Exploratory Analyses for Missing Data in Meta-Analyses and Meta-Regression: A Tutorial

Abstract

While systematic reviews of substance abuse interventions hold great promise for informing what works for whom and under what conditions, such reviews must contend with missing data. Missing data can limit the accuracy of statistical analyses or the relevance of the evidence base and deciding how to handle missing data depends on why it is missing in the first place. In this tutorial, we examine methods for exploring missingness in a dataset in ways that can help identify the sources and extent of missingness, as well as clarify gaps in evidence. We demonstrate these methods on a meta-analysis of substance abuse interventions for adolescents to highlight areas where the evidence less clear and where statistical adjustments taken by the original authors appear to have been necessary.

Introduction

Systematic reviews of substance abuse research hold great promise for better understanding effects of interventions (Tanner-Smith, Wilson, & Lipsey, 2013; Tanner-Smith *et al.*, 2016; White *et al.*, 2010; Newbury-Birch *et al.*, 2018; Ramsey *et al.*, 2019; Yuvaraj *et al.*, 2019). Traditional meta-analyses summarize the results of ensembles of studies of interventions, typically reporting the average impact or variation across impacts. Alternatively, meta-regression is a statistical model analogous to standard linear regression, wherein effect estimates are regressed on covariates pertaining to those effects, including study- and effect-level information (Hedges, 1982a, 1982b; Cooper, Hedges, & Valentine, 2019). For example, a meta-analyst may examine how the effectiveness of interventions is related to the type of treatment (e.g., type of therapy provided), how or on whom it was implemented, or the context in which it was studied (see Cooper *et al.*, 2019). However, meta-analyses and meta-regressions frequently contend with missing data (Pigott, 2019). A vast literature on missing data methodology highlights the ways missingness can bias statistical inferences, examines conditions under which these biases can be corrected, and proposes various statistical procedures to do so (Rubin, 1976; Pigott, 2019, 2001a; Schafer & Graham, 2002; Graham, 2009; Little & Rubin, 2002; van Buuren, 2018).

A key assumption of many analysis methods for incomplete data (i.e., data with missing values) is that analysts understand *what* data is missing and *why* (i.e., the mechanism that drives missingness). Much of the literature on missing data has focused on the implications of this assumption (Pigott, 2019; Little & Rubin, 2002). Yet, outside of some statistical tests (e.g., Little, 1988), there is little guidance for forming and examining theories about missingness (Tierney & Cook, 2018). Various researchers have suggested analysts can better understand missingness in their data through exploratory analyses, including visual and numerical summaries (Cheng, Cook, & Hofmann, 2015; Buja, Cook, & Swayne, 1996) akin to classical exploratory data analyses (Tukey, 1962). These explorations, which occur before running confirmatory statistical analyses, can shed greater light on key issues relevant to missingness. Tools for doing so are only now emerging in statistics, but these tools have yet to gain broader traction in quantitative disciplines (Tierney, 2017; Tierney & Cook, 2018). Nor has this approach garnered much discussion among meta-analysts, where missing data is a common problem. The focus in meta-analysis has instead been on methods to estimate models with incomplete data (see Pigott, 2019; Cheung, 2019; Higgins, White, & Wood, 2008; Ellington *et al.*, 2015).

Because a meta-analysis involves an ensemble of effects (i.e., intervention impacts) reported by primary studies, there are at least three different types of missing data in a meta-analysis, each of which would require its own exploratory analysis. First, data could be missing on individual participants within studies, including their outcomes in the study or other characteristics (e.g., their age, race, prior substance use) (e.g., Higgins *et al.*, 2008). Although this type of missingness cannot be controlled by a meta-analyst, it may impact the precision of each primary study's effect size estimates. Second, entire studies or effects may be missing from a meta-analytic dataset if they are not reported or published, including for reasons related publication selection and bias (Rosenthal, 1979; Hedges, 1984; Rothstein, Sutton, & Borenstein, 2005). Third, and the focus of this article, missingness may refer to information that could not be extracted from a completed study by a meta-analyst (Pigott, 2001a). This may occur if a study fails to report enough detail for analysts to back out effect estimates, standard errors, or study- and effect-level characteristics. When this occurs, it is often reasonable to reach out to the authors of primary studies with requests for missing information. However a recent investigation found that responses rates to such inquiries are low (around 12%) and responses that include the missing information are even less frequent (about 0.5%) (Polanin *et al.*, 2020)

This tutorial demonstrates exploratory analysis methods for studying missingness that arises from incomplete reporting in scientific research as described above. We refer to this type of exploration as an *exploratory missingness analyses* (EMA). The focus of the tutorial involves a dataset on substance abuse interventions typical of those used in meta-regression models (Tanner-Smith *et al.*, 2016). The following

section describes these data. We then discuss some relevant considerations for EMA and demonstrate an EMA on these data.

Data Used for Tutorial

Tanner-Smith *et al.* (2016) examined the impacts of substance abuse interventions for adolescents on subsequent substance use. Among a final analytic dataset of $k = 95$ effect estimates from 61 studies, Tanner-Smith et al. identified a range of intervention types that have been studied in different venues and on different types of adolescent substance users. These interventions focus on cognitive behavioral therapy (CBT), family therapy, and pharmacological therapy. Individuals in studies presented using marijuana, alcohol, and/or opioids, and were from diverse racial and socioeconomic backgrounds. The effect size used in the meta-analysis compares two groups of study participants (referred to here as *Group 1* and *Group 2*). Some reported effects contrasted a given intervention with a placebo or with a "usual care" condition where individuals received services but no explicit drug treatment (e.g., youth in residential care who receive standard residential services but not drug treatment). Others contrasted two alternative interventions or implementations. Tanner-Smith et al. documented the intensity (in hours per week), duration (in days), and location (in- our out-patient) of treatment for each group. Finally, Tanner-Smith et al. recorded effect estimates at different time points after intake. In all, their raw data totaled some $k = 328$ effect estimates and $p = 43$ variables for each effect, including the effect estimates and standard errors. We are grateful to Tanner-Smith et al., who furnished their raw data, which will be used to illustrate useful tools to exploring missingness in this tutorial.

Tanner-Smith et al. fit meta-regression models to their data in order to examine how treatment impacts varied according to the type of therapies and individuals studied. They found that assertive continuing care (ACC), behavioral therapy, (CBT), motivational enhancement therapy (MET), and family therapy tended to be more effective than generic "practice as usual" interventions that often involved referrals to community services. However, they found no association between characteristics of the adolescents in the studies and intervention effectiveness controlling for intervention type. In other words, intervention effectiveness was unrelated to the background characteristics of the adolescents in treatment.

A complicating factor in conducting these analyses was that some of the data were missing, especially for the covariates included in the meta-regression. For instance, not all studies reported treatment intensity (in hours per week) or the racial makeup of each group in the study. To address this issue, Tanner-Smith et al. opted to estimate their meta-regression models using the expectation-maximization (EM) algorithm, which has been an important tool for analyzing incomplete data (Dempster, Laird, &

Rubin, 1977; Graham, Cumsille, & Elek-Fisk, 2003). The EM algorithm has also been studied as a useful approach to estimation when missing covariates in a statistical model, which was primarily the issue facing Tanner-Smith et al. (Ibrahim, 1990; Ibrahim, Lipsitz, & Chen, 1999).

There are several alternatives to the EM algorithm. Pigott (2019) discusses methods applicable to meta-analyses such as multiple imputation and full information maximum likelihood (FIML), as well as complete- and available-case analysis methods. Each of these methods, including EM, rely on assumptions about the reason data are missing, and their effectiveness can be limited by how much data are missing (Rubin, 1976, 1987; Pigott, 2001b; Schafer & Graham, 2002; Graham, 2009; Little & Rubin, 2002; van Buuren, 2018).

Exploratory Missingness Analyses

The tools discussed in this article facilitate EMA to quantify and visualize missingness in a meta-analytic dataset (Cheng *et al.*, 2015; Buja *et al.*, 1996; Tierney & Cook, 2018). The point of conducting an EMA is to better understand the pattern and potential impact of the missing data in a meta-analysis, which can aid researchers to make appropriate choices about an analysis strategy. Researchers can explore whether assumptions about the missingness mechanism are defensible and can also highlight areas where evidence is sparse. For example, meta-analysts may hypothesize that average age of the study sample may relate to the effectiveness of an intervention but find that studies report average age in various ways. Looking closely at the data collected in a meta-analysis affords opportunities to create moderators (i.e., covariates in meta-regression model) based on information reported more frequently across studies (Pigott & Polanin, 2020). EMAs can also highlight gaps in the evidence base by showing what information (e.g., on treatment frequency or setting) is missing from a systematic review. However, it is worth noting two key aspects of EMA to better contexualize their use.

First, it will often be difficult to draw very strong conclusions about missingness mechanisms based on exploratory analyses. Even proposed tests for missingness mechanisms must be interpreted with caution (Molenberghs *et al.*, 2008; Seaman *et al.*, 2013; Rhoads, 2012). Instead, EMA can provide support for or help to generate theories that explain missingness in ways that are consistent with the assumptions required of incomplete data analysis methods. Some of these theories may require consultation with data curators and other individuals who extracted information from the studies reviewed.

Second, there is no single visualization or set of metrics guaranteed to provide a complete picture of missingness for all datasets. A plot that is tremendously useful for one dataset may be of less interest for others. Any EMA must rely on knowledge of how data were collected and extracted, and can help

leverage that knowledge to examine theories about missingness. Because this process is about exploration, rather than confirmation, it will be subject to the analysts' judgment.

In the following sections, we present and discuss an example EMA of Tanner-Smith's et al. data on substance abuse interventions for adolescents. To simplify presentation, we focus on $p = 20$ variables relevant to the analyses conducted by Tanner-Smith et al. (2016). This example serves to highlight some potential techniques, but it is not exhaustive. The supplementary material to this tutorial includes a vignette that presents and describes alternative visualizations and numerical summaries of missingness. Further, though the focus of these data and the resulting EMA pertain to meta-regression, the same general approaches can be used for other meta-analytic data. Both the demonstration presented in this article and the supplementary vignette are implemented in the `R` software language and draw heavily on the `visdat` and `naniar` libraries with some custom extensions developed specifically for meta-analysis (Tierney, 2017; Tierney & Cook, 2018). Code is included with the supplementary materials.

Notation

Assume a literature search reveals k effect estimates and data is collected on p variables regarding each effect (including the estimate itself). This can be summarized and stored in a $k \times p$ table where rows correspond to effect estimates and columns correspond to variables concerning those estimates. One column would contain the effect estimates themselves, and another would contain the standard error or estimation error variance of those estimates. The remaining $p - 2$ columns could contain effect- or study-level covariates, including summary demographics (e.g., the percent of a study's sample that were minorities), treatment type (e.g., behavioral therapy versus pharmacological interventions), or dosage/duration of an intervention. These tables are used by most standard meta-analysis software, including Comprehensive Meta-Analysis, the `metafor` library, or OpenMetaAnalyst (Viechtbauer, 2010; Borenstein *et al.*, 2012; Trikalinos, 2012).

The structure of such tables is shown in the matrix **D** below. In the matrix, T_i denote the effect size estimates and σ_i are their standard errors. The X_{ij} refer to additional variables collected that pertain to a given effect size and that might be used in an analysis. Note that X_{ij} may be effect- or study-level covariates. In a meta-regression, the effect estimates T_i are regressed on the covariates X_{ij} using a weighted least-squares approach (see Cooper *et al.*, 2019).

$$
\mathbf{D} = \begin{bmatrix}
T_1 & \sigma_1 & X_{11} & ... & X_{1,p-2} \\
T_2 & \sigma_2 & X_{21} & ... & \mathbf{NA} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
T_k & \sigma_k & \mathbf{NA} & ... & X_{k,p-2}
\end{bmatrix}
\qquad
\mathbf{R} = \begin{bmatrix}
1 & 1 & 1 & ... & 1 \\
1 & 1 & 1 & ... & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 1 & 0 & ... & 1
\end{bmatrix}
$$

Missing data in this context would refer to individual cells in the table that are missing values, and are denoted in $\mathbf{D}$ as $\mathbf{NA}$. To gain insight about missingness, we can augment the traditional dataset $\mathbf{D}$ with a matrix $\mathbf{R}$, which is comprised of response indicators R_{ij}. These indicators take a value of $R_{ij} = 1$ if entry i, j in the dataset $\mathbf{D}$ is observed, and $R_{ij} = 0$ if it is missing. For instance, $R_{21} = 1$ because the first element of row 2 (T_2) is observed, but $R_{2p} = 0$ because the last element of row 2 is missing ($\mathbf{NA}$).

Note that there are both *observed* and *unobserved* data. The observed data comprises all of the entries in $\mathbf{D}$ for which $R_{ij} = 1$, and we denote it as $\mathbf{D}_{obs}$. For instance, the observed data would include the effect estimates T_i and standard errors σ_i. The unobserved data comprises all of the entries of $\mathbf{D}$ for which $R_{ij} = 0$, including $X_{2,p-2}$ and $X_{k,1}$. Denote the unobserved data as $\mathbf{D}_{mis}$. EMA differ from traditional exploratory data analyses (EDA) because they focus on missingness indicators in $\mathbf{R}$, as well as the relationship between those indicators and the observed data $\mathbf{D}_{obs}$. Many software tools, including most graphics software used to conduct a standard EDA omit observations with missing values, which would eliminate information about missingness (Tierney, 2017).

Numerical Summaries of Missingness

One issue with missingness involves how much data is missing, which can affect the accuracy of incomplete data analysis methods like the EM algorithm, multiple imputation, or even complete-case analyses. Schafer (1999) suggests that if less than 5% of the data are missing, this can be inconsequential, and Bennett (2001) argues that a missingness rate of over 10% can induce non-negligible bias in statistical analyses. Statistical guidance regarding the effectiveness of procedures such as multiple imputation or FIML typically typically suggest that they are appropriate when missingness is less than 40% (Dong & Peng, 2013; Jakobsen *et al.*, 2017). However, multiple imputation that involves informative auxiliary variables to impute missing data can perform well at higher rates of missingness under certain assumptions (Madley-Dowd *et al.*, 2019).

Missingness rates could refer to several different quantities in the datasets described in the previous section. First, we could be interested in the total fraction of cells with missing values:

$$\sum_{i=1}^{k} \sum_{j=1}^{p} \frac{1 - R_{ij}}{kp} \tag{1}$$

In the subset of the Tanner-Smith et al. (2016) data, this value is 11.6%.

Second, we can compute the proportion of effects missing *any* variables:

$$\sum_{i=1}^{k} \frac{1 - \prod_{j=1}^{p} R_{ij}}{k} \tag{2}$$

In the data, 73.8% of rows are missing at least one value, Thus, a complete-case analysis of all variables would retain only about 26% of the rows in the data.

Third, and potentially more relevant for meta-regression, we may wish to know the percentage of effects missing a given variable (i.e., how much of each column is missing):

$$\sum_{i=1}^{k} \frac{1 - R_{ij}}{k} \tag{3}$$

As an example, just over 24% of the rows in the data are missing the intensity of Group 1's treatment (in hours per week). The proportion of effects missing other variables are displayed in Table 1.

An alternative to quantifying missingness in terms of raw percentages is to use weighted percentages. In a meta-analysis, parameters are often estimated by weighting effect size estimates in such a way that the most precisely estimated effects (i.e., that have the smallest standard errors) receive the most weight (Borenstein, 2009; Cooper *et al.*, 2019). The precision with which we estimate important quantities in a meta-analysis, including meta-regression coefficients, will therefore depend on the precision of the studies included in an analysis. That is, we will have better estimates of a meta-regression model if the precision of each effect $1/\sigma_i^2$ is large (Hedges & Olkin, 1985). Missing a variable for an effect estimate with a large standard error (and thus low precision) can potentially be less detrimental than missing the same variable for an effect estimate with a small standard error.

Thus, when standard errors of effect sizes are fully observed, an alternative way to quantify the extent of missingness is with a precision-weighted percentage. The weighted percentage

$$\sum_{i=1}^{k} \frac{\left(1 - \prod_{j=1}^{p} R_{ij}\right)/\sigma_i^2}{\sum_{i=1}^{k} 1/\sigma_i^2} \tag{4}$$

gives the fraction of information in a meta-analysis associated with effects that are missing any data. In the Tanner-Smith et al. example, effects that are missing any covariates make up roughly 74.4% of the total precision in the data.

The quantity

$$\sum_{i=1}^{k} \frac{(1 - R_{ij})/\sigma_i^2}{\sum_{i=1}^{k} 1/\sigma_i^2} \tag{5}$$

is the percentage of information associated with effects missing a specific variable. If the weighted percentage (4) is greater than the raw percentage in (2) or if (5) is greater than (3), that would indicate that a missingness problem may be more acute because data is missing from larger studies.

Raw percentages and precision-weighted percentages are presented in Table 1. We would discourage interpreting the size of the differences between the raw and weighted percentages, however comparing those columns can identify variables missing from larger studies, which typically receive more weight in a meta-analysis. For example, the raw percentage column in Table 1 would suggest missingness in the hours per week that Group 1 spent in treatment (24% missing) may be much less acute than missingness in the hours per week Group 2 spent in treatment (over 46% missing). But the weighted percentage indicates that the effects for which Group 1's hours per week variable is missing make up nearly 37% of the total precision of effect estimates. Hence, excluding those effects in a complete-case or available-case analysis would reduce how accurately the relationship between Group 1's treatment intensity and the intervention's impact can be assessed. This reduction in accuracy would likely be greater than what is indicated by the raw percentages.

[Table 1 about here.]

Visual Displays of Missingness

Visualizations of missingness can help provide context to the numerical summaries in the previous section. In addition, they can help analysts uncover key missingness patterns, as well as examine relationships between missingness and observed data. This section examines missingness in the Tanner-Smith et al. (2016) data using a series of plots.

Aggregation Plots

Aggregation plots provide a high-level picture of missingness in a dataset, and are a useful starting point for EMA visualizations. Figure 1 shows an aggregation plot for the Tanner-Smith et al. (2016) data. The plot is laid out exactly like the data structure: The columns correspond to variables in the data, and rows correspond to effect sizes. Dark areas correspond to cells that are missing values.

[Figure 1 about here.]

Aggregation plots can identify which columns do not have missing values, such as the columns corresponding to the effect size estimates, standard errors, or study designs (Figure 1). They also show which columns or groups of columns contain many missing values. In particular, Figure 1 appears to show three general kinds of missingness patterns. First, studies are missing information on the treatment in-

tensity (hours per week and duration) for Group 1 or Group 2. Note that occasionally this information is missing for both groups, as with the rows near the top of the plot. Second, studies are missing information on the demographic makeup (percent of the group that is white, black, Hispanic, or male) for Group 1 and Group 2 simultaneously. Finally, for a number of rows in the middle of the data, it appears that studies are missing information both on Group 2's treatment intensity and demographics.

Figure 1 also displays some numerical summaries regarding the extent of missingness in the data. In the legend, we see that over 11% of all cells are missing values in the data (equation (1)). Figure 1 also reports the percent of each column that is missing, which aligns with the raw percentages in Table 1.

Variable Missing Plots

While aggregation plots can provide a good overview, we typically want more detailed information about how many observations are missing a given variable. *Variable missing plots* display the overall missingness in each column of a dataset, and can be used as an analog to the numerical summaries in Table 1. From variable missing plots, it is often easy to identify variables that might be driving any missing data problems, and they can quantify the extent to which a given column has missing values on the scale of raw percentages.

Figure 2 shows a variable missing plot for the Tanner-Smith et al. (2016) data. Each variable is listed on the y-axis, and the line extending along the x-axis indicates the fraction of rows for which a variable is missing. Figure 2 suggests that the hours per week that Group 2 spent in their assigned treatment is missing for almost half of the effects in the data, while the hours per week that Group 1 spent in treatment was missing for far fewer rows—roughly half as many. Further, the percentage of each group that is black or Hispanic is missing more frequently than is the percentage of each group that is white. That is, studies reported the breakdown of white/non-white adolescents in their studies more frequently than they reported the percentage of adolescents identifying as a specific non-white race.

[Figure 2 about here.]

Plots for Missingness Patterns

Understanding missingness patterns can be crucial for statistical analysis methods of incomplete data. Such methods often rely on relationships between variables in the observed data as part of their adjustments for missing data. For instance, imputation models typically use many (if not all) available variables to impute missing values. If variables that are seemingly closely related are often missing together, this can limit the accuracy of such methods.

One useful plot for unpacking missingness patterns is to examine the correlation between miss-

ingness in one variable versus missingness in another variable (i.e., the correlation between $1 - R_{is}$ and $1 - R_{it}$ for $s \neq t$). Figure 3 shows a heatmap of pairwise correlations of missingness between variables. Each tile corresponds to a pair of variables, and tiles are shaded according to the correlation of missingness between those variables. Figure 3 shows a few clusters of variables that tend to be missing together. First, information about Group 2's treatment, including duration and intensity, are frequently missing from the same effects. Second, if information on the percentage of minorities is missing for one of the groups (e.g., Group 1), then it is very likely to be missing for the other group (e.g., Group 2). Finally, missingness in Group 2's treatment, including location, duration, and intensity, is positively correlated with missingness in Group 1's treatment duration (in number of sessions), and is negatively correlated with missingness in Group 1's treatment intensity (in hours per week).

[Figure 3 about here.]

A more fine-grained approach to studying missingness patterns uses an *upset plot*, which is shown in Figure 4 (Conway, Lex, & Gehlenborg, 2017). The bottom of Figure 4 presents different variables and indicates how many rows each of those variables is missing from. The dots along the bottom panel indicate different patterns of missingness, which means that a given set of variables are missing from the same row(s). The bars in the top panel of Figure 4 show the frequency with which these patterns occur.

The bottom four variables listed in Figure 4 involve the treatment duration and intensity for Group 2. Four of the top five variables listed involve the racial composition of the studies. Judging by the bottom panel in the figure, there are few different patterns worth noting. The first pattern, in which much of the information about the treatment condition in Group 2 is missing. However, that pattern also occurs in conjunction with other patterns. For instance, the sixth pattern involves rows that are missing information about Group 2's treatment, as well as the racial composition of the study.

In the top panel, we see how frequently each pattern occurs. The first pattern, which largely contains information about Group 2's treatment duration and intensity, occurs for 38 rows in the data. However, because that pattern is part of other patterns, it occurs frequently in the data. Judging from Figure 4, many rows are missing information about Group 2's treatment, and several of those rows are also missing other variables, including variables that describe the demographics of the study participants.

[Figure 4 about here.]

Plots to Relate Missingness to Observed Values

Explorations of missingness can also examine whether missingness in one variable is related to observed values in another variable. This can highlight potential biases in the observed data, and give insight into the mechanism of missingness. The *missingness mechanism* is an important consideration for incomplete

10

data analyses. Rubin (1976) classified three different possible types of mechanisms that relate the probability that a value is observed $\mathbf{R}$ to the observed and unobserved data $\mathbf{D}_{obs}$ and $\mathbf{D}_{mis}$. Data are missing completely at random (MCAR) if the probability that values are observed is independent of all observed and unobserved data $p(\mathbf{R}|\mathbf{D}_{obs}, \mathbf{D}_{mis}) = p(\mathbf{R})$. Data are said to be missing at random (MAR) if the probability that values are observed only depenends on observed data $p(\mathbf{R}|\mathbf{D}_{obs}, \mathbf{D}_{mis}) = p(\mathbf{R}|\mathbf{D}_{obs})$. Finally, data are missing not at random (MNAR) if the probability of observing a value is related to unobserved and observed data.

Pigott (2019) notes that some meta-analytic methods, such as analyses of complete cases, assume that data are MCAR. Other methods, such as multiple imputation or the EM algorithm, assume that data are MAR. If missingness in one column of the data is correlated with observed values in another column, that would be an indication that data are not MCAR. A similar logic underpins the test for MCAR proposed by Little (1988). While the visualizations presented in this section can be suggestive of MCAR or MAR mechanisms, they are not conclusive about the mechanism.

Heatmaps, like that in Figure 5 can help demonstrate the correlation between missingness in one column and observed values of another categorical variables. Figure 5 plots the missingness rate for each variable as it relates to Group 1's treatment location. The figure is grouped into tiles. The columns of the figure correspond to the different types of care provided to Group 1, shown along the bottom of the graph: inpatient, outpatient, or continuing care. The rows of the figure correspond to other variables in the dataset. Each tile is shaded according to how frequently those variables are missing for each level of care provided.

Figure 5 shows that missingness in other variables is more common for effects where Group 1 received inpatient or continuing care treatment. In particular, studies where Group 1 received outpatient treatment were more likely to report racial demographics, information about Group 1's treatment intensity (in hours per week) and duration (in days and number of sessions). Likewise, studies where Group 1 received outpatient care were more likely to report characteristics of Group 2's treatment type, duration, and intensity. We should note, however, that the *Inpatient* column should be interpreted with some caution, as only five (5) effects in the raw data involve Group 1 receiving inpatient treatment. Still, the differences in missingness for effects where Group 1 received outpatient versus continuing care interventions suggest that missingness in several variables is related to the venue of treatment.

[Figure 5 about here.]

Similar patterns emerge in Figure 6, which shows that missingness in several variables is also correlated with the venue of Group 2's treatment. Note that Figure 6 includes an *NA* column, which indicates that Group 2's treatment location is itself missing. As with Figure 5, the *Inpatient* column is

11

difficult to interpret as it represents only one (1) effect in the raw data. However, comparing the *Outpatient* and *Continuing Care* columns reveals that studies where Group 2 received outpatient care were more likely to report other variables, particularly those related to treatment intensity (in hours per week) and racial composition. In addition, the *NA* column in Figure 6, which indicates that Group 2's treatment location is missing, also confirms a notable missingness pattern identified in Figure 4. The relevant rows in Figure 6 indicate that aspects about Group 2's treatment, including location, duration, an intensity, are often missing together.

[Figure 6 about here.]

Two crucial variables that are in nearly all meta-analyses are the effect size estimates and their standard errors σ_i (or variances σ_i^2). If missingness in a covariate is correlated with effect sizes or standard errors, this is likely to impact the analytic results. One way to explore such relationships are with *comparative density plots* that present the distribution of effect estimates and standard errors among effects for which a covariate is missing versus when the covariate is observed. Figure 7 displays several density plot pairs for different variables. For each pair of plots we see the distribution of effect estimates (left) and standard errors (right), each colored according to whether a given covariate is missing.

Figure 7A, shows the relationship between the missingness in Group 1's treatment intensity (in hours per week) and the distribution of effect estimates and standard errors. Effects are on the scale of Cohen's *d*. From Figure 7A, we can see that effect estimates tend to be slightly smaller and have smaller standard errors when Group 1's treatment intensity is missing than when it is observed. This is consistent with the weighted percentages reported in Table 1, which found that missingness in Group 1's treatment intensity occurred with studies with greater precision.

Contrast that with Figure 7B, which focuses on whether Group 1's treatment duration (in days) is missing or not. Figure 7B shows that effect estimates for which Group 1's treatment duration is missing tend to be larger and have larger standard errors than effects for which duration of treatment is reported. Plots A and B in Figure 7 suggest that missingness of information about treatment dosage (hours per week and duration) will be related to the size of effects found and how precisely those effects were estimated.

[Figure 7 about here.]

Figure 8 shows an analogous set of plots comparing the effect size estimate and standard error distributions for when Group 2's treatment intensity (in hours per week) and duration (in days) are missing or observed. Figure 8A shows that when Group 2's treatment intensity is observed, effect sizes and standard errors are slightly smaller than when Group 2's treatment intensity is missing. As well, Figure 8B suggests that effect estimates and standard errors are slightly smaller when Group 2's treatment dura-

tion is observed compared to when it is missing. Comparing Figures 7 and 8 reveals that missingness in variables pertaining to Group 1's treatment has a stronger relationship with the effect size estimates and standard errors than does missingness in variables pertaining to Group 2's treatment. This suggests that omitting effects for which Group 1's treatment duration or intensity are missing would seemingly have a stronger impact on an analysis.

[Figure 8 about here.]

Discussion

Missing data is and will continue to be an issue with most meta-analyses that can affect what we can learn about substance abuse interventions from research syntheses. Attempts to recover missing information by contacting primary study authors should be encouraged, however empirical research suggests that such inquiries are unlikely to resolve missingness entirely (see Polanin & Terzian, 2019; Polanin *et al.*, 2020). While there are various potential approaches to handling missing data in meta-analysis, most of those approaches assume that the missingness mechanism is known to the analyst. This is almost never the case in practice, and instead analysts are often left to hypothesize possible mechanisms. This article argued that an exploratory analysis of missingness might help analysts better understand and form hypotheses about missingness. It also outlined and demonstrated some tools that can support exploratory analyses into the scale and correlates of missingness in a meta-analytic dataset. These numerical and visual tools are a useful first step to understanding why data is missing and how data is structured.

These tools were applied to data on a large meta-analysis conducted by Tanner-Smith *et al.* (2016) on substance abuse interventions for adolescents. We found 73.8% of the effect sizes were missing at least one of their corresponding covariates. This was driven by some variables that were missing frequently (e.g., Group 2 hours of treatment per week). Our analysis also revealed that missingness in some variables may be more severe than was obvious from first glance (e.g., Group 1 hours of treatment per week). Variables quantifying the intensity and duration of treatment in a study were frequently missing together. Finally, we identified some variables whose missingness appears to be related to the size and standard errors of effect size estimates (e.g., Group 1 hours of treatment per week and treatment duration in days), which suggests that missingness was not MCAR.

Conducting an EMA as outlined here provides insight into the evidence base in a meta-analysis. In the example, we have scant information about the duration and intensity of Group 2 interventions and thus can make limited inferences about how treatment effectiveness varies as a function of the comparison group's treatment. Presenting EMA results can highlight both the gaps and the areas where effect

size models are best supported by the data. The EMA results also demonstrate the complex relationships between covariates in a meta-regression and the potential for confounding as discussed by Tipton, Pustejovsky, and Ahmadi (2019a,b).

How to proceed from an EMA will depend on what is known about the data collection and missingness. Based on our analysis, we would be cautious of using analysis methods that assume data are MCAR, such as complete-case analysis or shifting units of analysis (Cooper, 2017). Tanner-Smith et al. (2016) used the EM algorithm to estimate their meta-regression models, which assumes data are MAR. This is consistent with our findings, and it is a common assumption made in analyses of incomplete data.

The analyses presented, while suggestive, do have several limitations. First, data curators and analysts who extract data for a meta-analysis can and perhaps should play a larger role in EMA. While our post-hoc analysis had limited input from these individuals, they will almost certainly have some insight about what made particular variables difficult to extract from the literature, and why that may have occurred. Data curators and analysts can also use the information from EMA to consider alternative ways to create variables from the data that is provided. For example, if average age is missing frequently from studies, analysts may be able to create a new variable that draws on the various ways age is recorded by included studies.

Second, it will be impossible to distinguish between data that are MAR or MNAR using the methods demonstrated in this article. This is because MNAR assumes that missingness is related to data that is not actually observed. Thus, in order to study or confirm whether data are MNAR would require some knowledge of the unobserved data. Analogous limitations have been noted in tests for whether data are MAR or MNAR (Molenberghs *et al.*, 2008; Rhoads, 2012).

The methodology discussed in this tutorial could be used to create different visualizations than were presented in this paper. Our complementary material develops on these results with a vignette that contains further visualizations and executable code implemented in the R computing language. Even though the data on substance abuse interventions for adolescents has a particular structure with information at the effect size level for each study, the tools exposed in this tutorial can be easily applied to other dataset structures.

References

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

Borenstein M. (2009) *Introduction to meta-analysis.* Chichester, U.K.: John Wiley & Sons.

Borenstein M, Hedges LV, Higgins JP, Rothstein H. (2012) Comprehensive Meta-Analysis Version 3.0. https://www.meta-analysis.com/downloads/Meta-Analysis%20Manual%20V3.pdf.

Buja A, Cook D, Swayne DF. (1996) Interactive high-dimensional data visualization. *Journal of Computational and Graphical Statistics* **5**: 78.

van Buuren S. (2018) *Flexible Imputation of Missing Data, Second Edition*, 2nd edn. Second edition. | Boca Raton, Florida : CRC Press, [2019] |: Chapman and Hall/CRC.

Cheng X, Cook D, Hofmann H. (2015) Visually exploring missing values in multivariable data using a graphical user interface. *J Stat Soft* **68**.

Cheung MW-L. (2019) Handling missing covariates in mixed-effects meta-analysis with full-information maximum likelihood. Chicago, IL. http://www.srsm.org/uploads/4/6/1/3/46138157/abstract_-_mike_cheung.pdf [accessed 8 September 2020].

Conway JR, Lex A, Gehlenborg N. (2017) UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**: 2938–2940.

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

Cooper HM, Hedges LV, Valentine JC (eds). (2019) *Handbook of research synthesis and meta-analysis*3rd edition. New York: Russell Sage Foundation.

Dempster AP, Laird NM, Rubin DB. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* **39**: 1–22.

Dong Y, Peng C-YJ. (2013) Principled missing data methods for researchers. *SpringerPlus* **2**: 222.

Ellington EH, Bastille-Rousseau G, Austin C, et al. (2015) Using multiple imputation to estimate missing data in meta-regression. *Methods Ecol Evol* **6**: 153–163.

Graham JW. (2009) Missing data analysis: Making it work in the real world. *Annu Rev Psychol* **60**: 549–576.

Graham JW, Cumsille PE, Elek-Fisk E. (2003) Methods for handling missing data. In I.B. Weiner (ed), *Handbook of Psychology*, p. wei0204. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Hedges LV. (1982a) Fitting continuous models to effect size data. *Journal of Educational Statistics* **7**: 245.

Hedges LV. (1982b) Fitting categorical models to effect sizes from a series of experiments. *Journal of Educational Statistics* **7**: 119.

Hedges LV. (1984) Estimation of effect size under nonrandom sampling: The effects of censoring studies yielding statistically insignificant mean differences. *Journal of Educational Statistics* **9**: 61.

Hedges LV, Olkin I. (1985) *Statistical methods for meta-analysis.* Orlando: Academic Press.

Higgins JP, White IR, Wood AM. (2008) Imputation methods for missing outcome data in meta-analysis of clinical trials. *Clinical Trials* **5**: 225–239.

Ibrahim JG. (1990) Incomplete data in generalized linear models. *Journal of the American Statistical Association* **85**: 765–769.

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

Jakobsen JC, Gluud C, Wetterslev J, Winkel P. (2017) When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. *BMC Med Res Methodol* **17**: 162.

Little RJA. (1988) A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association* **83**: 1198–1202.

Little RJA, Rubin DB. (2002) *Statistical Analysis with Missing Data.* Hoboken, NJ, USA: John Wiley & Sons, Inc.

Madley-Dowd P, Hughes R, Tilling K, Heron J. (2019) The proportion of missing data should not be used to guide decisions on multiple imputation. *Journal of Clinical Epidemiology* **110**: 63–73.

Molenberghs G, Beunckens C, Sotto C, Kenward MG. (2008) Every missingness not at random model has a missingness at random counterpart with equal fit. *J Royal Statistical Soc B* **70**: 371–388.

Newbury-Birch D, Ferguson J, Landale S, et al. (2018) A systematic review of the efficacy of alcohol interventions for incarcerated people. *Alcohol and Alcoholism* **53**: 412–425.

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

Pigott TD. (2001a) Missing predictors in models of effect size. *Eval Health Prof* **24**: 277–307.

Pigott TD. (2001b) A review of methods for missing data. *Educational Research and Evaluation* **7**: 353–383.

Pigott TD, Polanin JR. (2020) Methodological guidance paper: High-quality meta-analysis in a systematic review. *Review of Educational Research* **90**: 24–46.

Polanin JR, Espelage DL, Grotpeter JK, et al. (2020) Locating unregistered and unreported data for use in a social science systematic review and meta-analysis. *Syst Rev* **9**: 116.

Polanin JR, Terzian M. (2019) A data-sharing agreement helps to increase researchers' willingness to share primary data: results from a randomized controlled trial. *Journal of Clinical Epidemiology* **106**: 60–69.

Ramsey AT, Satterfield JM, Gerke DR, Proctor EK. (2019) Technology-based alcohol interventions in primary care: Systematic review. *J Med Internet Res* **21**: e10859.

Rhoads CH. (2012) Problems with tests of the missingness mechanism in quantitative policy studies. *Statistics, Politics, and Policy* **3**.

Rosenthal R. (1979) The file drawer problem and tolerance for null results. *Psychological Bulletin* **86**: 638–641.

Rothstein H, Sutton AJ, Borenstein M (eds). (2005) *Publication bias in meta-analysis: prevention, assessment and adjustments.* Chichester, England ; Hoboken, NJ: Wiley.

Rubin DB. (1976) Inference and missing data. *Biometrika* **63**: 581–592.

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

Schafer JL. (1999) Multiple imputation: a primer. *Stat Methods Med Res* **8**: 3–15.

Schafer JL, Graham JW. (2002) Missing data: our view of the state of the art. *Psychol Methods* **7**: 147–177.

Seaman S, Galati J, Jackson D, Carlin J. (2013) What is meant by 'missing at random'? *Statist Sci* **28**:

257–268.

Tanner-Smith EE, Steinka-Fry KT, Kettrey HH, Lipsey MW. (2016) Adolescent substance use treatment effectiveness: A systematic review and meta-analysis. Office of Justice Programs.

Tanner-Smith EE, Wilson SJ, Lipsey MW. (2013) The comparative effectiveness of outpatient treatment for adolescent substance abuse: A meta-analysis. *Journal of Substance Abuse Treatment* **44**: 145–158.

Tierney NJ. (2017) visdat: Visualising whole data frames. *JOSS* **2**: 355.

Tierney NJ, Cook DH. (2018) Expanding tidy data principles to facilitate missing data exploration, visualization and assessment of imputations. *arXiv:180902264 [stat]*.

Trikalinos T. (2012) OpenMetaAnalyst: Powerful open-source software for meta-analysis. https://effectiv ehealthcare.ahrq.gov/products/open-meta-analyst/abstract.

Tukey JW. (1962) The future of data analysis. *Ann Math Statist* **33**: 1–67.

Viechtbauer W. (2010) Conducting meta-analyses in R with the metafor package. *J Stat Soft* **36**.

White A, Kavanagh D, Stallman H, et al. (2010) Online alcohol Interventions: A systematic review. *J Med Internet Res* **12**: e62.

Yuvaraj K, Eliyas SK, Gokul S, Manikandanesan S. (2019) Effectiveness of workplace intervention for reducing alcohol consumption: A systematic review and meta-analysis. *Alcohol and Alcoholism* **54**: 264–271.

Table 1: *This table displays the total number, percentage, and precision-weighted percentage of effect sizes that are missing a given variable.*

Variable	# Missing	% Missing	Wt. % Missing
Group 2 Hrs Per Week	152	46.3	45.2
Group 1 Hrs Per Week	79	24.1	36.8
Group 2 Treatment Days	117	35.7	32.6
Group 2 No. Session	111	33.8	30.2
Group 2 Treatment Location	109	33.2	30.0
Group 2 Pct Black	75	22.9	17.5
Group 1 Pct Black	71	21.6	16.1
Group 1 Pct Hispanic	56	17.1	12.7
Group 2 Pct Hispanic	51	15.5	12.4
Group 1 No. Session	35	10.7	6.9
Group 2 Pct White	18	5.5	4.6
Group 1 Pct White	11	3.4	3.1
Group 2 Mean Age	14	4.3	3.0
Group 1 Mean Age	7	2.1	1.5
Group 1 Pct Male	9	2.7	1.3
Group 1 Treatment Days	2	0.6	0.1
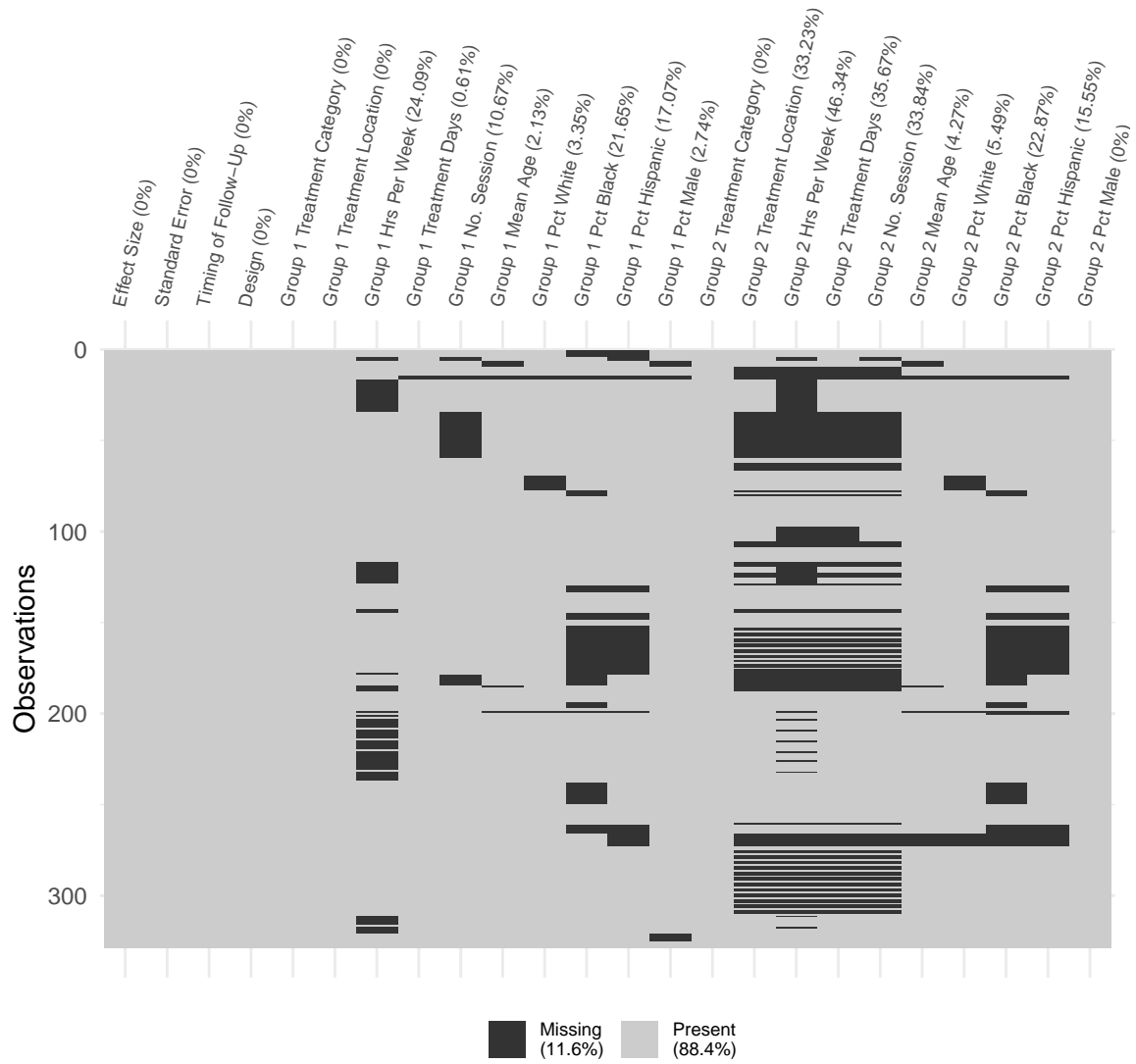
Figure 1: Aggregation plot. *This plot indicates the severity of missingness in the adolescent substance abuse intervention data (Tanner-Smith et al., 2016). Each row in the plot corresponds to a row in the data, and each column corresponds to a variable collected in the data. Missing cells in the data are indicated by a dark dash in plot. The legend shows the percent of cells in the data that contain missing values. The column labels show the precent of rows missing each variable in the data.*
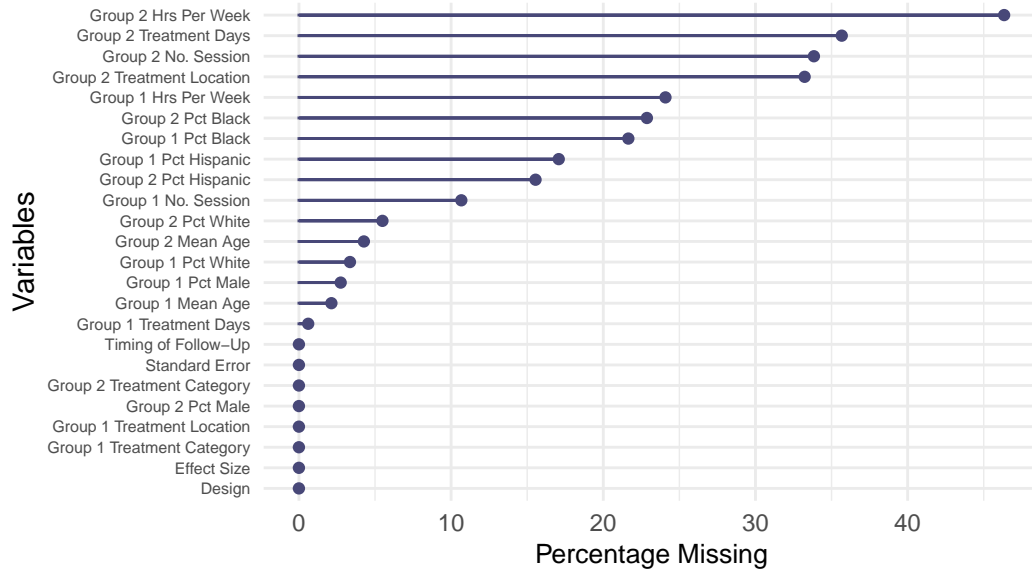
Figure 2: Variable missing plot. *This plot summaries missingness in variables, ordered by the percentage of missingness, in the adolescent substance abuse intervention data. Indicating that there are 10 variables with at least 10% of missing cases. This kind of visualization becomes relevant when deciding which variable to include in the analysis.*
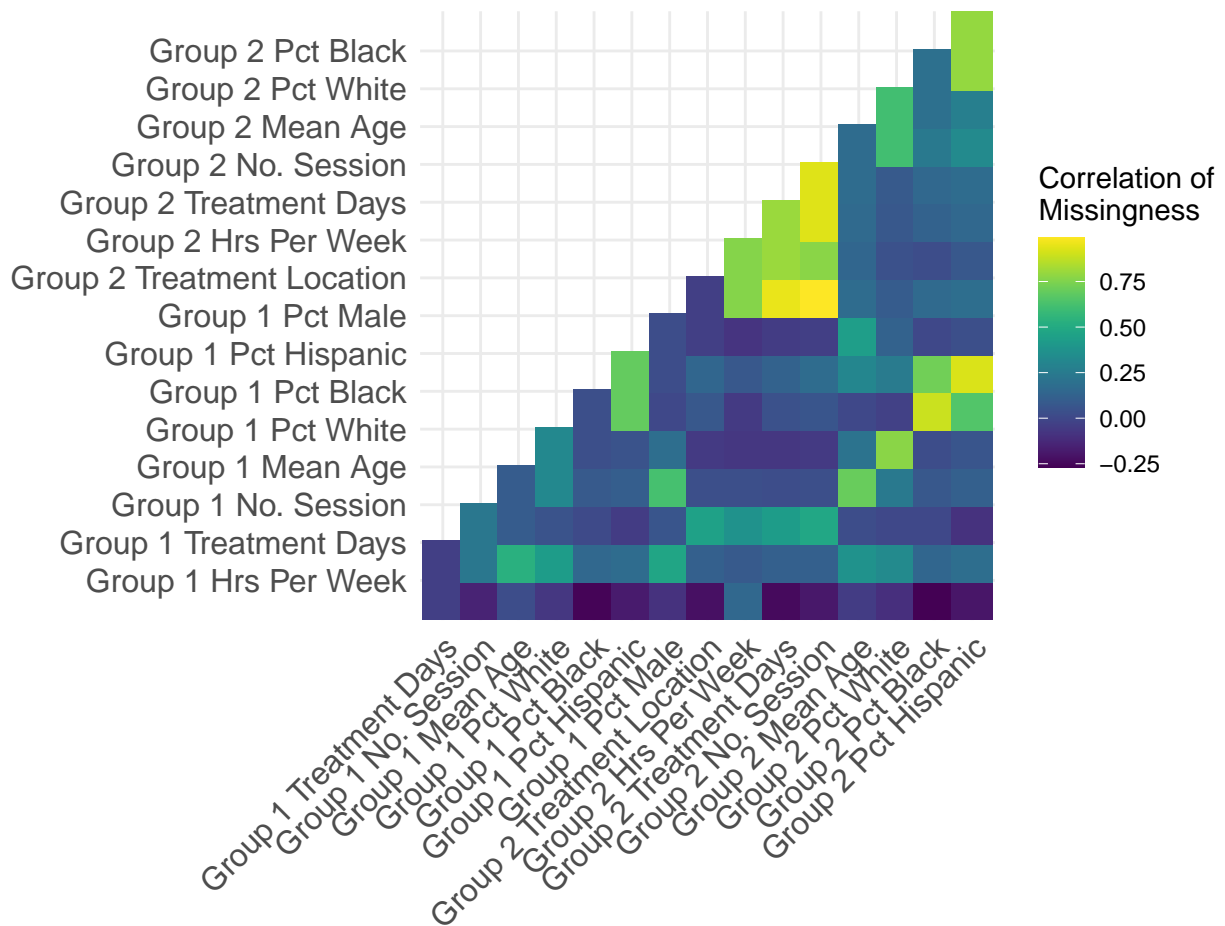
Figure 3: Missingness patterns plot. *This plot shows the pairwise correlation of missingnes for each variable in the data. Each tile refers to a pair of variables and is shaded according to the correlation.*
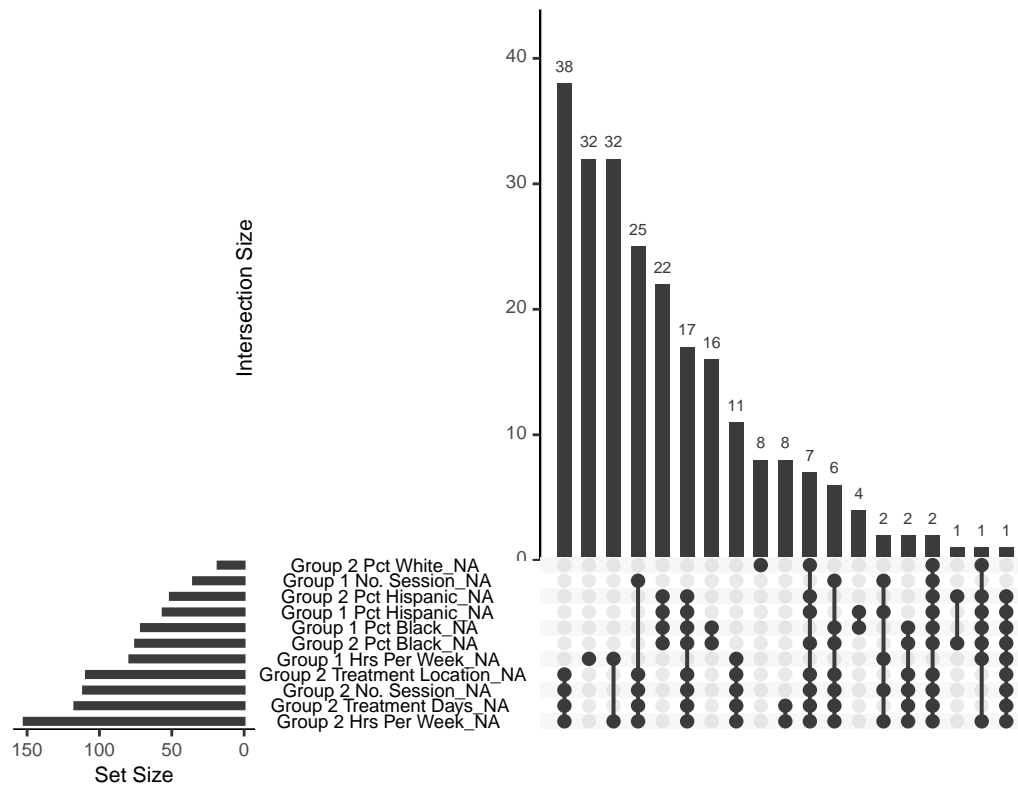
Figure 4: Upset plot. *This plot details those variables that are missing together. For instance, there are a large number of cases where group 2 level of care, number of sessions, treatment contact (hours per week) and duration of treatment (days) are missing together. This simple exploration provides valuable information for imputation.*
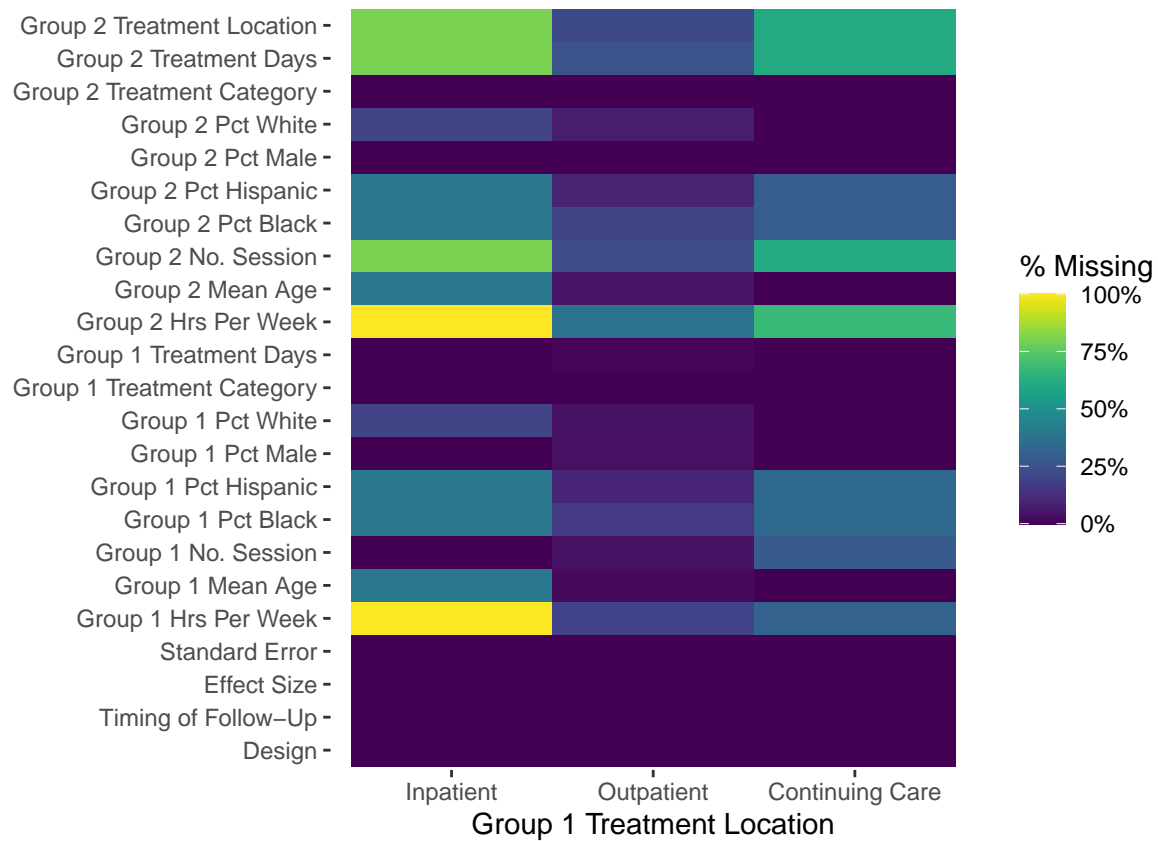
Figure 5: *This plot shows the rate of missingness for each variable as a function of Group 1's treatment location. Each column is broken down by where Group 1 received treatment (inpatient, outpatient, and continuing care). Each row represents another variable in the data. Tiles are shaded according the the fraction rows in the data for which each variable is missing for a given level of Group 1 treatment location.*
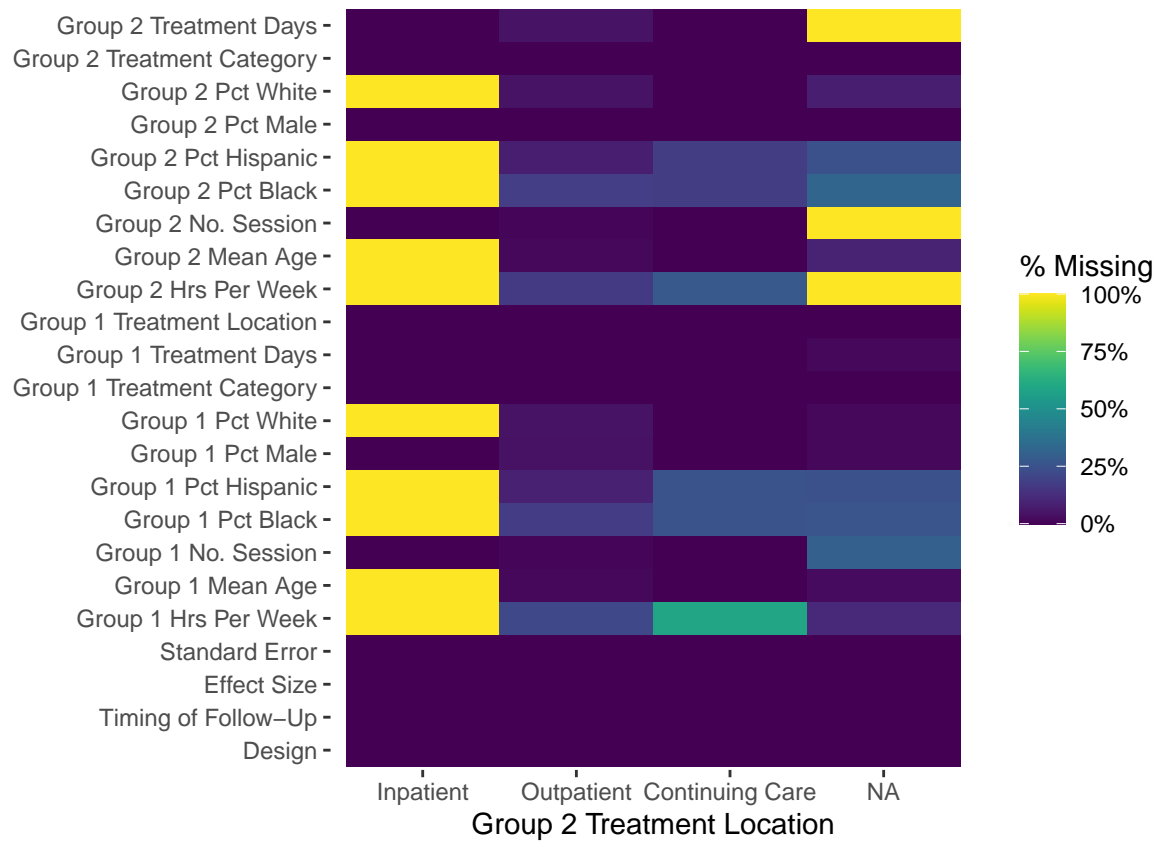
Figure 6: *This plot shows the rate of missingness for each variable as a function of Group 2's treatment location. Each column is broken down by where Group 2 received treatment (inpatient, outpatient, and continuing care). Each row represents another variable in the data. Tiles are shaded according the the fraction rows in the data for which each variable is missing for a given level of Group 2 treatment location.*
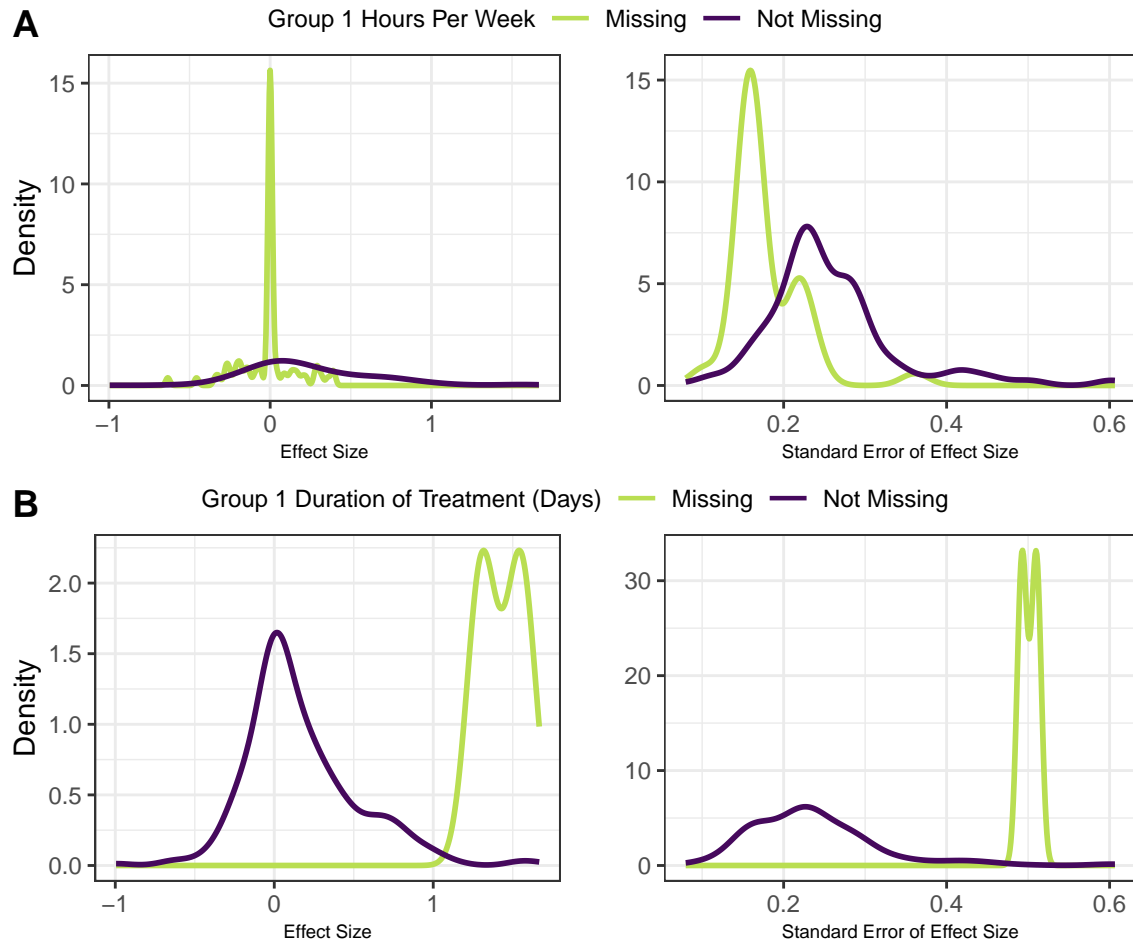
Figure 7: Comparative density plot. *This figure compares the distribution of effect size estimates and standard errors for when various covariates regarding Group 1 are observed versus missing. Plot (A) compares the distribution effect size estimates and standard erors for when Group 1 treatment intensity (hours per week) is missing versus observed. Plot (B) compares the distributions for when Group 1 treatment duration (in days) is observed versus missing.*
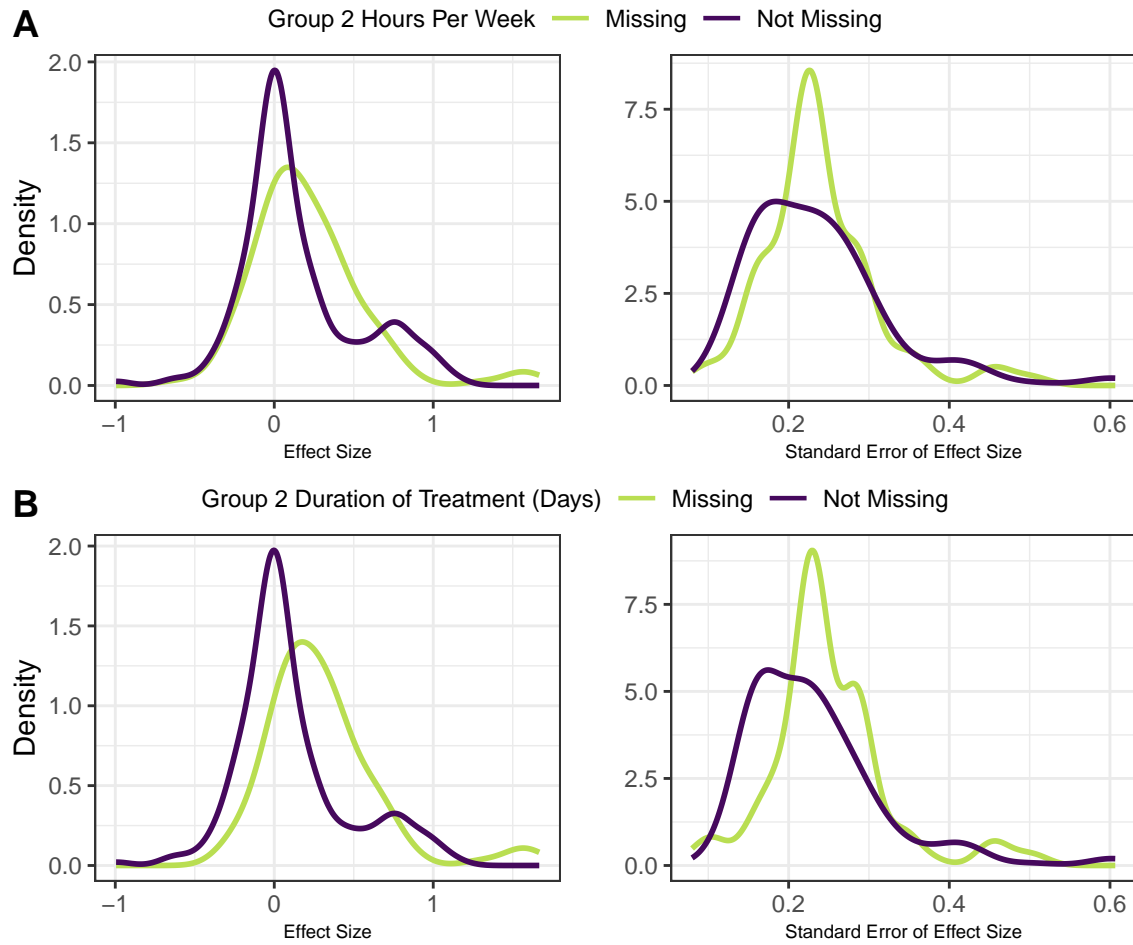
Figure 8: Comparative density plot. *This figure compares the distribution of effect size estimates and standard errors for when various covariates regarding Group 2 are observed versus missing. Plot (A) compares the distribution effect size estimates and standard erors for when Group 2 treatment intensity (hours per week) is missing versus observed. Plot (B) compares the distributions for when Group 2 treatment duration (in days) is observed versus missing.*