27). In the left plot, lines are colored according to $\pi_{01} = p(X_{i2} = 1|X_{i1} = 0)$. In the right plot, lines are colored according to $\rho_{12} = p(X_{i2} = 1|X_{i1} = 1) - p(X_{i2} = 1|X_{i1} = 0)$.
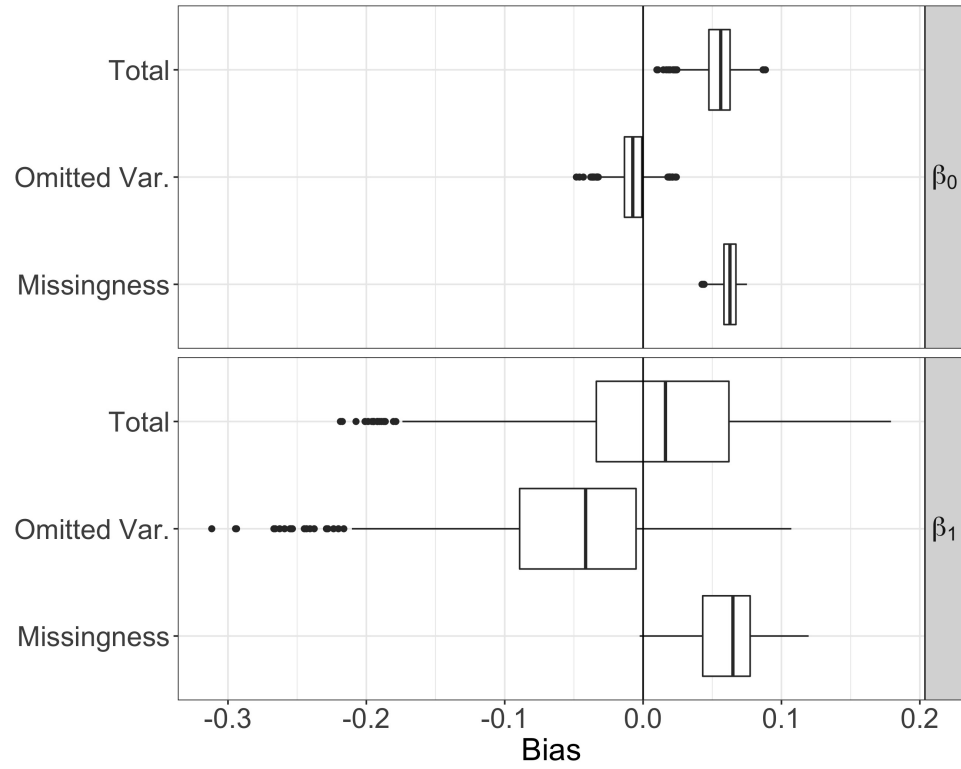
Figure 4: Bias in SCA regression of T_i on X_{i1}. For both the intercept β_0 and slope β_1, these boxplots show the total potential bias of the SCA estimators, as well as the omitted variable and missingness bias. Units are shown on the scale of Cohen's d.