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

Jacob M. Schauer, Northwestern University   Karina Dìaz, Columbia University
Jihyun Lee, University of Texas   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 ignore the fact that covariates are missing from the data. Ignoring missing data can potentially lead to biased estimates (see Little & Rubin, 2002; Graham, 2012). Some research has pointed out these issues in meta-analysis (Pigott, 2019). However, precisely how much bias can arise in a complete- or shifting-case analysis is not immediately clear, nor is there exhaustive guidance on when these analyses produce unbiased estimates.

This article examines the potential bias of complete- and available-case analyses. The following section provides a demonstration of these methods on [INSERT DATA SET]. We then introduce a statistical framework for studying bias for incomplete data meta-regressions that incorporates a model for whether or not an data point is observed. Using this framework, we describe conditions under which complete- and available-case analyses are unbiased. When these conditions are not met, we derive an approximation for the bias of complete- and available-case analyses using standard models for missingness and examine the magnitude of bias.

2 Case Study

[HOLD FOR ILLUSTRATION OF ANALYSIS METHODS WITH REAL DATA]

3 Model and Notation

Suppose a meta-analysis involves k effects estimated from 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 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 $V[e_i] = v_i$, which is true of some effect sizes, and is an accurate large-sample approximation for others (Cooper et al., 2019). The term u_i represents the random effect such that $u_i \perp e_i$ and $V[u_i] = \tau^2$.

This is the standard random effects meta-regression model, and it is also consistent with subgroup analysis models (Cooper et al., 2019; Hedges & Vevea, 1998). 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 (Hedges, 1983a; Hedges & Vevea, 1998; Laird & Mosteller, 1990; Viechtbauer, 2005):

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

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:

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

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* (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

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

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

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

This model can be expanded 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. While we focus on results for independent effect sizes in this article, the results hold for dependent effect sizes.

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, then $R_{i2} = 0$.

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.

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 $R_i = [1, \ldots, 1] = \mathbb{1}$.

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 complete-case analysis estimates 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 (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). A given regression in a shifting-case analysis conditions on a set of missingness patterns $R \in \mathcal{R}_j$ where $\mathcal{R}_j \subset \mathcal{R}$. These patterns are such that $R_{ij} = 1$ for those covariates X_{ij} that are observed *and* included in the analyses, but $R_{ik} \in \{0, 1\}$ for covariates X_{ik} 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 X_{i1} and X_{i2}. A shifting-case analysis might first regress T_i on observed values of X_{i1}. This regression would include 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., $R_i = [1, 0]$). This would imply that the regression involves effects i for which $R_i \in \mathcal{R}_j = \{[1, 1], [1, 0]\}$.

Consider a single regression in a shifting case analysis, and let S index the covariates included in that model $S = \{j : j = 0 \text{ or } X_{ij} \text{ in analysis}\}$ and let E be the complement of S. Note that this model is used to estimate β_S, which is a subset of the full vector of coefficients β. In the example above, where T_i is regressed on only X_{i1}, $\beta_S = [\beta_0, \beta_1]$. 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 R such that all the included covariates are observed, but any excluded covariates may be either observed or unobserved. Finally, let $U(S)$ be the set of effects for which X_{iS} are observed $U(S) = \{i : R_i \in \mathcal{R}_j\}$. 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 of $\mathbf{X}$ that pertain to the covariates that are included in the model, and the rows for which all of those covariates are observed. The matrix $\mathbf{W}_U$ contains the rows of $\mathbf{W}$ for which X_{iS} are observed (and similarly for $\mathbf{T}_U$).

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 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 (1976) 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:

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

MCAR implies that probability that a given value is missing is unrelated to anything observed or unobserved.

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

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

MAR differs from MCAR in that missingness might be related to observed values. As an example, if 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 MCAR 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 (Gelman, 2014; Graham, 2012; Little & Rubin, 2002; van Buuren, 2018). 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, complete- and available-case analyses proceed as if the complete-data and conditional models are equivalent. Doing so ignores the sources and impacts of missingness, and can lead to inaccurate 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 (Gelman, 2014; Little & Rubin, 2002; van Buuren, 2018). 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 complete-case analysis, $\mathcal{R}_j = \{\mathbb{1}\}$.

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 the probability of observing the missingness pattern r 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 wee that the expectation of T_i conditional on R_i can be written as the complete-data expectation $X_i\beta$ plus a bias term δ_{ij}. 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 complete-case estimator in (7) and the shifting-case estimator in (8) are weighted averages of the T_i, the following sections show that 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.

A standard approach for modeling missingness mechanisms for covariates is to assume R_i follows some log-linear distribution (Agresti, 2013). Various authors have described approaches to modelling R 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 $R_i \in \mathcal{R}_j \subset \mathcal{R}$:

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

where $f_{0j}(T_i, X_i, v_i) = 1$, so that ψ_{0j} would be the intercept term for the logit model for the set of missingness patterns $\mathcal{R}_j$. While log-linear models are not the only applicable or appropriate model for missingness, 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 $\hat{\beta}_C$ 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 missingness mechanism.

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

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

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 $T_i = X_i\beta$.

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 (36). 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 $\tau^2 + v_i$, which means that it will be larger when T_i vary more around the regression line. Finally, δ_{ij} depends on $\psi_{mj} f'_{mj}(X\beta, X, v)$. Note that f'_{mj} is the derivative of f_{mj} with respect to T. If f_{mj} does not depend on T, $f'_{mj} = 0$. Thus, δ_{ij} depends on the components of the selection model that are functions of T 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 $\hat{\beta}_C$ 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 complete case analysis will lead to unbiased coefficient estimates. First, if the covariates are MCAR, so that $R_i \perp T_i, X_i, v_i$, then

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

and hence

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

Thus, the complete-data and conditional models are equivalent. Assuming X is MCAR, then $\hat{\beta}_C$ in (7) will be unbiased for β. This is consistent with broader results on analyses of MCAR data.

A complete-case analysis is also unbiased under slightly less restrictive assumptions. Suppose that $R_i \perp (X_i, T_i)|v_i$, then the complete-data model and the conditional model are equivalent:

$$p(T_i|X_i, v_i, R_i = \mathbb{1}) = \frac{p(R_i = \mathbb{1}|v_i, \psi)p(T_i|x_i, v_i, \eta)}{p(R_i = \mathbb{1}|v_i, \psi)} = p(T_i|x_i, v_i, \eta) \tag{15}$$

Under this assumption, $\hat{\beta}_C$ will be unbiased.

The assumption that $R_i \perp (X_i, T_i)|v_i$ 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 $R_i \perp (T_i, X_i, v_i)$. Intuitively, if both X_i and T_i are conditionally independent of R_i, then so is the relationship between X_i and T_i.

For most effect size indices, variances v_i are functions of the sample sizes within studies n_i. Some effect sizes, such as the z-transformed correlation coefficient, have variances v_i 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 $R_i \perp$

$T_i|X_i, v_i$, we can write:

$$p(T_i|X_i, v_i, R_i = \mathbb{1}) = \frac{p(R_i = \mathbb{1}|X_i, v_i)p(T_i|X_i, v_i)}{p(R_i = \mathbb{1}|X_i, v_i)} = p(T_i|X_i, v_i) \tag{16}$$

Thus, if $R_i \perp T_i|X_i, v_i$ then the complete-case model will be the same as the complete-data model. Note that R_i can still depend on X_i, including X_{ij} that are not observed, which means that the data are MNAR.

However, when R_i is not independent of X_i or T_i (given v_i), then analyses can be biased. Let $\mathcal{R}_1 = \{\mathbb{1}\}$ so that the complete-case analysis 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

$$\text{Bias}[\hat{\beta}_C] = (\mathbf{X}_C^T \mathbf{W}_C \mathbf{X}_C)^{-1} \mathbf{X}_C^T \mathbf{W}_C \Delta_C \tag{17}$$

The bias in equation (17) is a weighted average of individual biases δ_{i1}. Hence, the bias will be larger if the δ_{i1} are larger.

Precisely how large the bias in (17) 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

$$\text{Bias}[\hat{\beta}_C] = (\mathbf{X}_C^T \mathbf{W}_C \mathbf{X}_C)^{-1} \mathbf{X}_C^T \mathbf{H}_{1C} \mathbf{f}_{1C} \psi_1 \tag{18}$$

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 (18) involves $\hat{H}_{1C}$ which contains the rows of $\mathbf{H}_1$ for which $R_i = \mathbb{1}$; similarly for $\mathbf{f}_{1C}$.

While (18) 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 observations are omitted due to missingness $\mathbf{H}_{1C}$, as well as some function of the components of the log-linear 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 covariate $X_{i1} \equiv X_i \in \{0, 1\}$, so that the complete data model is

$$T_i = \beta_0 + \beta_1 X_i + u_i + e_i \tag{19}$$

where β_0, 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 complete-case analysis 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 $R_i = 1$:

$$\hat{\beta}_{0C} = \frac{\sum_{i:X_i=0,R_i=1} w_i T_i}{\sum_{i:X_i=0,R_i=1} w_i} \tag{20}$$

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

$$\hat{\beta}_{1C} = \frac{\sum_{i:X_i=1,R_i=1} w_i T_i}{\sum_{i:X_i=1,R_i=1} w_i} - \hat{\beta}_{0C} \tag{21}$$

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

$$\text{logit}[p(R_i = 1|T_i, X_i, v_i)] = \psi_0 + \psi_1 T_i + \psi_2 X_i \tag{22}$$

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 (36), we can write

$$\delta_{i1} \approx H_1(X_i)(v_i + \tau^2)\psi_1 \tag{23}$$

Under this selection model, the bias of the complete-case estimator for β_0 is:

$$\text{Bias}[\hat{\beta}_{0C}] \approx H_1(0)(\bar{v}_0 + \tau^2)\psi_1 \tag{24}$$

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 (24) 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 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 $\bar{v}_0 + \tau^2$, the average variation of T_i for which $X_i = 0$; the greater the variation, the greater the bias. Finally, the bias depends on ψ_1, which is the relationship between an X_i being observed and the size of T_i.

To gain better insight into equation (24), 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 $\tau^2 = v$ (Hedges & Pigott, 2001, 2004; Hedges & Schauer, 2019). Thus, we can write $\tau^2 + v = 4(1 + r)/n$ from some constant r that ranges from 0 to 1.

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 $P[R = \mathbb{1}|T]$. 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 ψ_1 will depend on the scale of T_i. A difference of $T_i - T_j = 1$ is considered quite large for standardized mean differences. A less extreme difference $D_T = |T_i - T_j|$ for a standardized mean difference would be no larger than the size of an individual T_i. Conventions for standardized mean differences imply that a "small" effect would be about $T_i = 0.2$, a "medium" effect would be $T_i = 0.5$, and a "large" effect would be $T_i = 0.8$ (Cohen, 1988). Thus, meaningful values of D_T might feasibly range from 0.2 to 1.0. These conventions for odds ratios and D_T would imply that relevant values of $|\psi_1|$ might range from 0.4 to over 7.5.

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 0.2 or even 0.5 (Cohen's d).

It is worth noting that Figure 1 gives the bias for when T_i is positively correlated with X_i being fully observed, and hence $\psi_1 > 0$. This need not be the case, and $\psi_1 < 0$. 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.

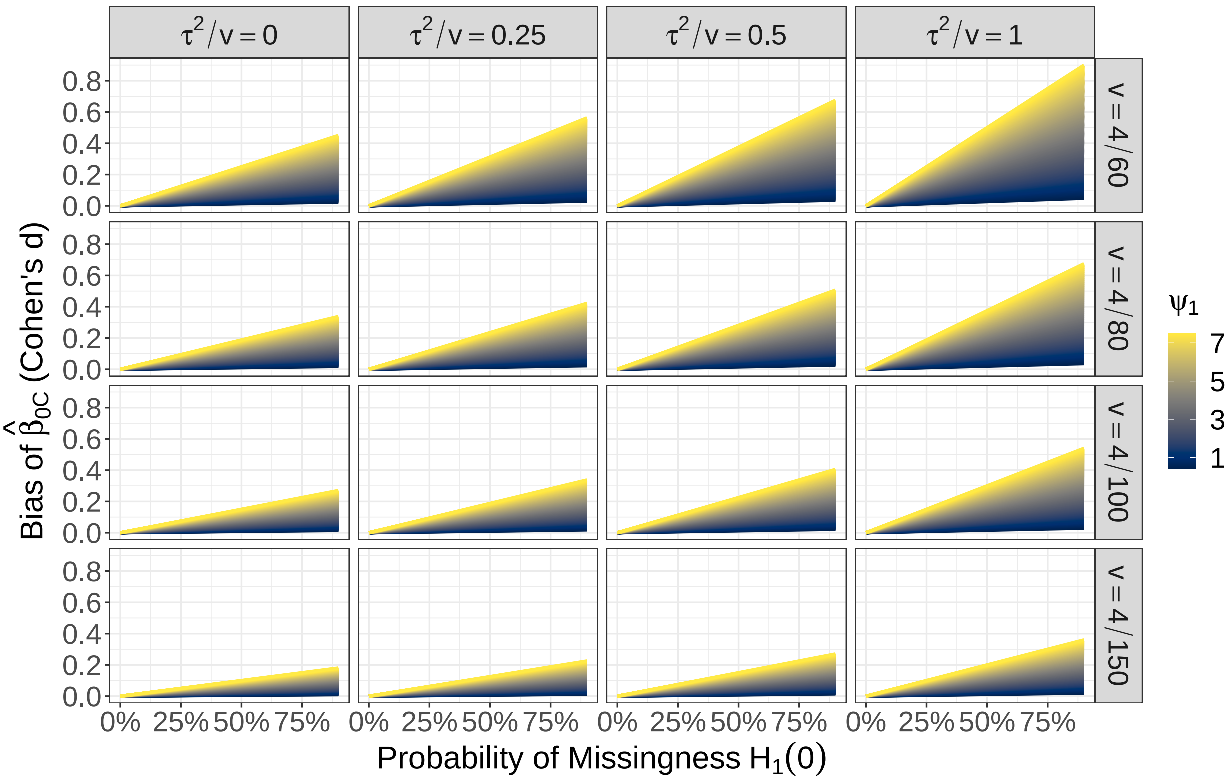

Figure 1: This figure plots the bias of the intercept estimate $\hat{\beta}_{0C}$ (y-axis) of the example. Bias is shown as a function of the average sampling variance v, residual heterogeneity τ^2, the probability of missingness $H_1(0)$ (x-axis), and the correlation between missingness and the effect 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.

The bias of $\hat{\beta}_{1C}$ under selection model (22) is given by:

$$\text{Bias}[\hat{\beta}_{1C}] \approx \left[H(1)(\bar{v}_1 + \tau^2) - H(0)(\bar{v}_0 + \tau^2)\right] \psi_1 \tag{25}$$

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:

$$\text{Bias}[\hat{\beta}_{1C}] \approx [p(R = 0|X = 1) - p(R = 0|X = 0)]\,(v + \tau^2)\psi_1 \tag{26}$$

The expression in (26) depends on three 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 $p(R = 0|X = 1) - p(R = 0|X = 0)$, which can be thought of as a difference in observation rates between $X_i = 1$ and $X_i = 0$. This difference will be greater if R and X are strongly correlated. Taken together, the bias of $\hat{\beta}_{1C}$ will be greatest when there are fewer complete cases, missingness is strongly related to the size of effects and the value of the covariate X.

To gain insight into the bias of $\hat{\beta}_{1C}$, consider the values of $\psi_1 \in [0.4, 7.5]$ and $\tau^2 + v = 4(1 + r)/n$ discussed above. The difference $p(R = 0|X = 1) - p(R = 0|X = 0)$ is just 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 and X (assuming equal marginals in a 2×2 table). Thus, $|p(R = 0|X = 1) - p(R = 0|X = 0)|$ might feasibly range from 0 to as large as 0.5.

Figure 2 shows the potential bias of $\hat{\beta}_{1C}$ for this example assuming the values 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 $p(R = 0|X = 1) - p(R = 0|X = 0)$ (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.

6 Bias in Shifting-Case Analyses

Shifting-case analyses are a common approach in meta-regression when there are very few complete cases. 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 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:

$$p(T_i|X_{iS}, v_i, R_i \in \mathcal{R}_j) = \frac{p(R_i \in \mathcal{R}_j|T_i, X_{iS}, v_i)p(T_i|X_{iS}, v_i)}{p(R_i \in \mathcal{R}_j|X_{iS}, v_i)} \tag{27}$$

The model in (27) 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 relevant covariates. The partial-data model $p(T_i|X_{iS}, v_i)$ need not be equivalent to the complete-

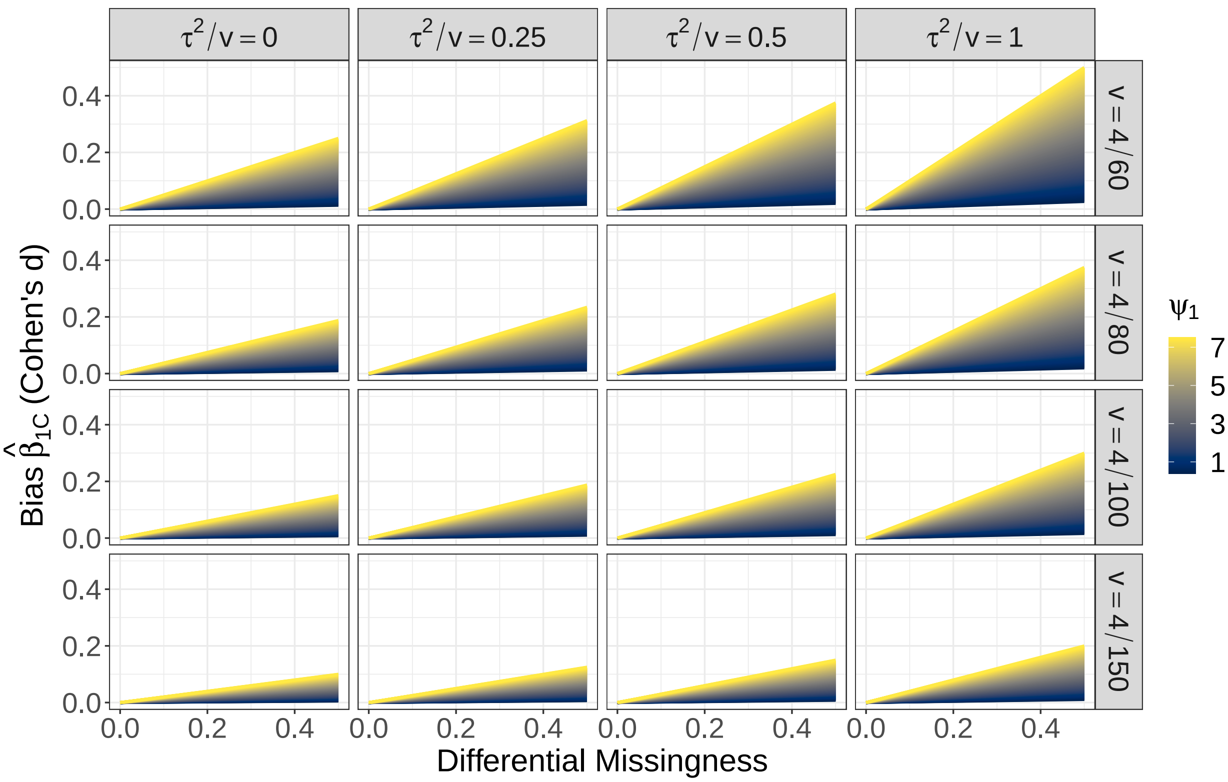

Figure 2: This figure plots the bias of $\hat{\beta}_{1C}$ (y-axis). Each panel corresponds to a given value of residual heterogeneity τ^2 and estimation error variance v. Within panels, the bias of $\hat{\beta}_{1C}$ is plotted as function of differential observation rates ($p(R = 0|X = 1) - p(R = 0|X = 0)$), 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.

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 no X_{iS} are missing.

The model in (27) suggests a very strict set of conditions for which $\hat{\beta}_S$ is unbiased. First, missingness must be independent of effect sizes. This arises if $R_i \perp T_i|X_{iS}, v_i$ or $R_i \perp (T_i, X_{iS})|v_i$. 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: $T_i \perp X_{iE}|X_{iS}, v_i$. This amounts to $\beta_j = 0$ for all $j \in E$. A related assumption is that $(T, X_{iS}) \perp X_{iE}|v$, 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. Note that conditions on omitted covariates *and* omitted observations must hold in order for $\hat{\beta}_S$ to be unbiased.

When assumptions about omitted variables and cases are not met, $\hat{\beta}_S$ 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:

$$\text{Bias}[\hat{\beta}_S] = (\mathbf{X}_{US}^T \mathbf{W}_U \mathbf{X}_{US})^{-1} \mathbf{X}_{US}^T \mathbf{W}_U \mathbf{X}_{UE} \beta_E + (\mathbf{X}_{US}^T \mathbf{W}_U \mathbf{X}_{US})^{-1} \mathbf{X}_{US}^T \mathbf{W}_U \Delta_{jU} \tag{28}$$

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 (36):

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

The expression in (28) shows that a shifting-case analysis suffers from two sources of bias. The first source, captured in the first term in (28), is a function of the coefficients for the excluded covariates β_E. 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 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 $\mathbf{X}_{UE}\beta_E$. 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 (28) captures the bias due to ignoring observations missing X_{iS}. This *missingness bias* is a function of ΔjU, 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 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 (36), δ_{ij} will be positive if T_i is positively correlated with whether $R_i \in \mathcal{R}_j$, and δ_{ij} 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 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

$$T_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_2 + u_i + e_i \tag{29}$$

If there is missingness in both X_1 and X_2, 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 shifting units analysis 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 $R \in \{[1, 1], [1, 0]\}$ and the excluded X_{i2} could be either 0 or 1. The shifting-case estimators follow from equation (**??**):

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

Assume that missingess follows the following log-linear model:

$$\text{logit}[p(R_i = [a, b]|T_i, X_i, v_i)] = \psi_{0ab} + \psi_{1ab}T_i + \psi_{2ab}X_{i1} + \psi_{3ab}X_{i2} \tag{30}$$

Here, the probability that an observation is included in the first SCA model depends on T_i, X_{i1}, and X_{i2}. It also depends on the parameters ψ_{mab} such that $a = 1$ (i.e., ψ_{010}, ψ_{011}, etc.). The first SCA model that includes only X_{i1} conditions on $R_i \in \mathcal{R}_1 = \{[1, 0], [1, 1]\}$. The observed and missing X_{i2} take values of either 0 or 1.

Given the selection model in (30), the bias of the SCA estimators can be written as:

$$\text{Bias}[\hat{\beta}_0] = \beta_2 \frac{\sum_{X_1=0,X_2=1} w_i}{\sum_{X_i=0} w_i} + \frac{\sum_{X_{i1}=0} w_i \tilde{\delta}_i}{\sum_{X_{i1}=0} w_i} \tag{31}$$

$$\text{Bias}[\hat{\beta}_1] = \beta_2 \left(\frac{\sum_{X_{i1}=1,X_{i2}=1} w_i}{\sum_{X_{i1}=1} w_i} - \frac{\sum_{X_{i1}=0,X_{i2}=1} w_i}{\sum_{X_{i1}=0} w_i} \right) + \left(\frac{\sum_{X_{i1}=1} w_i \tilde{\delta}_i}{\sum_{X_{i1}=1} w_i} - \frac{\sum_{X_{i1}=0} w_i \tilde{\delta}_i}{\sum_{X_{i1}=0} w_i} \right) \tag{32}$$

Note that 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. The second term in each expression is the missingness bias. To gain some insight into these expressions, consdier the omitted variables bias. When effects are estimated with roughly the same precision, so that $w_i \approx w$, then the omitted variable biases reduce to

$$\text{Omitted Var. Bias}[\hat{\beta}_{0S}] = \beta_2 P[X_2 = 1|X_1 = 0] \tag{33}$$

$$\text{Omitted Var. Bias}[\hat{\beta}_{1S}] = \beta_2 \left(P[X_2 = 1|X_1 = 1] - P[X_2 = 1|X_1 = 0]\right) \tag{34}$$

To get a better sense of these expressions, consider the bias for $\hat{\beta}_1$. This will depend on two quantities. The first is β_2, which is the contribution of X_2 to the complete-data model. The second is the difference $P[X_2 = 1|X_1 = 1] - P[X_2 = 1|X_1 = 0]$. Because both X_1 and X_2 are binary, this difference is equivalent to their Pearson correlation (assuming equal marginals). If $X_1 \perp X_2$, then their correlation is zero, and $\hat{\beta}_1$ will be unbiased. But if X_1 and X_2 are correlated, the bias of $\hat{\beta}_1$ will depend on how strongly correlated they are, and how big β_2 is.

Figure 3 shows the omitted variable bias of $\hat{\beta}_0$ (left plot) and $\hat{\beta}_1$ 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 $\pi_{01} = P[X_2 = 1|X_1 = 0]$ is the proportion of $X_2 = 1$ when $X_1 = 0$. In the right plot, $\rho_{12} = P[X_2 = 1|X_1 = 1] - P[X_2 = 1|X_1 = 0]$, which is roughly the correlation between X_1 and X_2. 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), 0.2 (medium correlation), and 0.5 (large correlation) (Cohen, 1988).

The figure shows that if $\beta_2 = 0$ so that X_2 is irrelevant given X_1, that both $\hat{\beta}_0$ and $\hat{\beta}_1$ will be unbiased. However, when β_2 is nonzero, both estimators will be biased. If X_1 and X_2 are highly correlated, or if $X_2 = 1$ when $X_1 = 0$ 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 $\hat{\beta}_1$ the bias will be less than about $d = 0.05$ when $|\beta_2| \leq 0.1$ or if $\rho_{12} < 0.5$.

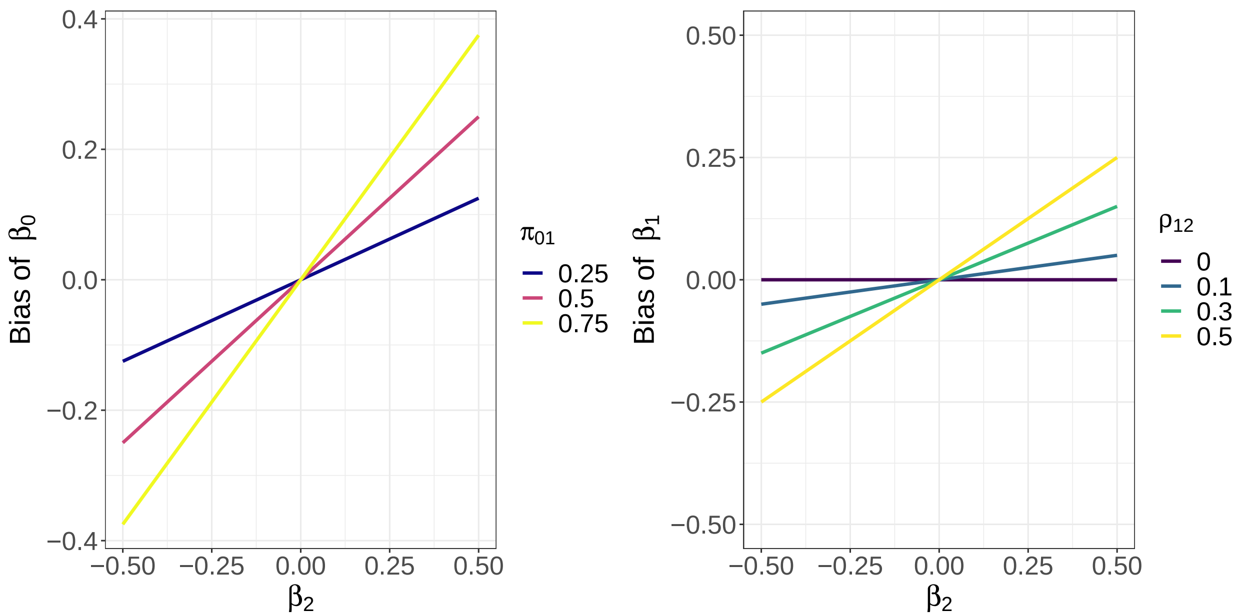

Figure 3: This figure shows the omitted variable bias of $\hat{\beta}_0$ and $\hat{\beta}_1$ for the model in (29) as a function of the omitted variable β_2. The bias (y-axis) and β_2 (x-axis) are on the scale of Cohen's d. The bias displayed is soley due to omitting X_2 from (29). In the left plot, lines are colored according to $\pi_{01} = P[X_2 = 1|X_1 = 0]$. In the right plot, lines are colored according to $\rho_{12} = P[X_2 = 1|X_1 = 1] - P[X_2 = 1|X_1 = 0]$.

Figure 3 does not take into account any bias induced by missingness. Note that the expression for missingness bias for $\hat{\beta}_{0S}$ is a weighted average of $\tilde{\delta}_i$ among effects for which $X_{i1} = 0$. Similarly, $\tilde{\delta}_i$ is a weighted average of δ_{ij}. This when $w_i \approx w$ and $\delta_{ij} \approx \delta$, the missingness bias for $\hat{\beta}_{0S}$ will be about the same as the missingness bias for $\hat{\beta}_{0C}$. Figure 1 shows possible values of this bias on the scale of Cohen's d.

Likewise, the missingness bias of $\hat{\beta}_{1S}$ is a difference of weighted averages of δ_{ij}. Assuming $\delta_{ij} \approx \delta$ and $w_i \approx w$, the missingness bias of $\hat{\beta}_{1S}$ is approximately the same as in **??**. Figure **??** shows possible values of that bias on the scale of Cohen's d.

[FIGURE IDEA: Panels for ψ_1 and β_2. Within panels, plot total bias vs. Prob. Missing].

7 Discussion

1. Both CCA and SCA will produce biased coefficient estimates unless certain conditions are met. While these sets of conditions are somewhat restrictive for CCA, they are considerably more restrictive for SCA.

2. The bias of CCA/SCA when these conditions are not met can be substantial.

3. An important part of the bias is that it will depend on unknown parameters and unobserved data. This means that for CCA/SCA to be unbiased, we would need to assume those conditions hold. Because it will be impossible to say if they hold, or if they don't, it will be difficult to say just how biased a given analysis is. [PERHAPS TIE IN EXAMPLE].

4. We would recommend avoiding SCA unless there are strong reasons to suspect both conditions are true. We would avoid using CCA unless there is very little missing data.

5. There are alternatives available. EM (Ibrahim), EM with MNAR (Lipsits), MI, FIML. Their application in MR is less broad due to a lack of easy-to-integrate software for most meta-analysts. For MI, the properties of various imputation techniques has not been fully studied for MR models.

6. There is still a large amount of work to be done on missing data in MA.

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:

$$E[T_i|X_i, v_i, R_i \in \mathcal{R}_j] \approx X_i\beta + H_j(X_i\beta, X_i, v_i)(\tau^2 + v_i) \sum_{m=0}^{m_j} \psi_{mj} f'_{mj}(X_i\beta, X_i, v_i) \tag{35}$$

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:

$$\delta_{ij} \approx H_j(X_i\beta, X_i, v_i)(\tau^2 + v_i) \sum_{m=0}^{m_j} \psi_{mj} f'_{mj}(X_i\beta, X_i, v_i) \tag{36}$$

Proof:

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

$$\begin{aligned} H_j(X\beta, X, v) \equiv H_j(X, v) &= P[R =\notin \mathcal{R}_j|T, X, v]|_{T=X\beta} \\ G_j(X, v) &= 1 - H_j(X, v) \\ g_j(X, v) &= P[R \in \mathcal{R}_j|X, v] \text{ as in (10).} \end{aligned}$$

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 $T = X\beta$

$$\begin{aligned} E[T|X, v, R \in \mathcal{R}_j] &= \int \frac{T \exp\left\{-\frac{(T-X\beta)^2}{2(\tau^2+v)} + \sum_i \psi_{ij} f_{ij}(T, X, v)\right\}}{g_j(X, v)\sqrt{2\pi(\tau^2 + v)}\left(1 + e^{\sum_i \psi_{ij} f_{ij}(T,X,v)}\right)} dT \\ &= \int \frac{T}{g_j(X, v)\sqrt{2\pi(\tau^2 + v)}} \exp\left\{-\frac{(T - X\beta)^2}{2(\tau^2 + v)} + \sum_i \psi_{ij} f_{ij}(X\beta, X, v)\right. \\ &\qquad + \sum_i \psi_{ij} f'_{ij}(X\beta, X, v)(T - X\beta) - \log\left(1 + e^{\sum_i \psi_{ij} f_{ij}(X\beta,X,v)}\right) \\ &\qquad \left. - G_j(X, v)(\sum_i \psi_{ij} f'_{ij}(X\beta, X, v))(T - X\beta) + O(T^2)\right\} dT \\ &\approx \int \frac{T}{g_j(X, v)\sqrt{2\pi(\tau^2 + v)}} \exp\left\{-\frac{1}{2(\tau^2 + v)}\left(T^2 - 2TX\beta - 2T(\tau^2 + v)\sum_i \psi_{ij} f'_{ij}(X\beta, X, v)\right.\right. \\ &\qquad \left.\left. + 2T(\tau^2 + v)G_j(X, v)(\sum_{i\in\mathcal{D}} \psi_{ij} f'_{ij}(X\beta, X, v)) + \ldots\right)\right\} dT \\ &= X\beta + H_j(X, v)(\tau^2 + v)\sum_i \psi_{ij} f'_{ij}(X\beta, X, v) \qquad \blacksquare \end{aligned}$$

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 Issues with Conditional Inference on Incomplete Data

Both complete- and shifting-case analyses ignore information, which can lead to biased meta-regression estimates. Complete-case analyses omit effects with missing covariates, so that $\mathbf{X}_A \equiv \mathbf{X}_C$ is a matrix containing the rows of $\mathbf{X}$ that are not missing any covariates. Shifting-case analyses omit covariates *and* missing data, so that $\mathbf{X}_A \equiv \mathbf{X}_S$ contains a subset of the columns of $\mathbf{X}$ and only rows from that subset that are not missing any variables.

Neither analytic approach typically makes any adjustments for the information they exclude. They therefore estimate regression coefficients by

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

where A corresponds to the relevant subset of $\mathbf{X}$ as described above, and $\mathbf{T}_A$ and $\mathbf{W}_A$ are the relevant subsets of $\mathbf{T}$ and $\mathbf{W}$ used in the analysis.

These estimates will be biased under many conditions, and that bias can be expressed by:

$$(\mathbf{X}_A^T \mathbf{W}_A \mathbf{X}_A)^{-1} \mathbf{X}_A^T \mathbf{W}_A E[\mathbf{T}_A | \mathbf{X}_A, \mathbf{v}_A, R \in \mathcal{R}_j] - \beta \tag{37}$$

where $\mathcal{R}_j$ is the missing data pattern(s) upon with the analytic approach conditions.

Equation (37) shows that the bias induced by complete- or shifting-case analyses will be a function of the conditional expectation $E[T_i | X_i, v_i, R_i]$, and how it differs from $E[T_i | X_i, v_i] = X_i \beta$. One way to convceive of this is to note that the conditional expectation $E[T_i | X_i, v_i, R_i \in \mathcal{R}_j]$ can be expressed as

$$E[T_i | X_i, v_i, R_i \in \mathcal{R}_j] = X_i \beta + \delta_i \tag{38}$$

where δ_i corresponds to the bias induced for observation i by conditioning on R_i. If we denote $\delta = [\delta_1, \ldots, \delta_k]^T$, and δ_A as the relevant subset of δ, then we can express the bias in conditional estimates of regression coefficients as:

$$(\mathbf{X}_A^T \mathbf{W}_A \mathbf{X}_A)^{-1} \mathbf{X}_A^T \mathbf{W}_A \delta_A \tag{39}$$

The conditional expectation $E[T_i|X_i, v_i, R_i]$, and hence δ will depend on the selection model given in equation (9), which in turn depends on the missingness mechanism $p(R_i|T_i, X_i, v_i)$. To compute just how large the bias is requires assumptions about missingness.

For the remainder of this article, we will denote

$$(1 - G_j(X_i, v_i)) = H_{ij}, \qquad \mathbf{f}_{ij}^T \psi_j = \sum_{m=0}^{m_j} \psi_{mj} f'_{kj}(X_i^T \beta, X_i, v_i) = \sum_{m=1}^{m_j} \psi_{mj} f'_{kj}(X_i^T \beta, X_i, v_i)$$

The bias δ_i under the log-linear selection model in equation (36) depends on a variety of quantities. First, H_{ij} is essentially the probability that $R_i \neq r_j$, which means that it can loosely be understood as the probability that a given observation is excluded from the analysis (conditional on X, v). For example, if $r_j = \nVdash$ as in a complete-case analysis, $H_{ij} \equiv H_i$ would give the probability that a covariate in for the ith effect is missing a covariate. Note that δ_i is increasing in H_{ij} which means that the greater the probability of missingness, the greater the bias induced by conditioning on missingness.

The second quantity that δ_i depends on is $v_i + \tau^2$. When effects are estimated with low precision or if there is a large amount of residual heterogeneity, then $v_i + \tau^2$ will be larger, and hence so will the bias. However, if effects have very small standard errors, then this can reduce the size of δ_i.

Finally, δ_i depends on the selection model. Note that if $f_{mj}(T_i, X_i, v_i)$ does not depend on T_i, then its derivative is 0. Therefore, δ_i will be a function of the ψ_{mj} such that f_{mj} depends on T_i.

As an example, consider the following log-linear selection model for complete cases (i.e., $R_i = [1, \ldots, 1] = \mathbb{1}$):

$$\text{logit}[p(R_i = \mathbb{1}|T_i, X_i, v_i)] = \psi_0 + \psi_1 T_i \tag{40}$$

Then $f_1(T, X, v) = T$, and the bias δ_i can be expressed as

$$\frac{1}{1 + \exp\{\psi_0 + \psi_1 X\beta\}}(\tau^2 + v_i)\psi_1 \tag{41}$$

To gain some insight into equation (41), suppose T_i is on the scale of a standardized mean difference so that $v_i \approx 4/n_i$ where n_i is the total sample size used to compute T_i. The expression in (41) depends on three key quantities. The first term is essentially the probability that a case is missing any covariates. The second term is the total variation of T_i around the regression lines, which depends on τ^2. Various researchers have described conventions for the magnitude of τ^2 that range from $\tau^2 = v_i/4$ to $\tau^2 = v_i$ (Hedges & Pigott, 2001, 2004; Hedges & Schauer, 2019). Thus, we can write $\tau^2 + v_i = 4(1+r)/n_i$ from some constant r that ranges from 0 to 1.

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 $P[R = \mathbb{1}|T]$. 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 ψ_1 will depend on the scale of T_i. A difference of $T_i - T_j = 1$ is considered quite large for standardized mean differences. A less extreme difference $\Delta_T = |T_i - T_j|$ for a standardized mean difference would be no larger than the size of an individual T_i. Conventions for standardized mean differences imply that a "small" effect would be about $T_i = 0.2$, a "medium" effect would be $T_i = 0.5$, and a "large" effect would be $T_i = 0.8$ (Cohen, 1988). Thus, meaningful values of Δ_T might feasibly range from 0.1 to 1.0. These conventions for odds ratios and Δ_T would imply that relevant values of $|\psi_1|$ might range from 0.4 to over 15.

Based on these conventions, Figure 1 shows the potential bias δ_i for this example. Each panel corresponds to a given within-study sample size n and residual heterogeneity τ^2. Within plots, the bias contributed by a single case δ_i is plotted as a function of the probability of missingness H_i (x-axis) and ψ_i (color). The plots 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 0.2 or even 0.5 (Cohen's d).

It is worth noting that Figure 1 gives the bias for when T_i is positively correlated with X_i being fully observed, and hence $\psi_1 > 0$. This need not be the case, and $\psi_1 < 0$. in that case, the biases δ_i would be a mirror image of those in Figure 1. Larger, more negative values of ψ_1 would lead to a greater downward bias δ_i.

C References

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

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

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

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. (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

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

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