Some Properties of Complete- and Available-Case Meta-Regressions with Missing Covariates

Writing Notes

- Add introduction

- Add section with real dataset that illustrates complete-case analysis and shifting-units

- There's some dodgy notation that needs to be fixed.

- Structure for results: General form followed by concrete example.

  - Derive results for distributions, which highlight conditions for unbiasedness.
  - Derive conditional expectations of $T|X, v, R$.
  - Show general matrix-form biases as a function of conditional biases
  - Walk through simple example

Conceptual Notes

- How do we relate coefficients in log-linear selection models to coefficients in meta-regression models?

1   Introduction

[HOLD FOR INTRO]

This article examines the potential bias of two common meta-regression methods for incomplete data: complete-case analysis and available-case analysis. 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 assumptions under which complete- and available-case analyses are un-

biased. We also study the bias of these approaches when these assumptions are not met using standard models for missingness.

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 term e_i is the estimation error for effect i and $V[e_i] = v_i$, and u_i is 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, 1983; 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

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

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 given by

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

Here, $\mathbf{W} = diag(1/(v_i + \tau^2))$ 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.

Let R_i be a vector of response indicators for effect i. Each element R_{ij} of R_i corresponds to a variable in a meta-analytic dataset. The R_{ij} take a value of either 0 or 1: $R_{ij} = 1$ indicates the corresponding variable in the meta-analysis is observed and $R_{ij} = 0$, indicates a that the corresponding variable is not observed. For the data $[T_i, v_i, X_i]$, $R_i \in \mathcal{R} \equiv \{0, 1\}^{p+2}$ is a vector of 0s and 1s of length $p + 1$. If v_i were missing, then $R_{i2} = 0$.

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, we amend the notation so that $R_i \in \mathcal{R} \equiv \{0, 1\}^p$ and $R_{ij} = 1$ if X_{ij} is observed and $R_{ij} = 0$ if X_{ij} is missing. The set $\mathcal{R}$ contains any possible missingness pattern for missing covariates. This notation omits the intercept term in X_i, since that will never be missing. As an example, if $X_i = [1, X_{i1}]$ with $X_{i1} \in \mathbb{R}$, then R_i is a scalar such that $R_i = 1$ if X_{i1} is observed, and $R_i = 0$

if X_{i1} is missing. Note that R_i can take one of many values of $r \in \mathcal{R} \equiv \{0, 1\}^p$. Denote the matrix $\mathbf{R}$ of all missingness indicators, where the ith row of $\mathbf{R}$ is the vector R_i.

The mechanism by which missingness arises is typically modelled through the distribution of R. Let ψ denote the parameters that index the distribution of R so that the probability mass function of R can be written as $p(R|T, X, v, \psi)$. Analyses of incomplete data usually depend on some assumptions about $p(R|T, X, v, \psi)$, which are discussed in the following sections.

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.

In the context of a complete meta-analytic dataset, $\mathbf{X}$ would describe the complete matrix of predictors, and $\mathbf{X}_O$ would indicate the entries of $\mathbf{X}$ that are observed and not missing. As argued below, complete- and available-case meta-regressions typically rely on a subset of the $\mathbf{X}$ matrix generated by omitting rows and/or columns of $\mathbf{X}$. Denote $\mathbf{X}_A$ as the matrix of covariates used in the analysis, where $\mathbf{X}_A$ is generated by removing rows and/or columns from $\mathbf{X}$ so that no entry of $\mathbf{X}_A$ is missing. Because the anlaysis methods discussed in this article use a subset of the observations, they will not use the full vector of effects $\mathbf{T}$, but rather some subset $\mathbf{T}_A$. Likewise, denote $\mathbf{W}_A$ denote the relevant subset of $\mathbf{W}$ for the analysis in question.

3.1 Conditional Incomplete Data Meta-Regression

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, 2001, 2019; Tipton et al., 2019b, 2019a). A complete-case analysis (CCA) includes only effects for which all covariates of interest are observed, which means that $R_i = [1, \ldots, 1] = \mathbb{1}$ for all effects included in the analysis (van Buuren, 2018). Therefore, complete-case analyses provide inferences for the conditional distribution of $T_i|X_i, v_i, R_i = \mathbb{1}$.

Available-case meta-regressions typically amount to including a subset of relevant covariates that are completely observed (Pigott, 2019; Tipton et al., 2019b). For instance, if two covariates X_1 and X_2 are of interest, an available-case analysis could involve regressing effects T_i on X_{i1} (excluding X_{i2}) and then fitting a separate model that regresses T_i on X_{i2} (excluding X_{i1}). 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).

A shifting-case analysis is equivalent to including observations for which R_i falls into some set of missingness patterns $\mathcal{R}_j \subset \mathcal{R}$. For instance, including effects for which X_{i1} is observed and omitting X_{i2} from the model means that the analysis incorporates effects for which both X_{i1} and X_{i2} are observed (i.e., $R_i = [1, 1]$), as well as effects for which X_{i1} is observed, but X_{i2} is missing (i.e., $R = [1, 0]$). This would imply that $\mathcal{R}_j = \{[1, 1], [1, 0]\}$. Seen this way, inferences for complete-case analyses are based on the conditional distribution $T_i | X_i, v_i, R_i \in \mathcal{R}_j$.

Because both complete- and available-case analyses condition on the value of R_i, they can be viewed as *conditional on missingness.* Models that are conditional 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, the analytic approaches of 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 a selection model as:

$$p(T_i | X_i, v_i, R_i = r, \eta, \psi) = \frac{p(R_i = r | T_i, X_i, v_i, \psi) p(T_i | X_i, v_i, \eta)}{p(R_i = r | X_i, v_i, \psi, \eta)} \tag{7}$$

where ψ indexes the distribution of $R | T, X, v$. This describes the conditional model as a function of the complete-data model $p(T_i | X_i, v_i, \eta)$ and a selection model $p(R_i = r | T_i, X_i, v_i, \psi)$ that gives the probability that a given set of covariates are observed. The denominator on the right hand side of (7) 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

$$p(R_i = r | X_i, v_i, \psi, \eta) = \int p(R_i = r | T_i, X_i, v_i, \psi) p(T_i | X_i, v_i, \eta) dT_i \tag{8}$$

A standard approach for modelling missingness in 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 pattern $R_i = r_j \in \mathcal{R}$:

$$\text{logit}P[R_i = r_j | T_i, X_i, v_i] = \sum_{m=0}^{n_j} \psi_{mj} f_{mj}(T_i, X_i, v_i) \tag{9}$$

where $f_{0j}(T_i, X_i, v_i) = 1$, so that ψ_{0j} would be the intercept term for the logit model for missingness pattern 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   Missingness Mechanisms

Analyses of incomplete data often require some assumption about why data are missing, which is referred to as the missingness *mechanism.* 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

6

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.

5   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 $\mathbf{X}_C \equiv \mathbf{X}_A$ 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 $\mathbf{X}_S \equiv \mathbf{X}_A$ 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 $\mathbf{X}_A = \mathbf{X}_C$ for the complete case estimator and $\mathbf{X}_A = \mathbf{X}_S$ for the shifting case analysis.

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

$$(\mathbf{X}_A^T \mathbf{W}_A \mathbf{X}_A)^{-1} \mathbf{X}_A^T \mathbf{W}_A E[\mathbf{T}_A | \mathbf{X}_A, \mathbf{v}_A, R \in \mathcal{R}_j] - \beta \tag{10}$$

where $\mathbf{T}_A$ is a subset of effects from $\mathbf{T}$ that are included in the analysis, $\mathbf{W}_A$ is an analogous subset of $\mathbf{W}$, and $\mathcal{R}_j$ is the missing data pattern(s) upon with the analytic approach conditions.

Equation (10) 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

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

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

as:

$$(\mathbf{X}_A^T \mathbf{W}_A \mathbf{X}_A)^{-1} \mathbf{X}_A^T \mathbf{W}_A \delta \tag{12}$$

The conditional expectation $E[T_i|X_i, v_i, R_i]$, and hence δ will depend on the selection model given in equation (7), 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.

6   Approximate Bias for Log-linear Selection Models

It is possible to derive an approximation for $E[T_i|X_i, v_i, R_i = r_j]$ when $R_i|T_i, X_i, v_i$ follows a log-linear distribution as described in equation (9).

Proposition: Suppose $p(T_i|X_i, v_i)$ is the standard fixed- or random effects meta-regression model in equation (2), and suppose $p(R_i = r_j|T_i, X_i, v_i) = G_j(T_i, X_i, v_i)$ follows the log-linear model in (9). Then:

$$E[T_i|X_i, v_i, R_i = r] \approx X_i\beta + (1 - G_j(X_i\beta, X_i, v_i))(\tau^2 + v_i) \sum_{m=0}^{n_j} \psi_{mj} f'_{mj}(X_i\beta, X_i, v_i) \tag{13}$$

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:

$$\delta_i \approx (1 - G_j(X_i\beta, X_i, v_i))(\tau^2 + v_i) \sum_{m=0}^{n_j} \psi_{mj} f'_{mj}(X_i\beta, X_i, v_i) \tag{14}$$

Proof:

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

$$G_j(X\beta, X, v) \equiv G_j(X, v) = P[R = r_j|T, X, v]|_{T=X\beta}$$

8

Then an approximation for $E[T|X, v, R = r_j]$ is as follows:

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

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

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

For the remainder of this article, we will denote

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

The bias δ_i under the log-linear selection model in equation (14) 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 included in the analysis. For example, if $r_j = \not{R}$ 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. 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., $R_i = [1, \ldots, 1] = \mathbb{1}$):

$$\text{logit} P[R_i = \mathbb{1} | T_i, X_i, v_i] = \psi_0 + \psi_1 T_i \tag{15}$$

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

$$\frac{1}{1 + \exp\{\psi_0 + \psi_1 X \beta\}} (\tau^2 + v_i) \psi_1 \tag{16}$$

To gain some insight into equation (16), 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 (16) 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 $\tau^2 = v_i$. Thus, we can write $\tau^2 + v_i = v_i(1 + r)$ 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 on the scale of 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. 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 feasible values of ψ_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).
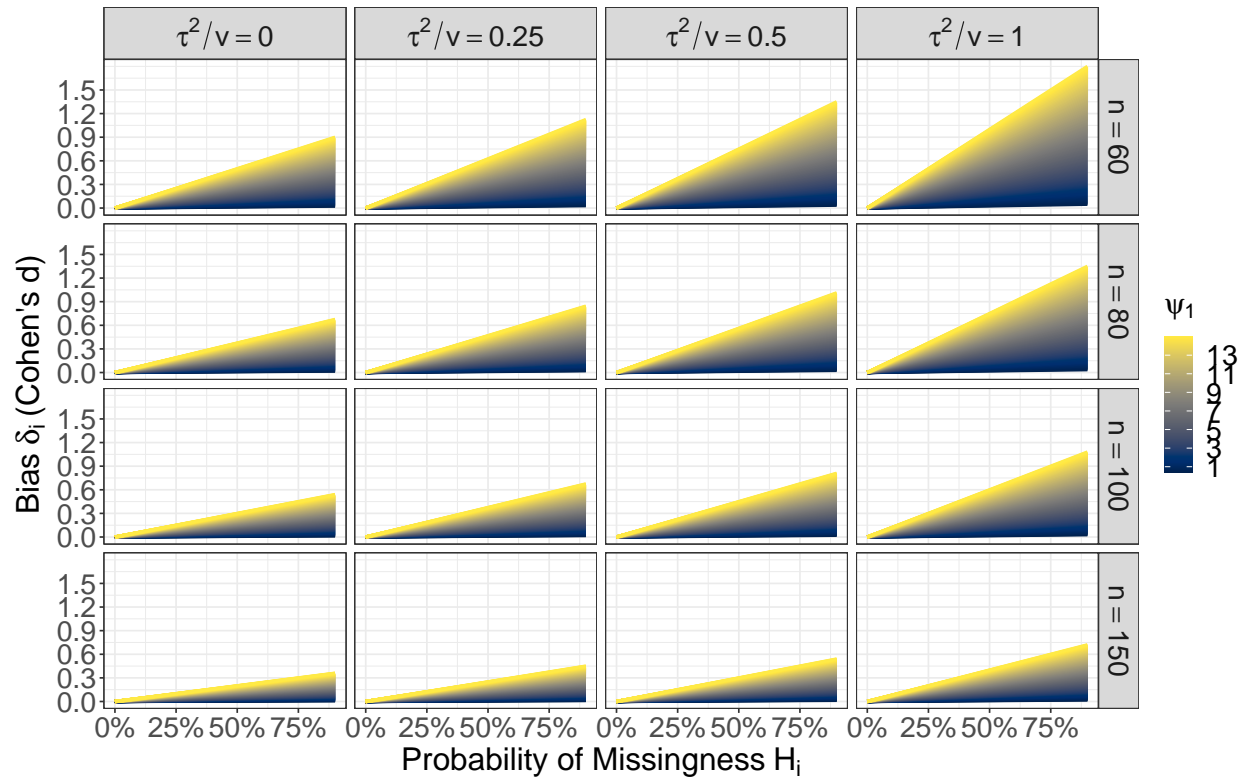
Figure 1: This figure plots the bias δ_i (y-axis) of a single case as a function of the sampling variance v, residual heterogeneity τ^2, the probability of missingness H_i (x-axis), and the correlation between missingness and the effect size as measured by ψ_1 (color).

7 Bias in Complete-Case Analyses

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

There are conditions under which the complete case analysis will lead to unbiased 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

$$
\begin{aligned}
p(T_i|X_i, v_i, R_i = \mathbb{1}, \eta) &= \frac{p(R = \mathbb{1}|T_i, X_i, v_i, \psi)p(T_i|X_i, v_i, \eta)}{p(R = \mathbb{1}|X_i, v_i, \psi)} \\
&= \frac{p(R_i = \mathbb{1}|\psi)p(T_i|x_i, v_i, \eta)}{p(R_i = \mathbb{1}|\psi)} \\
&= p(T_i|x_i, v_i, \eta)
\end{aligned}
$$

Thus, likelihood-based estimation should be consistent, assuming it can be done when X is MCAR. This is consistent with broader results on analyses of MCAR data.

However, a complete-case analysis is also unbiased under slightly less restrictive assumptions. Suppose that $R_i \perp (X_i, T_i)|v_i$, then

$$
\begin{aligned}
p(T_i|X_i, v_i, R_i = \mathbb{1}) &= \frac{p(R_i = \mathbb{1}|T_i, X_i, v_i, \psi)p(T_i|X_i, v_i, \eta)}{p(R_i = \mathbb{1}|X_i, v_i, \psi)} \\
&= \frac{p(R_i = \mathbb{1}|v_i, \psi)p(T_i|x_i, v_i, \eta)}{p(R_i = \mathbb{1}|v_i, \psi)} \\
&= p(T_i|x_i, v_i, \eta)
\end{aligned}
$$

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 $R \perp T|X, v$, we can write:

$$
\begin{aligned}
p(T|X, v, R = \mathbb{1}) &= \frac{p(R = \mathbb{1}|T, X, v)p(T|X, v)}{p(R = \mathbb{1}|X, v)} \\
&= \frac{p(R = \mathbb{1}|X, v)p(T|X, v)}{p(R = \mathbb{1}|X, v)} \\
&= p(T|X, v)
\end{aligned}
$$

Thus, if $R \perp T|X, v$ then the complete-case model will be the same as the complete-data model. Note that R can still depend on X, which means that the data are MNAR.

This result is consistent with the bias approximation in equation (14). There, the bias was a function of f'_{ij}, which is the derivative of $f_{ij}(T, X, v)$ with respect to T. If f_{ij} does not depend on T, so that $f_{ij}(T, X, v) \equiv f_{ij}(X, v)$ then the derivative will be zero, and hence ψ_{ij} will not contribute to the bias. Thus, only parts of the missingness model that depend on T will contribute to bias.

However, when R_i is not independent of X_i or T_i (given v_i), then analyses can be biased. Precisely how biased will depend on the distribution of R_i and its relationship to effect estimates T_i and their covariates X_i. A result based on the approximation in the previous section follows from the fact that a complete-case analysis involves $R_i = \mathbb{1}$. Let C 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 $\mathbf{W}_C = diag(1/(v_i + \tau^2)) : R_i = \mathbb{1}$ is the corresponding subset of weights. The complete-case analysis estimates coefficients β with:

$$
(\mathbf{X}_C^T\mathbf{W}_C\mathbf{X}_C)^{-1}\mathbf{X}_C^T\mathbf{W}_C\mathbf{T}_C
$$

The expectation of this estimator assuming the seleciton model in (9) is given by:

$$
\begin{aligned}
E[\hat{\beta}] &= (\mathbf{X}_C^TW_R\mathbf{X}_C)^{-1}\mathbf{X}_C^T\mathbf{W}_C E[\mathbf{T}_C|\mathbf{X}_C, \mathbf{v}_C, R = \mathbb{1}] \\
&= (\mathbf{X}_C^T\mathbf{W}_C\mathbf{X}_C)^{-1}\mathbf{X}_C^T\mathbf{W}_C \left[\mathbf{X}^T\beta + \mathbf{W}_C^{-1}\mathbf{H}_C\mathbf{f}_C\psi\right]
\end{aligned}
$$

where $\mathbf{H}_C = diag(H_{ij})$ and $\mathbf{f}_C = [f'_{i1}, \ldots, f'_{in}]$. Thus, the bias of the regression coefficients for a complete-

13

case analysis is given by

$$(\mathbf{X}_C^T \mathbf{W}_C \mathbf{X}_C)^{-1} \mathbf{X}_C^T \mathbf{H}_C \mathbf{f}_C \psi$$

7.1 Example: Complete-Case Analysis with a Single Binary Covariate

Suppose there is only one covariate $X_{i1} \equiv X_i \in \mathbb{R}$, so that the complete data model is

$$T_i = \beta_0 + \beta_1 X_i + u_i + e_i$$

where β_0, β_1 are the regression coefficients of interest. Assume that the model for missingness is log-linear:

$$p(R_i = 1 | T_i, X_i, v_i) \propto \frac{\exp\{\psi_0 + \psi_1 T_i + \psi_2 X_i + \psi_3 T_i X_i\}}{1 + \exp\{\psi_0 + \psi_1 T_i + \psi_2 X_i + \psi_3 T_i X_i\}}$$

Under this model, $H_{ij} = H_i$ depends only on X_i and not v_i, so we can write H_{ij} as:

$$H_{ij} = H(X_i) = p(R \neq \mathbb{1} | T_i, X_i, v_i)|_{T_i = X_i T \beta}$$

Note that $f_{1j}(T_i, X_i, v_i) = X_i$, $f_{2j}(T_i, X_i, v_i) = T_i$, and $f_{3j}(T_i, X_i, v_i) = T_i X_i$. Thus, we can write

$$E[T_i | X_i, v_i, R_i = 1] = \beta_0 + \beta_1 X_i + H(X_i)(v_i + \tau^2)(\psi_1 + \psi_3 X_i)$$

Suppose X_i is a binary variable, so that $X_i \in \{0, 1\}$, and that among the observed covariates, $X_i = 0$ (but $R_i = 1$) for $i = 1, \ldots m_0$ and $X_i = 1$ (and $R_i = 1$) for $m_0 + 1, \ldots, M$ where $M - m_0 = m_1$. Then then the complete-case estimator for β_0 is given by

$$\hat{\beta}_0 = \frac{\sum_{i=1}^{m_0} w_i T_i}{\sum_{i=1}^{m_0} w_i}$$

This would imply that the complete-case estimator of β_0 under the missing data model specified would have expectation:

$$E[\hat{\beta}_0] = \beta_0 + \frac{H(0) m_0}{\sum_{X_i=0} w_i} \psi_1$$

14

The standard estimator of β_1 is given by

$$\hat{\beta}_1 = \frac{\sum_{X_i=1} w_i T_i}{\sum_{X_i=1} w_i} - \frac{\sum_{X_i=0} w_i T_i}{\sum_{X_i=0} w_i}$$

$$= \frac{\sum_{X_i=1} w_i T_i}{\sum_{X_i=1} w_i} - \hat{\beta}_0$$

The expectation of this estimator is approximately

$$E[\hat{\beta}_1] = \beta_0 + \beta_1 + \frac{H(1)m_1}{\sum_{X_i=1} w_i}(\psi_1 + \psi_3) - \beta_0 - \frac{H(0)m_0}{\sum_{X_i=0} w_i}\psi_1$$

$$= \beta_1 + \frac{H(1)m_1}{\sum_{X_i=1} w_i}(\psi_1 + \psi_3) - \frac{H(0)m_0}{\sum_{X_i=0} w_i}\psi_1$$

$$= \beta_1 + \left[\frac{H(1)}{\bar{w}_{1.}} - \frac{H(0)}{\bar{w}_{0.}}\right]\psi_1 + \frac{H(1)}{\bar{w}_{1.}}\psi_3$$

When all studies have approximately the same estimation error variance so that $w_i \approx w$, then the bias of these estimators is approximately:

$$\text{Bias}[\hat{\beta}_0] = B_0 \approx \frac{P[R=0|X=0]}{w}\psi_1$$

$$\text{Bias}[\hat{\beta}_1] = B_1 \approx \frac{P[R=0|X=1] - P[R=0|X=0]}{w}\psi_1 + \frac{P[R=0|X=1]}{w}\psi_3$$

The bias of $\hat{\beta}_0$, B_0, is nearly identical to expression (**??**), and hence its magnitude is roughly the same as the values displayed in Figure **??**. The bias of $\hat{\beta}_1$ is a slightly more complex expression that depends on overall missingness, as well as differences in the relationship between missingness and variables in the data. Like B_0, B_1 is larger when there is more missingness, which occurs when $P[R=0|X]$ is greater. The bias of $\hat{\beta}_1$ also increases as a function of $P[R=0|X=1] - P[R=0|X=0]$, which will be greater if R and X are strongly correlated. Further ψ_1 will be larger if missingness is more strongly related to T, and ψ_3 will be larger if missingness has a stronger relationship to T when $X = 1$. Thus, the bias of $\hat{\beta}_1$ will be greatest when there are fewer complete cases, missingness is strongly related to the size of effects, and that relationship is even stronger when $X = 1$.

To gain insight into the bias of $\hat{\beta}_1$, suppose the average $w = (4/150 + 1/150)^{-1}$ which amounts to all studies having about 150 subjects with $\tau^2 = v/4$. Consider the values of ψ_1 from the previous section that ranged from 0.4 to over 15. Note that ψ_3 is a difference in log-odds ratios, which means that the size of ψ_3 should be no larger than the values of ψ_1. Thus, consider values that range from -2 to 2. Figure 2 shows the potential bias of $\hat{\beta}_1$ for this example. Each panel corresponds to a given value of ψ_3 and level of differential missingness $P[R=0|X=1] - P[R=0|X=0]$ (denoted $\pi_1 - \pi_0$). Within panels, the

shaded regions show the bias of $\hat{\beta}_1$ as a function of the probability that an $X = 1$ is missing (x-axis) and the log-odds ratio ψ_1 (color).
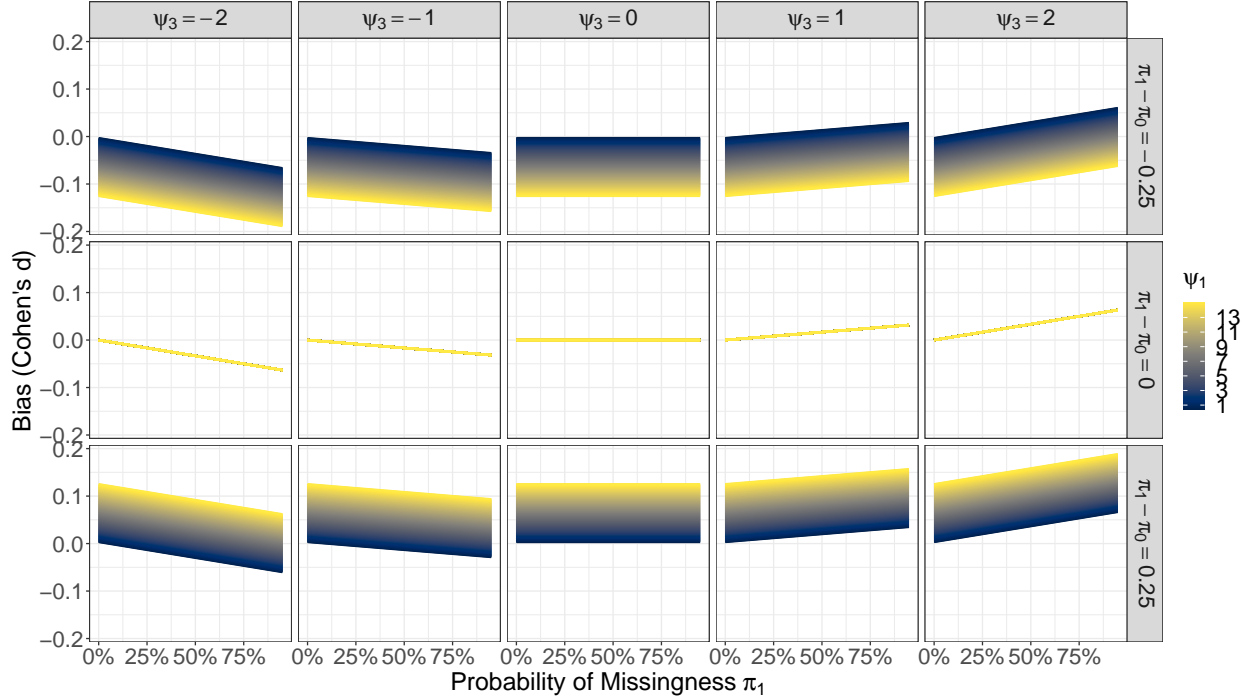

Figure 2: This figure plots the bias of $\hat{\beta}_1$ (y-axis) as a function of the proportion of missingness for cases where $X = 1$ (π_1, shown on x-axis), differential missingness rates between groups of effects where $X = 1$ versus $X = 0$ (denoted $\pi_1 - \pi_0$), the correlation between missingness and the effect size ψ_1 (color) and ψ_3. Note that bias is shown on the scale of Cohen's d, ψ_1 is on the scale of a log-odds ratio for the same scale, and ψ_3 is a difference of log-odds ratios. Bias was computed assuming all estimation variances were $v_i = 4/150$ and $\tau^2 = v_i/4 = 1/150$.

Figure 2 reveals a number of patterns about $\hat{\beta}_1$. First, its bias can plausibly range has high as 0.2 or as low as -.2. Though bias of this size arise in more extreme cases in the figure (i.e., larger values of π_1 and ψ_3), less extreme biases are still largely on the scale of $|d| = 0.1$. Second, in contrast to $\hat{\beta}_0$, a weaker relationship between missingness and T as measured by ψ_1 can actually lead to greater bias, which occurs if that relationship depends in large parge on X (i.e., when ψ_3 is larger than ψ_1). Third, when a greater proportion of missing X values are 1 than 0 so that $\pi_1 - \pi_2$ is positive, the bias will often be positive, but when the opposite is true the bias will often be negative.

Finally, it is worth noting that the biases displayed in Figure 2 assume that $v_i = 4/150$ and $\tau^2 = v_i/4 = 1/150$, so that $v_i + \tau^2 = 5/150$. The bias will be larger if the $v_i + \tau^2$ is larger, and will be smaller if $v_i + \tau^2$ is smaller. Suppose $v_i + \tau^2 = 5c/150$ for some $c \in \mathbb{R}$, then the bias in Figure 2 will be scaled by c. For instance, suppose studies are half as large, so that $v_i = 4/75$ and $\tau^2 = 0$. This would imply $c = 1.6$,

and hence the bias of $\hat{\beta}_1$ would largely on the scale of $|d| = 0.16$. Conversely, if $v_i = 4/250$ and $\tau^2 = 1/250$, then $c = 0.6$ and hence biases would largely be on the order of $|d| = 0.06$.

- Continuous covariates?

  For a single continuous covariate, note that

  $$\hat{\beta}_1 = \frac{\sum w_i(T_i - \bar{T}.)(X_i - \bar{X}.)}{\sum w_i(X_i - \bar{X}.)^2}$$

Given the selection model above, the bias of $\hat{\beta}_1$ can be expressed as:

$$\psi_1 \frac{\sum(1 - G(X_i))(X_i - \bar{X}.)}{\sum w_i(X_i - \bar{X}.)^2} + \psi_3 \frac{\sum(1 - G(X_i))X_i(X_i - \bar{X}.)}{\sum w_i(X_i - \bar{X}.)^2}$$

  It is not immediately clear that these simplify. Perhaps we leave this result out?

8 Bias in Shifting-Case Analysis

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 meta-analysis, an *available-case analysis* typically takes the form of fitting several meta-regression models, each including a subset of the covariates of interest (Cooper, 2017; Tipton et al., 2019b). Sometimes this even takes the form of regressing effect estimates on one covariate at a time. Referred to as "shifting units of analysis"—and referred to in this article as a shifting-case analysis—this approach inherently conditions on a set of missingness patterns $R \in \mathcal{R}_j$ where $\mathcal{R}_j \subset \mathcal{R}$. To see this, note that a shifting-case analysis amounts to a complete-case analysis on a subset of covariates. Thus $R_{ij} = 1$ 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; that is, excluded covariates may be observed or unobserved.

Let S index the covariates included in a SCA model $S = \{j : X_j \text{ in model}\}$ and let E be the complement of S. Denote $\mathcal{R}_j$ as the set of missingness patterns such that all included covariates are observed: $\mathcal{R}_j = \{R \in \mathcal{R} : R_S = 1\}$. 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:

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

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 $\mathbf{T}_U$).

To understand the conditions under which this estimator is unbiased, we can write the SCA model as:

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

The model above is slightly different from the models in the previous sections. There are two sources of bias that arise in SCA models. The first arises from omitting relevant variables from the analysis. Note that all of the functions involved condition only on the included covariates X_{iS}. Thus, the function $p(T_i | X_{iS}, v_i)$ 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 of $T_i \perp X_{iE} | X_{iS}, v_i$. That is, unless the excluded covariates are completely unrelated to effect size (given the covariates included in the SUA model), then even if none of the included covariates had any missingness, there would be bias in an SUA model.

The second source arises from the fact that SCA approaches ignore effects for which any of the included covariates are missing. That is, if any X_{iS} are missing, then a SCA omits the effect T_i and its associated covariates from the analysis.

Taken together, these two facets of the conditional SUA model in the expression above suggest a very strict set of conditions for which SCAs are unbiased. First, $R_i \perp (T_i, X_{iS}) | v$, so that missingness must be independent of effect size estimates and any included covariates conditional on the (completely observed) estimation error variances. This is a similar, though slightly weaker assumption as that made for unbiased complete-case analyses. Second, $T_i \perp X_{iE} | X_{iS}, v_i$, which means that any excluded covariates must be completely irrelevant given the included covariates. This amounts to $\beta_j = 0$ for all $j \in E$. Alternatively, we may write $(T, X_{iS}) \perp X_{iS} | v$, 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 both the conditions on omitted covariates and omitted observations must hold in order for a SCA to be unbiased.

When assumptions about both the missingness mechanism and the relevance of X_E do not hold, then SCAs 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.

When no data are missing from X_{iS}, then all effects are included in the analysis and the bias is

given by:

$$E[\hat{\beta}_S] - \beta_S = (\mathbf{X}_{US}^T \mathbf{W}_U \mathbf{X}_{US})^{-1} \mathbf{X}_{US}^T \mathbf{W}_U \mathbf{X}_U \beta - \beta_S$$

$$= (\mathbf{X}_{US}^T \mathbf{W}_U \mathbf{X}_{US})^{-1} \mathbf{X}_{US}^T \mathbf{W}_U (\mathbf{X}_{US} \beta_U + \mathbf{X}_{UE} \beta_E) - \beta_S$$

$$= \beta_S + (\mathbf{X}_{US}^T \mathbf{W}_U \mathbf{X}_{US})^{-1} \mathbf{X}_{US}^T \mathbf{W}_U \mathbf{X}_{UE} \beta_E - \beta_S$$

$$= (\mathbf{X}_{US}^T \mathbf{W}_U \mathbf{X}_{US})^{-1} \mathbf{X}_{US}^T \mathbf{W}_U \mathbf{X}_{UE} \beta_E$$

This bias arises even if no data are missing, and is referred to in the statistical and econometric literature as *omitted variable bias* (e.g., Farrar & Glauber, 1967; Mela & Kopalle, 2002). It is related to the issue of multicollinearity in generalized 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 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 $\mathbf{X}_{US}$.

When data are missing, there is a second source of bias that can arise from ignoring missingness. This bias will depend on $p(T_i | X_i, v_i, R_i \in \mathcal{R}_j)$, which is the conditional distribution of T_i given *all* relevant covariates (i.e., not just the covariates X_{iS}) and a set of missingness patterns. Note that we can write:

$$p(T_i | X_i, v_i, R_i \in \mathcal{R}_j) = \frac{\sum_{r \in \mathcal{R}_j} p(R_i = r | T_i, X_i, v_i) p(T_i | X_i, v_i)}{\sum_{r \in \mathcal{R}_j} p(R_i = r | X_i, v_i)}$$

$$= p(T_i | X_i, v_i) \frac{\sum_{r \in \mathcal{R}_j} p(R_i = r | T_i, X_i, v_i)}{\sum_{r \in \mathcal{R}_j} p(R_i = r | X_i, v_i)}$$

Using the approximation from the proposition above, we can write:

$$E[T | X, v, R \in \mathcal{R}_j] = \frac{\sum_{r \in \mathcal{R}_j} p(R = r | X, v) \left[X\beta + (1 - G_j(X, v)) \sum_{k=0}^{n_j} \psi_{kj} f'_{kj}(X\beta, X, v) \right]}{\sum_{r \in \mathcal{R}_j} p(R = r | X, v)}$$

Therefore, the total bias of an SCA estimator can be written as

$$
E[(\mathbf{X}_{US}^T \mathbf{W}_U \mathbf{X}_{US})^{-1} \mathbf{X}_{US}^T \mathbf{W}_U \mathbf{T}_U] - \beta_S
$$

$$
= (\mathbf{X}_{US}^T \mathbf{W}_U \mathbf{X}_{US})^{-1} \mathbf{X}_{US}^T \mathbf{W}_U E[\mathbf{T}_U | \mathbf{X}_U, \mathbf{v}, \mathbf{R}] - \beta_S
$$

$$
= (\mathbf{X}_{US}^T \mathbf{W}_U \mathbf{X}_{US})^{-1} \mathbf{X}_{US}^T \mathbf{W}_U \left(\mathbf{X}_U \beta + \left[\frac{\sum_{j:R_j \in \mathcal{R}} \pi_{ij} (1 - G_j(X_i^T \beta, X_i, v_i))(\tau^2 + v_i) \sum_{\mathcal{D}_j} \psi_{kj} f'_{kj}}{\sum_{j:R_j \in \mathcal{R}} \pi_{ij}} \right] \right) - \beta_S
$$

$$
= (\mathbf{X}_{US}^T \mathbf{W}_U \mathbf{X}_{US})^{-1} \mathbf{X}_{US}^T \mathbf{W}_U \left(\mathbf{X}_E \beta_E + \sum_j \mathbf{W}_U^{-1} \mathbf{P}_j \mathbf{H}_j \mathbf{f}_j \psi_j \left(\sum_j \mathbf{P}_j \right)^{-1} \right)
$$

$$
= (\mathbf{X}_{US}^T \mathbf{W}_U \mathbf{X}_{US})^{-1} \mathbf{X}_{US}^T \mathbf{W}_U \mathbf{X}_E \beta_E
$$

$$
+ (\mathbf{X}_{US}^T \mathbf{W}_U \mathbf{X}_{US})^{-1} \mathbf{X}_{US}^T \sum_j \mathbf{P}_j \mathbf{H}_j \mathbf{f}_j \psi_j \left(\sum_j \mathbf{P}_j \right)^{-1}
$$

Note that this bias decomposes neatly into the two sources of bias. The first term is simply the omitted variable bias from equation (). The second term is the weighted average of biases across missingness patterns, where the weights are equivalent to the conditional distribution of R given X, v: $p(R = r | X, v)$ for $r \in \mathcal{R}_j$.

8.1 Example: Shifting-Cases Analysis with Two Binary Covariates

Suppose $X = [1, X_1, X_2]$ and X_1 and X_2 are binary covariates such that

$$
T = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u_i + e_i
$$

If there is missingness in both X_1 and X_2, then $R \in \{0, 1\}^2$ so that $R = [1, 1]$ indicates both covariates are observed, and $R = [1, 0]$ indicates only X_1 is observed. If missingness is such that $R = [1, 1]$ for very few effect estimates, then a shifting units analysis might involve regressing T on the observed values of X_1 and then on the observed values of X_2.

The first regression would take only rows for which X_1 is observed. It would use the following estimates:

$$
\hat{\beta}_0 = \frac{\sum_{X_1=0} w_i T_i}{\sum_{X_1=0} w_i} = \bar{T}_{10}, \qquad \hat{\beta}_1 = \frac{\sum_{X_1=1} w_i T_i}{\sum_{X_1=1} w_i} - \hat{\beta}_0 = \bar{T}_{11} - \bar{T}_{10}
$$

Note that if no X_1 values were missing then the bias of these estimates is given by

$$\text{Bias}[\hat{\beta}_0] = \beta_2 \frac{\sum_{X_1=0,X_2=1} w_i}{\sum_{X_i=0} w_i} = \beta_2 \frac{w_{10}}{w_{0.}}$$

$$\text{Bias}[\hat{\beta}_1] = \beta_2 \left(\frac{\sum_{X_1=1,X_2=1} w_i}{\sum_{X_i=1} w_i} - \frac{\sum_{X_1=0,X_2=1} w_i}{\sum_{X_i=0} w_i} \right) = \beta_2 \left(\frac{w_{11}}{w_1} - \frac{w_{10}}{w_{0.}} \right)$$

When effects are estimated with roughly the same precision, so that $w_i \approx w$, then these expressions reduce to

$$\text{Bias}[\hat{\beta}_0] = \beta_2 P[X_2 = 1 | X_1 = 0] \qquad \text{Bias}[\hat{\beta}_1] = \beta_2 \left(P[X_2 = 1 | X_1 = 1] - P[X_2 = 1 | X_1 = 0] \right)$$

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 $P[X_2 = 1 | X_1 = 1] - P[X_2 = 1 | X_1 = 0]$. When X_1 and X_2 are positively correlated, then X_2 is much more likely to equal 1 when $X_1 = 1$, and much less likely to equal 1 when $X_1 = 0$. Thus, when the covariates are highly correlated, $P[X_2 = 1 | X_1 = 1] - P[X_2 = 1 | X_1 = 0]$ will be larger and hence the bias will be larger. Conversely, if X_1 and X_2 are negatively correlated, $P[X_2 = 1 | X_1 = 1] - P[X_2 = 1 | X_1 = 0]$ is likely to be larger in magnitude by negative, which means the bias will be larger and negative. Finally, if $X_1 \perp X_2$, then this difference, and hence the bias, will be zero.

[IDEA FOR PLOT: β_2 on the x-axis, bias on the y-axis, different lines for the correlation between X_1 and X_2]

However, suppose that some of the X_1 and X_2 are missing, and that R follows a log-linear model:

$$P[R = r_j | T, X, v] = \frac{\exp\{\psi_{0r} + \psi_{1r}X_1 + \psi_{2r}X_2 + \psi_{3r}T + \psi_{4r}X_1 T + \psi_{r5}X_2 T\}}{1 + \exp\{\psi_{0r} + \psi_{1r}X_1 + \psi_{2r}X_2 + \psi_{3r}T + \psi_{4r}X_1 T + \psi_{r5}X_2 T\}}$$

Note that with two covariates, $R \in \{[1,1],[1,0],[0,1],[0,0]\}$. The shifting-case analysis would condition in the first two patterns $r_1 = [1,1], r_2 = [1,0]$. Further, because both covariates are binary, we can write $H_j(X_1, X_2)$ for $j = 1, 2$ corresponding to r_j. Finally, let $\pi_{kj}(x) = P[R = r_j | X_1 = k, X_2 = x, v]$. Then, we can write the bias in $\hat{\beta}_0$ as

$$\beta_2 \frac{\sum_{X_0=1,X_2=1} w_i}{\sum_{X_1=0} w_i} + \left(\sum_{X_1=0} \sum_{j=1}^{2} \frac{H_j(0, X_{i2})(\psi_{3j} + \psi_{5j}X_{i2})\pi_{0j}(X_{2j})}{\sum_{j=1}^{2} \pi_{0j}(X_{2j})} \right) \bigg/ \left(\sum_{X_1=0} w_i \right)$$

It is not immediately clear how much this simplifies. We could (instead of graphs) point out that this the first term is exactly the omitted variables bias, and the second term is the missingness

bias. We could also point out that these two terms can be correlated. If they are positively correlated, then the total bias will be greater when there is missing data. If they are negatively correlated, then the total bias will be less when there is missing data. If that's the route we go, what are those conditions for positive/negative correlation?

9   Discussion

10 References

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

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. (1983). 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., & 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

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. (2001). 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