Putting the pieces together: Accounting for missing data in aggregate and Individual Participant Data Meta-Analysis

University of Ioannina
School of Education
Department of Primary Education

A Thesis
by
Christos Christogiannis

A thesis submitted in fulfilment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

Supervisor

Dimitris Mavridis	Department of Primary Education, University of Ioannina

Thesis committee

Ioannis Ntzoufras	Department of Statistics, Athens University of Economics & Business
Michail Tsagris	Department of Economics, University of Crete

External examination committee

Stavros Nikolakopoulos	Department of Psychology, University of Ioannina Department of Biostatistics, University Medical Center Utrecht
Georgios Ntritsos	Department of Economics, University of Ioannina
Georgios Markozannes	Department of Hygiene and Epidemiology, University of Ioannina Department of Epidemiology and Biostatistics, Imperial College London
Adriani Nikolakopoulou	Department of Hygiene, Aristotle University of Thessaloniki

Acknowledgement

I would like to start by expressing my deepest gratitude to my supervisor, Prof. Dimitris Mavridis. Thank you for your unwavering guidance, your patience during the difficult phases of this research, and for always pushing me to look at the data from a different perspective. Your mentorship has been the compass that kept this project on track; so, I could say nothing less than: "O Captain! My Captain!"

I also want to thank Prof. Ioannis Ntzoufras and Prof. Michail Tsagris for their time and valuable feedback throughout this process. I would also like to express my gratitude to the members of the external examination committee: Prof. Nikolakopoulos, Prof. Ntritsos, Prof. Nikolakopoulou and Prof. Markozannes. I am incredibly thankful for their thorough feedback and the wisdom they shared, which truly elevated this work. A special shout-out to Stavros Nikolakopoulos; thank you for the exceptional cooperation and for being a constant source of help and advice whenever I needed it.

Lastly, but certainly not least, I want to thank my family for their support all these years, the friends I shared so many moments with, and my co-workers for making the office such a nice place to be. There is a beautiful overlap between these groups, as these past five years have been truly unforgettable. To everyone who played a role—large or small—in this journey, I would like them to know that: I would choose you all again every day, and twice on Sunday.

Περίληψη

Η παρούσα διατριβή στοχεύει στη μελέτη και ανάπτυξη μεθοδολογίας για μοντέλα μετα-ανάλυσης σε περιπτώσεις ελλιπών τιμών, τόσο σε συγκεντρωτικά όσο και σε ατομικά δεδομένα. Τα ελλιπή δεδομένα είναι ένα συχνό φαινόμενο στη στατιστική και κατ' επέκταση στη μετα-ανάλυση. Η αντιμετώπισή τους προϋποθέτει τη διατύπωση συγκεκριμένων υποθέσεων, οι οποίες επηρεάζουν άμεσα τη διαδικασία μοντελοποίησης. Στην εισαγωγή παρουσιάζονται οι βασικές αρχές της μετα-ανάλυσης, η θεωρία των ελλιπών δεδομένων και οι διαθέσιμες μέθοδοι διαχείρισής τους στο πλαίσιο της μετα-ανάλυσης. Στην συνέχεια αναπτύσσεται η στατιστική μεθοδολογία για την εκτίμηση των ελλιπών τιμών σε ατομικά δεδομένα, με κατάλληλη τροποποίηση της διαδικασίας των πολλαπλών εκτιμήσεων. Η βασική ιδέα πίσω από τη συγκεκριμένη μεθοδολογία σχετίζεται με την ετερογένεια που παρατηρείται μεταξύ των μελετών οι οποίες συντίθενται στη μετα-ανάλυση. Δεδομένου ότι η ετερογένεια αυτή μπορεί να είναι σημαντική, κρίθηκε αναγκαίο να αναπτυχθεί μεθοδολογία σύμφωνα με την οποία οι ελλιπείς τιμές εκτιμώνται αξιοποιώντας κυρίως τις «κοντινές» ως προς τα χαρακτηριστικά τους μελέτες. Τέλος, όσον αφορά τις ελλιπείς τιμές σε συγκεντρωτικά δεδομένα, τόσο στη συμβατική μετα-ανάλυση όσο και στη μετα-ανάλυση δικτύου, αναπτύχθηκε πακέτο στη γλώσσα R. Το πακέτο στηρίζεται σε υπάρχουσες μεθοδολογίες οι οποίες χρησιμοποιούν pattern-mixture μοντέλα, με στόχο την αξιοποίηση πληροφοριών από ειδικούς του εκάστοτε πεδίου κατά την αντιμετώπιση των ελλιπών τιμών. Σκοπός του πακέτου είναι η διάδοση της σχετικής μεθοδολογίας και η διευκόλυνση των χρηστών στην εφαρμογή των συγκεκριμένων αναλυτικών διαδικασιών.

Abstract

This dissertation aims to study and develop methodology for meta-analytic models in the presence of missing data, both at the aggregate and individual-participant levels. Missing data are a common phenomenon in statistical analysis and, consequently, in meta-analysis. Their handling requires the formulation of specific assumptions, which directly influence the modelling process. The introduction presents the fundamental principles of meta-analysis, the theory of missing data, and the available methods for handling missingness within the meta-analytic framework. Subsequently, statistical methodology is developed for the imputation of missing values in individual participant data, through an appropriate modification of the multiple imputation procedure. The central idea underlying this methodology relates to the heterogeneity observed across studies synthesized in a meta-analysis. Given that such heterogeneity may be substantial, it was deemed necessary to develop an approach in which missing values are primarily imputed by borrowing information from studies that are 'close' in terms of their characteristics. Finally, with regard to missing data in aggregate-level analyses, both in pairwise meta-analysis and in network meta-analysis, an R package was developed. The package builds upon existing methodologies that employ pattern-mixture models and allows the incorporation of expert knowledge from the relevant field when addressing missing data. The aim of the package is to disseminate the existed methodology and to facilitate its practical implementation by researchers and analysts.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

In this chapter, we will present briefly the statistical methods of meta-analysis, presenting the basic models (common effect and random effects models), the different types of meta-analysis (aggregate data and individual participant data meta-analysis) and its extension to multiple interventions (network meta-analysis). We will also focus on the problem of missing data, the reasons that may lead to missing data, the types of missing data we may encounter in an Individual Participant Data Meta-Analysis (IPD MA) (sporadically and systematically missing), and the main methods we use to handle them along with the available packages in *R*.

1.1 Meta-analysis

Systematic reviews (SRs) of health technologies are crucial for informing clinical practices and healthcare decision-making, forming an essential part of comparative effectiveness research. For that reason, SRs must be exhaustive and comprehensive. For ensuring these two key characteristics, the following steps are suggested:

- defining the research question
- searching for published and unpublished literature in several databases using certain inclusion and exclusion criteria
- extracting the necessary data
- assessing the quality of studies and potential risk of bias sources
- summarizing the evidence either quantitatively or narratively
- interpreting and presenting the results.

Global entities such as the World Health Organization (WHO) and the National Institute of Clinical Excellence (NICE) consider them the foundation for evidence-based practices and policies. Detailed steps for conducting a systematic review can be found in the Cochrane handbook [Chandler et al. (2019)].

Different types of studies, including cohort studies, animal studies, observational studies, case-control studies, and randomized controlled trials (RCTs), can be synthesized in a systematic review to answer a research question. In this thesis, we focus on the synthesis of RCTs for evaluating the efficacy of interventions. It is widely agreed that RCTs are the most valid (gold standard) type of clinical trials [Stewart and Tierney (2002)]. In an RCT, individuals are randomly assigned to either an experimental group, which receives the intervention, or a control group, which typically receives a placebo. Any observed differences between the two groups can be attributed solely to the interventions received, as randomization and probability theory ensure that the groups do not differ with respect to any characteristic, either observed or unobserved, but the intervention received.

There are numerous individual studies addressing health, social, or educational issues, but deriving conclusions from these studies can be misleading due to potential biases or conflicting results across studies. SRs integrate multiple

studies to offer clearer conclusions for research questions, ensuring that clinical decisions are based on the entirety of the available evidence rather than a single study. The quality of systematic reviews is contingent upon the quality of the primary studies and the rigor of the review process. Poorly conducted systematic reviews may result in misleading conclusions.

Typically, the research question of each systematic review should be defined by some prespecified criteria, also known as PICO (Patient or Population – Intervention – Comparison – Outcome) criteria. Where the first component refers to patients' or population characteristics or problem been investigated, the second refers to the intervention(s) under consideration, the third to the control intervention, such as placebo, and the last one concerns the outcome(s) of interest, such as efficacy or safety outcomes.

Each study generates one or more effect sizes, which quantify the differences in outcomes between two or more groups. Classical examples of effect sizes include the Mean Difference and the Odds/Risk Ratio. These effect sizes are typically expressed with a confidence or credible interval, usually 95%, to reflect the uncertainty of the estimate. Meta-analysis is a statistical approach that synthesizes evidence extracted during the systematic review/process. One of the primary goals of meta-analysis is to synthesize the results from multiple studies to produce an overall, more precise effect size. Figure 1.1 shows how meta-analysis works conceptually.

It is common for several studies to produce contradictory results. However, a meta-analysis can quantify and investigate the reasons for the disagreement and resolve the conflicting results. The results of meta-analysis provide more precise results compared to those reported in individual studies, especially when there are many small trials. For example, there could be a meta-analysis with several small negative trials, but the overall effect can be clearly positive as it will be much more precise. Additionally, meta-analysis aims to identify how this effect size varies with certain study or individual characteristics.

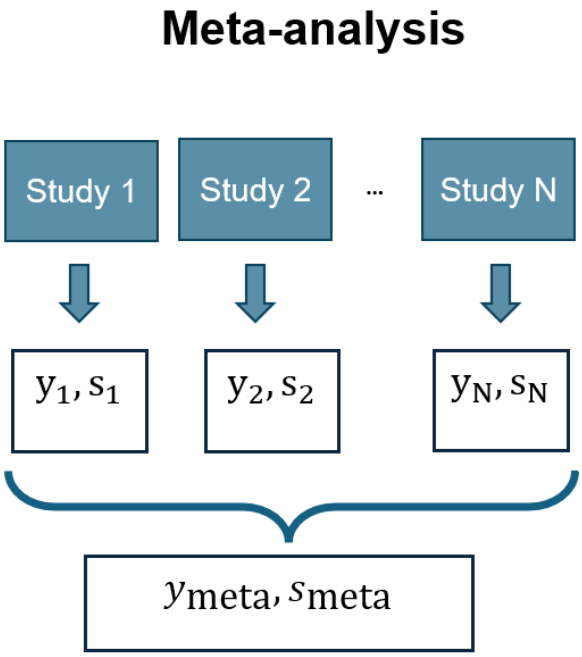

Figure (1.1) How data from different studies would be analyzed, assuming a meta-analysis model. Each study j gives an effect size y_j and a standard error s_j with an aim to get pooled estimates.

Meta-analysis originated from the seventeenth century when astronomers and geologists intuitively combined data rather than choosing among them. In 1861, Airy published his "textbook" on "meta-analysis" for astronomers, introducing the first formulation of a random effects model to account for heterogeneity in the results. In 1904, Pearson Karl Pearson combined data from different studies, arguing that pooling studies is crucial because many groups are too small to form definite opinions due to large probable errors; in 1976, psychologist Glass coined the term "meta-analysis" and three years later, epidemiologist Archie Cochrane highlighted the need for reliable reviews of evidence to inform healthcare decisions [Egger et al. (2022)].

Statistical models for meta-analysis

We describe in brief the two most popular approaches to meta-analysis in the following paragraphs: these are the common effect and the random effects models.

Let us start by considering a collection of N studies, each comparing two interventions for a specific healthcare problem measuring a specific outcome. Note that when only two interventions are involved, the meta-analysis model is called pairwise meta-analysis.

In a meta-analysis, we may have arm-level data, where arms refer to single treatment groups and the data needed are either mean values, standard deviations, sample size for continuous outcomes or number of events and sample size for dichotomous outcomes. Alternatively, we may have contrast-based data in which each study provides an estimate of the relative treatment effect (y_j) and the corresponding measure of uncertainty for this estimate, such as the estimated standard error (s_j). Relative effects or effect sizes can be expressed in various forms, such as mean difference, standardized mean difference (for continuous outcomes), odds ratio, risk ratio, risk difference (for binary outcomes), or hazard ratios (for time-to-event outcomes) etc. Note that odds, risk and hazard ratios are typically modeled in the logarithmic scale.

Common Effect Model

The common effect model assumes a single true treatment effect across all studies in the analysis. Differences observed between study estimates are attributed solely to random (sampling) error. This means that the studies are similar in every aspect that could potentially alter the relative treatment effect, such as population characteristics and study design. Denoting the true treatment effect in study j as ψ_j (where $j = 1, 2, \ldots, N$), the common effect assumption posits that all ψ_j are equal, i.e. $\psi_j = \mu$. The observed effects in

each study are given by

$$\boldsymbol{y_j = \mu + \epsilon_j, \quad \epsilon_j \sim \mathcal{N}(0, s_j^2)} \tag{1.1}$$

where ϵ_j represents the random error. In the literature, this is also known as the fixed effect model, but according to Rice et al. (2018) those two models are distinct. In their paper, they present the fixed effects model as a model that assumes the true effects sizes are different and unrelated among studies (are not produced from a single distribution). For a clearer explanation of those terms, see Figure 2 on their paper.

Random Effects Model

Typically, studies are expected to have some variability in terms of the PICO criteria, leading to different true effect sizes in each study. Random effects models assume that the true effect size ψ_j varies across studies and is drawn from a single distribution, conventionally a normal one. More specifically, it is assumed that $y_j \sim \mathcal{N}(\psi_j, s_j^2)$ where $\psi_j = \mu + \delta_j$ and $\delta_j \sim \mathcal{N}(0, \tau^2)$. In a random effects model, the true effect sizes are considered a random sample of an underlying distribution. The variability within this distribution is termed heterogeneity. Therefore, in a random effects meta-analysis, observed differences in study estimates can be attributed to two sources: random (sampling) error and random effects [Riley et al. (2011)]. The observed effects in each study are modeled as

$$\boldsymbol{y_j = \mu + \delta_j + \epsilon_j, \quad \epsilon_j \sim \mathcal{N}(0, s_j^2), \quad \delta_j \sim \mathcal{N}(0, \tau^2)} \tag{1.2}$$

where ϵ_j is the random error. The standard deviation of the random effects is denoted by τ. Setting $\tau = 0$ implies no variation in study-specific effects, degenerating the random effects model to a common effect model.

Mathematically, the two models differ in the way they assign weights to studies with the fixed effect giving most weight to large studies and the random effect distributing weights more uniformly across studies. In most cases, it is

unrealistic to assume a common effect size. The Cochrane handbook suggests conducting a random effect meta-analysis as the primary analysis and apply the common effect model as a sensitivity analysis [Chandler et al. (2019)]. When both models yield similar results, it is advised to follow the random effects model. When they give different results, there is a phenomenon called small-study effects, in which smaller studies give systematically larger effect sizes compared to large studies. In such a case, one should consider how much confidence can be placed in smaller studies and whether the larger effects in smaller studies are genuine or due to poor quality/publication and other biases. The meta-analytic model should be denoted a-priori the analysis in the study's protocol.

1.1.1 Aggregate and IPD meta-analysis

A meta-analysis can be performed using summary estimates, also known as aggregate data (AD), or using the raw data of each individual participant, also known as individual participant data (IPD), or sometimes by combining both AD and IPD from those studies that provide them.

In the AD meta-analysis, the data are retrieved from published papers, conference abstracts or study authors and are synthesized with a meta-analytic model. There are several limitations when performing this kind of meta-analysis. One cannot fully explore heterogeneity using aggregate data. Although seemingly this can be done through a subgroup meta-analysis (explore for effect modification in treatment effect) or meta-regression, these approaches lack power and are prone to bias because these methods fail to account for individual-level characteristics (ecological fallacy) [Hirst et al. (2025), Kadlec et al. (2023), Schmidt (2017), Spineli and Pandis (2020) and Thompson and Higgins (2002)]. Also, with AD meta-analysis we are restricted in the analysis that has been applied in each study (model, covariates, treatment effect, missing data dealing, interactions, etc.).

Due to these reasons, IPD meta-analysis has become increasingly popular [Veroniki et al. (2023) and Riley et al. (2010)]. With the term IPD we refer to the fact that we have the raw data for every individual in each study. With this meta-analysis model, we can explore for effect modifiers and adjust for clinical/statistical covariates to improve patient outcomes. Also, access to IPD makes our model more flexible since the same statistical approaches can be performed for each study (e.g. deal with outliers the same for each study or apply the same imputation approach). IPD meta-analysis has been described as the "gold standard" of systematic reviews.

Although IPD meta-analysis looks more appealing, since it is more reliable and can answer more questions, data acquisition is a cumbersome procedure. Most probably data are not publicly available and for their acquisition one has to contact the authors or companies that planned the studies. It is very likely that authors and/or companies will not share their data. Maxwell and colleagues propose a toolkit for planning and managing IPD meta-analysis [Maxwell et al. (2024)]. In many cases, one may obtain IPD for some of the studies. In such a case, there will be a two-stage meta-analysis in which we estimate the effect size in the first stage using IPD if available and, in the second stage, effect sizes (aggregate data and estimated effect sizes from the first stage) are pooled just like in the common/random effect(s) model. When IPD are available for all studies, the analyst should choose between one-stage or two-stage approaches [Riley et al. (2023), Kontopantelis (2018) and Burke et al. (2017)].

In the so-called one-stage approach, all the IPD are analyzed together in a single statistical model (e.g. a hierarchical regression model with random effects). Various modelling choices that account for clustering of participants within studies and, potentially, between-study heterogeneity can be used. Additionally, this prevents the need for continuity corrections, often used in the first stage of a two-stage approach when studies have zero events in one of the groups. It is therefore recommended to use one-stage approaches in cases

where there are few participants or outcome events in most of the included studies [Riley et al. (2023)].

In a two-stage approach, the IPD are analyzed separately for each study to produce study-specific estimates (aggregate data). In the second stage a meta-analytic model is applied to produce a combined estimate. So, the modelling options for the two-stage model are mainly around the choice of the second step model (random or common-effect(s) options). An advantage of the two-stage procedure is that it is not constricted in datasets that are on the same server (data sharing) but can use data that are coming from private servers [Riley et al. (2023) and Veroniki et al. (2023)].

According to Riley et al. (2023), except in the cases of few participants or rare events, both approaches yield similar results. When differences arise in the results of the two approaches, this is due to different model assumptions or estimation methods used by the analysts. Both methods are useful in the IPD meta-analysis, and it is in the analyst's judgment on which one to use.

1.1.2 Network Meta-Analysis (NMA)

The term network meta-analysis (NMA), an extension of pairwise meta-analysis, was introduced by Lumley more than twenty years ago [Ades et al. (2024) and Lumley (2002)]. NMA has become an established statistical method for evidence synthesis. It allows us to compare multiple interventions (more than two), even if these have never been compared head to head [Salanti (2012) and Seitidis et al. (2022)]. NMA produces consistent estimates of the relative effects of all interventions compared to each other within a single analysis, utilizing both direct and indirect evidence, resulting usually in more precise effect estimates. It allows incorporating the relative effects from trials with more than two arms, while also avoiding the double counting of patients. This approach allows also for a coherent ranking of all interventions.

A key concept in NMA is that of indirect comparison. If interventions B

and C have both been compared to A, then we have indirect evidence for their relative effectiveness as one can go from B to C via A (this is an indirect comparison). If there is also direct evidence from studies directly comparing B and C, NMA proceeds by synthesizing both direct and indirect evidence. A key assumption in NMA is that direct and indirect evidence are in agreement.

The similarity assumption means that for the indirect estimate between B and C to be valid, we would like this indirect path not to be confounded with any effect modifier. For example, if age is an effect modifier and AB studies have older participants compared to AC studies, we won't be able to tell if the result for BC is genuine or is observed due to differences in age. It is impossible to test transitivity statistically (unless there are multiple studies per treatment comparison), but its plausibility can be evaluated conceptually.

Another very important assumption is that of consistency, which is the statistical counterpart of transitivity. Statistical consistency can only be assessed within closed loops of evidence. Generally, there are two ways to check for consistency: either locally (by comparing specific treatments) or globally (by checking across the entire network). Several methods and models have been proposed for both approaches [Lu and Ades (2006), Dias et al. (2010), and Seitidis et al. (2023)].

1.2 Missing data

By missing or unobserved data, we refer to the case in which the value of something we measure is missing. There are several factors that contribute to missing values. For example, there might be variables that are prone to huge missingness due to their sensitive nature, an example of such a variable is income, where most of the questionees do not feel comfortable declaring their income. Also, there might be the case that in long questionnaires the questionees get tired and might skip some questions. It is also possible for

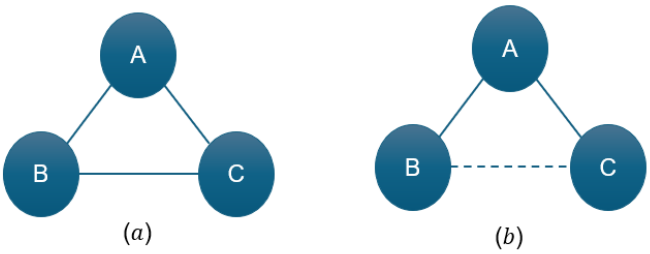

Figure (1.2) Two networks with ABC interventions. The solid lines denote direct evidence, while the dashed line denotes indirect evidence. In (a): Studies that compare B versus C are present, both direct and indirect evidence are available for the comparison of BC. In (b): There are no studies that compare B versus C, only indirect evidence for the comparison of BC is available.

missing data to occur when the person who acquires the data makes an error when passing it in an electronic format.

There are some main implications of missing data in the analysis.

i) Statistical methods do not work.

The known statistical methods do not work when missing data are present. This occurs because the data matrix shape is no longer rectangular and most of the known statistical approaches were designed to handle data in matrices with a rectangular shape e.g. regression modeling.

ii) Biased estimates.

The representativeness of the sample might no longer exist, as omitting the questionees might make our sample less representative, which will lead to bias and loss of efficiency. The bias in parameter estimates will make the study's results less generalizable from the sample to the population.

iii) Reduction in power.

Missing data reduce statistical power. Statistical power is the probability

of rejecting the null hypothesis when it is not true. This is due to the ability of statistical power to go in par with the sample size, and since with missing data we lose sample size the statistical power is reducing.

1.2.1 Complete Case Analysis

To deal with the previous problems several methods have been proposed in the literature with the most known been Complete Case Analysis (CCA). This method deletes every questionee that has not responded to a single question and keeps only those who provide full data. With the deletion of questionees the shape of data matrix is rectangular, and we can apply any statistical analysis. This method does not prevent the second and third implications presented above, though, on some occasions, might overcome the restrictions of the second.

1.2.2 Single Imputation Methods

Under the umbrella term of "single imputation", we place all methods using a technique to impute the missing values to result in a full data set for our analysis. Some of those methods are:

1. Mean Imputation: In mean imputation, the missing values of variables are replaced with their mean values from the observed ones. This approach reduces the variability of the data and affects the relation between variables. Instead of replacing missing values with their mean we could use whichever point estimate matches our beliefs such as median or mode. Figure 1.3 shows how mean imputation handles missing data in the dataset of air quality from the mice package [Van Buuren and Groothuis-Oudshoorn (2011) and Van Buuren (2018)]. The figure shows the problem of altering the relationship between the variable's ozone and solar radiation, by forming a cross of the imputed values from each

variable.

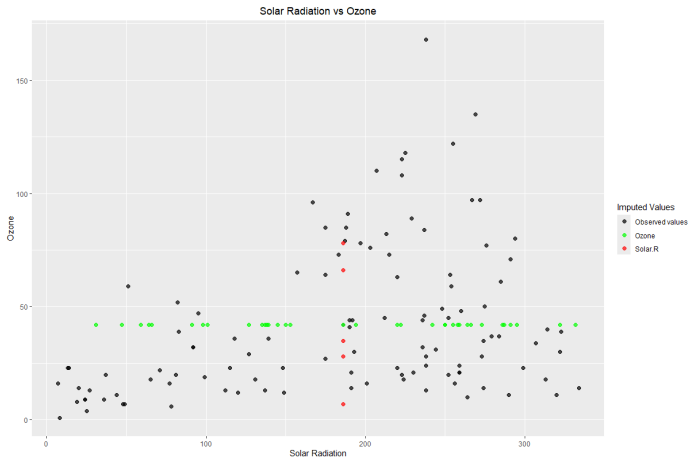

Figure (1.3) Performance of mean imputation in air quality dataset. In the scatter diagram, the vertical axis represents the solar variation variable, and the horizontal axis represents the ozone variable. The black dots indicate observed values. Green dots represent values where the solar radiation variable was observed but the ozone variable was missing and imputed with its mean. Red dots represent values where the ozone variable was observed but the solar radiation variable was missing and imputed with its mean.

2. Last Observation Carried Forward (LOCF): LOCF imputation is an approach that is commonly used in longitudinal studies. This approach is taking the last measurement of a participant and projects this value to the end of the study. If, for example a study observes patients for ten weeks and a patient drops out of the study in the third week, then, we use this value until the end of the study. In some fields, we expect disease to progress over time, e.g., in dementia or in others, we expect people to get better e.g., depression. Assuming no progress after dropout leads to biased conclusions. A good example of the manipulation of the results when applying LOCF can be seen in Figure 1.4. This figure is

depicted in the paper of Rostain et al. (2017) and shows the change of the ranking of drugs when using LOCF instead of CCA.

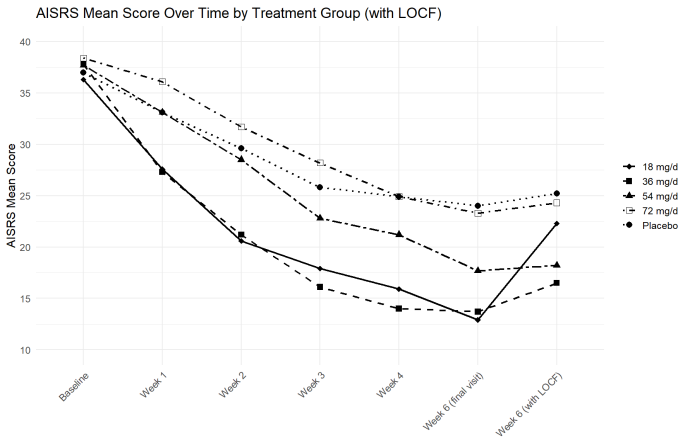

Figure (1.4) Mean AISRS [scale for ADHD (attention deficit hyperactivity disorder)] in a duration of six weeks. The treatments are placebo and Osmotic-Release Oral System (OROS) in different doses.

3. $k-$NN (k-Nearest Neighbor): $k-$NN is a technique that fills missing values by borrowing values from similar patients. This technique estimates the distances between variables and takes as a donor the value from the patient that has the smallest distance from the one that we want to impute.

4. Regression Imputation: A method that uses a regression model to predict values for any variable with missing data based on the rest available variables of the dataset.

A disadvantage of single imputation methods is that they treat imputed values as they were observed, inflating spuriously the sample size. In practice, we would not want to increase the certainty in the results since we do not know the true values.

1.2.3 Multiple Imputation

Multiple imputation (MI) is a popular strategy to tackle the problem of missing values. Initially, for each variable with missing data, several imputations (M) are drawn from its posterior predictive distribution given the observed data, ending up with M different completed datasets. Then, the same statistical analysis (regression model) is performed on the M datasets resulting in M regression estimates for each of the covariates of interest. As a final step, Rubin's rules, which are based on asymptotic theory in a Bayesian framework, are used to pool the estimates [Rubin (1987), White et al. (2011), Azur et al. (2011), Wulff and Jeppesen (2017) and Van Buuren (2018)]. MI is attractive as it generates multiple replicas (not a single) of the dataset, each incorporating different imputed values, and thus taking into consideration the uncertainty that exists around missing data. Keep in mind that MI imputation is not a panacea and there are cases that can go wrong. To avoid its pitfalls, it is suggested to explore different graphs depending on the type of imputed variable and compare the results with those of CCA. The results of MI and CCA are not expected to be identical, but if there are substantial differences, there must be a logical explanation [White et al. (2022b)].

1.2.4 Rubin's missing mechanisms

To select the method that is appropriate to use for analysis when dealing with the problem of missing data, we should make some assumptions about the mechanism of missingness. Rubin (1976) divided the mechanism in three categories, those are Missing At Random (MAR), Missing Completely At Random (MCAR) and Missing Not At Random (MNAR). All these mechanisms relate the probability of missing data to observed or unobserved data. Suppose that we divide the data into two categories, the ones that are observed ($\boldsymbol{z_{obs}}$) and those that are missing ($\boldsymbol{z_{mis}}$), then the full data will be $\boldsymbol{z = (z_{obs}, z_{mis})}$. Also, we will denote with $r = (r_1, \ldots, r_A)'$ the missing indicator vector with

values 0 and 1 ($\boldsymbol{r_i}$= 1 if the i_{th} data point is missing and 0 otherwise).

- MCAR: The mechanism causing missing data depends neither on observed ($\boldsymbol{z_{obs}}$) nor on missing ($\boldsymbol{z_{mis}}$) data

$$f(\boldsymbol{r} \mid \boldsymbol{z}) = f(\boldsymbol{z})$$

  Under MCAR assumption, the observed data representative of the original sample, and CCA where we analyze only the observed values is valid.

- MAR: The mechanism causing missing data depends only on the observed data ($\boldsymbol{z_{obs}}$) and is independent of unobserved data ($\boldsymbol{z_{mis}}$)

$$f(\boldsymbol{r} \mid \boldsymbol{z_{mis}}, \boldsymbol{z_{obs}}) = f(\boldsymbol{r} \mid \boldsymbol{z_{obs}})$$

  Under MAR assumption the observed data cannot be viewed as representative of the original sample, and multiple imputation methods should be used to address the analysis.

- MNAR: The mechanism causing missing data depends on both observed ($\boldsymbol{z_{obs}}$) and missing ($\boldsymbol{z_{mis}}$) data. Under MNAR assumption, most statistical approaches that address missing data are invalid. Some options that make hypotheses on how missing data differ from observed data are pattern-mixture and selection models can be used to overcome the MNAR challenges [Muñoz et al. (2024) and Little and Wang (1996)].

We cannot verify the assumptions of MAR and MNAR based on observed data alone, however, exploring the observed data might be very useful. A hypothesis testing procedure can be explored by using Pearson's Chi-squared test or a logistic regression model to determine if data are MCAR or not in a single variable [Pham et al. (2022) and Cro et al. (2020)]. For example, Pham et al. (2022), divide age in two groups (age $>$ 25 and age $\leqslant$ 25) and make a 2×2 table with those that the mean probing depth has been observed or not. In that table they performed a Pearson's Chi-squared test and based on its value

they conclude that the data there is evidence against MCAR. Although we can distinguish between MCAR and not MCAR, we cannot easily distinguish if the missing mechanism is MNAR or MAR.

There is no strict rule for the proportion of missingness for a variable to be discarded from the analysis model. What is changing is the model that we could use for the encounter of missing data. A single imputation method or a CCA may be a good option for a small percentage of missingness (around 15%). But for a large proportion of missingness we better use MI methods. Even for a variable with missing values over 60%, if the other variables of the model have a small amount of missingness, we expect MI to give credible results [Harrell (2012)]. Our most pressing concern should be whether the imputation model is appropriate or not, as even a small percentage of missing data (e.g. 10%) may bias the results.

As one cannot solve a crime with certainty without all the evidence, one has to resort to assumptions when there is missing evidence. As the true values for the missing values are unknown, it is impossible to assess how close we got to the truth, and we treat any method for handling missing values as a sensitivity analysis.

1.3 Missing data in IPD meta-analysis

In the meta-analysis literature, there are two broad classes of missing covariate data. The first one is called sporadically missing and is the case where covariates are partly reported in the included studies. In a nutshell, we do not have information about all patients in all covariates. The second one is systematically missing and is the case where the set of covariates is not the same across all studies. In this type of category, the problem arises because some studies did not intend, from their design, to measure some covariates. This is not a problem when analyzing the results of the study, but it is a

problem in an IPD meta-analysis in which we plan to account for a specific set of covariates.

The two most popular MI classes used in IPD meta-analysis are a) Fully Conditional Specification (FCS), also known as Multiple imputation using chained equation (MICE) and sequential regression and b) Joint modeling (JM) imputation [Riley et al. (2021)]. These classes differ in how they approach the problem, and several methods have been suggested for both of them.

1.3.1 Model Formulation

Suppose that we have an outcome y and p covariates x_k where $k = 1, \ldots, p$. Also, suppose that we have N studies with each study $j = 1, \ldots, N$ having n_j participants each.

We denote with $\boldsymbol{\beta}^{'} = (\beta_1, \ldots, \beta_p)^{'}$ the set of regression coefficients and β_0 the constant term.

$$y_{ij} = \beta_0 + \gamma \times I(treat) + \sum_{k=1}^{p} \beta_k \times x_{ijk} + \epsilon_{ij} \tag{1.3}$$

Where $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$, in which i refers to the participants of a study $j, i = 1, \ldots, n_j$. The $I(treat)$ denotes the treatment variable, while γ is the conditional average treatment effect. Suppose that X_j^{miss} is the set of covariates with missing data in the jth study. Generally, we have $\boldsymbol{X_j = (X_j^{obs}, X_j^{miss})}$.

1.3.2 Fully Conditional Specification (FCS) method

The FCS implementation is based on three steps. Note that just like one can do an IPD (N)MA in one or two steps, the same holds for multiple imputation. We can do it either in each study separately or, in one step, using the total sample size from all N trials. For clarity, we will divide the first step of FCS in sub-steps and we will present it assuming the multiple imputation is done in one stage.

Step 1: Produce M imputed data sets

Step 1.1

Multiple imputed data sets (M) are generated by replacing the missing values for each $x_k \in X^{miss}$ with values from the posterior predictive distribution $\boldsymbol{f(x_k \mid y, x_{(-k)}, r)}$. For simplicity, we consider that the outcome variable has no missing values and $\boldsymbol{x_{(-k)}}$ denotes the set of covariates with the exception of covariate k. Typically, a multivariate normal distribution is used as the predictive distribution for the initial imputation; however, alternative methods, such as mean imputation, $k-$nearest neighbors ($k-$NN) or simple regression models for each covariate can also be applied.

Step 1.2

We choose randomly one of the covariates, say x_k, and its marked/imputed values are returned to their initial condition (become missing) for the first imputed data set ($M = 1$). While there is no strict rule on which covariate to impute first, a reasonable start would be to start with the covariate which initially has the smallest amount of missing data. We regress this variable on the other covariates that are full (with all missing values imputed in step 1.1) and impute its missing values based on the predictions of the regression model. The form of the model depends on the nature of the outcome variable (e.g., for continuous variables a usual choice is a linear regression model, while for dichotomous variables a usual choice is a logistic regression model)

$$x_{ijk} = \beta_0^{(k)} + \gamma^{(k)} \times I(treat) + \sum_{h=1, h\neq k}^{p} \beta_h^{(k)} \times x_{ijh} + \delta^{(k)} \times y_{ij} + \epsilon_{ij}^{(k)},$$

$$\epsilon_{ij}^{(k)} \sim \mathcal{N}(0, \sigma^{2(k)})$$

Where super index (k) denotes that the model is estimated by using x_{ijk} as dependent variable. Furthermore, constant $\beta_0^{(k)}$ can be either study-specific or shared across studies, depending on the variation we aim to induce between them (heterogeneity) [Burgess et al. (2013)].

Step 1.3

Repeat the procedure using as dependent variable each one of the remaining covariates with missing data, say k^*. Note that the variable that has already been imputed, say $x_{ijk^*}^{imp}$, is now used as a covariate without missing data

$$x_{ijk^*} = \beta_0^{(k^*)} + \gamma^{(k^*)} \times I(treat) + \sum_{h=1, h\neq(k^*,k)}^{p} \beta_h^{(k^*)} \times x_{ijh} + \beta_k^{(k^*)} \times x_{ijk}^{imp} + \delta^{(k^*)} \times y_{ij} + \epsilon_{ij}^{(k^*)}, \epsilon_{ij}^{(k^*)} \sim \mathcal{N}(0, \sigma^{2(k^*)})$$

Step 1.4

Repeat the previous steps until there is no large dependence between the starting imputation values and the final imputation values. The number of times is usually set to five. This is considered a full cycle. This cycle is then repeated $M-1$ times for the rest of the sets that we initially generated. The choice of M is defined a-priori and a common value in the literature is to set it equal to 10. By theoretical perception the higher the M the better, but we have to restrict ourselves as the process is both storage and time consuming. White et al. (2011) suggests that the M should be at least as large as the percentage of incomplete cases of the variable with the highest rate of missingness.

Step 2: Analyze the M imputed data sets

The second step is to analyze the M (pseudo-)imputed data sets, that is to estimate the regression coefficients for each one of the imputed data sets. A suggestion when using MI, is that the imputation model should be compatible with the proposed analysis model, this is called congeniality; meaning that analysis and imputation model are the same. Another approach, in order to improve the quality of the imputation might be to include in the imputation model variables that are highly correlated with the imputed variables. Those variables are called auxiliary variables [White et al. (2022a)]. This means that

in (1.3) we keep only a subset of the variables k.

Step 3: Synthesize the M (pseudo-)imputed data sets using Rubin's rules

In the final step of MI, we combine the derived M regression estimates and acquire an overall estimate by applying Rubin's Rules.

Suppose β_m (for $m = 1, ..., M$) is an estimate (e.g. a regression coefficient) derived from the m imputed dataset and $var(\beta_m)$ represents its estimated variance. The overall estimate is determined by computing the average across the individual m estimates:

$$\hat{\boldsymbol{\beta}} = \frac{1}{M} \sum_{m=1}^{M} \beta_m \tag{1.4}$$

The computation of the total variance considers variability both within and between imputations ($\boldsymbol{Var_{Within}}$ and $\boldsymbol{Var_{Between}}$, respectively):

$$\boldsymbol{Var}(\hat{\boldsymbol{\beta}}) = \boldsymbol{Var_{Within}} + \left(1 + \frac{1}{M}\right) \boldsymbol{Var_{Between}} \tag{1.5}$$

For more details about the calculations of total variance, see the work of Rubin. A graphical representation of the procedure can be seen in Figure 1.5.

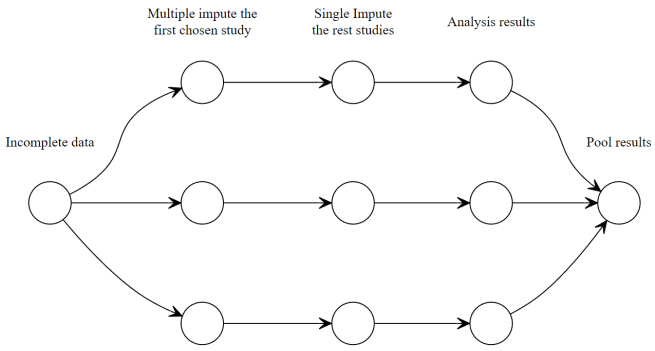

Figure (1.5) MICE steps as an acyclic graph.

Several modifications of FCS have been developed for tackling with problems of sporadically, systematically or both in IPD meta-analysis. The most recent ones are one-stage models since they are more appealing and can comfort systematically missing data. This is because one-stage models have the advantage of borrowing strength across studies, which can be important for estimating covariate and subgroup effects. Initially, Resche-Rigon et al. (2013) proposed a linear mixed effects model to deal with missingness in models with continuous outcomes, and then Jolani et al. (2015) extended this model to deal with other types of outcomes. Those two methods were proposed for dealing only with systematically missing covariates. To deal with both sporadically and systematically missing data Resche-Rigon and White (2018) proposed a variation of their previous method, while Jolani (2018) used a hierarchical imputation by using Bayesian approaches. A recent method for IPD meta-analysis has been proposed by Muñoz et al. (2024) which utilizes Heckman models, to deal with MNAR mechanism.

The two-stages due to their nature impute each study separately with standard imputation routines and hence can only be used for sporadically data. The logic behind this model is that there will be considerable between-study heterogeneity which should not be reflected in the imputation process. Burgess et al. (2013), considered a two-stage model where they compared two methods for combining the estimates. Those methods have to do with the sequence in which meta-analysis and Rubin's rules should be applied. In their study, they found that Rubin's rules should be applied before the meta-analysis was conducted rather than the other way around.

1.3.3 Joint Modeling (JM) imputation

Joint Modeling (JM) imputation is an approach that uses a multivariate joint model for all variables that are of interest. In this model, all variables with missing data are simultaneously imputed. In the multivariate joint models,

the incomplete variables are regressed on the complete variables.

The method was originally described by Schafer (1997) and Schafer and Yucel (2002). After that, Yucel (2011) proposed a joint (multivariate) imputation technique that models cluster-specific associations by using random within-cluster covariance matrices. This method was further extended using a latent normal model by Quartagno and Carpenter (2016), to handle categorial variables and applied the method in the context of IPD meta-analysis.

For clustered data, such as an IPD meta-analysis, these methods adopt a random effect model that accounts for the relationships between variables, clustering of patients within studies and allows for potential heterogeneity between them (covariance matrices are different across studies). The models that this approach proposes are the random intercept and random slopes models. This is a natural approach for imputing systematically missing variables in IPD meta-analysis.

Once the imputation model is selected, Bayesian techniques, specifically Gibbs sampling (an iterative Markov chain Monte Carlo (MCMC) algorithm), are used to fit the model and impute missing data via a data augmentation algorithm [Tanner and Wong (1987)]. This involves iteratively drawing new parameter values from their conditional distributions. Initial parameter values must be chosen carefully to ensure rapid convergence. The number of iterations required to reach convergence is often referred to as burn in period. As a Bayesian method, prior distributions are also employed. The Gibbs sampler is run until it reaches a stationary distribution, after which the current draws of missing values are combined with the observed data to create the first imputed dataset. After that, a few MCMC iterations are discarded before obtaining the next draws of the second dataset in order for the subsequent datasets to be independent. This number of discarded values is known as the burn between. This procedure is repeated M times. A limitation of this approach is that in order to perform the imputations, it needs at least one variable to be fully

observed. For technical details of the MCMC approach see Schafer (1997).

1.3.4 Comparison of the methods

Plenty of simulations have been conducted for the comparison of the methods of the two classes [Kunkel and Kaizar (2017), Enders et al. (2018), Audigier et al. (2018), Mistler and Enders (2017) and Grund et al. (2018)]. Different scenarios explored, so an overall rule about which method to choose, does not exist. The two approaches can diverge in small samples (individuals and/or clusters) due to the influence of priors. Specifically, the JM approach tends to give biased results due to the Inverse-Wishart prior distributions which becomes very informative.

The FCS approach is slightly more flexible in terms of substantive model since it can incorporate in the models' interactions or nonlinear terms. Also, while the multivariate model of JM does a strong assumption of normality, Schafer shows via simulation the JM is robust to moderate skewness. With FCS, the assumption of normality can be avoided.

1.3.5 Packages in *R*

The most common packages that are used when dealing with missing data and includes most of the aforementioned methods are *mice* [Van Buuren and Groothuis-Oudshoorn (2011)], *micemd* Audigier and Resche-Rigon (2023)], *pan* [Quartagno and Carpenter (2023)], and *jomo* [Zhao and Schafer (2023)]. The first two packages are using the FCS approach in their imputation routines. The mice package has a big arsenal of imputation methods, tests, and graphs, while micemd is focused on imputation methods for multilevel data only. On the other hand, jomo and pan are using the JM approach. They are very similar in terms of algorithm with jomo having the advantage that it allows for heteroscedacity (variance of the error term can be different between studies),

while pan does not. Also, jomo allows for the imputation of categorical variables in contrast with pan. In terms of speed, the pan is more efficient than the jomo package. Lastly, a package named bipd, which is a package that uses Bayesian methodology to perform IPD meta-analysis, allows for some of the methods of section 1.3.2 to be used [Seo (2022)].

A limitation of these packages occurs with one-stage models involving systematically missing covariates. Although more appealing than two-stage models, they require all data to be on the same server [Veroniki et al. (2023) and Riley et al. (2010)]. This is rarely the case. Data typically comes from various companies, and access is restricted to their private servers, making it impossible to integrate information from multiple studies, meaning that IPD across studies cannot be combined in a one-stage analysis.

1.4 Objectives & outline of the Thesis

The aim of this dissertation is twofold, first create a new methodology for missing data in IPD meta-analysis and second create an R-package that uses and extends some missing data methodologies for AD meta-analysis.

For the first cause, a one-stage FCS missing data model has been proposed in which the proximity of studies is considered in the imputation stage. This model is based on the FCS approach, but the intermediate steps are slightly different. More specifically, instead of focusing on covariate imputation in each study, we focus on imputing covariates separately for each study. Once the first study has been multiple imputed, we reduce our imputation method to a single impute approach. We conduct a simulation study to evaluate the performance of our proposed method (MI with $k-$nearest studies) across a wide range of scenarios including the number of studies and probability of missingness.

As part of the second cause of the dissertation, we will develop an R-package that considers informative missingness (reasons for dropout are associated with

the outcome measured, given all observed data) parameters, that cannot be estimated from observed data, but can be specified, with associated uncertainty, using evidence outside of the meta-analysis, such as expert opinion. We created routines both for pairwise and network meta-analysis.

The Thesis is structured as follows. Chapter 2 introduces the one-stage IPD method and also includes the simulation scenarios that we explore. Chapter 3 gives an introduction to the already known methodology and describes in detail the R functions that are used. Finally in Chapter 4, we end by discussing the results, next steps, and providing concluding remarks.

Chapter 2

AN ONE-STAGE IMPUTATION METHOD FOR HETEROGENEOUS STUDIES USING DISTANCE METRICS

In systematic reviews, there is a clear research question with certain PICO criteria. Though all studies address the same research question, there is usually substantial variation in the PICO criteria. Especially in the populations involved, as studies are conducted in different populations that share different individual characteristics. To explore the heterogeneity of treatment effect, we typically conduct subgroup and meta-regression analyses e.g., on baseline risk, average age. When using multiple imputations, if we focus on one study, we may get biased and imprecise results because this study has a lot of missing data or is very small. Additionally, we cannot deal with the problem of systematically missing covariates. However, if we borrow for all available studies, we may use information from populations that are substantially different than the study that we want to impute data for.

In this chapter, we propose a new, one-stage, imputation approach for IPD meta-analysis that, unlike standard multiple imputation approaches, does not borrow information from all available studies in the systematic review, but

borrows instead from those trials that are similar in terms of the distribution of baseline covariates. Our method is based on the FCS approach, but the intermediate steps are slightly different. Initially, we compute distance metrics among studies to identify clusters of studies that we will subsequently use to impute data not only from the relevant study but from its class of studies. We will allow the borrowing of information within the cluster but not between clusters. Once the first study has been imputed multiple times, let's assume M times, we proceed to the second study and continue until all studies have their missing values imputed M times. Then, we take the M imputed datasets and proceed using Rubin's rules. We conduct a simulation study to evaluate the performance of the proposed method across a wide range of scenarios with varying number of studies and probability of missingness. Lastly, we conclude by discussing the results.

2.1 Methods

In our approach we propose an imputation method that is based on borrowing information only from those studies that are close to each other e.g., in terms of prognostic variables and effect modifiers, instead of borrowing either from a single study or from all studies. As a result, we keep the advantage of one-stage approaches, rely on other studies' strengths, while respecting the advantages of two-stage approaches, taking into account studies' diversity. There is a trade-off in this approach, and we have a crucial question to consider: "Is it better to have only homogeneous studies (less sample size and less heterogeneity) or keep all studies (more sample size and more heterogeneity)?". Though we will base our definition of similarity of studies on mathematical grounds and distance metrics, one could consider experts' opinion and a broader consideration of the PICO criteria.

2.1.1 Similarity among studies

Similarity between any pair of studies is expressed by a numerical value called a similarity measure that is a real-valued function that quantifies how closely data points or distributions are.

2.1.1.1 Distance and similarity metrics

The similarity matrix is a square $N \times N$ matrix $S = [s_{jj'}]$,$(j, j^{'} \in \mathcal{N}$ and $\mathcal{N}$ is the set of the studies), that shows how similar studies are to each other. We have $s_{jj} = 1, \forall j \in \mathcal{N}$. The similarity is usually the inverse of the distance. We can estimate the distance between studies by $d(\boldsymbol{x_j}, \boldsymbol{x_{j'}}) = \sqrt{1 - s_{jj'}}, \forall j, j^{'} \in \mathcal{N}$.

The higher the similarity value, the closer the two studies are whereas a similarity value close to zero is indicative of studies that differ considerably. Since the similarity metrics do not handle missing data by default, we exclude the missing values using listwise deletion before calculating similarity metrics (this is equivalent to a CCA). For a high percentage of sporadically missing or systematically missing data, the use of those methods might give us biased or no results at all, respectively. As an alternative we could use single imputation and calculate the distance of the studies or employ a multiple imputation method and take the average of the distances of each imputed dataset using Rubin's rules. We explore some of the most common similarity and distance metrics; Euclidean, and Manhattan distances and Jensen-Shannon divergence, which are addressed in different similarity measures (data points and distributions).

2.1.1.2 Similarity measures for data points

We calculate the arithmetic averages of each covariate x_k ($k \in \mathcal{P}$, where $\mathcal{P}$ is the set of the p a-priori defined covariates), in each study. Let us consider two studies $(j, j^{'} \in \mathcal{N})$ with mean covariate values $\boldsymbol{x_j} = (\bar{x}_{j1}, .., \bar{x}_{jp})^{'}$ and

$\boldsymbol{x_{j'}} = (\bar{x}_{j'1}, .., \bar{x}_{j'p})'$, respectively, in $p-$dimensional Euclidean space.

Euclidean distance

The most common similarity metric is the Euclidean distance. It works on the principle of the Pythagorean theorem to find the shortest distance between two points. However, it can be sensitive to outliers and can be used only for continuous variables.

$$\begin{aligned} d(\boldsymbol{x_j}, \boldsymbol{x_{j'}}) &= \sqrt{(\boldsymbol{x_j} - \boldsymbol{x_{j'}})'(\boldsymbol{x_j} - \boldsymbol{x_{j'}})} \\ &= \sqrt{(\bar{x}_{j1} - \bar{x}_{j'1})^2 + (\bar{x}_{j2} - \bar{x}_{j'2})^2 + \cdots + (\bar{x}_{jp} - \bar{x}_{j'p})^2} \end{aligned}$$

It is also possible to standardize the mean difference by dividing with the pooled standard deviation (S) and calculate the similarity measures from the standardized mean difference.

$$d(\boldsymbol{x_j}, \boldsymbol{x_{j'}})^2 = (\boldsymbol{x_j} - \boldsymbol{x_{j'}})' S^{-1} (\boldsymbol{x_j} - \boldsymbol{x_{j'}})$$

Manhattan distance

Another common metric is the Manhattan distance also known as taxi-cab metric or city block distance. The calculation of this metric is done by defining the distance between two points as the sum of absolute differences between points in the Cartesian coordinates.

$$d(\boldsymbol{x_j}, \boldsymbol{x_{j'}}) = |\bar{x}_{j1} - \bar{x}_{j'1}| + |\bar{x}_{j2} - \bar{x}_{j'2}| + \cdots + |\bar{x}_{jp} - \bar{x}_{j'p}|$$

2.1.1.3 Similarity measures for distributions

Jensen-Shannon divergence

Jensen-Shannon divergence (JSD) is one of the many methods that compute the distance between two probability distributions. For finding distances

between distributions, it employs the Kullback-Leibler divergence (denoted as D_{KL}) formula.
The mathematical definition of the JSD between two probability distributions $\boldsymbol{P}$ and $\boldsymbol{Q}$ is:

$$\text{JSD}(\boldsymbol{P} \parallel \boldsymbol{Q}) = \frac{1}{2}\left(D_{KL}(\boldsymbol{P} \parallel \boldsymbol{W}) + D_{KL}(\boldsymbol{Q} \parallel \boldsymbol{W})\right)$$

where $\boldsymbol{W}$ is the average of the two distributions:

$$\boldsymbol{W} = \frac{1}{2}\left(\boldsymbol{P} + \boldsymbol{Q}\right)$$

2.1.2 Finding studies' divergence

In this subsection we are going to propose four methods of how we could calculate the divergence between studies. Those methods are:

First method: For each study $j \in \mathcal{N}$, we calculate the distances $d(\boldsymbol{x_j}, \boldsymbol{x_{j'}})(j \in \mathcal{N}, j' \in \mathcal{N} \backslash j)$.

Then, we define a criterion as to which studies we consider to be "close" to each other. For example, for each study j, we may consider those studies whose distance is smaller than the mid-range $\frac{\max(d(\boldsymbol{x_j},\boldsymbol{x_{j'}}))-\min(d(\boldsymbol{x_j},\boldsymbol{x_{j'}}))}{2}$.
For example, suppose that we have $N = 10$ studies and we have ordered, for the first study, the distances from the smallest to the largest. Let us suppose the order is

$$d(\boldsymbol{x_1}, \boldsymbol{x_3}), d(\boldsymbol{x_1}, \boldsymbol{x_5}), d(\boldsymbol{x_1}, \boldsymbol{x_7}), d(\boldsymbol{x_1}, \boldsymbol{x_9}), d(\boldsymbol{x_1}, \boldsymbol{x_2}),$$
$$d(\boldsymbol{x_1}, \boldsymbol{x_4}), d(\boldsymbol{x_1}, \boldsymbol{x_6}), d(\boldsymbol{x_1}, \boldsymbol{x_8}), d(\boldsymbol{x_1}, \boldsymbol{x_{10}})$$
$$\text{mid-range for study } j = MR_j = \frac{d(\boldsymbol{x_1}, \boldsymbol{x_{10}}) - d(\boldsymbol{x_1}, \boldsymbol{x_3})}{2}$$

For each study $j \in \mathcal{N}$, we create a set $\mathcal{C}_j$ of studies that includes any study for which we have $d(\boldsymbol{x_j}, \boldsymbol{x_{j'}}) < MR_j$.

Second method: Use multidimensional scaling which is a visual representation of distances in a lower-dimensional plane, usually in a two-dimensional plane, so that we can see how studies cluster with each other.

Third method: Use hierarchical clustering analysis and plot the results on a dendrogram. We can determine the proximity between any pair of studies by measuring how close their branches are.

Fourth method: Set an a-priori fixed number of studies to borrow from. For example, order by proximity and borrow only from the three closest studies to each study.

That results in N sets that are not mutually exclusive or do not necessarily create clusters of the same studies. For example, study 1 could be close to studies 2 and 3 but these studies are not necessarily close to each other.

2.1.3 Proposed method for imputation

Our method called Similar Studies Imputation Method (SSIM) is a variation of MICE [Van Buuren and Groothuis-Oudshoorn (2011) and Van Buuren (2018)]. Instead of employing the typical MI that creates multiple imputed datasets (M) using all studies at once, we create those M datasets by borrowing information only from neighboring studies (defined by one of the methods presented in Sections 2.1.1 and 2.1.2) at multiple steps (equal to the number of studies that have missing values). First, we create multiple imputed datasets, say M, for the first study with missing data imputation based on neighboring studies using standard multiple imputation techniques. Then we single impute those multiple imputed datasets for the remaining studies, one by one, following the same procedure until all have been imputed. Finally, we end up with M imputed datasets that we pool together using Rubin's formulae.

1) Suppose that we have N studies (all or some of them have sporadically missing values) with $p+1$ candidate variables (p covariates and an outcome). For each study, we calculate the average missing rate among all variables included, $\boldsymbol{mr_j} = \frac{\sum_{k=1}^{p+1} \boldsymbol{m_{jk}}}{p}$, where $\boldsymbol{m_{jk}}$ is the missing rate in study $j = 1, .., N$ and variable $k = 1, \ldots, p+1$.

2) Calculate the distance according to the methods described in Sections 2.1.1 and 2.1.2, $d(\boldsymbol{x_j}, \boldsymbol{x_{j'}}), j, j' = 1, \ldots, N, j < j'$, among all studies. We end up with a total of $\binom{N}{2} = \frac{N \times (N-1)}{2}$ distances.

3) We create clusters of studies with the aim to maximize similarity of studies within a cluster and minimize it across studies. This can be done based either on expert opinion and content knowledge or statistically by calculating some distance metric. Denote by $\boldsymbol{\mathcal{C}}$ the set of N created clusters, $\boldsymbol{\mathcal{C}} = (\boldsymbol{\mathcal{C}_1}, \boldsymbol{\mathcal{C}_2}, \ldots, \boldsymbol{\mathcal{C}_N})$. Note that some of these clusters may be identical, as certain studies may have the same set of closely related studies.

4) For each cluster, we compute the average missing rate in the included studies ($\boldsymbol{mrC_j} = \frac{\sum mr_{C_j}}{n_{C_j}}$,where n_{C_j} is the number of studies in each cluster). Then, we select from the cluster with the lowest missing rate ($\boldsymbol{mrC_j}$), the study with the lowest missing rate ($\boldsymbol{mr_j}$). Then, we use one of the MICE algorithms to impute this variable using information from the other studies in the cluster creating a plethora (e.g., $M = 10$) of full data sets.

   For example, suppose that the first cluster has the lowest missing rate ($\boldsymbol{mrC_1}$) and contains five studies (1st, 3rd, 4th, 5th, 9th) with the third study having the lowest missing rate ($min(\boldsymbol{mrC_1}) = \boldsymbol{mr_3}$). The model for this imputation will be

$$x_{ij3} = \alpha_{0C_1} + \sum_{k=1,4,5,9} \alpha_{kC_1} x_{ijk} + \delta_{C_1} y_{ij} + \epsilon_{ijC_1}$$

where i is an index denoting the patient, j denotes the covariate and δ the effect of outcome in the imputed covariate x_3 adjusted for the effect of the other covariates in the cluster. Sub-index C_1 is used to denote that regression coefficients are estimated using only variables included in cluster $\boldsymbol{\mathcal{C}_1}$.

In the end, 10 new datasets will have been created (those datasets will not have the same dimensions as the original one, since the number of studies in $\boldsymbol{\mathcal{C}_1}$ is five and not ten). All missing values for studies one, three, four, five, and nine will have been imputed. At this stage, we want to impute only the third study. Hence, we keep the imputed values for the third study and remove the imputed values for the remaining studies (1st, 4th, 5th, and 9th).

5) We keep only the study we intended to impute (in our case the third study) and substitute it on the original data set. Now we will have 10 data sets with the third study being imputed and all the other studies in their initial condition with their sporadically missing values. In the absence of missing values in other studies, SSIM looks similar to traditional MICE (the only difference is that imputation is based only on neighboring studies). If this is the case, we skip steps 6 and 7.

6) Move on to the next study with the lowest missing rate, single impute the M datasets based on its close studies and repeat step 5. E.g., suppose that we intend to impute study 5 whose close studies are 2, 3, 8, and 10; cluster $\boldsymbol{\mathcal{C}_2}$. From our previous step we have that in the ten created datasets study 3 is imputed. The model for the single imputation of the

10 data sets will be

$$x_{ij5}^{(1)} = \alpha_{0C_2}^{(1)} + \sum_{k=2,3,8,10} \alpha_{kC_2}^{(1)} x_{ijk}^{(1)} + \delta_{C_2}^{(1)} y_{ij}^{(1)} + \epsilon_{ij5}^{(1)}$$

$$\vdots$$

$$x_{ij5}^{(10)} = \alpha_{0C_2}^{(10)} + \sum_{k=2,3,8,10} \alpha_{kC_2}^{(10)} x_{ijk}^{(10)} + \delta_{C_2}^{(10)} y_{ij}^{(10)} + \epsilon_{ij5}^{(10)}$$

where i is the number of patients, j is the number of covariates, δ the estimate of outcome variable and the number in parenthesis indicates the dataset to be imputed.

7) We repeat step 6 until there are no studies with missing data left. Hence, we have ended up with 10 imputed datasets.

8) Apply Rubin's rule.

For a better understanding of the SSIM method, we will present graphically the differences between SSIM and the traditional MICE approach (Figure 2.1). Also, we present the flowchart of SSIM with all the necessary steps (Figure 2.2).

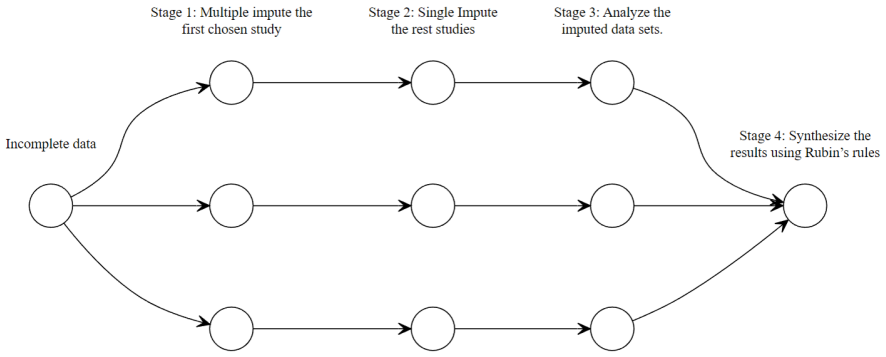

Figure (2.1) SSIM steps as an acyclic graph.

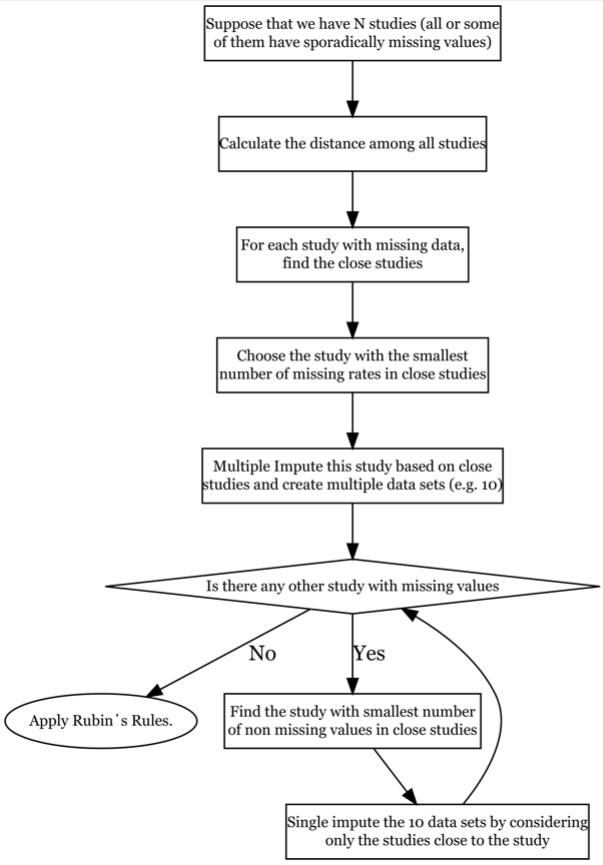

Figure (2.2) Flowchart of the SSIM method.

2.2 Simulation

We will follow the ADEMP (aims, data-generating mechanisms, estimands, methods, and performance measures) framework to present our simulation study [Morris et al. (2019)].

2.2.1 Aims

Our aim is to develop a method for imputing sporadically missing data in IPD meta-analysis. The proposed method will account for differences in

population characteristics across studies, thereby avoiding the use of studies that are substantially different among them for imputation purposes.

2.2.2 Data generating mechanism

In the simulation, we considered J studies, indexed by $j = 1, \ldots, J$, where J was set to 7 or 10 depending on the specific scenario. For each study, we generated equal number of patients in each scenario. In our simulation, we chose to have either $n_j = 100$, 300 or 500 individuals per study in each scenario. The covariates were generated from a multivariate normal distribution. The number of covariates was fixed in two. The generating mechanism for the two covariates was:

$$x_i = (x_{i1}, x_{i2})' \sim \text{MVN}\left(\boldsymbol{\mu} = \begin{pmatrix} \mu_{\mathbf{1}} \\ \mu_{\mathbf{2}} \end{pmatrix}, \boldsymbol{\Sigma}\right), i = 1, \ldots, n_j$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} s^2 & r \\ r & s^2 \end{pmatrix}, \quad \boldsymbol{\mu_j} = \begin{cases} 1, & \text{if Bin}(0.5) = 1 \\ -1, & \text{if Bin}(0.5) = 0 \end{cases}$$

An overall study-specific mean of one or minus one was assumed to create heterogeneous studies. To this aim, we assume a Binomial distribution with probability 0.5 that assumes value 1 or -1. The elements of the principal diagonal were set equal ($s1 = s2 = s$), but with different values (1, 1.5, or 2) depending on the scenario. The elements of the secondary diagonal of the variance covariance matrix were generated from a Uniform distribution with upper and lower values depending on the correlation magnitude that we simulate. Those values were randomly drawn for each repetition in each scenario.

The outcome variable was generated as $Y_{ij} = inter + \sum_k \beta_k x_{ijk} + d * treat + \epsilon_i$, where $inter$ is the intercept and β's the covariates coefficient with corresponding constant values -1.280, 1.456, and 1.842. The values of d were generated from normal distribution with mean zero and standard deviation equal to tau, were

the values of *tau* were equal to 0 or 0.2 depending on the scenario. The *treat* binary variable was generated from a Binomial distribution with a success probability equal to 0.5. The random error was simulated for each patient separately as $\epsilon_i \sim \mathcal{N}(0, 1)$.

Finally, we assumed that the first covariate ($\boldsymbol{x_{i1}}$) was missing while the second covariate was always reported in all studies. We explored scenarios where the probability of missingness was either 0.3 or 0.5.

2.2.3 Estimand/target of analysis

We will evaluate and compare methods for estimating the adjusted coefficients (a conditional estimand) of the covariates of the analysis model.

2.2.4 Methods

The methods that we compared were CCA, FCS, SSIM and the full model as it would be if no missing data imputation was conducted. That way we can see how our method performs in comparison to the full model which is the reference method, the simplest approach which is CCA and the FCS which is the most common approach when it comes to missing data.

2.2.5 Performance measures

We measured the performance of the method in three different ways: relative bias (RB), root mean squared error (RMSE), and confidence interval coverage rate (CR). Relative bias is calculated as the difference between the average estimate and the true population value divided by the true value (that is, bias as a percentage of the true value). Several simulation studies have declared relative bias of 10% as acceptable [Kaplan (1988)]. So, relative bias is given by

the type

$$\text{RB} = \frac{\text{Aver. Estimate} - \text{True Value}}{\text{True Value}}$$

The Root Mean Square Error (RMSE) can be conceptualized as the square root of two components: one quantifying the variability of the estimator (precision), and the other assessing its bias (accuracy). This outcome is an overall measure of accuracy that considers both squared bias and sampling variance.

$$\text{RMSE}(Estimate) = \sqrt{\underbrace{\text{Var}(Estimate)}_{\text{Precision}} + \underbrace{\big(\text{Bias}(Estimate)\big)^2}_{\text{Accuracy}}}$$

We computed the 95% confidence interval CR as the proportion of estimated values in which the 95% symmetric confidence interval includes the true parameter value (i.e., the estimate plus or minus 1.96 standard error units). Type I error inflation is reflected in lower rates than the nominal 95%, whereas conservative inference is expressed at higher rates. Confidence interval coverage is an indicator of standard error quality when estimates are unbiased. The type for calculating this quantity is

$$\text{CR} = P(\text{Estimate} - 1.96 \cdot \text{SE} \;\leq\; \text{True Value} \;\leq\; \text{Estimate} + 1.96 \cdot \text{SE})$$

To sum up, the estimator will be considered trustworthy if the RB is smaller than 10%, the RMSE is as low as possible compared to the other methods, and the CR is near 95% (non-lower than 90% and not close to 100%).

2.2.6 Overview of scenarios explored

We compared in simulations the performance of the four approaches in terms of i) Relative Bias (RB), ii) Root mean squared error (RMSE), and iii) Coverage Rate (CR). We generated data across 216 distinct scenarios, and for each scenario we simulated 1000 independent datasets. Within these scenarios, we examined diverse configurations, including variations in the number of

studies, number of patients per study, probability of covariate to be sporadically missing in the studies (number of covariates: 2; the second covariate is always reported), the degree of heterogeneity, the covariates standard deviation, and the correlation of the covariates to each imputed study. The missing mechanism was set to be MAR.

MAR was created using a simple cut-off rule: 80% of the missing values of x_1 were created for observations where x_2 was above its median. We chose this approach not to mimic any particular real-world dataset, but to provide a clear MAR scenario of how the method behaves under missingness.

We explored the following configurations for the data generating mechanisms:

1. Number of studies: $N = 7$ or 10.
2. Probability of sporadically missing for first covariate (p_{x_1}) in each study: 0.3 or 0.5.
3. Number of patients per study: $n = 100$, 300 or 500.
4. Standard deviation of the size of the treatment effect (tau) equal to 0 or 0.2.
5. Standard deviation of the covariates (s) was set to 1, 1.5, or 2.
6. Correlation of covariates (r): small, medium, and large; small is generated from $U(0.15, 0.25)$, medium from $U(0.45, 0.55)$ and large from $U(0.65, 0.75)$.

We calculate the similarity among studies, by using CCA, Euclidean distance and first method of finding studies divergence.

2.2.7 Results

Due to the similarity of many of the simulation results, only a part of them is presented here. Full results can be found in the appendix. We give in the

appendix 3 tables with 72 scenarios each which differ in the variance-covariance matrix.

Table S1 shows summaries of parameter estimates, RB, RMSE and CR. In almost all scenarios with a 50% missing rate, the SSIM methodology outperforms the other methods, whereas in 30% missingness, all methods perform similarly, with CCA having the lowest coverage rate in the second covariate (the one with missing data). In psychiatry, missingness rates around 50% are not uncommon. Interestingly, Wahlbeck et al. (2001) support that dropout rates in mental health trials can exceed 50% in some cases, and the same is true for psychotherapy dropouts in the meta-analysis of Wierzbicki and Pekarik (1993) where the average dropout rate was 46.86%. Additionally, in a meta-analysis of palliative interventions, the authors provide a forest plot showing that the missingness in several studies is close to or more than 50% [Hussain et al. (2016)].

The pattern we observed for 50% missingness is that the SSIM methodology has better statistical properties as the sample size increases. It is slightly better in scenarios with small correlations and sample sizes of 100 and performs better in scenarios with larger sample sizes (scenarios 1-3, 10-12, 19-21, 28-30). The medium correlation scenarios showed characteristics similar to those of the small correlation scenarios, but the differences were smaller (scenarios 4-6, 13-15, 22-24, 31-33). For high correlation values, all methods were nearly identical in most cases for 100 and 300 patients per study, and the SSIM methodology was superior only when the sample size was 500 (scenarios 7-9, 16-18, 25-27, 34-36). When missingness equals 30%, there are no differences across methods.

Also, from Table S1 we can see that as the number of studies increases from seven to ten, SSIM performs better than the traditional approach. As *tau* increases, this difference in the performance is slightly reduced.

In Figure 2.3 we present the results of scenarios 19-21 when missingness is

50% missingness and in Figure 2.4 we present the same scenarios but with 30% missingness (scenarios 55-57). Scenarios 19-21 were the triplet of scenarios that the SSIM methodology outperformed by far the traditional FCS approach (Figure 2.3). Especially in scenarios 20 and 21, the traditional approach CR was smaller for both covariates than all other methods, and with a low bottom for traditional approach with a CR around 0.2 in scenario 21 for the second covariate.

The main difference in the results that lies in the CR, illustrates a well-known property of CR in the presence of bias. When an estimator is biased, increasing the sample size does not eliminate bias; instead, it only reduces the width of the confidence interval. In response, the confidence interval is concentrated around the biased estimate and since it is narrower, it fails to include the true effect size frequently leading to a decline in CR as sample size increases.

In the results, we also report the intercept term although it is an estimate of less interest in IPD MA. In Figure 2.4, the results are almost identical between our approach and the traditional approach (scenarios 55-57), except of the intercept. In general, FCS tends to work better for estimating relative effects among covariates than for estimating absolute effects such as the intercept. Meaning that the relationships between variables are well preserved, whereas the overall level of an outcome can be more difficult to recover accurately. In this sense, FCS can be thought of as a method that accurately estimates the slope of a mountain, while being less reliable at estimating its absolute height.

The sharp differences observed as missingness increases from 30% to 50% may appear disproportionate at first glance. However, these differences should be interpreted in terms of relative data loss rather than absolute percentages. With 500 participants across 10 trials (5,000 observations), 30% missingness corresponds to 3,500 observed values, while 50% missingness corresponds to only 2,500. The loss of 1,000 additional observations represents a 40% reduction relative to the remaining data, which explains the substantial degradation in

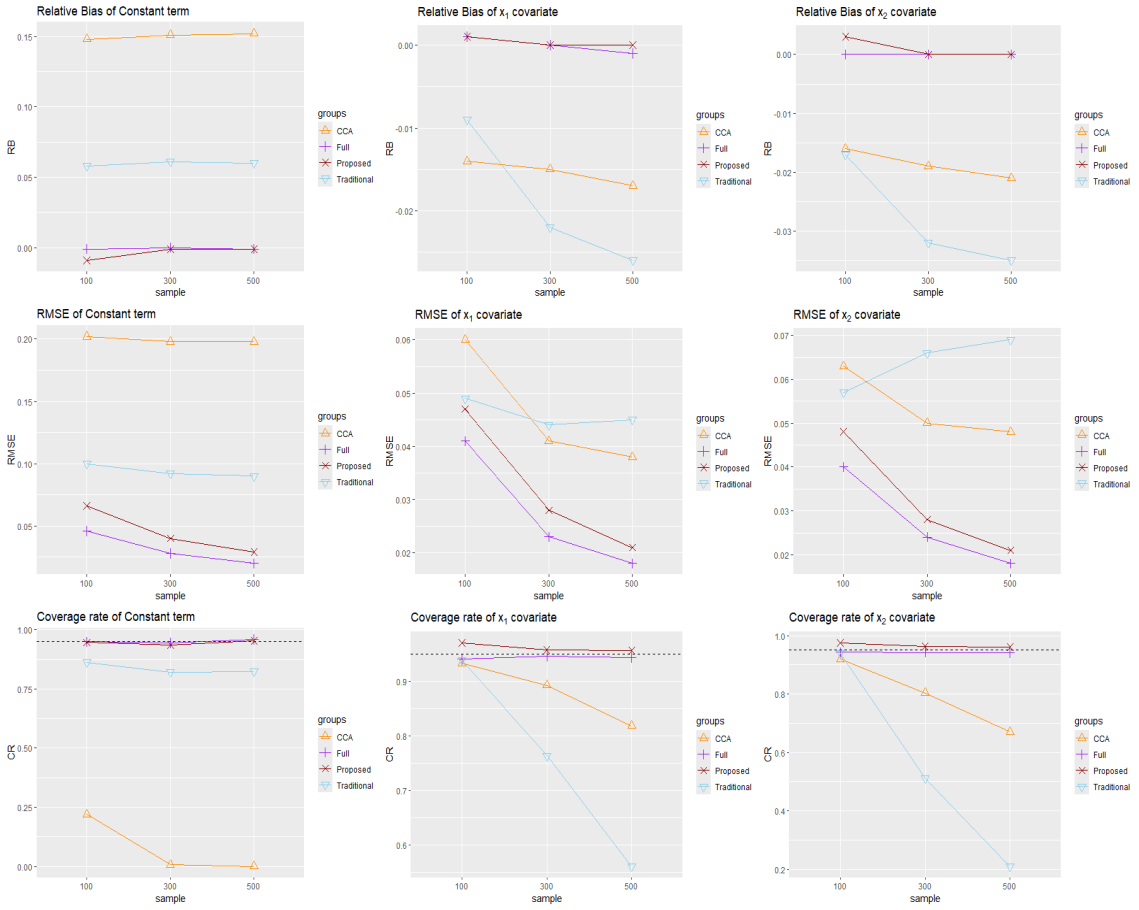

Figure (2.3) Scenario 19 to 21 (common characteristics for those are missingness = 0.5, studies = 10, $tau = 0$, r = small, $s = 1$). Scenario 19 has sample size 100, scenario 20 has sample size 300, and scenario 21 has sample size 500. The top horizontal boxes present the RMSE, the middle presents the CR, and the bottom boxes present the RB. The left vertical boxes present the results for the constant term, the middle for the first covariate, and the right side of the plot the second covariate. The yellow, purple, brown and light blue colors are the CCA, full data regression, SSIM and traditional FCS methods, respectively.

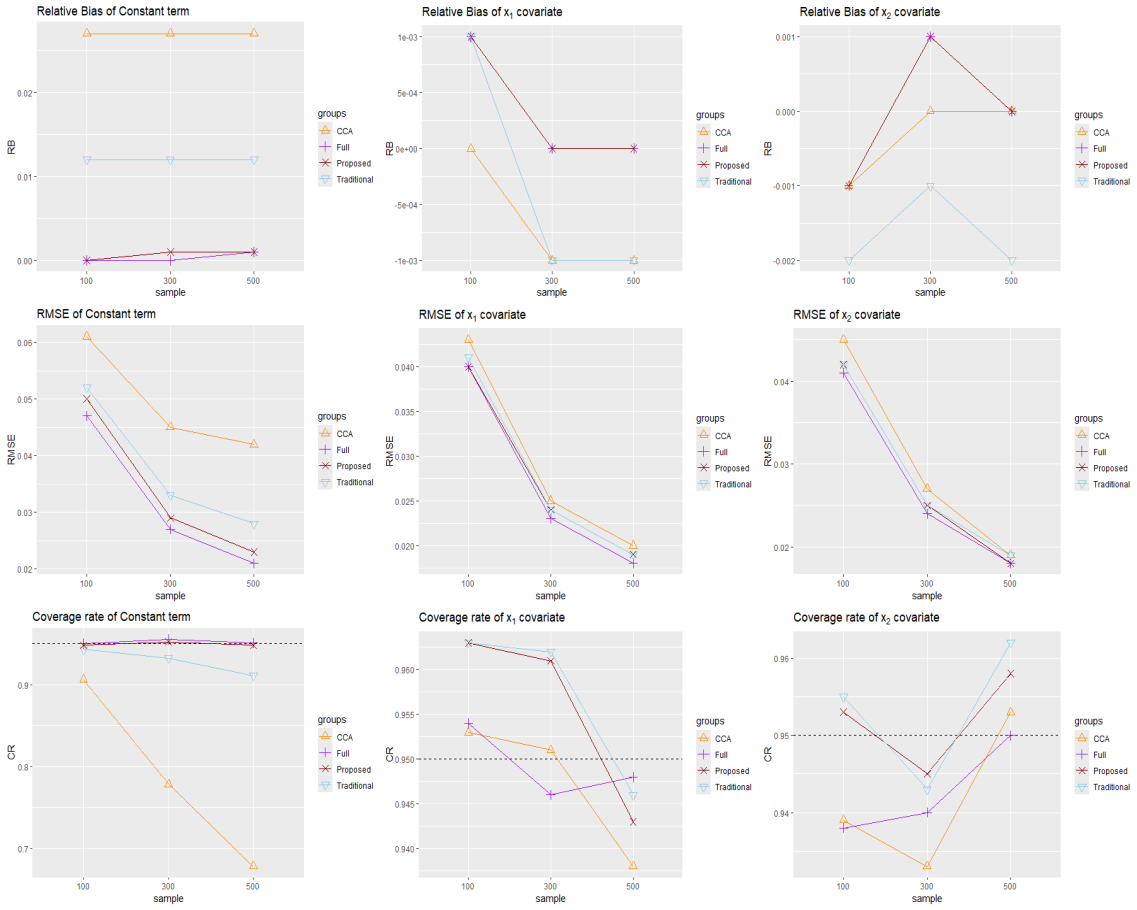

Figure (2.4) Scenario 55 to 57 (common characteristics for those are missingness = 0.3, studies = 10, $tau = 0$, r = small, $s = 1$). Scenario 55 has sample size 100, scenario 56 has sample size 300, and scenario 57 has sample size 500. The top horizontal boxes present the RMSE, the middle presents the CR, and the bottom boxes present the RB. The left vertical boxes present the results for the constant term, the middle for the first covariate, and the right side of the plot the second covariate. The yellow, purple, brown and light blue colors are the CCA, full data regression, SSIM and traditional FCS methods, respectively.

performance observed in the simulation.

Furthermore, we investigated the properties of the covariates when we increased s in variance-covariance matrix. In Tables S2-S3, we see the results of $s = 1.5$ and $s = 2$, respectively. In those cases, the differences between the methods with missingness 50%, were slightly in favor of SSIM for $s = 1.5$ and almost equal for $s = 2$.

2.3 Discussion

We have developed and evaluated techniques for multiple imputation of clustered datasets with variables that are incomplete due to MAR mechanisms. Existing imputation methods are effective for imputing incomplete MAR variables within individual studies, but their applicability is limited in contexts such as IPD MA. This thesis aims to propose a new one-stage multiple imputation method for incomplete continuous MAR variables with sporadic missingness that is based in FCS methodology. This method is specifically designed for clustered datasets, accommodating heterogeneous effects and error variances.

In our simulation we compared the performance of four methods (CCA, Full model, FCS, and SSIM method). We investigated various scenarios with variations in the number of studies, number of patients per study, probability of covariate to be sporadically missing, degree of heterogeneity, the standard deviation of the covariates, and the correlation of the covariates in the data generation mechanism.

We found that the CCA method was overall the worst approach in cases where the missingness of the studies was 50 percent. We also found that SSIM performed best in most of the scenarios that the missingness was 50 percent and was significantly better when the heterogeneity was different from

0. Especially in the cases with large studies and medium to small correlations, the traditional FCS approach was even worse than the CCA.

Our new method was inspired by the divergence that exists among studies in IPD MA. For that reason, we took advantage of known similarity measures for quantifying their distance. Despite the simplicity of the idea, to the best of our knowledge, such methodology has not been previously used in the field of IPD MA. The advantage of our approach is based that if studies that are dissimilar among them are used in the same imputation model, then the most dissimilar one should be excluded from this model but keep them in the analysis model. In that way, we also do not violate the congeniality assumption, since the two models contain the same covariates. Based on this advantage maybe SSIM is more suitable in observational studies than RCT's where the differences among studies are more negligible.

Limitations

This work has several limitations that should be noted. The major limitation is that no real data were used to test SSIM. Due to restrictions on data sharing, we were unable to obtain IPD data to compare the SSIM method with the other methods that were used in the simulation.

Furthermore, in our simulations, we only used simple data-generating mechanisms. We could have explored more complex and potentially more realistic mechanisms, such as non-linear associations between covariates and outcomes, or interactions between covariates. Non-normal continuous variables are potentially problematic for multiple imputation.

Another issue is that if the number of studies is small and there are significant differences between studies or no differences at all, SSIM might perform worse compared to other methods. In our generation mechanism, data are generated in a way that ensures a-priori dissimilarities between studies.

In addition, in all one-stage models in the literature, Rubin's rule is applied

in simulation studies to summarize the random-effects parameters by averaging the estimates from multiple imputations. However, due to the highly skewed distribution of these parameters, using the arithmetic mean may not be the optimal method for summarizing them across multiple imputed datasets. Further research is necessary to systematically evaluate this approach.

Moreover, using a single imputation model for each study can be time-consuming, especially when dealing with a large number of studies. In those cases, SSIM will be slower in comparison with other method, due to fitting several random effects models. It is, however, important to note that this should not be a significant problem nowadays due to fast processors.

Additionally, although we mentioned alternative methods in the introduction to impute missing variables sporadically while accounting for clustering at the study level, we did not consider any of those methods in our simulations. This is something that should be explored further in future simulation studies.

Lastly, although we explored a large number of scenarios, no safe conclusions can be drawn from a single simulation. However, we believe that our study provided a good setting for comparing different methods. There is no doubt that future simulation studies and further exploration of methods will be useful in a number of ways.

Chapter 3

ALLOWING FOR INFORMATIVE MISSINGNESS IN AGGREGATE DATA META-ANALYSIS

In clinical trials, it is common for missing data to be informatively missing (IM) (reasons for dropout are associated with the outcome measured). This is often ignored in aggregate data meta-analysis where we assume that the missing data problem has been handled at the trial level. In this chapter, unlike Chapters 1 and 2, we consider the case that we are at the meta-analysis level of aggregate data without access to the IPD of the trials. For example, for each arm in each trial we have mean values, standard deviations, and sample size for continuous outcomes, whereas, for dichotomous outcomes, we have number of events and sample size. There is no access to covariate values and methods such as multiple imputation cannot be employed. Unlike Chapters 1 and 2, we consider that we have missing outcomes for some of the individuals of each study and not missing covariate values.

Since the outcome is missing, we can only make hypotheses about how the outcome in the missing participants is related to that in the participants who provided it. For example, we may assume that patients in the active

treatment group are likely to drop out due to severe adverse events or those in the placebo group might retreat due to lack of efficacy having a much worse outcome compared to the observed participants (e.g., the outcome in the missing participants is 20-40% worse than that of the observed participants). Based on the context of the problem, one may think of several scenarios of what is happening to the outcome in the missing participants. Methods described in this chapter (that take into account IM) provide advantages over naive approaches (e.g. mean imputation, LOCF), which consider imputed values as fully observed without accounting for the uncertainty induced by imputation.

In the presence of missing data, doing a sensitivity analysis, and exploring how robust results are to the assumptions made is a sensible step. In this chapter, we present the development of an R package, called metamiss, which provides methods for adjusting meta-analytic estimates based on pre-specified scenarios regarding how missing outcomes relate to the observed ones. This document makes an introduction to the set of tools available in the metamiss package and illustrates their use in pairwise and network meta-analysis.

3.1 Theoretical background

We assume that we have $m_{ij} + n_{ij}$ participants randomized, with i index denoting the study and j the arm of the trial while m denotes the number of missing participants and n the number of observed participants. For brevity, we keep the notation to refer to two groups only and we specify j to be equal to T (Treatment) or equal to C (Control). For an extension to multi-arm trials see Mavridis et al. (2015). Initially, we consider two patterns for the data since by dividing our data ($\boldsymbol{y}$) in two patterns; observed($\boldsymbol{y^{obs}}$) and missing ($\boldsymbol{y^{miss}}$). More specifically, we assume ($\boldsymbol{y = (y^{obs}, y^{miss})}$). In our notation, π denotes the probabilities at the population level for a participant. The symbols χ and σ represent the population outcome means (in binary case represents the population proportion) and standard deviations, respectively; and their

sample-based counterparts are denoted by p for probabilities, x for the mean, and s for the standard deviation.

We have the following data for each study i and arm j depending on the nature of the outcome

- Binary outcome: the number of observed successes (r_{ij}) and the number of observed failures ($f_{ij} = n_{ij} - r_{ij}$).
- Continuous outcome: the true mean of the observed data χ_{ij}^{obs} and its variance σ_{ij}^2/n_{ij}, where σ_{ij} is the standard deviation among observed participants.

The relative treatment effect in each study i is defined as the difference

$$\boldsymbol{\beta_i} = f(\chi_{iT}^{tot}) - f(\chi_{iC}^{tot})$$

where χ_{ij}^{tot} represents the true mean (proportion) outcome of both observed and missing outcomes in the jth group of the ith study and f is a link function that depends on the nature of the relative effect.

We consider randomized clinical trials in which there are no important differences across arms. According to Little and Rubin (2002), under the MAR assumption the true mean outcome equals to mean (proportion) of the observed data ($\chi_{ij}^{tot} = \chi_{ij}^{obs}$). Under MNAR assumption, the true mean (proportion) outcome can be thought as a mixture of means in the observed and missing data. Under this speculation, we can write the true mean (proportion) as:

$$\chi_{ij}^{tot} = \pi_{ij}\chi_{ij}^{obs} + (1 - \pi_{ij})\chi_{ij}^{miss} \tag{3.1}$$

where for continuous data, we consider that the probability of a participant completing the study (being observed) follows a Beta distribution with parameters n_{ij} and m_{ij}, $\pi_{ij} \sim Beta(n_{ij}, m_{ij})$ and its expected value is equal to $p_{ij} = \frac{n_{ij}}{m_{ij}+n_{ij}}$ and χ_{ij}^{obs} is also directly estimated from the data, while χ_{ij}^{miss} cannot. For binary data, the π_{ij} can be estimated as $p_{ij} = \frac{m_{ij}}{m_{ij}+f_{ij}+r_{ij}} = \frac{m_{ij}}{m_{ij}+n_{ij}}$

and χ_{ij}^{obs} is also directly estimated from the data as the sample proportion $\frac{r_{ij}}{f_{ij}+r_{ij}}$.

In the case of continuous outcomes, the link function is the identity function ($f(x) = x$), if the measure of interest is the Mean Difference (MD), is the function $f(x_i) = \frac{x_i}{s_{i,pooled}}$ where $s_{i,pooled}$ is the pooled standard deviation in the ith study, if the measure of interest is the Standardized Mean Difference (SMD), and finally the logarithmic function ($f(x) = logx$), if the measure of interest is the logarithm of Ratio of Means (logROM). In contrast, for binary outcomes the link function is the identity function, when the measure of interest is the Risk Difference (RD)), the logarithmic function when the measure of interest is the logarithm of Risk Ratio (RR), and the logit function ($f(x) = logitx$), when the measure of interest is the logarithm of Odds Ratio (OR).

The two patterns (observed and missing participants) are linked through a parameter lambda that sets the difference in the outcome between these two patterns and quantifies the IM. We set

$$\lambda_{ij} = g(\chi_{ij}^{miss}) - g(\chi_{ij}^{obs})$$

such that the mean value of the missing data is connected to the mean of the observed data (in study i and arm j) by an unknown parameter λ_{ij}. Depending on the nature of the outcome different function are attributed in the $g(.)$ function. According to White et al. (2008), informative missingness (IM) in binary outcomes is quantified by comparing the odds of the outcome between participants with unobserved outcomes and those with observed outcomes. This comparison is expressed as informative missingness odds ratios (IMORs), which can vary between the two trial arms. The logit function $g(x) = logit(x)$ is used, and the IMOR is referred to as logIMOR (logarithm of informative missingness odds ratios).

In continuous outcomes, $g(.)$ has the possibility of two choices, that of identity function $g(x) = x$ and that of the logarithmic function $g(x) = log(x)$,

which are called IMDoM (informative missingness difference of means) and logarithm of the informative missingness ratio of means (IMRoM) respectively. The choice of the link functions of $f(.)$ and $g(.)$ in the continuous outcomes are independent of each other. However, logRoM is expected to be used spontaneously when we consider IMRoM, while MD and SMD are to be used with IMDoM, as they are intuitively related.

Since, $\boldsymbol{\lambda}$'s are not identified by the observed data, we can set them to assume a range of values (random variable) and conventionally we set them to follow a normal distribution e.g., $\boldsymbol{\lambda \sim N(\mu_\lambda, \sigma^2_\lambda)}$.
This λ can be specified to be:

a) Separate for each arm j and study i, λ_{ij}

b) Study specific λ_i

c) Correlated across arms

Expert opinion can be used to inform the joint distribution of lambda in the treatment and control groups. When $\lambda_{ij} = 0$, that is equivalent to assuming MAR mechanism and this is equal to the complete case analysis, while to account for deviations from the MAR assumption we allow λ_{ij} to take nonzero values.

Since the aim of this approach is to estimate the treatment effect β_i and its variance in each study, given the prior distribution of λ_{ij} and the observed data, a two-stage procedure is used. In the first step, an adjustment of the effect sizes and their variances in each study is done to account for the IM data, while in the second, the adjusted effect sizes are pooled. Two proposed methods for estimating β_i and its uncertainty are employed. The first method employs full Bayesian inference and Monte Carlo integration and is considered the gold standard. However, due to its potential computational complexity and time consuming, an approximate method based on a Taylor series expansion is also performed. Details on the methods can be found in White et al. (2008),

Higgins et al. (2008), and Mavridis et al. (2015). Note that a one-stage pattern mixture model has also been suggested White et al. (2008).

3.2 The metamiss package

Below we describe in detail the arguments for the meta-analytic models. The user must give the arm-based data for each treatment group. In the commands of the *metamiss* package we can apply routines from the *meta* package for the presentation of the results, e.g. *forest* command for plotting a forest plot [Balduzzi et al. (2019)]. Also, both random and common effects model(s) are presented.

Separated by outcomes

1) Continuous outcomes arguments

- ***meanC, sdC, meanT, sdT***: are the mean and standard deviation of the observed data of the control (C) and treatment (T) group, respectively.
- ***nC, mC, nT, mT***: are the number of observed and missing participants in the control and treatment group, respectively.
- ***sm***: specify the measure of interest with plausible values: ***SMD*** (Standardized Mean Difference), ***MD*** (Mean Difference), and ***ROM*** (Ratio of Means - ROM is back transformed and is not in logarithmic scale).
- ***pool***: if the measure is set to SMD, the user can also specify the *pool* argument. The plausible values are ***yes*** and ***no***, where ***yes*** sets the standard deviations of both groups equal to the pooled standard deviation, while ***no*** leaves their original values. The default option is ***yes***.

2) Binary outcomes arguments

The arguments for both binary and proportional outcomes are the same except that for proportional outcomes we define only the treatment group (***rT, fT*** and ***mT***, see below).

- ***rC, fC, rT, fT***: are the number of successes and failures in the observed data in control and treatment groups, respectively.
- ***mC, mT***: are the number of missing participants in control and treatment groups, respectively.
- ***sm***: specify the measure of interest with plausible values: ***logOR, logRR***, and ***RD***.
- ***incr***: a numerical value that denotes the continuity correction if needed to apply in zero-cell counts.

Informative missingness arguments

The remaining arguments are the same for both continuous and binary outcomes. The only exception is the argument *type*, which is only used for continuous outcomes.

- ***type***: specify the IM pattern for continuous outcomes. The plausible values are **IMDoM** and **IMRoM**.
- ***mlT, sdlT, mlC, sdlC***: The ***mlT*** and ***sdlT*** are the mean value and standard deviation of the distribution of lambda in the treatment group, while ***mlC*** and ***sdlC*** are the respective values for the distribution of lambda in the control group. One could use the same value of lambda across studies by writing down only a single value e.g. $\boldsymbol{mlT} = 1$. If we want to make lambda study specific, then a vector should be given c(#a vector equal to the number of studies). If those values are left unspecified then we are in CCA (all equal to zero). Those values can be informed by

expert opinion. For proportions we have to specify only the values of the treatment group.

- ***corl***: argument for the correlation of the two lambdas, no need for correlation in proportional outcomes. The default value is zero.

Estimation options and meta-analytic model

- ***method***: Define the method to use for integration over the distribution of the IM. Two options are available, **Bootstrap** and **Taylor**. With the **Bootstrap** argument, bootstrap is performed to estimate the variance of the effect size while **Taylor** uses a Taylor-series expansion to estimate the variance of the effect size without bootstrapping.

- ***studlab***: An optional argument in which a vector with the studies names is contained.

- ***iterations***: An optional argument with the number of bootstrap repetitions (default is 1000)

- ***method.tau***: An optional argument with the available methods to estimate the between-study variance τ^2. The default option is the "REML" (Restricted Maximum Likelihood), the alternative methods are the ones used in *meta* package.

For synthesizing the effect sizes, we use the inverse variance meta-analytic model.

3.3 Pairwise meta-analysis

3.3.1 Continuous outcome

We consider a dataset with eighteen placebo-controlled trials aiming to evaluate an antipsychotic for the treatment of schizophrenia. This dataset is included in the *metamiss* package.

```
library(metamiss)
data("Antipsychotic")
```

In this data set, we have a continuous outcome, and we consider a random effects meta-analysis model using the standardized mean difference (SMD) as an effect size. In this data set, we have the mean outcomes, standard deviations, number of completers, number of missing patients, and number of LOCF (Last Observation Carried Forward) imputed for each study and treatment. Data are shown in Table 3.1, and Figure 3.1 presents the forest plot.

Pairwise meta-analysis can be applied in this dataset by using the R-package *meta* [Balduzzi et al. (2019)]. Applying the ***metacont*** function of *meta* package, we can perform a CCA, which is equivalent to setting $\lambda = 0$ in our model.

3.3.1.1 Sensitivity analysis for missing data of continuous outcomes with cont_cor

cont_cor performs the adjusting estimating method for meta-analysis using a pattern mixture model with either a Taylor series approximation or Bootstrap resampling for estimating the adjusted standard error of the effect size. To run the function, we must specify the mean values, standard deviations, number of completers, and number of missing values for each treatment and study. We also must specify the ***sm*** (effect size), and the ***type*** argument that depends

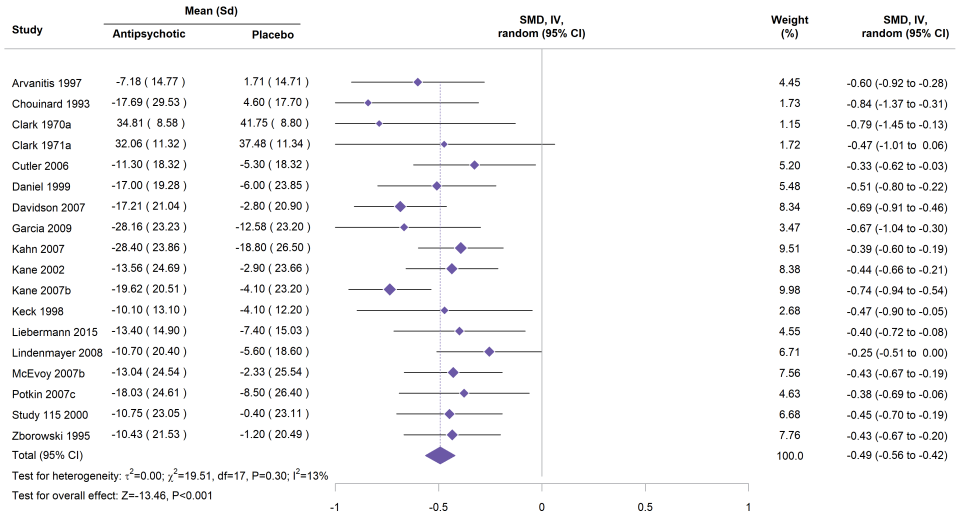

Figure (3.1) Forest plot of the antipsychotic versus placebo treatments.

on the g function that quantifies the degree of departure from MAR by the IMDoM or by the IMRoM.

The mathematical details of this method can be found in Mavridis et al. (2015).

We can run the function ***cont_cor*** for applying complete case analysis by

cont_cor(meanC = Antipsychotic$mean.plac, sdC = Antipsychotic$sd.plac, nC = Antipsychotic$n.plac, mC = Antipsychotic$m.plac, meanT = Antipsychotic$mean.antips, sdT = Antipsychotic$sd.antips, nT = Antipsychotic$n.antips, mT = Antipsychotic$m.antips, sm = "SMD", type = "IMDOM", method = "Taylor")

Suppose we want λ distribution values to be treatment specific equal across studies, such as λ of the treatment $\sim N(0, 4)$ and λ of the control group

Study	Antipsychotic					Placebo				
	mean	sd	mis	compl	impu	mean	sd	mis	compl	impu
Arvanitis 1997	-7.18	14.77	3	71	83	1.71	14.71	0	16	35
Chouinard 1993	-17.69	29.53	0	25	18	4.60	17.70	0	6	16
Clark 1970a	34.81	8.58	1	28	1	41.75	8.80	0	13	1
Clark 1971a	32.06	11.32	8	56	1	37.48	11.34	3	17	1
Cutler 2006	-11.30	18.32	0	53	41	-5.30	18.32	3	44	41
Daniel 1999	-17.00	19.28	1	67	36	-6.00	23.85	1	45	46
Davidson 2007	-17.21	21.04	4	166	83	-2.80	20.90	3	47	73
Garcia 2009	-28.16	23.23	2	49	9	-12.58	23.20	3	39	22
Kahn 2007	-28.40	23.86	12	361	97	-18.80	26.50	3	85	30
Kane 2002	-13.56	24.69	0	190	118	-2.90	23.66	0	58	48
Kane 2007b	-19.62	20.51	1	357	145	-4.10	23.20	1	58	68
Keck 1998	-10.10	13.10	6	24	17	-4.10	12.20	1	24	23
Liebermann 2015	-13.40	14.90	7	67	8	-7.40	15.03	5	66	14
Lindenmayer 2008	-10.70	20.40	15	117	135	-5.60	18.60	6	29	49
McEvoy 2007b	-13.04	24.54	6	69	131	-2.33	25.54	1	30	77
Potkin 2007c	-18.03	24.61	6	52	62	-8.50	26.40	2	21	39
Study 115 2000	-10.75	23.05	9	135	105	-0.40	23.11	3	27	53
Zborowski 1995	-10.43	21.53	8	101	123	-1.20	20.49	10	44	62

Table (3.1) Antipsychotic meta-analysis: mean scores, standard deviations, numbers of non-completers, completers and LOCF imputed for the antipsychotic and placebo arms.

$\sim N(1, 1)$.

cont_cor(meanC = Antipsychotic$mean.plac, sdC = Antipsychotic$sd.plac, nC = Antipsychotic$n.plac, mC = Antipsychotic$m.plac, meanT = Antipsychotic$mean.antips, sdT = Antipsychotic$sd.antips, nT = Antipsychotic$n.antips, mT = Antipsychotic$m.antips, mlT = 0, sdlT = 2, mlC = 1, sdlC = 1, sm = "SMD", type = "IMDOM", method = "Taylor")

We could also give values in the ***corl*** argument for the correlation of those two λs. Moreover, to give different values of λ for each arm and study, we could give as argument a vector for each mean and sd of lambdas by using c(#a vector equal to the number of studies). If we choose not to give values in the arguments related to the lambda parameter or if we set $\lambda = 0$ (mean and variance of λ equal to zero), then we are performing a CCA, equivalent to assuming MAR.

3.3.1.2 Sensitivity analysis for LOCF imputed and missing data of continuous outcomes with bilocf

The LOCF method is a statistical approach for imputing missing data which is typically used in the context of longitudinal repeated measure studies. In LOCF, any missing follow-up visit value is replaced by the last observed value for that subject. This means that the most recent available data point is carried forward to fill in the gaps. Since this is a single imputation approach, the combined dataset (including both observed and imputed values), is then analyzed as if there were no missing data. If there is progress in the disease, LOCF results have questionable validity and may even be outright misleading [Lachin (2016)]. Consequently, meta-analyses that include clinical trials using LOCF imputation are also likely to be biased.

The pattern described in the theoretical background section is extended when we know how many patients were LOCF imputed. The new pattern now is

($\boldsymbol{y' = (y^{obs}, y^{miss}, y^{imp})}$), where $\boldsymbol{y^{imp}}$ is the outcome for those participants that were imputed. Throughout this section, we will consider LOCF imputed values as this is the most common method, but y^{imp} could refer to any other single imputation method (e.g., mean imputation). To account for the LOCF imputed patients, equation 3.1 is extended in

$$\chi_{ij}^{tot} = \pi_{ij}^{rep}\chi_{ij}^{rep} + \pi_{ij}^{miss}\chi_{ij}^{miss} = \pi_{ij}^{obs}\chi_{ij}^{obs} + \pi_{ij}^{imp}\chi_{ij}^{imp} + \pi_{ij}^{miss}\chi_{ij}^{miss} \tag{3.2}$$

where *rep* means that the patients are reported (either observed or imputed). In counterpart with λ in the simple case, we introduce a δ parameter that quantifies the bias in the imputed values as the difference between the true outcome and the imputed outcome in early dropouts. More about the theoretical concept of this method can be found in Mavridis et al. (2019).

The ***bilocf*** function is another function that is used for continuous outcome data, but this time we have the information on how many of these participants were imputed with the LOCF approach. It is an extension of the model that we use in the ***cont_cor*** function. The arguments from the ***cont_cor*** function apply in this function, but the available measures are now reduced to two, the SMD and MD, and the *type* argument is no longer available since the only plausible choice is IMDOM. However, extra arguments come from a parameter delta that is the bias in the LOCF outcome (it relates the outcome of the LOCF participants to that of the observed participants). Those extra arguments for the ***bilocf*** function are:

- ***mgT, sdgT, mgC, sdgC***: The ***mgT*** and ***sdgT*** are the mean value and standard deviation of the distribution of delta in the treatment group, while ***mgC*** and ***sdgC*** are the respectfully value for the distribution of delta in the control group. We could use the same value of deltas across studies, specify only a value e.g. 1. If we want to make delta study specific, then a vector should be given c(#a vector equal to the number of studies). If those values are left unspecified or set equal to zero, then

we are in the LOCF case. Ideally, those values should be informed from historical data and/or expert opinion.

- ***corg***: argument for the correlation of the two deltas.

We make again some assumptions about the distribution of delta as we did with lambda, and we give values to those normal distributions accounting ideally using an expert's opinion. If we set the values of delta equal to zero, then we are back in the case of the LOCF analysis. Again, if both λ and δ are zero, then we are in the CCA case.

We can run the function ***bilocf*** for applying complete case analysis by

```
bilocf(meanC = Antipsychotic$mean.plac, sdC = Antipsychotic$sd.plac, nC = Antipsychotic$n.plac, mC = Antipsychotic$m.plac, meanT = Antipsychotic$mean.antips, sdT = Antipsychotic$sd.antips, nT = Antipsychotic$n.antips, mT = Antipsychotic$m.antips, lT = Antipsychotic$l.antips, lC = Antipsychotic$l.plac, sm = "SMD", method = "Taylor")
```

Suppose we want λ distribution values to be equal across studies and treatment specific, such as λ of the treatment $\sim N(0, 4)$ and λ of the control group $\sim N(1, 1)$, while δ is common both treatment and control groups $\sim N(0, 1)$.

bilocf(meanC = Antipsychotic$mean.plac, sdC = Antipsychotic$sd.plac, nC = Antipsychotic$n.plac, mC = Antipsychotic$m.plac, meanT = Antipsychotic$mean.antips, sdT = Antipsychotic$sd.antips, nT = Antipsychotic$n.antips, mT = Antipsychotic$m.antips, lT = Antipsychotic$l.antips, lC = Antipsychotic$l.plac, sm = "SMD", method = "Taylor", mlT = 0, sdlT = 2, mlC = 1, sdlC = 1, mgT = 0, sdgT =1, mgC =0, sdgC = 1)

Again, if we want δ's to be study specific, we can give them as an argument a vector of length equal to the number of studies.

3.3.2 Binary outcome

Load the Haloperidol data of the *metamiss* package. The Haloperidol data set includes 17 trials comparing the effectiveness of haloperidol with placebo for the treatment of schizophrenia.

data("Haloperidol")

In this data set the outcome is binary and we consider a random effects meta-analysis model using the Risk Ratio (RR) as an effect size. The variables contained in this data set are the number of successes, the number of failures, and the number of missing observations for each study and treatment. Two extra columns exist with the author of the study's name and the year of conduct. Data are shown in Table 3.2, and Figure 3.2 presents the forest plot.

Pairwise meta-analysis can be applied to this dataset by using the *R*-package *meta* [Balduzzi et al. (2019)]. Applying the ***metabin*** function of the *meta* package, we perform a CCA.

Study	Haloperidol			Placebo		
	Successes	Failures	Missing	Successes	Failures	Missing
Arvanitis 1997	25	25	2	18	33	0
Beasley 1996	29	18	22	20	14	34
Becheli 1983	12	17	1	2	28	1
Borison 1992	3	9	0	0	12	0
Chouinard 1993	10	11	0	3	19	0
Durost 1964	11	8	0	1	14	0
Garry 1962	7	18	1	4	21	1
Howard 1974	8	9	0	3	10	0
Marder 1994	19	45	2	14	50	2
Nishikawa a 1982	1	9	0	0	10	0
Nishikawa b 1984	11	23	3	0	13	0
Reschke 1974	20	9	0	2	9	0
Selman 1976	17	1	11	7	4	18
Serafetanidis 1972	4	10	0	0	13	1
Simpson 1967	2	14	0	0	7	1
Spencer 1992	11	1	0	1	11	0
Vichaiya 1971	9	20	1	0	29	1

Table (3.2) Haloperidol vs Placebo for the treatment of Schizophrenia: number of successes, failures, and non-completers for the haloperidol and placebo arms.

3.3.2.1 Sensitivity analysis for missing data of binary outcomes with bin_cor

bin_cor performs the adjusting estimating method for meta-analysis by using a pattern mixture model with either a Taylor series approximation or Bootstrap resampling for estimating the adjusted standard error of the effect size. The arguments we need to specify are number of successes, the number of failures and the number of missing values for each treatment and study.

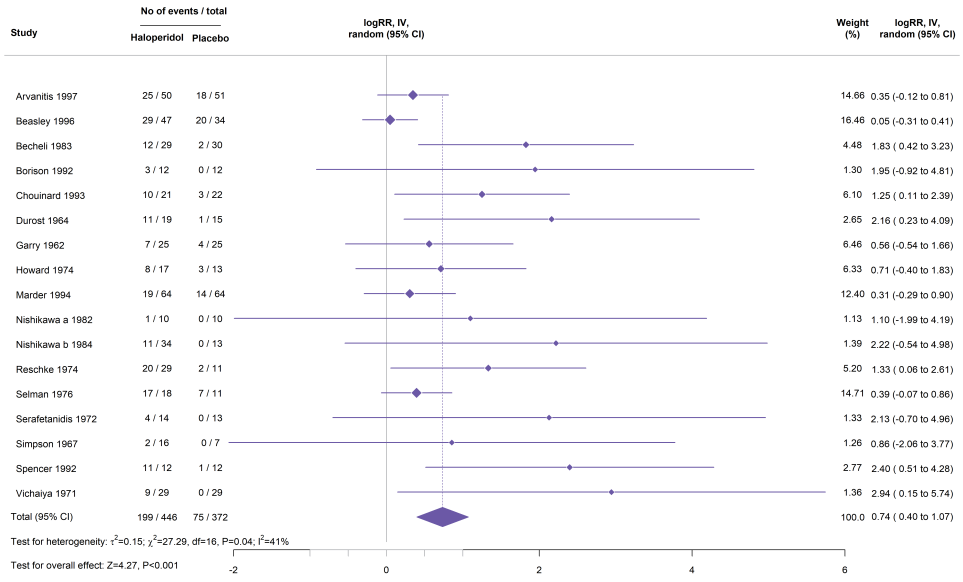

Figure (3.2) Forest plot of the haloperidol versus placebo treatments. A continuous correction was applied in zero cell studies.

As its continuous counterpart, we can specify the distribution of lambda according to our beliefs and ideally using expert opinion. The correlation between those two lambdas can be specified by the ***corl*** argument. Similarly, to the continuous case, different lambdas could be assigned to each arm and study.

Available measures for specification in the binary case are the logarithm of Odds Ratio (logOR), the logarithm of Risk Ratio (logRR), or the Risk Difference (RD). Because of the nature of binary data, there might be studies with no events in one or both groups in a study. Since, this can create computational problems, we used some values to create a fictitious event. Our intention is to avoid using large numbers for not disturbing the results. The usual value is 0.5 (***incr*** argument).

The presentation and mathematical details are presented in the articles of White et al. (2008) and White et al. (2008). The authors prefer to express IM via the odds ratio between missingness and success, which they call IMOR, and via the logarithm transformation, which they call logIMOR.

We can run the function bin_cor for applying complete case analysis by

```
bin_cor(rC = Haloperidol$rc, fC = Haloperidol$fc, mC = Haloperidol$mc, rT = Haloperidol$rt, fT = Haloperidol$ft, mT = Haloperidol$mt, sm = "logRR", studlab = paste(Haloperidol$author, Haloperidol$year), method = "Taylor")
```

Suppose we want λ distribution values to be equal across studies and treatment specific, such as λ of the treatment $\sim N(0, 4)$ and λ of the control group $\sim N(1, 1)$.

```
bin_cor(rC = Haloperidol$rc, fC = Haloperidol$fc, mC = Haloperidol$mc, rT = Haloperidol$rt, fT = Haloperidol$ft, mT = Haloperidol$mt, sm = "logRR", studlab = paste(Haloperidol$author, Haloperidol$year), mlT = 0, sdlT = 2, mlC = 1, sdlC = 1, method = "Taylor")
```

3.3.3 Single proportional outcome

3.3.3.1 Sensitivity analysis for missing data of proportional outcomes with prop_cor

prop_cor performs the adjusting estimating method for meta-analysis of proportions by using a pattern mixture model with either a Taylor series approximation or Bootstrap resampling for estimating the adjusted standard error of the effect size. The arguments we need to specify are the number of

successes, the number of failures, and the number of missing values for the treatment and study. In ***prop_cor*** we can specify the distribution of lambda's treatment ideally by an expert's opinion or our beliefs. Like in the previous functions, λ can be either a number (the same for each study) or a vector (different for each study). Since no control arm exists there is no need for ***corl*** argument. If no argument is specified for lambda, then the CCA model will perform.

Like in binary outcomes data, there might be studies with no events in a group of a study. A zero-cell correction can also be made with the ***incr*** argument, just like the ***bin_cor***.

The equation, mathematical formulation, and proof of this method cannot be found anywhere since it is a result that is derived from the binary outcome method. The logit transformation is the only one available in our method, unlike ***metaprop***, which offers a wider range of transformations.

We can run the function ***prop_cor*** for applying complete case analysis by

```
prop_cor(rT = Haloperidol$rt, fT = Haloperidol$ft, mT =
Haloperidol$mt, studlab = paste(Haloperidol$author, Haloperi-
dol$year), method = "Taylor")
```

Suppose we want λ distribution values to be equal across studies, such as λ of the treatment $\sim N(0, 4)$.

```
prop_cor(rT = Haloperidol$rt, fT = Haloperidol$ft, mT =
Haloperidol$mt, studlab = paste(Haloperidol$author, Haloperi-
dol$year), mlT = 0, sdlT = 2, method = "Taylor")
```

To take the same estimates with those of ***metaprop*** function we should apply in the results of our function the back logit transformation which is **exp(.)/(1 + exp(.))**.

3.4 Network meta-analysis

We consider the Antidepressant dataset, which includes twelve studies aiming at evaluating nine antidepressants within the context of network meta-analysis. We can load those data from the *metamiss* package. We have only included continuous data for the context of the NMA in the *metamiss* package.

```
data("Antidepressant")
```

In this data set the outcome is continuous and consider a random effects NMA model using the standardized mean difference (SMD) or ROM as an effect size. This data set includes the mean outcomes, standard deviations, number of completers and number of missing patients for each study and treatment. The twelve studies were selected from a large network of antidepressants for illustrative purposes and are not indicative of treatment effectiveness [Cipriani et al. (2009)]. We focused on this subset because the observed mean values for each treatment group, prior to any imputation strategies (e.g., last observation carried forward), are available exclusively for these studies. Data are shown in Table 3.3, and Figure 3.3 presents the network plot.

NMA can be applied in this dataset by using the *R*-package *netmeta* [Balduzzi et al. (2023)]. Applying the ***netmeta*** function of *netmeta* package, we can perform a CCA, which is equivalent to setting $\lambda = 0$ in our model.

Most of the methodologies that was discussed in the pairwise meta-analysis have been extended in the framework of network meta-analysis. Specifically, for the binary and continuous data, and not for LOCF and proportional data. The arguments for those functions are proportional to those of pairwise functions with extra arguments for picking the reference treatment and denoting the treatment arms. The argument that specifies the reference treatment is called ***ref*** and can be specified as seen below for binary and continuous data. The

Study	Comparisons		Treat A				Treat B			
	Treat A	Treat B	mean	sd	n	m	mean	sd	n	m
Kato 2006	Fluxovamine	Paroxetine	9.00	8.65	41	8	4.60	8.65	39	13
Rossini 2002	Fluxovamine	Sertaline	7.56	12.31	39	1	11.27	11.33	45	3
Annseau 1994	Milnacipran	Fluoxetine	17.20	11.10	74	23	14.20	11.10	75	18
Yang 2003	Milnacipran	Sertaline	13.50	2.10	12	15	13.80	1.80	15	11
Eker 2005	Reboxetine	Sertaline	6.55	5.23	20	5	7.76	2.89	21	3
Chen 2001	Sertaline	Venlafaxine	7.00	5.00	45	0	6.00	5.00	43	1
Shelton 2006	Sertaline	Venlafaxine	9.60	6.20	63	19	8.40	5.40	67	11
Ekselius 1997	Citalopram	Sertaline	10.40	9.12	37	163	11.00	9.12	55	145
Khanzode 2003	Citalopram	Fluoxetine	15.80	3.30	30	3	18.70	5.10	32	2
Diaz/Martinez 1998	Fluoxetine	Venlafaxine	8.20	7.52	55	20	8.70	7.52	55	15
Amini 2005	Mirtazapine	Fluoxetine	7.40	7.52	16	2	10.40	7.52	15	3
Silverston 1999	Fluoxetine	Venlafaxine	12.00	8.65	89	32	11.30	8.65	91	37

Table (3.3) Antidepressant network meta-analysis: study name, treatments compared, observed mean change in depression score, standard deviations, number of completers, and noncompleters for each arm.

treat*1** and ***treat*2** arguments are given as character vectors with the treatment names. If the network contains studies with more than two arms, then we should obligatorily give the values in ***studlab argument for the function to understand this feature.

If the outcome is continuous then in accordance with our previous continuous example, we specify the same arguments for the model and tease only the arguments ***ref***, ***treat*1**, and ***treat*2**.

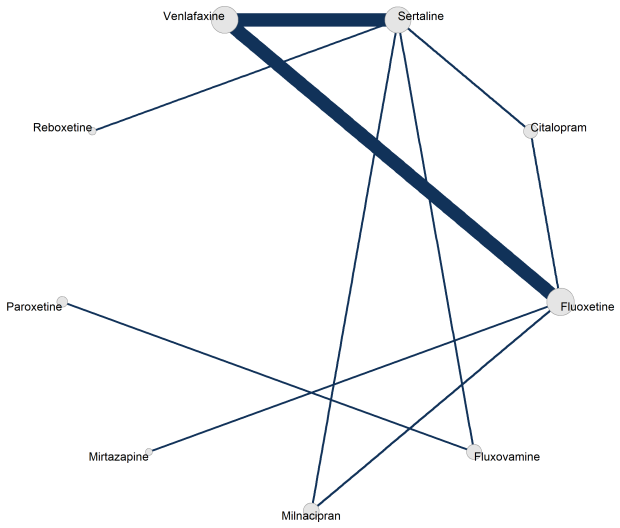

Figure (3.3) Network plot for comparison of antidepressants.

```
nmacont_cor (meanC = Antidepressant$meanC, sdC = Antide-
pressant$sdC, mC = Antidepressant$mC, nC = Antidepres-
sant$nC, meanT = Antidepressant$meanT, sdT = Antidepres-
sant$sdT, mT = Antidepressant$mT, nT = Antidepressant$nT,
treat1 = Antidepressant$Treat1, treat2 = Antidepressant$Treat2,
studlab = Antidepressant$Study, sm = "SMD", ref ="Flux-
ovamine", mlT = 0, sdlT = 2, mlC = 1, sdlC = 1, type =
"IMDOM", method = "Taylor")
```

else if binary we could use the **nmabin_cor** and specify the treatment that we want as reference.

3.5 Sensitivity of the results

Since the results that we acquired by using the previous methodology are sensitive to changes of λ and the correlations that we choose to use, we have made a plot that gives the meta-analysis diamond for different scenarios that we choose. For example, let's say that we want to compare the results for different values of mean, variance and correlation of λ's for the Antipsychotic data.

First model: $\lambda_T \sim N(0,1), \quad \lambda_C \sim N(0,1)$ and $\rho = 0$

Second model: $\lambda_T \sim N(1,4), \quad \lambda_C \sim N(4,6)$ and $\rho = 0.3$

Third model: $\lambda_T \sim N(2,3), \quad \lambda_C \sim N(5,2)$ and $\rho = -0.8$

To begin with, we must run the three models separately and store their results separately

model1 ← cont_cor(meanC = Antipsychotic$mean.plac, sdC = Antipsychotic$sd.plac, nC = Antipsychotic$n.plac, mC = Antipsychotic$m.plac, meanT = Antipsychotic$mean.antips, sdT = Antipsychotic$sd.antips, nT = Antipsychotic$n.antips, mT = Antipsychotic$m.antips, mlT = 0, sdlT = 1, mlC = 0, sdlC = 1, sm = "SMD", type = "IMDOM", method = "Taylor")
model2 ← cont_cor(meanC = Antipsychotic$mean.plac, sdC = Antipsychotic$sd.plac, nC = Antipsychotic$n.plac, mC = Antipsychotic$m.plac, meanT = Antipsychotic$mean.antips, sdT = Antipsychotic$sd.antips, nT = Antipsychotic$n.antips, mT = Antipsychotic$m.antips, mlT = 1, sdlT = 2, mlC = 4, sdlC = $\sqrt{6}$, corl = 0.3, sm = "SMD", type = "IMDOM", method = "Taylor")

model3 ← cont_cor(meanC = = Antipsychotic$mean.plac, sdC = Antipsychotic$sd.plac, nC = Antipsychotic$n.plac, mC = Antipsychotic$m.plac, meanT = Antipsychotic$mean.antips, sdT = Antipsychotic$sd.antips, nT = Antipsychotic$n.antips, mT = Antipsychotic$m.antips, mlT = 2, sdlT = $\sqrt{3}$ mlC = 5, sdlC =$\sqrt{2}$, corl = -0.8, sm = "SMD", type = "IMDOM", method = "Taylor")

Then, we have to run the function **sensit.an**, by defining its arguments. The first one is the models that we ran in the previous step, the second one is the limits in our graph, and the third one specifies the λ's distribution and correlation that will be plotted in the graph.

sensit.an(models = list(model1, model2, model3), limsvec = c(-0.63, -0.35), lambdas = list(c(0, 1, 0, 1, 0), c(1, 2, 4, 6, 0.3), c(2, 3, 5, 2, -0.8)))

Based on the Figure 3.4, we can see how sensitive our results are to lambda specifications. Considering the plot results, we can claim that the data are not sensitive to different lambdas. In addition, we could use CCA model in the sensitivity plot to see how the results would differ without experts' opinions.

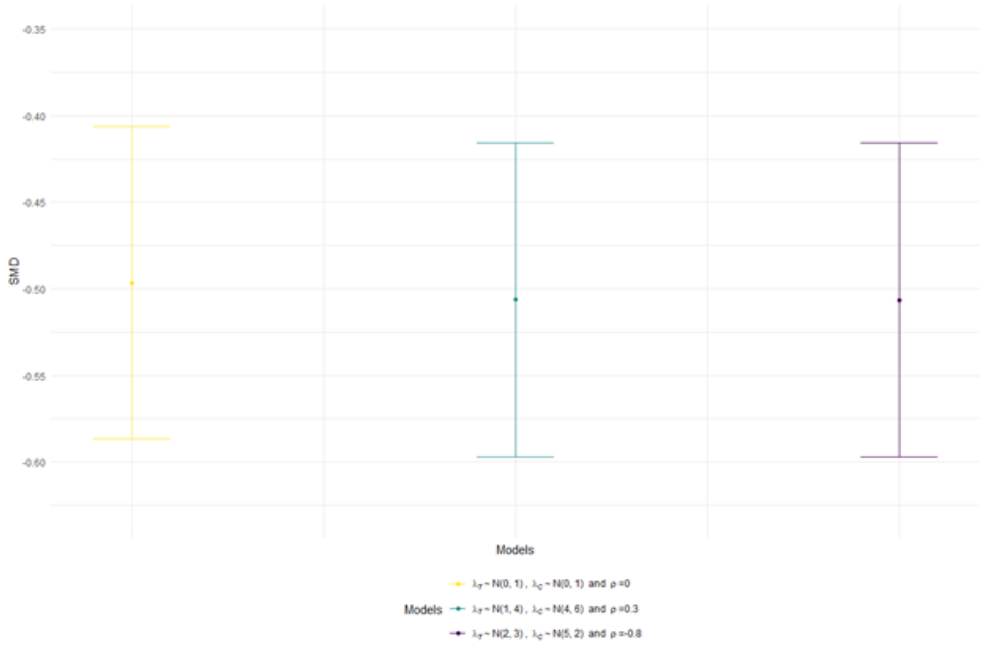

Figure (3.4) Sensitivity analysis plot in the Antipsychotic dataset. The yellow, green and purple error bars depict the First, Second and Third models, respectively. The error bar represents the diamond of the forest plots depending on the assumptions for the lambda's in each model.

Chapter 4

CONCLUSIONS AND FURTHER RESEARCH

Missing data is a challenging issue in meta-analysis and statistics more broadly. In this work, we focus on missing data in both AgD and IPD, and we provide potential solutions along with code implementations to address these challenges.

4.1 Conclusions

So far, the imputation of missing studies in IPD MA is done by borrowing power from all the available studies. This raises some questions about the quality of the imputation process when studies are not very similar. We built based on FCS a one-stage imputation approach that successfully deals with sporadically missing values when multiple individual datasets are being used such as in IPD MA.

We recommend its use and believe that offers a potentially powerful alternative to researchers, when performing an IPD MA with sporadically missing covariates, rather than applying traditional imputation methods that do not (fully) account for between-study heterogeneity. Resulting from the extensive simulation studies, the SSIM imputation approach also has favorable statistical

properties in terms of bias, RMSE and CR. Finally, the code for using SSIM is freely accessible in GitHub.

Another important issue in MA of missing data is the handling of missing data when only AgD are available. The standard way of presenting the results in cases of missing data is CCA. This is equivalent to making the assumption that missing data are MAR. We present an *R* library that can adjust effect estimates and their standard errors by making assumptions on how missing outcomes are related to observed outcomes. Since this is a missing data problem and we rely on hypotheses, we can only make assumptions about the missing data and the method should be used as a sensitivity analysis.

There are two commands in STATA that do similar analysis, the *metamiss* and its extension *metamiss2* command. For a step-by-step demo for the usage of *metamiss* and *metamiss2* commands see the articles of White and Higgins (2009) and Chaimani et al. (2018), respectively. Our package adds to the arsenal of meta-analysis packages the function for the LOCF missing data approach and the meta-analysis of proportions which is a straight abridgment of the binary methods.

4.2 Further research

In terms of further research, there exist various intriguing research topics that could be relevant to our work. One of our primary goals for future extensions is to consider a different type of outcome for SSIM.

Another interesting line of research is to extend the SSIM methodology to other type of missingness assumptions. In this study, we restricted our attention to data that are MAR, but the performance of SSIM should also be explored under MCAR and MNAR assumption. When more than one covariate is available in a model, the type of missingness is complex and it is difficult to assume one of those three types. By using acyclic graphs, we can

investigate how missing variables are influenced by other variables and relax those assumptions.

Also, SSIM since it is a one-stage method can be extended to systematically missing covariates. A problem would arise if at least one of the subclasses includes a covariate that is systematically missing in each study, as the FCS approach cannot estimate an imputation model for a variable with no observed values within that subclass. If that were the case, our model would not work. Furthermore, calculating distance metrics when studies have systematically missing covariates is problematic, but we propose multiple imputations or single imputations in Chapter 2.1.1 for those reasons. Keep in mind that we expect to encounter studies that also vary in their PICO criteria. So, another extension of the method that we could avoid checking the closeness of studies based on the covariates, would be to check how studies vary in terms of PICO, reducing reliance on purely data-driven criteria.

In addition, more simulations should be done with different types of covariates (dichotomous, ordered, mix), which are widely measured in health sciences. For this to work, we should use other distance metrics such as Gower's distance to account for the different types of variables, which can be done easily.

As it concerns the metamiss package, this is the first *R* package that uses those methods, as far as we know. Someone can view our video in the Cochrane training course to better understand the issues addressed in this article by clicking on this link: Cochrane Training Video. Finally, we are planning to upload the package metamiss to CRAN in order to increase its popularity.

BIBLIOGRAPHY

Ades, A., N. J. Welton, S. Dias, D. M. Phillippo, and D. M. Caldwell (2024). Twenty years of network meta-analysis: Continuing controversies and recent developments. *Research Synthesis Methods 15*(5), 702–727.

Airy, G. B. (1861). *On the algebraical and numerical theory of errors of observations and the combination of observations*. Macmillan.

Audigier, V. and M. Resche-Rigon (2023). micemd: Multiple imputation by chained equations with multilevel data. R package version 1.9.0.

Audigier, V., I. R. White, S. Jolani, T. P. Debray, M. Quartagno, J. Carpenter, S. Van Buuren, and M. Resche-Rigon (2018). Multiple imputation for multilevel data with continuous and binary variables. *Statistical Science 33*(2), 160–183.

Azur, M. J., E. A. Stuart, C. Frangakis, and P. J. Leaf (2011). Multiple imputation by chained equations: what is it and how does it work? *International Journal of Methods in Psychiatric Research 20*(1), 40–49.

Balduzzi, S., G. Rücker, A. Nikolakopoulou, T. Papakonstantinou, G. Salanti, O. Efthimiou, and G. Schwarzer (2023). netmeta: an R package for network

meta-analysis using frequentist methods. *Journal of Statistical Software 106*, 1–40.

Balduzzi, S., G. Rücker, and G. Schwarzer (2019). How to perform a meta-analysis with R: a practical tutorial. *BMJ Mental Health 22*(4), 153–160.

Burgess, S., I. R. White, M. Resche-Rigon, and A. M. Wood (2013). Combining multiple imputation and meta-analysis with individual participant data. *Statistics in Medicine 32*(26), 4499–4514.

Burke, D. L., J. Ensor, and R. D. Riley (2017). Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. *Statistics in Medicine 36*(5), 855–875.

Chaimani, A., D. Mavridis, G. Salanti, J. P. Higgins, and I. R. White (2018). Allowing for informative missingness in aggregate data meta-analysis with continuous or binary outcomes: extensions to metamiss. *The Stata Journal 18*(3), 716–740.

Chandler, J., M. Cumpston, T. Li, M. J. Page, and V. Welch (2019). *Cochrane Handbook for Systematic Reviews of Interventions*. Hoboken: Wiley.

Cipriani, A., T. A. Furukawa, G. Salanti, J. R. Geddes, J. P. Higgins, R. Churchill, N. Watanabe, A. Nakagawa, I. M. Omori, H. McGuire, et al. (2009). Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *The Lancet 373*(9665), 746–758.

Cochrane, A. L. (1979). 1931-1971: a critical review, with particular reference to the medical profession. *Medicines for the year 2000. London: Office of Health Economics*, 1–11.

Cro, S., T. P. Morris, M. G. Kenward, and J. R. Carpenter (2020). Sensitivity analysis for clinical trials with missing continuous outcome data using con-

trolled multiple imputation: a practical guide. *Statistics in Medicine 39*(21), 2815–2842.

Dias, S., N. J. Welton, D. M. Caldwell, and A. E. Ades (2010). Checking consistency in mixed treatment comparison meta-analysis. *Statistics in Medicine 29*(7-8), 932–944.

Egger, M., D. Buitrago-Garcia, and G. D. Smith (2022). *Systematic Reviews of Epidemiological Studies of Etiologyand Prevalence*, Volume 1. Wiley Online Library.

Enders, C. K., T. Hayes, and H. Du (2018). A comparison of multilevel imputation schemes for random coefficient models: fully conditional specification and joint model imputation with random covariance matrices. *Multivariate Behavioral Research 53*(5), 695–713.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher 5*(10), 3–8.

Grund, S., O. Lüdtke, and A. Robitzsch (2018). Multiple imputation of missing data for multilevel models: Simulations and recommendations. *Organizational Research Methods 21*(1), 111–149.

Harrell, F. E. (2012). Chapter 3. In *Regression modeling strategies*, Second Edition, pp. 45–59. Springer.

Higgins, J. P., I. R. White, and A. M. Wood (2008). Imputation methods for missing outcome data in meta-analysis of clinical trials. *Clinical Trials 5*(3), 225–239.

Hirst, L., I. Ntaga, J. Seehra, D. Mavridis, and N. Pandis (2025). Reporting and interpretation of subgroup analyses in orthodontic meta-analysis; a meta-epidemiological study. *European Journal of Orthodontics 47*(4), cjaf053.

Hussain, J. A., I. R. White, D. Langan, M. J. Johnson, D. C. Currow, D. J. Torgerson, and M. Bland (2016). Missing data in randomized controlled trials

testing palliative interventions pose a significant risk of bias and loss of power: a systematic review and meta-analyses. *Journal of Clinical Epidemiology 74*, 57–65.

Jolani, S. (2018). Hierarchical imputation of systematically and sporadically missing data: an approximate bayesian approach using chained equations. *Biometrical Journal 60*(2), 333–351.

Jolani, S., T. P. Debray, H. Koffijberg, S. van Buuren, and K. G. Moons (2015). Imputation of systematically missing predictors in an individual participant data meta-analysis: a generalized approach using MICE. *Statistics in Medicine 34*(11), 1841–1863.

Kadlec, D., K. L. Sainani, and S. Nimphius (2023). With great power comes great responsibility: common errors in meta-analyses and meta-regressions in strength & conditioning research. *Sports Medicine 53*(2), 313–325.

Kaplan, D. (1988). The impact of specification error on the estimation, testing, and improvement of structural equation models. *Multivariate Behavioral Research 23*(1), 69–86.

Kontopantelis, E. (2018). A comparison of one-stage vs two-stage individual patient data meta-analysis methods: A simulation study. *Research Synthesis Methods 9*(3), 417–430.

Kunkel, D. and E. E. Kaizar (2017). A comparison of existing methods for multiple imputation in individual participant data meta-analysis. *Statistics in Medicine 36*(22), 3507–3532.

Lachin, J. M. (2016). Fallacies of last observation carried forward analyses. *Clinical Trials 13*(2), 161–168.

Little, R. J. and D. B. Rubin (2002). *Statistical analysis with missing data*. John Wiley & Sons.

Little, R. J. and Y. Wang (1996). Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics 52*(1), 98–111.

Lu, G. and A. Ades (2006). Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association 101*(474), 447–459.

Lumley, T. (2002). Network meta-analysis for indirect treatment comparisons. *Statistics in Medicine 21*(16), 2313–2324.

Mavridis, D., G. Salanti, T. A. Furukawa, A. Cipriani, A. Chaimani, and I. R. White (2019). Allowing for uncertainty due to missing and LOCF imputed outcomes in meta-analysis. *Statistics in Medicine 38*(5), 720–737.

Mavridis, D., I. R. White, J. P. Higgins, A. Cipriani, and G. Salanti (2015). Allowing for uncertainty due to missing continuous outcome data in pairwise and network meta-analysis. *Statistics in Medicine 34*(5), 721–741.

Maxwell, L., P. Shreedhar, M. Carabali, and B. Levis (2024). How to plan and manage an individual participant data meta-analysis. an illustrative toolkit. *Research Synthesis Methods 15*(1), 166–174.

Mistler, S. A. and C. K. Enders (2017). A comparison of joint model and fully conditional specification imputation for multilevel missing data. *Journal of Educational and Behavioral Statistics 42*(4), 432–466.

Morris, T. P., I. R. White, and M. J. Crowther (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine 38*(11), 2074–2102.

Muñoz, J., O. Efthimiou, V. Audigier, V. M. de Jong, and T. P. Debray (2024). Multiple imputation of incomplete multilevel data using Heckman selection models. *Statistics in Medicine 43*(3), 514–533.

Pearson, K. (1904). Report on Certain Enteric Fever Inoculation Statistics. *The British Medical Journal 2*(2288), 1243–6.

Pham, T. M., N. Pandis, and I. R. White (2022). Missing data, part 3. How to explore missing data. *American Journal of Orthodontics and Dentofacial Orthopedics 162*(2), 282–283.

Quartagno, M. and J. Carpenter (2016). Multiple imputation for IPD meta-analysis: allowing for heterogeneity and studies with missing covariates. *Statistics in Medicine 35*(17), 2938–2954.

Quartagno, M. and J. Carpenter (2023). jomo: A package for Multilevel Joint Modelling Multiple Imputation. R package.

Resche-Rigon, M. and I. R. White (2018). Multiple imputation by chained equations for systematically and sporadically missing multilevel data. *Statistical Methods in Medical Research 27*(6), 1634–1649.

Resche-Rigon, M., I. R. White, J. W. Bartlett, S. A. Peters, S. G. Thompson, and P.-I. S. Group (2013). Multiple imputation for handling systematically missing confounders in meta-analysis of individual participant data. *Statistics in Medicine 32*(28), 4890–4905.

Rice, K., J. P. Higgins, and T. Lumley (2018). A re-evaluation of fixed effect (s) meta-analysis. *Journal of the Royal Statistical Society Series A: Statistics in Society 181*(1), 205–227.

Riley, R. D., J. Ensor, M. Hattle, K. Papadimitropoulou, and T. P. Morris (2023). Two-stage or not two-stage? That is the question for IPD meta-analysis projects. *Research Synthesis Methods 14*(6), 903–910.

Riley, R. D., J. P. Higgins, and J. J. Deeks (2011). Interpretation of random effects meta-analyses. *BMJ Research Methods and Reporting 342*(7804), d549.

Riley, R. D., P. C. Lambert, and G. Abo-Zaid (2010). Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ Clinical Research Edition 340*(7745), 1–7.

Riley, R. D., J. F. Tierney, and L. A. Stewart (2021). Individual participant data metaanalysis: a handbook for healthcare research, Chapter 18.

Rostain, A. L., D. W. Goodman, H. L. Starr, Y.-W. Ma, S. Ascher, and R. B. Armstrong (2017). Randomized, 6-week, placebo-controlled study of treatment for adult attention-deficit/hyperactivity disorder: individualized dosing of osmotic-release oral system (oros) methylphenidate with a goal of symptom remission. *The Journal of Clinical Psychiatry 78*(1), 105–114.

Rubin, D. (1976). Inference and missing data. *Biometrika 63*(3), 581–592.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. John Wiley & Sons Inc., New York.

Salanti, G. (2012). Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Research Synthesis Methods 3*(2), 80–97.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. CRC press.

Schafer, J. L. and R. M. Yucel (2002). Computational strategies for multivariate linear mixed-effects models with missing values. *Journal of Computational and Graphical Statistics 11*(2), 437–457.

Schmidt, F. L. (2017). Statistical and measurement pitfalls in the use of meta-regression in meta-analysis. *Career Development International 22*(5), 469–476.

Seitidis, G., S. Nikolakopoulos, E. Hennessy, E. Tanner-Smith, and D. Mavridis (2022). Network meta-analysis techniques for synthesizing prevention science evidence. *Prevention Science 23*(3), 415–424.

Seitidis, G., S. Nikolakopoulos, I. Ntzoufras, and D. Mavridis (2023). Inconsistency identification in network meta-analysis via stochastic search variable selection. *Statistics in Medicine 42*(26), 4850–4866.

Seo, M. (2022). bipd: Bayesian Individual Patient Data Meta-Analysis using 'JAGS'. R package version 0.3.

Spineli, L. M. and N. Pandis (2020). Problems and pitfalls in subgroup analysis and meta-regression. *American Journal of Orthodontics and Dentofacial Orthopedics 158*(6), 901–904.

Stewart, L. A. and J. F. Tierney (2002). To ipd or not to ipd? advantages and disadvantages of systematic reviews using individual patient data. *Evaluation & the Health Professions 25*(1), 76–97.

Tanner, M. A. and W. H. Wong (1987). The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association 82*(398), 528–540.

Thompson, S. G. and J. P. Higgins (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine 21*(11), 1559–1573.

Van Buuren, S. (2018). *Flexible imputation of missing data*. CRC press.

Van Buuren, S. and K. Groothuis-Oudshoorn (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software 45*, 1–67.

Veroniki, A. A., G. Seitidis, G. Tsivgoulis, A. H. Katsanos, and D. Mavridis (2023). An introduction to individual participant data meta-analysis. *Neurology 100*(23), 1102–1110.

Wahlbeck, K., A. Tuunainen, A. Ahokas, and S. Leucht (2001). Dropout rates in randomised antipsychotic drug trials. *Psychopharmacology 155*, 230–233.

White, I. R. and J. P. Higgins (2009). Meta-analysis with missing data. *The Stata Journal 9*(1), 57–69.

White, I. R., J. P. Higgins, and A. M. Wood (2008). Allowing for uncertainty due to missing data in meta-analysis—part 1: two-stage methods. *Statistics in Medicine 27*(5), 711–727.

White, I. R., N. Pandis, and T. M. Pham (2022a). Missing data, part 6. Choices in doing multiple imputation. *American Journal of Orthodontics and Dentofacial Orthopedics 162*(5), 795–797.

White, I. R., N. Pandis, and T. M. Pham (2022b). Missing data, part 7. Pitfalls in doing multiple imputation. *American Journal of Orthodontics and Dentofacial Orthopedics 162*(6), 975–977.

White, I. R., P. Royston, and A. M. Wood (2011). Multiple imputation using chained equations: issues and guidance for practice. *Statistics in Medicine 30*(4), 377–399.

White, I. R., N. J. Welton, A. M. Wood, A. Ades, and J. P. Higgins (2008). Allowing for uncertainty due to missing data in meta-analysis—part 2: hierarchical models. *Statistics in Medicine 27*(5), 728–745.

Wierzbicki, M. and G. Pekarik (1993). A meta-analysis of psychotherapy dropout. *Professional Psychology: Research and Practice 24*(2), 190–195.

Wulff, J. N. and L. E. Jeppesen (2017). Multiple imputation by chained equations in praxis: guidelines and review. *Electronic Journal of Business Research Methods 15*(1), 41–56.

Yucel, R. M. (2011). Random covariances and mixed-effects models for imputing multivariate multilevel continuous data. *Statistical Modelling 11*(4), 351–370.

Zhao, J. H. and J. L. Schafer (2023). pan: Multiple imputation for multivariate panel or clustered data. R package version 1.9.