

# 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
- Nomenclature: *complete-case analysis* (CCA) is pretty clearly defined. *Available-case analysis* is a broad term, especially given the application to meta-analysis where *available-case analysis* has become synonymous with *shifting units of analysis*. But, *shifting units of analysis* is kind of burdensome for writing. How about *shifting-case analysis* (SCA)?
- 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 unbiased. 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

Let  $T_i$  be the estimate of the effect parameter  $\theta_i$ , and let  $v_i$  be the estimation error variance of  $T_i$ . Denote a vector of covariates that pertain to  $T_i$  as  $X_i = [1, X_{i1}, \dots, X_{i,p-1}]^T$ . Note that the first element of  $X_i$  is a 1, which corresponds to an intercept term in a meta-regression model, and that model can be expressed as:

$$T_i|X_i, v_i, \eta = X_i^T \beta + u_i + e_i \quad (1)$$

Here,  $\beta \in \mathbb{R}^p$  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 parameter  $\eta = [\beta, \tau^2]$  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  $\tau^2$  (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

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

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

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

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  $\mathbf{X} \in \mathbb{R}^{k \times p}$  is the matrix of covariates. Note that the functions in both (2) and (3) assume that *all* of the  $p - 1$  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

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

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

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

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  $[T_i, v_i, X_i]$ ,  $R_i \in \mathcal{R} \equiv \{0, 1\}^{p+1}$  is a vector of 0s and 1s of length  $p + 1$ . If  $v_i$  were missing, then  $R_{i2} = 0$ .

Our focus in this article is on 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  $R_i \in \mathcal{R} \equiv \{0, 1\}^{p-1}$  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. Note that this omits the intercept term described above. For instance 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  $r \in \mathcal{R} \equiv \{0, 1\}^{p-1}$ .

The mechanism by which missingness arises is typically modelled through the distribution of  $R$ . Let  $\psi$  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

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

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,  $\mathbf{X} = [X_1, X_2, \dots, X_k]^T$  would describe the complete matrix of predictors, where  $\mathbf{X}_O$  would indicate the entries of  $\mathbf{X}$  that are observed and not missing. As argued below, common approaches in meta-regression with incomplete data involve using a matrix that omits rows and/or columns of the complete covariate matrix  $\mathbf{X}$ . 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.

### 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  $R_i = [1, \dots, 1] = \mathbf{1}$  for all effects included in the analysis (van Buuren, 2018). Therefore, complete-case analyses provide inferences for the conditional distribution of  $T_i|X_i, v_i, R_i = \mathbf{1}$ .

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 often involves regressing effects on  $X_1$  (excluding  $X_2$ ) and then on  $X_2$  (excluding  $X_1$ ). 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-cases analysis* (SCA).

A shifting-cases 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  $X_1$  is observed and omitting  $X_2$  from the model means that the analysis incorporates effects for which both  $X_1$  and  $X_2$  are observed (i.e.,  $R = [1, 1]$ ), as well as effects for which  $X_1$  is observed, but  $X_2$  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 iden-

tical 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:

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

where  $\psi$  indexes the distribution of  $R|T, X, v$ . This describes the conditional model as a function of the complete-data model  $p(T|X, v, \eta)$  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

$$p(R_i = r|X_i, v_i, \psi, \eta) = \int p(R_i = r|T_i, X_i, v_i, \psi)p(T_i|X_i, v_i, \eta)dT \quad (8)$$

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  $R_i = r_j \in \mathcal{R}$ :

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

where  $f_{0j}(T_i, X_i, v_i) = 1$ , so that  $\psi_{0j}$  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  $R$  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 itself.

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)  $\psi$  in order to conduct inference on  $\eta$  (Gelman, 2014; Graham, 2012; Little & Rubin, 2002; van Buuren, 2018). In practice, missing data are ignorable if they are MAR and if  $\psi$  and  $\eta$  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  $\mathbf{X}_R \equiv \mathbf{X}_A$  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  $\mathbf{X}_U \equiv \mathbf{X}_A$  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

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

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

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

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 depend on the conditional expectation  $E[T_i | X_i, v_i, R_i]$ , and how it differs from  $E[T_i | X_i, v_i] = X_i^T \beta$ . The expectation  $E[T_i | X_i, v_i, R_i]$  will depend on the selection model given in equation (7), which in turn depends on the missingness mechanism  $p(R_i | T_i, X_i, v_i)$ .

## 6 Approximate Bias for Log-linear Selection Models

It is possible to derive an approximation for  $E[T_i | X_i, v_i, R_i = r]$  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  $p(R_i = r | T_i, X_i, v_i) = G_j(T_i, X_i, v_i)$  follows the log-linear model in (9). If we denote  $\mathcal{D}_j = \{m : f_{mj} \text{ depends on } T_i\}$ , then

$$E[T_i | X_i, v_i, R_i = r] \approx X_i^T \beta + (1 - G_j(X_i^T \beta, X_i, v_i))(\tau^2 + v_i) \sum_{m \in \mathcal{D}_j} \psi_{mj} f'_{mj}(X_i^T \beta, X_i, v) \quad (11)$$

**Proof:**

For  $i \in \mathcal{D}_j$ , we can write  $f_{ij}(T, X, v)$  and we can write  $f_{ij}(T, X, v) = f_{ij}(X, v)$  for  $i \notin \mathcal{D}$ . Further, denote

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

Then an approximation for  $E[T_i | X_i, v_i, R_i = r_j]$  is as follows:

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

Note that this uses a first order Taylor expansion of the exponentiated terms of the log-linear model, and thus assumes the  $f_j$  are differentiable. This approximation will be more accurate if the  $f_j$  are linear in  $T$ . In that case only a Taylor expansion of the denominator of the log-linear model is required.

For the remainder of this article, we will denote

$$(1 - G_j(X_i, v_i)) = H_{ij}, \quad \psi_j^T \mathbf{f}_{ij} = \sum_{k \in \mathcal{D}_j} \psi_{kj} f'_{kj}(X_i^T \beta, X_i, v_i) = \sum_{k=1}^n \psi_{kj} f'_{kj}(X_i^T \beta, X_i, v_i)$$

## 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  $R = [1, \dots, 1] = \mathbf{1}$ .

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



First, if the covariates are MCAR, so that  $R \perp T, X, v$ , then

$$P(R|T, X, v, \psi) = \psi$$

and hence

$$\begin{aligned} p(T|X, v, R = r, \eta) &= \frac{p(R = r|T, X, v, \psi)p(T|X, v, \eta)}{p(R = r|X, v, \psi)} \\ &= \frac{\psi p(T|x, v, \eta)}{\psi} \\ &= p(T|x, v, \eta) \end{aligned}$$

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, it would appear that a complete-case analysis is also valid under slightly less restrictive assumptions. Suppose that  $R \perp (X, T)|v$ , then

$$\begin{aligned} p(T|X, v, R = r) &= \frac{p(R = r|T, X, v, \psi)p(T|X, v, \eta)}{p(R = r|X, v, \psi)} \\ &= \frac{p(R = r|v, \psi)p(T|x, v, \eta)}{p(R = r|v, \psi)} \\ &= p(T|x, v, \eta) \end{aligned}$$

The assumption that  $R \perp (X, T)|v$  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  $R \perp (T, X, v)$ . Intuitively, if both  $X$  and  $T$  are conditionally independent of  $R$ , then so is the relationship between  $X$  and  $T$ .

For most effect size indices, variances  $v$  are functions of the sample sizes within studies  $n$ . Some effect sizes, such as the  $z$ -transformed correlation coefficient, have variances  $v$  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).

- Note about reduction in precision.

However, when  $R$  is not independent of  $X$  or  $T$  (given  $v$ ), then analyses can be biased. Precisely

how biased will depend on the distribution of  $R$  and its relationship to effect estimates  $T$  and their covariates  $X$ . A general result based on the approximation in the previous section follows from the fact that a complete-case analysis involves  $R = \mathbf{1}$ . Further, denote  $\mathbf{X}_R$  as the matrix of covariates such  $R_i = \mathbf{1}$ . If we denote  $\mathbf{W}_R = \text{diag}(v_i + \tau^2) : R_i = \mathbf{1}$ , the complete-case analysis will estimate coefficients as:

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}_R^T \mathbf{W}_R \mathbf{X}_R)^{-1} \mathbf{X}_R^T \mathbf{W}_R E[T|X, v, R = r] \\ &= (\mathbf{X}_R^T \mathbf{W}_R \mathbf{X}_R)^{-1} \mathbf{X}_R^T \mathbf{W}_R \left[ X_i^T \beta + \left( (1 - G(X_i, v_i))(\tau^2 + v_i) \sum_{j \in \mathcal{D}} \psi_j f'_j(X_i^T \beta, X_i, v_i) \right) \right] \\ &= \beta + (\mathbf{X}_R^T \mathbf{W}_R \mathbf{X}_R)^{-1} \mathbf{X}_R^T \mathbf{W}_R \left[ (1 - G(X_i, v_i))(\tau^2 + v_i) \sum_{j \in \mathcal{D}} \psi_j f'_j(X_i^T \beta, X_i, v_i) \right]\end{aligned}$$

Thus, the bias of the regression coefficients is given by

$$(\mathbf{X}_R^T \mathbf{W}_R \mathbf{X}_R)^{-1} \mathbf{X}_R^T \mathbf{W}_R \left[ (1 - G(X_i, v_i))(\tau^2 + v_i) \sum_{j \in \mathcal{D}} \psi_j f'_j(X_i^T \beta, X_i, v_i) \right]$$

- Refine for matrix notation. Let  $\mathbf{I}$  be the identity matrix and  $\mathbf{H} = \mathbf{I} - \text{diag}(G(X_i, v_i))$ . Further, let  $\mathbf{f}_j = [f'_{ij}(X_i^T \beta, X_i, v_i)]$  and  $\psi_j = [\psi_{0j}, \dots, \psi_{nj}]^T$ , then  $\mathbf{f}_j \psi_j = [\sum_{k \in \mathcal{D}} \psi_{kj} f'_{kj}(X_i^T \beta, X_i, v_i)]^T$ . Hence, we can write:

$$(\mathbf{X}_R^T \mathbf{W}_R \mathbf{X}_R)^{-1} \mathbf{X}_R^T \mathbf{W}_R \mathbf{W}_R^{-1} \mathbf{H}_j \mathbf{f}_j \psi_j = (\mathbf{X}_R^T \mathbf{W}_R \mathbf{X}_R)^{-1} \mathbf{X}_R^T (\mathbf{I} - \mathbf{G}) Y$$

- If the likelihood for conditional inference can be written free of  $R$  then conditional inference may be appropriate.

## 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  $\beta_0, \beta_1$  are the regression coefficients of interest. Assume that the model for missingness is log-linear:

$$p(R = 1|T, X, v) \propto \frac{\exp\{\psi_0 + \psi_1 X + \psi_2 T + \psi_3 TX\}}{1 + \exp\{\psi_0 + \psi_1 X + \psi_2 T + \psi_3 TX\}}$$

Note that  $G(X, v) \equiv G(X)$ . Further,  $\mathcal{D} = \{2, 3\}$  and  $f_j(T, X, v) = T f_j(X, v)$ , and hence  $f'_j(T, X, v) = f'_j(X, v)$  for  $j \in \mathcal{D}$ . Thus,

$$E[T_i|X_i, v_i, R_i = 1] = \beta_0 + \beta_1 X_i + (1 - G(X_i))(v_i + \tau^2)(\psi_2 + \psi_3 X_i)$$

Suppose  $X$  is a binary variable, so that  $X \in \{0, 1\}^{p-1}$ , and that  $X$  is observed for  $M \leq k$  effects. Denote  $m_0$  as the number of observed  $X_i = 0$  and  $m_1$  be the observed  $X_i = 1$  so that  $M = m_0 + m_1$ . Further, assume that  $X_i = 0$  (but  $R_i = 1$ ) for  $i = 1, \dots, m_0$  and  $X_i = 1$  (and  $R_i = 1$ ) for  $m_0 + 1, \dots, M$ . Then then the ML and least squares estimator for  $\beta_0$  is given by

$$\hat{\beta}_0 = \frac{\sum_{i=1}^{m_0} T_i / (v_i + \tau^2)}{\sum_{i=1}^{m_0} 1 / (v_i + \tau^2)}$$

This would imply that the complete-case estimator of  $\beta_0$  under the missing data model specified would have expectation:

$$\begin{aligned} E[\hat{\beta}_0] &= \beta_0 + \frac{(1 - G(0)) m_0}{\sum_{i=1}^{m_0} 1 / (v_i + \tau^2)} \psi_2 \\ w_j &= \left[ \sum_{i: X_i=j, R_i=1} 1 / (v_i + \tau^2) \right] / m_j \\ E[\hat{\beta}_0] &= \beta_0 + \frac{1 - G(0)}{w_0} \psi_2 \end{aligned}$$

Note that the second term on the right hand side constitutes the bias of  $\hat{\beta}_0$ . The bias depends on a variety of quantities. First, it depends on the selection model, solely through  $\psi_2$ . When  $\psi_2 > 0$  and hence when larger effect estimates are more likely to be missing covariates, the bias is negative. The bias is positive when  $\psi_0 < 0$ , which occurs when larger effect estimates are less likely to be missing covariates.

The ML estimator of  $\beta_1$  is given by

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{i=m_0+1}^M T_i / (v_i + \tau^2)}{\sum_{i=m_0+1}^M 1 / (v_i + \tau^2)} - \frac{\sum_{i=1}^{m_0} T_i / (v_i + \tau^2)}{\sum_{i=1}^{m_0} 1 / (v_i + \tau^2)} \\ &= \frac{\sum_{i=m_0+1}^M T_i / (v_i + \tau^2)}{\sum_{i=m_0+1}^M 1 / (v_i + \tau^2)} - \hat{\beta}_0 \end{aligned}$$

The expectation of this estimator is approximately

$$\begin{aligned}
E[\hat{\beta}_1] &= \beta_0 + \beta_1 + \frac{(1 - G(1))m_1}{\sum_{i=m_0+1}^M 1/(v_i + \tau^2)}(\psi_2 + \psi_3) - \beta_0 - \frac{(1 - G(0))m_0}{\sum_{i=1}^{m_0} 1/(v_i + \tau^2)}\psi_2 \\
&= \beta_1 + \frac{(1 - G(1))m_1}{\sum_{i=m_0+1}^M 1/(v_i + \tau^2)}(\psi_2 + \psi_3) - \frac{(1 - G(0))m_0}{\sum_{i=1}^{m_0} 1/(v_i + \tau^2)}\psi_2 \\
&= \beta_1 + \left[ \frac{1 - G(1)}{w_1} - \frac{1 - G(0)}{w_0} \right] \psi_2 + \frac{1 - G(1)}{w_1} \psi_3
\end{aligned}$$

- Plots? [NOTE: How do relate the  $\psi$  parameters to the  $\beta$  parameters?]
- 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:

$$\psi_2 \frac{\sum (1 - G(X_i))(X_i - \bar{X})}{\sum w_i(X_i - \bar{X})^2} + \psi_3 \frac{\sum (1 - G(X_i))X_i(X_i - \bar{X})}{\sum w_i(X_i - \bar{X})^2}$$

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  $U = \{j : X_j \text{ in SCA model}\}$  index the covariates included in a SCA model. Let  $E$  be the complement of  $U$  so that it indexes all of the covariates excluded from the SCA model. Let  $O(U) = \{i : R_{ij} = 1, j \in U\}$  be the effect sizes for which  $X_{iU}$  are observed, and  $M(U)$  be the effect sizes for which  $X_{iU}$  are missing. Further, let  $\mathcal{R}_j = \{R \in \mathcal{R} : R_{ij} = 1, \forall j \in U\}$ . 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. Then, the shifting-case model can be written as:

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

The model above is slightly different from the models in the previous sections. Note that all of the functions involved condition only on the included covariates  $X_{iU}$ . Thus, the function  $p(T_i|X_{iU}, v_i)$  is analogous to the complete-data likelihood  $p(T|X, v)$ , however it is important to note that the two functions are not necessarily equivalent because the former conditions only on  $X_{iU}$  and not the full set of covariates  $X_i$ .

There are two sources of bias that arise in SCA models. The first is due to the fact that  $p(T_i|X_{iU}, v_i)$  need not be equivalent to  $p(T_i|X_i, v_i)$ . These models would only be equivalent if  $T \perp X_E|X_U, v$ . 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 omit effects for which any of the included covariates are missing. That is, if any  $X_{ij}$  is missing for  $j \in U$ , then a SCA ignores the effect  $T_i$  and its associated covariates.

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,  $R \perp (T, X_U)|v$ , 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,  $T \perp X_E|X_U, v$ , 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  $(T, X_U) \perp X_E|v$ , which would imply that the complete data likelihood involves no interactions between  $X_U, X_E$  and that  $X_U$  and  $X_E$  are completely orthogonal. Note that both of these conditions 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  $X_U$ , then all effects are included in the analysis and the bias is given by:

$$\begin{aligned}
E[(\mathbf{X}_U^T W \mathbf{X}_U)^{-1} \mathbf{X}_U^T \mathbf{W} \mathbf{T}] - \beta_U &= (\mathbf{X}_U^T W \mathbf{X}_U)^{-1} \mathbf{X}_U^T \mathbf{W} \mathbf{X} \beta - \beta_U \\
&= (\mathbf{X}_U^T W \mathbf{X}_U)^{-1} \mathbf{X}_U^T \mathbf{W} (\mathbf{X}_U \beta_U + \mathbf{X}_E \beta_E) - \beta_U \\
&= \beta_U + (\mathbf{X}_U^T W \mathbf{X}_U)^{-1} \mathbf{X}_U^T \mathbf{W} \mathbf{X}_E \beta_E - \beta_U \\
&= (\mathbf{X}_U^T W \mathbf{X}_U)^{-1} \mathbf{X}_U^T \mathbf{W} \mathbf{X}_E \beta_E
\end{aligned}$$

This bias arises even if no data are missing. When data are missing, there is a second source of bias that can arise due to missingness. This bias will depend on  $E[T|X, v, R \in \mathcal{R}_j]$ . Note that we can write:

$$\begin{aligned}
p(T|X, v, R \in \mathcal{R}_j) &= \frac{\sum_{r \in \mathcal{R}_j} p(R = r|T, X, v) p(T|X, v)}{\sum_{r \in \mathcal{R}_j} p(R = r|X, v)} \\
&= p(T|X, v) \frac{\sum_{r \in \mathcal{R}_j} p(R = r|T, X, v)}{\sum_{r \in \mathcal{R}_j} p(R = r|X, v)}
\end{aligned}$$

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

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

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

$$\begin{aligned}
E[(\mathbf{X}_{UO(U)}^T W_O \mathbf{X}_{UO(U)})^{-1} \mathbf{X}_{UO(U)}^T \mathbf{W}_O \mathbf{T}_O] - \beta_U &= (\mathbf{X}_{UO(U)}^T W_O \mathbf{X}_{UO(U)})^{-1} \mathbf{X}_{UO(U)}^T \mathbf{W}_O E[\mathbf{T}_O | \mathbf{X}_O, \mathbf{v}, \mathbf{R}] - \beta_U \\
&= (\mathbf{X}_{UO(U)}^T W_O \mathbf{X}_{UO(U)})^{-1} \mathbf{X}_{UO(U)}^T \mathbf{W}_O \left( X_O \beta + \left[ \frac{\sum_{j: R_j \in \mathcal{R}} \pi_{ij} (1 - G_j(X_i^T \beta, X_i, v_i)) (\tau^2 + v_i) \sum_{\mathcal{D}_j} \psi_{kj} f'_{kj}}{\sum_{j: R_j \in \mathcal{R}} \pi_{ij}} \right] \right) - \beta_U \\
&= (\mathbf{X}_{UO(U)}^T W_O \mathbf{X}_{UO(U)})^{-1} \mathbf{X}_{UO(U)}^T \mathbf{W}_O \left( X_E \beta_E + \left[ \frac{\sum_{j: R_j \in \mathcal{R}} \pi_{ij} (1 - G_j(X_i^T \beta, X_i, v_i)) (\tau^2 + v_i) \sum_{\mathcal{D}_j} \psi_{kj} f'_{kj}}{\sum_{j: R_j \in \mathcal{R}} \pi_{ij}} \right] \right)
\end{aligned}$$

## 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

$$T = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + u_i + e_i$$

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:

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

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

$$\begin{aligned} Bias(\hat{\beta}_0) &= \beta_2 \frac{\sum_{X_1=0, X_2=1} w_i}{\sum_{X_i=0} w_i} = \frac{w_{10}}{w_0} \\ Bias(\hat{\beta}_1) &= \beta_2 \left( \frac{\sum_{X_1=1, X_2=1} w_i}{\sum_{X_i=1} w_i} - \frac{\sum_{X_1=0, X_2=1} w_i}{\sum_{X_i=0} w_i} \right) = \beta_2 \left( \frac{w_{11}}{w_1} - \frac{w_{10}}{w_0} \right) + \beta_3 \frac{w_{11}}{w_1} \end{aligned}$$

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

[

]

Then the bias is given by [

]

## 9 Discussion

## 10 References

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