UNIVERSITY OF IOANNINA
SCHOOL OF MEDICINE
FACULTY OF HEALTH SCIENCE

DEPARTMENT OF HYGIENE AND EPIDEMIOLOGY

STUDIES WITH MISSING VALUES IN NETWORK META-ANALYSIS

Loukia Spineli
Statistician

Dissertation Prepared for the Degree of
DOCTOR OF PHILOSOPHY

IOANNINA 2014

ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΥΓΕΙΑΣ
ΤΜΗΜΑ ΙΑΤΡΙΚΗΣ

ΕΡΓΑΣΤΗΡΙΟ ΥΓΙΕΙΝΗΣ ΚΑΙ ΕΠΙΔΗΜΙΟΛΟΓΙΑΣ

ΜΕΛΕΤΕΣ ΜΕ ΕΛΛΙΠΟΥΣΕΣ ΤΙΜΕΣ ΣΤΗ ΜΕΤΑ-ΑΝΑΛΥΣΗ ΔΙΚΤΥΩΝ

Λουκία Σπινέλη
Στατιστικός

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΙΩΑΝΝΙΝΑ 2014

Η έγκριση της διδακτορικής διατριβής από την Ιατρική Σχολή του Πανεπιστημίου Ιωαννίνων δεν υποδηλώνει αποδοχή των γνωμών του συγγραφέα Ν. 5343/32, άρθρο 202, παράγραφος 2 (νομική κατοχύρωση του Ιατρικού τμήματος).

Ημερομηνία αίτησης της κ. Σπινέλη Λουκίας: 28-1-2011

Ημερομηνία ορισμού Τριμελούς Συμβουλευτικής Επιτροπής: 708^{α}/8-3-2011

Μέλη Τριμελούς Συμβουλευτικής Επιτροπής:

Επιβλέπων

Ντζάνη Ευαγγελία Επίκουρη Καθηγήτρια Υγιεινής με ιδιαίτερη έμφαση στην Επιδημιολογία

Μέλη

Σαλαντή Γεωργία Λέκτορας Επιδημιολογίας

Julian Piers Thomas Higgins, MRC (Medical Research Courial) του Ινστιτούτου Δημόσιας Υγείας του Cambridge

Ανασύσταση Τριμελούς Συμβουλευτικής Επιτροπής: Αριθμ Συνεδρ. 718^{α}/12-7-2011

Επιβλέπων

Σαλαντή Γεωργία Επίκουρη Καθηγήτρια Επιδημιολογίας

Μέλη

Σκαπινάκη Πέτρο Επίκουρο Καθηγητή Ψυχιατρικής

Julian Piers Thomas Higgins, MRC (Medical Research Courial) του Ινστιτούτου Δημόσιας Υγείας του Cambridge

Ημερομηνία ορισμού θέματος: 1-9-2011

«Μελέτες με ελλειπούσες τιμές στη μετα-ανάλυση δικτύων».

ΔΙΟΡΙΣΜΟΣ ΕΠΤΑΜΕΛΟΥΣ ΕΞΕΤΑΣΤΙΚΗΣ ΕΠΙΤΡΟΠΗΣ : 763^{α}/10-4-2014

1. Higgins Julian Piers Thomas, Professor of Evidence Synthesis, University of Bristol and University York, H.B.
2. Μαυρέας Βενετσάνος Καθηγητής Ψυχιατρικής του Τμήματος Ιατρικής του Παν/μίου Ιωαννίνων
3. Δημολιάτης Ιωάννης Επίκουρος Καθηγητής Υγιεινής του Τμήματος Ιατρικής του Παν/μίου Ιωαννίνων
4. Σαλαντή Γεωργία Επίκουρη Καθηγήτρια Επιδημιολογίας του Τμήματος Ιατρικής του Παν/μίου Ιωαννίνων
5. Τσιλίδης Κωνσταντίνος Λέκτορας Υγιεινής με έμφαση στην Επιδημιολογία του Τμήματος Ιατρικής του Παν/μίου Ιωαννίνων
6. Υφαντής Θωμάς Αναπληρωτής Καθηγητής Ψυχιατρικής του Τμήματος Ιατρικής του Παν/μίου Ιωαννίνων
7. Σκαπινάκης Πέτρος Επίκουρος Καθηγητής Ψυχιατρικής του Τμήματος Ιατρικής του Παν/μίου Ιωαννίνων

Έγκριση Διδακτορικής Διατριβής με βαθμό **«ΑΡΙΣΤΑ»** στις **22-5-2014**

ΠΡΟΕΔΡΟΣ ΤΟΥ ΤΜΗΜΑΤΟΣ ΙΑΤΡΙΚΗΣ

Φωτόπουλος Ανδρέας

Καθηγητής Πυρηνικής Ιατρικής

Η Γραμματέας του Τμήματος

ΚΑΠΙΤΟΠΟΥΛΟΥ ΜΑΡΙΑ

Acknowledgements

I would like to sincerely thank my supervisor and mentor Dr. Georgia Salanti for her understanding, guidance and patience during my PhD studies. This collaboration fulfilled a desire I had as a postgraduate student; to write and publish scientific articles. I also feel grateful and very lucky that I had the opportunity through this collaboration to meet and collaborate with successful and respectful scientists in the field of statistics and health. All the experiences I lived during these three years, no matter how difficult and stressful they were, proved to be a valuable lesson for my evolution as a person and as a scientist. This is something I realized in my new life in Hannover.

I would also like to sincerely thank Dr. Nikolaos Pandis for his assistance, support, advice and endless patience in our collaboration. During my collaboration with Dr. Pandis I had the opportunity to contribute to the writing of a chapter in his book and this brought me closer to the fulfillment of a childhood dream; to write a book. I honestly hope this collaboration to keep up in the future and give rise to more articles and chapters along with assistance, support, advice and endless patience from his side.

I would like to thank Dr Stefan Leucht for our fruitful and enjoyable collaboration. I feel grateful for having the opportunity to meet Dr. Leucht in person during my visit in Munich for a project which gave rise to a publication. Dr. Leucht and his wife helped me a lot during my accommodation in Munich and I am very thankful for their support. I truly hope for more fruitful and enjoyable collaborations with Dr. Leucht in the future.

Finally, and most importantly, I would like to thank my family, my friends and my co-workers in the institute for Biometry in the Medical School of Hannover for their support, patience and understanding during the writing of the present thesis.

Abstract

Premature discontinuation is common in mental health trials and the dropout rate has been reported to be considerable high. Despite the high dropout out, the reporting on the extent of missing outcome data, the analytical strategies undertaken and the acknowledgment of the implication of missing outcome data on the results have been particularly insufficient. Another importance consideration in the missing outcome data issue is the minimization of their occurrence. In order to take the necessary measures to prevent as much as possible the missing outcome data, it is important to investigate and to understand the reasons why patients terminated the trials early. In addition, in the field of psychiatry, where missingness is usually informative, it is important to investigate the extent to which informative missing outcome data may have distorted the conclusions of the randomized controlled trials and meta-analyses.

The present study aims to provide empirical evidence about the methodology reported to address and to describe the extent of the missing outcome data as well as the acknowledgement of their impact in the findings of the systematic reviews. Moreover, this study aims to detect trial characteristics that may be associated with discontinuation rates in a large collection of randomized controlled trials in schizophrenia. It also investigates the characteristics associated with the likelihood that a patient falls into different categories of completing the study, being included in a LOCF analysis, or being excluded from the analysis altogether. Last, the present study extends the idea of IMOR into a network meta-analysis setting in order to explore the impact of missing outcome data on the inferences about the relative effectiveness of several competing treatments in psychiatric trials. All statistical analyses have been implemented using the WinBUGS software.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction to missing outcome data

1.1 Missing outcome data in psychiatric trials: an overview

Randomized controlled trials are the "gold-standard" of the methodological designs for minimizing bias in the assessment of treatment effectiveness and safety by balancing known and unknown prognostic factors between the compared treatment arms, with any imbalance attributed merely to chance (1-4). Nevertheless, despite a well-designed study protocol, some of the randomized participants may leave the trial early without providing any or part of the data or be excluded after randomization (5-7). Premature discontinuation results in unavailable values, known as missing outcome data, which could have contributed to the analysis results if they were observed (8). In literature, the terms dropouts, attrition, losses to follow-up and withdrawals are also used synonymously to missing outcome data (9).

Missing outcome data are ubiquitous, especially in psychiatric trials (1;4;10). Elie et al. found that 60-89% of the randomized controlled trials published between 2005 and 2007 in five top general medical journals have reported missing outcome data (11). Wood et al. investigated 71 randomized controlled trials published in four top medical journals of which 63 (89%) reported having missing outcome; of these, 13 (21%) had more than 20% dropout rates (12). Depending on the extent of the risk of bias, the reliability and the generalizability of the results on the population of interest may be affected (1;4;13). Schulz and Grimes suggest that dropout rate of 5% or lower is considered too low to introduce bias, whereas dropout rate of 20% or greater is a warning that readers should be concerned about the possibility of bias; although any intermediate dropout rates may still be a threat of bias (14).

A review of the literature revealed that the prevalence of missingness in antidepressant trials was considerable. As stated by Nemeroff, almost one-third of the patients left the trial early, within the first month of antidepressant treatment, whereas nearly 44% of patients dropped out within the first 3 months (15). Khan et al investigated 45 trials of seven antidepressants using the Food and Drug

Administration (FDA) database and found a mean dropout rates of 37% (16). A recent FDA review of 24 randomized controlled trials of pediatric antidepressants reported that the mean dropout rate ranged from 24% to 44% (17). Heo et al. reviewed 68 antidepressant randomized controlled trials for the depressed elderly published from 1990 to 2003 and reported an overall 27.3% dropout rate (18).

The prevalence of early treatment discontinuation in antipsychotic drug trials was also found to be considerably high ranging from 25% - 75% (10;19). Liu-Seifert et al studied the discontinuation of treatment in four randomized trials of 1627 schizophrenic patients and found that 866 (53%) patients discontinued mostly either due to poor response along with symptom worsening or due to intolerability to medication (10;19). As mentioned by Rabinowith and Davidov, the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) trial reported that 74% of the participants discontinued the trial medication (20). Wahlbeck et al investigated 18,585 participants randomized in 163 antipsychotic drug trials identified in the Schizophrenia Module of the Cochrane Library and reported that one-third of the participants dropped out of the trials (21). In a recent meta-analysis of 163 trials of antipsychotics the average dropout rate was 33%, whereas dropout rates as high as 42% have been reported by other authors (22).

Dropout rate may be associated with patient characteristics, such as psychiatric diagnosis, efficacy and side-effects as well as trial characteristics, such as treatment assignment (i.e., placebo, psychopharmacological), duration of trial, dosing schedule and number of treatment arms. In a placebo-controlled randomized trial of 684 patients with major depressive disorder comparing a serotonin-noradrenalin reuptake inhibitor (SNRI) with a selective serotonin reuptake inhibitor (SSRI) the dropout rate was estimated at 21.2%, 28.6% and 27% in SSRI, SNRI and placebo, respectively, and it was lower compared to the 35% dropped out rate reported by the US FDA summary basis of approval reports (13;23). All seven pivotal trials of three antipsychotics approved by FDA that were compared with placebo and/ or haloperidol reported missing participants with mean dropout 50% for those assigned to the antipsychotics, 63% for placebo and 56% for haloperidol (16). Khan et al reviewed 103 placebo-controlled clinical trials reported in the FDA summary basis of approval reports the evaluated 20 psychotropics approved in the United States between 1985 and 2004 for the treatment of depression, schizophrenia, obsessive compulsive disorder, generalized anxiety disorder and panic disorder (13). They found that patients with schizophrenia had the highest dropout rates (51%), followed by patients with depression (35%), generalized anxiety disorder (31%), panic disorder (26%) and obsessive compulsive disorder (22%). Khan et al also found that psychiatric patients assigned to placebo had higher dropout rates than patients assigned to psychotropics in trials for antipsychotic and OCD agents. Additionally, they reported that there was significant association between the early dropouts and the interaction of the treatment with the outcome; premature discontinuation among the psychiatric patients assigned to placebo was mainly attributed to inefficacy, where early terminations among patients assigned to psychotropics were mostly related to side-effects. Duration of

trial, dosing schedule and number of treatment arms had weak association with the dropout rates.

Additionally, it has been stated that atypical drugs (also known as *newer antipsychotic agents*) can achieve greater improvement in psychotic symptomatology compared to typical drugs (also known as *conventional antipsychotic agents*), since the first are linked with increased acceptability and compliance due to lower extrapyramidal effects or prolactin levels (24;25). Martin et al compared atypical with atypical antipsychotic drugs in the treatment of schizophrenia, based on the dropout rates for any reason of 8 randomized controlled trials by performing a meta-analysis (26). In these trials, the dropout rate was ranging from 22% to 74% in atypical treatments and from 19% to 65% in the typical treatments (haloperidol or chlorpromazine). They found that there was less risk of all-cause dropout from atypical medications with flexible dosage, in both the short and long term, compared to the typical medications. However, no difference was observed when dosage was fixed. In Chapter 3 we assess and discuss in depth the impact of patient and trial characteristics on earlier termination of antipsychotic trials for schizophrenia.

1.2 Categories of participants who do not complete a study

Participants may not complete the study for reasons related to the design and conduct of the study (e.g. lack of improvement, side effects, unpleasant study procedures, protocol violations) or completely unrelated to the study (e.g. moving away, death unrelated to the study) and they raise many difficulties in the statistical analysis and the decision making process (27;28). These participants are usually referred in the literature as dropouts, attrition, withdrawals, losses to follow-up, missing participants or missing outcome data, while participants who complete a study are referred as completers (9). Furthermore, missing participants can be distinguished into three mutually exclusive classes (29;30): (i) participants who were excluded due to protocol violations after randomization, (ii) participants who left the trial early but provided at least one response (they were present in at least one assessment) and (iii) participants who left the trial early but did not provide any response (apart from their baseline measurements). The first class is known as exclusions or non-adherent participants and several analytic methods have been proposed to address the exclusions (1;21). The second class is observed in trials with continuous outcome(s) measured in multiple follow-up times and it is often included in the analysis of the original trials by manipulating the observed outcome(s) under a specific statistical methodology (30). The baseline observation carried forward (BOCF), the last observation carried forward (LOCF) and the worst observation carried forward (WOCF) methods belong to this class. Under LOCF, all subsequent missing responses of the participants are set equal to their last observed responses regardless the time or the reason of dropout (9;31). BOCF uses baseline measurements to replace missing responses, while WOCF replaces the missing responses with the worst observed responses assuming that the reason of dropout is symptom worsening and those symptoms will not worsen further (4;27). Finally, the

third class comprises the completely missing outcome data which can be manipulated at the meta-analysis level using a specific imputation assumption about the outcome that the missing participants could have provided had they stayed in the trial. If the assumptions are reasonable, the results will be unbiased but overprecise (5;32;33). Note that the definitions of missing outcome data presented here may not be universally accepted, but they are necessary in order to add clarity.

Trial reports provide completely missing outcome data very often. On the contrary, trial reports rarely distinguish between the observed and imputed outcomes (e.g., LOCF), and as a result the meta-analysts treat the trial outcomes as if estimated from observed data. Note that the analytic strategies of these three classes vary; the first class is analyzed either in trial or meta-analysis level, the second class is analyzed in trial level, whereas the third class is analyzed in meta-analysis level (1;5;9;32;34;35). In the following paragraphs, the first two classes are described in depth. Completely missing outcome data are described in detail in paragraph 1.4.

1.2.1 Exclusions or non-adherent participants

Exclusions, also known as *non-adherent participants*, occur very frequently in clinical trials (35). Trial characteristics, such as longer duration, placebo as control and strict inclusion criteria are often associated with higher likelihood of exclusions (22;35). Exclusions reduce the sample size, may break the balance of the prognostic factors and lead to biased results (1). Participants are usually excluded due to protocol violations after randomization, resulting often in missing outcome data (1;35). Note that we should not confuse participants who were excluded due to protocol violations with those excluded because of mistaken randomization. The latter group refers to the *mistakenly randomized* participants, that is, participants who after randomization were found not to have the disease of interest or did not undergo a procedure that is necessary in order to receive the intervention (such as, a surgery or routine medical tests)(1).

Protocol violations are a broad category that reflects any kind of deviations from the assigned treatment, such as switch or interruption of the assigned treatment, dose modification and reception of a proscribed treatment or deviations from the eligibility criteria, such as wrong diagnosis (35;36). It has been found that in depression trials the number of participants who do not adhere to their assigned antidepressant is quite high; Mulsant et al compared nortriptyline with paroxetine in the treatment of depression in elderly patients and found that the percentage of treatment non-compliance was 56% and 45%, respectively (35;37).

Data from excluded participants are usually either not published in the trial reports or are inadequately reported due to logistical difficulties, lack of follow-up the excluded patients or because intervention data after exclusion are not considered to be useful (1;35). However, for some excluded participants complete data may be available. In this case, it is best advised to include these participants in the analysis in order to reduce as much as possible the impact of exclusions on the reliability of the results (1;35;38). The literature for modeling longitudinal outcomes of non-adherent

participants jointly with longitudinal outcomes of adherent participants is very limited. Chen et al developed a Bayesian method to jointly model longitudinal treatment measurements from non-adherent participants with those from adherent participants and they investigated the various properties of the proposed model. They also implemented their method using real data from a randomized placebo-controlled trial that examined the effect of Exelon on subjects with mild cognitive impairment (38). Leuchs et al reviewed and compared different strategies that have been developed to analyse longitudinal data from non-adherent participants to the assigned protocol treatments (35). A common characteristic of these strategies is that they divide the data into two parts: (i) data collected before non-compliance (known as *on-treatment* data) and (ii) data collected after non-compliance (known as *off-treatment* data).

Depending on the aims of the study, the availability of data after non-compliance has different matter of importance in the analysis; off-treatment data are very useful to assess the effectiveness of the treatments, that is, the treatment effect irrespective of the compliance status of the participants, whereas on-treatment data are sufficient to assess the efficacy of the treatments, the effect of the treatments if all participants fully adhered to the treatment assigned (35). However, focusing only on the on-treatment data may increase the risk of selection bias, especially when the exclusion rates differ between the compared treatment arms (35). Therefore, it is of great importance for the adequate estimation of effectiveness and efficacy of the study treatments to put effort into following-up all randomized participants irrespective of their compliance to the protocol, to include all available data into an appropriate analytic strategy and to conduct sensitivity analysis in order to assess the robustness of the results (1;35).

1.2.2 Last observation carried forward (LOCF)

Last observation carried forward (LOCF) is a commonly used approach in the psychiatric trials to minimize the impact of attrition in the analysis, because it preserves most of the randomized sample and is also computationally easy (7;20;36;39;40). LOCF is implemented when longitudinal continuous observations are available for each patient (31;41). Several published psychiatric trial reports have investigated the performance of the LOCF strategy. For example, Rosenheck argued that the efficacy superiority of the new generation antipsychotics (also known as atypical) against the conventional antipsychotics (also known as typical) may have been an artifact of the use of LOCF, because conventional antipsychotics induce more extrapyramidal symptoms than new generation antipsychotics and hence, more patients may have terminated the trial early in the first group than in the second (42). Leucht et al collected data from pivotal trials that compared amisulpride and olanzapine with conventional antipsychotics in order to assess whether the results based on LOCF were different from those based on other analytic strategies of the dropouts. They found that the results were consistent irrespective of the strategy. They concluded that the impact of dropouts due to adverse events on the results of the

LOCF method depends both on the dropout rates due to adverse events and on whether the dropout rates differed between the compared groups (43).

The combination of LOCF with ANOVA analysis is another common strategy applied in psychiatric trials, and especially in depression trials (23;35). Under this strategy, an analysis of variance is implemented for the difference between baseline and endpoint measurements, while missing outcome data have been imputed using LOCF. However, it has been shown that this combination may increase treatment effect and, hence, it may inflate type I error, partly because the imputed values are treated as observed (35). Leberman et al assessed the efficacy and safety of olanzapine versus haloperidol in a randomized double-blind trial of patients with first-episode psychosis using both an LOCF of ANOVA model and a mixed random coefficient model. These two models gave opposite results; olanzapine was not found to be more efficacious than haloperidol when the LOCF model was performed, whereas the contrary was found under the mixed-effects model (44). Prakash et al came also to the same conclusion; the results using LOCF were in the opposite direction compared to the results using mixed model for repeated measures (23). In addition, Lane P found that the mixed model tends to minimize the bias in the results and it rarely inflates type I error rates compared to LOCF (6). Likelihood-based, mixed-effects models for repeated measurements have received considerable attention in the analysis of longitudinal trials, because they are considered to be more reliable and better grounded statistically because they are robust to biases from early dropouts and they provide better control of type I and type II errors than LOCF (6;45-48).

The use of LOCF has been heavily criticized for yielding biased treatment estimates in terms of safety and efficacy either in favor of or against the study treatment, especially when early dropouts occur and when the response variable is expected to change over time (4;9;23;35;36;39;40;49-53). For instance, if an intervention improves a health condition over time, then LOCF will underestimate the efficacy of this intervention, while the opposite will occur if an intervention deteriorates a health condition over time. The weaknesses of this approach lie on its unrealistic assumptions. First, LOCF assumes that the responses remain constant after dropping out early of the trial and hence, it ignores the natural history of the disease (4;9;20;23;39-41;43;53;54). In many disorders, such as depression, it has been observed improvement in the symptoms over time, even in the absence of medication (53). Second, it does not make optimal use of all the data before discontinuation, and hence, it overlooks the history of responses of the participants (40;41;45;53). Third, LOCF tends to provide spuriously increased precision, since it treats the imputed values as observed and hence, it ignores uncertainty about these imputed values (9;23;36;39;41;54). Fourth, LOCF tends to provide biased results downwards when the outcome of interest is negative (e.g., side-effects), since it assumes no worsening over time (45;53). Last, in case of equal dropout rate in both treatment arms, LOCF may underestimate the treatment effects, whereas in case of unequal dropout rates, the bias may be larger in either direction (6). Despite these shortcomings, LOCF is frequently chosen as the primary analysis in the clinical trials (45).

1.2.3 Baseline observation carried forward (BOCF)

Another method to handle missing longitudinal continuous outcome data from early discontinuation is the baseline observation carried forward (BOCF). Compared to LOCF, there are only a few published articles on the description and evaluation of BOCF. Despite the criticisms, some regulatory agencies often suggest BOCF as a primary method of handling missing outcome data in confirmatory phase III clinical trials for some treatments, because they relate early withdrawal mainly with treatment-related reasons, such as inefficacy and adverse-events (27;38;55). Since these patients do not benefit from the treatment due to intolerable side-effects, their measurement should be the same as that in baseline. In practice, patients tend to dropout early from a trial for various reasons and hence, it is best advised to take into account both treatment-related and non-treatment-related reasons (e.g., protocol violations, lost to follow-up, eligibility criteria not met) (27). Another reasonable approach is to impute the missing outcome data based on the reasons of dropouts (55).

There have been proposed two modifications of the BOCF approach that account for the reasons of discontinuation; the modified BOCF (mBOCF) and the adverse-events BOCF (aeBOCF) (27). In the mBOCF approach the observation at the time of withdrawal is used for patients who dropped out due to non-treatment-related reasons, while the baseline observation is used for those who dropped out due to treatment-related reasons. Under the aeBOCF approach the observation at the time of withdrawal is used for patients who dropped out due to treatment-related reasons, while the baseline observation is used for those who dropped out due to any other reason (27;56). The concept of using mBOCF is that these patients provide actual treatment outcomes and including their outcomes in the analysis, any potential bias in the results due to dropout for non-treatment-related reasons is prevented. The rationale of employing aeBOCF is that if a treatment is ineffective or intolerable and it leads to early dropouts, then it should be considered as not effective (27).

BOCF ignores the exact time the dropout occurs during a trial (early in the trial, in the middle or before the end of the trial) and hence, it may mislead the inferences on the treatment effect by overlooking useful information. It is reasonable to believe that a patient who stopped the treatment after the first week of an 11-week trial may not benefit from the treatment, but it is intuitively absurd to ignore the treatment benefit received by a patient who dropped out near the end of the trial. In order to preserve most of the measurements of the patients, Shao et al suggested the combination of LOCF and BOCF; BOCF is applied for dropout due to treatment-related reasons, while LOCF is applied for non-treatment-related reasons, assuming that there is available information on the dropout reasons (55). They also developed statistical tests for asymptotical inferences after applying the mixed LOCF and BOCF method. Both LOCF and BOCF cause standards statistical tests, such as t-test, ANOVA or ANCOVA, to be invalid and/ or underpowered, because imputed data under LOCF or BOCF are treated as observed. It has been demonstrated by simulations that when the dropout rate is similar in both treatment arms, the power of the test decreases as the

dropout rate increases, whereas when then dropout rate in unequal, the power increases as the dropout rate increases for the control arm (55).

1.2.4 Worst observation carried forward (WOCF)

WOCF is the most conservative approach compared to LOCF and BOCF (57). The literature on the description and evaluation of WOCF is very scarce. It is an 'extreme' type of imputation which is usually performed as a sensitivity analysis. Following the idea discussed in BOCF, a combination of LOCF, BOCF and WOCF, when deemed appropriate, may be employed to handle dropouts (55).

In a quite similar concept stands the *worst case analysis*. This analysis uses the worst observed value to impute the missing outcome data in the experimental arm, while it uses the best observed value to impute the missing outcome data in the control arm (31;58). Worst case analysis can increase the overall variability, bias the experimental treatment effect downward and bias the placebo effect upward (31).

1.3 Missing outcome data mechanisms

Analysis of data from patients who terminated early a clinical trial is a real challenge since it involves untestable assumptions about the reasons these participants dropped out early from the study (27;28). The reasons leading to missing outcome data are called *missing outcome data mechanisms* or simply *missingness mechanisms* (29). The missingness mechanism actually reflects how the probability of an outcome being missing is associated with the baseline covariates and the unobserved outcomes (12). Therefore, the missingness mechanism affects the distribution of the missing outcome data and hence, the estimated treatment effect (9). To investigate the effect of dropout in the effectiveness of a study treatment, it is important to understand the missingness mechanisms (30). Little and Rubin classified the missing outcome data mechanisms into three categories: (i) missing completely at random (MCAR), (ii) missing at random (MAR) and (iii) missing not at random (MNAR) or informatively missing (IM) (31). The distinction among these categories is important, because the different techniques for dealing with missing data depend on specific mechanisms. As a result, bias in the estimate of effect sizes depends strongly both on the assumptions about the missing outcome data mechanism and the technique for addressing missingness (30).

1.3.1 Missing completely at random (MCAR)

The process in which the probability of dropout (also known as *missingness*) is independent of both the observed measurements (e.g. baseline covariates, observed responses) and unobserved measurements (those that would have been observed if the patient had stayed in the study) is defined as *missing completely at random (MCAR)* (5;32). This means that the reasons for dropout have nothing to do with the study in general. Thus, the effect of an intervention will be the same on average among those who remained and those who dropped out (9). For instance, if the outcome of a

participant is missing due to a car accident or relocation without informing his doctor, then the mechanism is MCAR. Under MCAR, every patient has the same probability to dropout. It is as if, after randomizing the patients to the interventions, we further randomly decide who to observe (9).

MCAR data are a special case of MAR (5). Although MCAR assumption is inherently untestable, it is possible to explore it by simply comparing the distribution of the variables (e.g. baseline, outcome) between dropouts and non-dropouts (5;32). If MCAR seems to hold, we can safely analyze the data *ignoring* missing participants without the risk of bias, but with the risk of statistical power loss, though, due to diminished sample size (20;32;33). This approach is known as *available case analysis (ACA)* and it is implemented when missing data are considered *ignorable* (34). However, this assumption is rarely plausible in clinical trials and it is generally recommended to be avoided (9;20;32;35).

1.3.2 Missing at random (MAR)

Under MAR assumption the probability of dropout is dependent of the fully observed covariates (e.g. intervention, baseline), but not of the unobserved ones. This category is more restrictive than MCAR (32). Due to the dependence of the missingness mechanism on the observed covariates, it is also called as *covariate-depended dropout* (36). If we identify fully observed covariates that are associated with high risk of dropout (as well with the outcome), then conditional on these covariates, we assume that the distribution of the response is the same between those who dropped out and those who remained in the study (9). For instance, elder patients seem to have lower response to a particular intervention than younger patients and hence, they are in higher risk of leaving the study early.

MAR data are considered *ignorable*, too (5;9;20;33). The starting point of many statistical techniques for missing data is the assumption that the data are MAR (35). Similar to MCAR, MAR can be assessed by exploring possible relationships between missingness and observed covariates (5;9).

1.3.3 Missing not at random (MNAR)

If the missingness mechanism is neither MCAR nor MAR, then it is MNAR (9). Under this assumption the probability of dropout depends on some unobserved covariates or on the outcome, conditional on the observed variables (5;9). For instance, patients without health improvement are in higher risk of dropout early from the study (here, missingness depends on the outcome). Similarly, patients with side-effects are in higher risk of leaving the study early (here, missingness depends on the side-effects which usually are not measured is some studies).

MNAR data are considered non-ignorable (9;20;32). In this case, the missingness mechanism must be modeled in order to yield unbiased estimates (32). The most prominent models are the *selection model* and the *pattern-mixture model* (9;27;33). A selection model examines the distribution of dropouts and completers in the outcome

responses. If among the participants with health improvement the majority has remained in the study, while among the participants with health stability or deterioration the majority has left the study early, then missing data may be MNAR. A pattern-mixture model examines the distribution of the outcome between dropouts and completers. If among the participants who left the study early the majority had health stability or deterioration, while among the participants who completed the study the majority had health improvement, then missing data may be MNAR.

MNAR data are often referred to as *informative* because the probability of dropout is related to unobserved measurements (5;9). However, distinguishing among ignorable and non-ignorable missingness is rarely possible (20;32;35). In this case, a versatile strategy could be to start the analysis by assuming MAR and then evaluate the robustness on these results under various scenarios about the outcome of the missing participants using a sensitivity analysis. Any inconsistency in the conclusions is a strong indication that missing data may be MNAR (9;28).

1.4 Handling missing outcome data in a meta-analysis

Missing outcome data reduce the power to detect any significant differences in the compared treatments due to diminished sample size and cancel out the benefits of randomization by creating imbalances in the distribution of the prognostic factors between the trial arms and hence, increasing the risk of bias in the results (1;4;12). Missing outcome data may also reflect lack of treatment tolerability, side-effects or lack of compliance (20). Depending on their extent and handling, missing outcome data may affect analysis and inferences (1;20). Numerous methods have been proposed and examined for handling missing outcome data. There is no universally best method to handle missingness and hence, the choice of a suitable method can be very difficult (4;45). The analysis must be tailored both to the amount of missing outcome data and the mechanisms leading to missing outcome data (45).

In an clinical trial where individual patient data are available for each time point, the missing outcome data mechanism can be explored using auxiliary data and plausible assumptions (e.g., baseline severity of disease may be related with early withdrawal) in order to apply an appropriate analytic strategy (5;9;32;34). The armamentarium of statistical techniques to handle missing outcome data within a clinical trial is large and includes simple approaches, such as, LOCF and complete case analysis, as well as more sophisticated approaches, such as, multiple imputations and maximum likelihood methods (4;5). It is important that the researcher is familiar with the advantages, the disadvantages, the assumptions and the complexity of these techniques in order to select the most appropriate. Analytic strategies to handle missing outcome data in trial-level are out of the scope of the thesis. The focus of the thesis is on the investigation and address of attrition within a meta-analysis framework.

Trials with missing outcome data may be a great threat to the generalizability of the inferences when they are combined in a meta-analysis (3;32). In theory, discontinuation may introduce selection bias in the randomized trials as a result of

distorting the balance in the baseline characteristics between those remaining in the trial and those dropping out early (3). If discontinuation is systematic and related to the outcome, the meta-analysis is likely to yield biased results (3). However, in a meta-analysis of trials the data are available in a summary form for each trial (i.e. mean responses with their standard deviation or number of successes and failures for each intervention) and hence, it is difficult to explore the missing outcome data mechanisms (34). The patient-level data are not always available to the meta-analysts for various reasons (e.g., unreported data from original trialists, limited information in published reports) and as a result, the meta-analysts analyze the data based on untestable assumptions about the missing outcome data mechanism. Therefore, handling missing outcome data in a meta-analysis level is a real challenge (5;34).

There is a misconception that the bias in the results due to premature discontinuation is associated mainly with the amount and the distribution of the dropout rate between the trial arms. Indeed, large dropout rates reduce the sample size and hence, the power of the results, but they do not necessarily lead to biased effect sizes. Similarly, whether dropout rates are similar or not between the trial arms cannot assure that the results will be unbiased or biased, respectively. The missing outcome data mechanism and the statistical analysis are the two key factors that affect the integrity of the inferences (59). Higgins et al suggest exploring the missingness mechanism by performing meta-regression of the effect estimates against the dropout rates (5). Any statistically significant relationship may be an indication of informative missingness, that is, MNAR data. However, the results should be interpreted with great caution, since meta-regression suffers the drawbacks of simple regression; confounding, low power to detect a statistically significant relationship and high false-positive error (5). Therefore, a better and more realistic approach to explore the missingness mechanism is to make sensible assumptions about whether data are informatively missing, and if so, what would be the outcomes if the participants had never dropped out, and then, to examine the sensitivity of the results to these assumptions (5). Usually, a starting point of the analysis is to assume MAR data and explore any deviations from this assumption by performing a series of sensitivity analyses. However, MAR is very strong assumption that unobserved outcomes do not differ systematically from the observed outcomes in the same trial arm. This assumption is implausible when we deal with a set of trials of various sizes and degree of missingness (5;34).

Complete missing outcome data constitute trial-level data that are analyzed within a meta-analysis framework. The available techniques to address completely missing outcome data in meta-analysis are very limited and the majority of these techniques have been developed for binary outcome (1;5;32;34). This is an indication that complete missing outcome data have received relatively little attention in the meta-analysis context (5). These methods may be applied either as a primary or a sensitivity analysis (1;5). Below, these methods are presented briefly along with their advantages and disadvantages.

1.4.1 Available case analysis (ACA, or complete case analysis, CCA)

Available case analysis limits the analysis only to completers, that is, participants who have observed responses at each scheduled time point and ignores (1;31). This approach is probably the most common in the meta-analysis particularly because of its simplicity and it is usually applied as a starting point (5;31). However, restricting the analysis only to individuals with complete observations may lead to imprecise estimates due to reduced sample size and it may yield biased results unless the data are MCAR (49).

1.4.2 Imputed case analysis (ICA)

Under this analysis, the missing responses in each intervention are filled in using specific assumptions about the outcomes that the missing participants could have provided if they had never left the study (1;5). If the assumptions are reasonable, the ICA tends to yield unbiased estimates (5). There are two conceptual approaches to implement ICA; either impute (fill in) an outcome (success or failure) for each missing participant or impute risks of events for the missing participants based on the observed responses (5). In case of low dropout rate, the former approach may suffer from high rounding error leading to unnecessary error in the effect estimates. However, rounding error is not a concern when treatment effects are estimated using standard methods, such as odds ratio and risk ratio, because these methods allow for numbers of participants in decimals (5). Note that both approaches yield identical estimates of treatment effect and provide a complete dataset in order to estimate the meta-analysis summary effect size with great precision (5).

When applying ICA, a great variety of different assumptions about the outcome of the missing participants can be undertaken. Then, a meta-analysis is implemented for each assumption. The range of the meta-analysis effect sizes indicates the robustness of the conclusions, while the overlap of their confidence intervals implies the relevance of the MAR assumption. Inconsistent results are a strong indication that data may be MNAR. The following assumptions may be implemented under ICA (5):

1. All missing outcome data are non-events (ICA-0) assumes that all missing participants in both interventions have not experienced the event;

2. All missing outcome data are events (ICA-1) assumes that all missing participants in both interventions have experienced the event;

3. Best case scenario for the experimental intervention (ICA-b) assumes that all missing participants in the experimental intervention have experienced the event, whereas all missing participants in the control intervention have not experienced the event;

4. Worst case scenario for the experimental intervention (ICA-w) assumes that all missing participants in the experimental intervention have not

experienced the event, whereas all missing participants in the control intervention have experienced the event;

5. Same risk as in the experimental intervention (ICA-pE) assumes that both interventions have the same risk of event as calculated in the experimental intervention;

6. Same risk as in the control intervention (ICA-pC) assumes that both interventions have the same risk of event as calculated in the control intervention;

7. Intervention-specific risk (ICA-p) where the missing outcome in the experimental intervention is imputed by using the estimated risk of event in the experimental intervention, whereas the missing outcome in the control intervention is imputed by using the estimated risk of events in the control intervention. This approach corresponds to the MAR assumption.

ICA-0 and ICA-1 are extreme assumptions that are appropriate when there is a rational relationship between these missingness mechanisms and the outcome (5). For example, it is argued that conventional neuroleptic and old atypical antipsychotic drugs are linked with extrapyramidal side-effects, and hence, it is rational to assume that patients who dropout may have experienced intolerable side-effects (ICA-0) (21). Or, new atypical antipsychotic drugs are linked with less extrapyramidal side-effects, and hence, it is rational to assume that patients who dropout may have experienced symptom improvements (ICA-1) (21;22). However, it is false to believe that all these patients in the conventional neuroleptic drugs, who did not experience side-effects, or all these patients in the new atypical antipsychotics, who experienced symptom deterioration or stability, actually completed the study.

ICA-b and ICA-w are the most extreme imputations; they provide the most extreme effect sizes in either direction (favoring the experimental or the control intervention), often yielding conflicting inferences due to these extremes (5;63). These extremes reflect the total potential uncertainty due to missingness (63). Gamble and Hollis (2005) proposed an approach that incorporates the results both under ACA and these two most extreme imputations. First, the ACA is performed to estimate the treatment effects in each trial. Then, ICA-b and ICA-w are implemented in each trial separately and the most extreme lower and upper confidence interval limits from these imputations are used to form a so-called *uncertainty interval* for each trial (63). As a result the standard errors that are extracted from the uncertainty intervals are inflated leading to reduced weights. These weights reflect the added uncertainty we might expect due to missing outcome data; the higher the missing rate the smaller the weights (5;63).

ICA-pE and ICA-pC are assumptions based on event risks (1). For instance, if the missing outcomes correspond to patients excluded from experimental (control) treatment and it is assumed that none of them received the experimental (control) treatment but they switched to the control (experimental) treatment, then ICA-pC

(ICA-pE) may be applicable (5). However, these strategies have two caveats. First, they imply that missing and observed participants have no systematic differences in the characteristics that may be related to the unobserved data. This is equivalent to assume that data are MAR which is rarely possible in psychiatric trials. Second, by imputing same risks of events in the two arms, the estimates of the treatment effects are pulled toward the null hypothesis (5). MAR is also implied by employing ICA-p which yields the same results with ACA. However, these two approaches have different impact on sample size and the uncertainty of the results; ICA-p preserves the randomized sample and wrongly reduces the standard errors, whereas ACA reduces the sample size and increases the uncertainty (5).

A better and more rational imputation scheme is the combination of the above seven schemes. This is plausible if and only if the reasons for dropout are available for all trials included in the meta-analysis (5). For instance, we could assign ICA-0 to those trials that reported withdrawals due to side-effects and ICA-1 to those trials that reported withdrawals due to symptom improvement. Or, we could assign ICA-p in trials that reported withdrawals for reasons unrelated to the treatment (e.g., moving away, death unrelated to the study).

The ICA method treats the imputed values as observed. As a result, it tends to provide spuriously increased precision because it fails to account for uncertainty about these imputed values (5;58). Higgins et al. developed a number of weighting schemes to calculate the standard errors adjusting for the uncertainty about the imputation (5). Another caveat of this method is that only a limited number of assumptions are usually applied as sensitivity analyses (61).

1.4.3 A graphical sensitivity analysis for missing outcome data

Hollis (2002) studied a great range of possible imputations of missing outcome cases in a clinical trial with two interventions and a binary outcome and collectively presented these results in a contour plot (61). This graphical sensitivity analysis can be extended into a set of trials for two or many interventions. The brief interventions data on the reduction of excessive alcohol consumption as presented in the article of White et al. were used to perform the graphical sensitivity analysis in figure 1.4.1 (34). The dataset comprises of seven randomized controlled trials that compared intervention (defined as, a few minutes of feedback, information and advice provided by general practitioners) with control (defined as, no or less intervention). The outcome was defined as a success if the individual was drinking below recommended levels after one year.

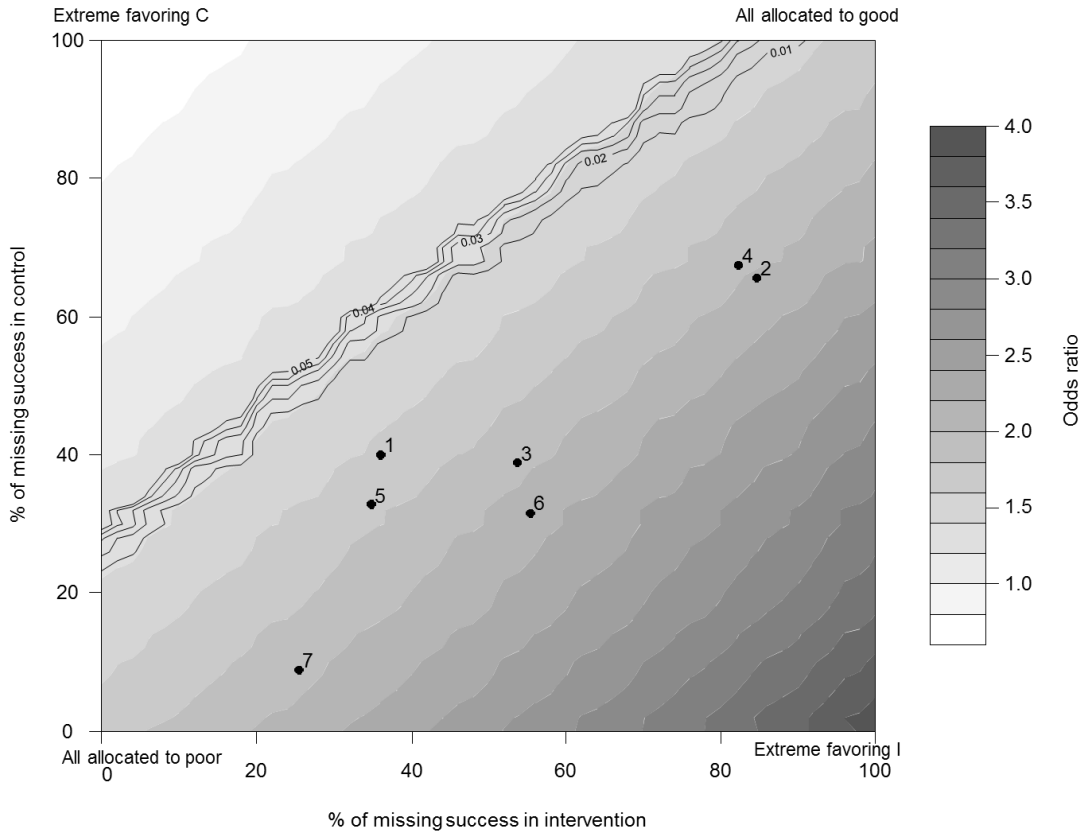

Figure 1.4.1 Graphical display of sensitivity analysis for the reduction of excessive alcohol consumption trial showing odds ratio of intervention (I) against control (C) under all possible allocation of cases with missing outcome. Contour lines show boundaries of statistical significance. The range of the magnitude of the effect is represented by the grey-scale color scheme as it is defined according to the plot-key; the darker the grey color is the higher the superiority of the intervention compared to control. The number dots correspond to trials and reflect the percentage of success in the intervention and the control.

Both axes represent the percentage of missing cases allocated as successes within each group to aid interpretation of particular scenarios about the missing cases; the bottom left (coordinates 0%,0%) and the top right (100%, 100%) correspond to allocation of all missing cases as failures or successes, respectively, whereas the top left (0%, 100%) and the bottom right (100%, 0%) correspond to allocations of all missing cases as successes (extreme favoring) in the control and the intervention, respectively. The black circles indicate the observed percentages of success within the observed cases in the trial denoted by a number on the respective circle, while the diagonal line corresponds to equal percentages of missing cases allocated as successes (this is equivalent to the MAR assumption). The boundaries of the statistical significance of the effect sizes under specific allocations of the missing cases as successes in each group are indicated by contour lines. The value $p = 0.05$ has been considered as the significance threshold. For each coordinate of the plot a meta-

analysis odds ratio has been estimated. The range of the magnitude of the effect is represented by the grey-scale color scheme as defined by the plot-key. Each black circle indicates the success rates within the observed cases in the trial denoted by a number on the circle.

According to this contour plot, substantially higher percentages of missing successes in intervention compared to those in control across a wide range of missing cases correspond to higher intervention effectiveness and hence, there is evidence in favor of intervention for the reduction of alcohol consumption. This contour plot has incorporated information both for the magnitude of the effect sizes and their significance under variant distributions of the percentage of missing successes in each group.

1.4.4 Informative missingness odds ratio (IMOR)

IMOR is a parameter that reflects whether and how much the completely missing outcome data are informative (5;34). It is calculated for each intervention in the form of the odds ratio of the event among missing participants relative to the event among the observed participants; $IMOR_E$ and $IMOR_C$ correspond to the IMOR in the experimental and control intervention, respectively (5;34). By denoting with p_E the risk of event in the experimental intervention and with p_C the risk of event in the control intervention, the IMOR in each intervention is defined as

$$IMOR = \frac{odds\left(\frac{missing}{success}\right)}{odds\left(\frac{missing}{failure}\right)}$$

or, equivalently

$$IMOR = \frac{P(missing|success)/[1 - P(missing|success)]}{P(missing|failure)/[1 - P(missing|failure)]}$$

$\boldsymbol{IMOR_E = 1}$, indicates that the odds of missing an event is equal to the odds of missing a non-event in the experimental intervention and hence, it implies the MAR assumption. IMOR values larger than one suggest that missing participants are more likely to experience the event, whereas the opposite holds for IMOR values lower than 1 (34). The IMOR method is a special case of the imputation methods described above; under the ICA-0 and ICA-1 assumptions it holds that $\boldsymbol{IMOR_E = IMOR_C}$, while under the assumptions ICA-b and ICA-w it holds that $\boldsymbol{IMOR_E = 1/IMOR_C}$ and $\boldsymbol{IMOR_C = 1/IMOR_E}$, respectively (5). The equivalence of these two methods is presented in Table 1.4.1:

Equivalence of informative missingness odds ratio with imputed case analysis

ICA	Description	$IMOR_E$	$IMOR_C$
ICA-0	All missing outcome data are non-events	0	0
ICA-1	All missing outcome data are events	∞	∞
ICA-b	Best case scenario for the experimental intervention	∞	
ICA-w	Worst case scenario for the experimental intervention	$\frac{p_C(1-p_E)}{p_E(1-p_C)}$	∞
ICA-pC	Same risk as in the control intervention		1
ICA-pE	Same risk as in the experimental intervention	1	$\frac{p_E(1-p_C)}{p_C(1-p_E)}$
ICA-p	Intervention-specific risk	1	1

Note: ICA: imputed case analysis; IMOR: informative missingness odds ratio; $IMOR_E$: informative missingness odds ratio in the experimental intervention; $IMOR_C$: informative missingness odds ratio in the control; ICA-0; all missing outcome data are non-events; ICA-1; all missing outcome data are events; ICA-b; best case scenario for the experimental intervention; ICA-w; worst case scenario for the experimental intervention; ICA-pC; same risk as in the control intervention; ICA-pE; same risk as in the experimental intervention; ICA-p; intervention-specific risk.

A sophisticated extension of the meta-analysis model combines the IMOR with the intervention effects derived from the observed individuals to obtain a 'missingness-adjusted' meta-analysis result for the entire randomized population. A prior assumption can be made for the IMOR parameter in each study and each intervention (32;34). For instance, we may assume that IMORs differ between the interventions if there is evidence that participants allocated to a more intensive intervention tend to provide worse outcomes and leave the trial early. Or, we may assume that IMORs differ among the trials if there is evidence that trials with longer follow-up duration tend to have higher dropout rate.

Contrary to ICA methods, the IMOR approach may account for the uncertainty due to missing outcome (5). This approach will be described in detail in chapter 4. Note that when the IMOR approach accounts for the uncertainty due to missing data, the meta-analysis standard error will be larger compared to that under ACA and ICA. The IMOR approach could be extended to assess missingness when the outcome is continuous. By contrast, the extension of the ICA methods to the continuous outcome is not straightforward. The IMOR approach can be implemented either in a frequentist or in a Bayesian framework (5;32;34). The latter requires that the researcher is familiar with the notion of the Bayesian analysis.

1.4.5 Response probability ratio (RPR)

Magder introduced the response probability ratio (RPR) which is a parameter similar to the IMOR, in order to quantify departures from the MAR assumption (64). RPR is defined as the ratio between the response probability (i.e., the probability of observing an outcome) among those who experience the event and the response probability among those who do not experience the event

$$RPR = \frac{P(observed|success)}{P(observed|failure)}$$

or, equivalently

$$PRP = \frac{1 - P(missing|success)}{1 - P(missing|failure)}$$

Note that the response probability may vary depending on the trial arm and the outcome (event or non-event) (64) Similar to IMOR, RPR is also allowed to be different in the two trial arms. When RPRs of both arms are equal to one, the probability of missing an event is the same with the probability of missing a non-event, and hence, the data are MAR whereas RPRs with values larger or lower than 1 indicate departures from the MAR assumption (64).

RPR is constrained by the observed dropout rate; very low dropout rate will pull RRPR towards one. A modification of RPR was suggested by Greenland. Greenland quantified the departure from MAR in observational trials by calculating the ratio of the treatment-specific RPRs

$$\frac{RPR_E}{RPR_C}$$

where RPR_E corresponds to RPR in the experimental arm and RPR_C corresponds to RPR in the control arm. However, it is hard to interpret and understand intuitively the results from this approach (32).

1.5 The importance of sensitivity analysis in addressing attrition within a meta-analysis.

It has been already mentioned that analysis of missing outcome data is based on untestable assumptions (32;34). Therefore, it is necessary to perform sensitivity analysis based on alternative assumptions about the missing outcome data mechanism, unless information on the reasons of dropout is available to safely model the missing outcome data. This strategy provides a range of estimates in order to infer on the robustness of the meta-analysis effects in the primary analysis after handling the missing outcome data (1;61). It is of great importance to clearly state, justify and assess any assumptions about the missing outcome data. Sensitivity analysis is the ideal tool to explore these assumptions (12).

Sensitivity analysis is a useful analytical tool that is necessary in every statistical analysis in order to assess the degree of certainty in the findings. It is actually a repeat of the primary analysis after substituting or removing some specific characteristics of

the dataset (65). For instance, we implement again the meta-analysis after having imputed missing outcome data under a specific assumption (substitution) or having removed studies with percentage of missingness above a specific threshold (removal). The goal of a sensitivity analysis is to understand whether the meta-analysis findings are robust to the assumptions and decisions made during the process of obtaining them (65;66). In the previous example, we would like to know whether the inferences change after imputing the missing data or after removing studies with percentage of missingness above a specific threshold, respectively. If the results overall are not different from those obtained from the primary analysis (e.g., by performing ACA), it is a strong indication that the results can be regarded with high certainty. By contrast, any inconsistency in the meta-analysis results between the primary and sensitivity analysis is a warning that further investigation should be employed (e.g., obtain individual patient data, contacting the author of the studies) before reporting the results (65).

Generally, the sensitivity analysis should be pre-specified in the meta-analysis protocol (65;67). However, many issues suitable for sensitivity analysis are usually identified during the review process of the analysis (65). In terms of missing outcome data, sensitivity analysis should focus on two issues: (i) the robustness of the estimated treatments effects and (ii) the extent of uncertainty reflected by the assumptions, because the results of a meta-analysis are affected both by the magnitude of the treatment effects and their standard errors (5). Proposal for sensitivity analyses that address the above mentioned two issues are presented explicitly in section 1.7.

It is recommended to present the results from a sensitivity analysis in a summary table or in a figure along with the results from the primary analysis. Forest plots and long reporting of the results for each sensitivity analysis should be avoided because they are time-consuming and not informative at all (65).

1.6 Analytic strategies to determine the analyzed sample size of the trials included in a meta-analysis

1.6.1 Intention-to-treat analysis

As mentioned by Alshurafa et al., trial methodology experts, systematic review organizations, and authorities including the Consolidated Standards of Reporting Trials (CONSORT), the Cochrane Collaboration, the US Food and Drug Administration, the Nordic Council on Medicine in Europe, and the American Statistical Associations Group have recommended the intention-to-treat (ITT) analysis as the ideal approach for the analysis of the primary outcome in the randomized controlled trials (RCTs) (68). According to this principle, all the participants in a trial should be analysed according to the intervention to which they were originally randomly allocated, whether they received it or not (33;68;69). ITT is desirable because it increases statistical power and precision, it preserves the aims of randomization and hence, it minimizes selection bias and confounding (33;68;69). Moreover, ITT is favored in the superiority trials for the assessment of the treatment

effectiveness where noncompliance may lower the impact of effective treatments, because it provides a conservative estimate of the treatment effectiveness (33;68). An implication of ITT is that investigators should put an extra effort to collect outcome data on all randomized participants. This can be achieved by careful trial design and conduct, considering appropriate eligibility criteria and remaining in contact with participants who terminated the trial early (69).

By definition, full application of ITT is possible only when complete outcome data are available for all randomized participants. In the presence of missing outcome data, there is no consensus on how missing outcome data should be addressed and hence, it is difficult for the analyst to adequately apply an ITT analysis, because ITT requires all randomized participants to be analyzed irrespective of whether they received the prescribed intervention or completed the study (33). Therefore, the actual method of analysis may actually deviate from the reported ITT method (68). This indicates that there is not a universally standard definition of ITT in relation to missing outcome data that is applied in clinical trials and methodological articles; investigators hold heterogeneous views regarding the relationship between missing outcome data and ITT and how to handle missing outcome data under ITT (68). Alshurafa et al. implemented a systematic review of methodology articles on the proposed definitions of ITT in relation to missing outcome data and in how missing outcome data should be addressed in RCTs under ITT (68). They demonstrated the inconsistency in the definition of ITT in the methodological articles; 36 out of the 66 articles included in their systematic review provided a definition for the ITT and discussed the relationship of ITT with missing outcome data. Particularly, 53% of these articles gave a sole definition of ITT (mostly that ITT requires a specific strategy, such as ACA, LOCF, sensitivity analysis) and the remaining mentioned multiple definitions. Hollis and Campbell surveyed all reports of RCTs published in 1997 in four major general medical journals in order to investigate the methodological quality of ITT as reported (33). They found that almost half the reports mentioned ITT. Of these, 13% applied incorrectly ITT either by excluding participants who did not receive the treatment randomized or by not analyzing all randomized participants as allocated, while 55% seemed to have performed correctly ITT by explicitly stating to have analyzed all randomized participants as allocated. In terms of the strategies employed to handle missing outcome data, the reporting was judged to be poor leading to an inadequate description of ITT and hence, increasing the risk of biased treatment effects.

The unavoidable presence of missing outcome data in the RCTs and the favor of ITT from the analysts have led to the need of solution to the problem of the multiple definitions of ITT and its relationship with missing outcome data (68). CONSORT advocates against the use of the widely misused term ITT and supports the use of a more explicit request for information about keeping participants in their randomized treatments (36). Therefore, no randomized participant should be excluded from the analysis. The Cochrane handbook also provided its own point of view on the definition of the ITT by describing three principles that underline ITT: (i) randomized participants shall retain in their original assigned interventions regardless of deviation

from the protocol or withdrawal, (ii) the outcome shall be measured for all randomized participants and (iii) all randomized participants shall be included in the analysis (70). While the first principle is attainable to be achieved, the second is often impossible and the third is argumentative. In the presence of missing outcome data, the Cochrane handbook suggests the application of complete case analysis or imputation and it recommends a sensitivity analysis in either case (70). However, only imputation assures the use of ITT, because it preserves the randomized sample irrespective the imputation scenario. Alshurafa et al. proposed also their solution to the problem; participant with complete data shall be analyzed to the intervention they were randomized, while participants with missing outcome data shall endure imputation under extreme assumptions (e.g., all experienced the outcome, none experienced the outcome) (68). Hollis and Campbell provided also recommendations in detail for ITT that complement the CONSORT guidelines to address ITT (33). As proposed by White et al., the investigators should focus on the following four points: (i) follow-up all randomized participants, even if they withdrew from the trial, (ii) employ a primary analysis under a plausible assumption about the missing outcome data, (iii) explore any departures from the assumption made in the primary analysis by performing sensitivity analyses and (iv) account for all randomized participants, at least in the sensitivity analyses (69).

There is not an optimal analytic strategy to handle missing outcome data in an ITT (33). Therefore, it is of great importance that both the RCTS and the systematic reviews of the RCTs report explicitly how missing outcome data were addressed in the analysis. Imputation methods have been recommended as they yield conservative results, whereas ACA should be avoided since it violates the principle of ITT and it may yield biased results unless data are MAR – which is rarely the case in psychiatric trial. Furthermore, a set of sensitivity analyses that investigate the impact of various strategies to address missing outcome data on the robustness of the primary analysis shall be always implemented (68).

1.6.2 Modified Intention-to-treat analysis

A variant of ITT, known as modified ITT (mITT), has been increasingly used by the RCTs the last 16 years (71;72). Deviations from the definition of the ITT analysis result in the mITT analysis. According to Abraha and Montedori, the four most common types of deviation from ITT are (i) treatment related (e.g., mITT includes patients who received at least X doses of the assigned treatment), (ii) baseline assessment (e.g., mITT consists of participants with baseline assessment), (iii) post-baseline assessment (e.g., mITT comprises of participants with at least one post-baseline measurement) and (iv) target related (e.g., mITT includes all randomized participants minus those without the diagnosis of interest) (71). Each type of deviation may overlap with at least one type.

In the medical literature there is not a clear explanation for the use of the mITT analysis. A plausible justification may be the difficulty of the investigators to manage in their analysis both the exclusions due to protocol deviations and the missing

outcome data that make impossible the application of ITT. Montedori et al. found that the use of mITT was associated both with sponsorship from a for-profit agency and with conflicts of interest from at least one author who participated on behalf of a for-profit agency, but it was not associated with favorable results (72). Additionally, mITT trials appeared to be methodologically of similar quality with trials where ITT was properly implemented. However, mITT trials were more likely to report exclusions due to protocol violations (72).

Since any deviation from the ITT approach may introduce bias in the results, it is best advised that, when mITT is performed, it should only supplement the primary analysis (71). Due to low proportion of mITT trials being conducted, further research is needed to investigate whether ITT is a potential source of bias (72).

1.6.3 Per protocol analysis

Similar to the missing outcome data, post-randomization exclusions are also a common problem in the studies. Patients who did not meet the inclusion/exclusion criteria or deviated from the protocol (e.g., switch or interruption of the assigned treatment, dose modification) are often excluded from the study. The analysis that is restricted only to those patients who completed the study without deviating from the assigned intervention protocol is known as *per protocol analysis* (PP) or *on-treatment* analysis (36). ACA is often mistaken with per protocol analysis (63). A true PP analysis restricts the analysis to those participants who strictly complied with the protocol, whereas ACA restricts the analysis to those participants who completed the trial irrespective of their adherence to the protocol. Note that patients who were excluded due to protocol violations should not be confused with those who were excluded because of mistaken randomization. The latter group is known as *mistakenly randomized* and it includes patients who after randomization were found not to have the disease of interest or did not undergo a procedure that is necessary in order to receive the intervention (such as, a surgery or routine medical tests) (1).

PP has the advantage that it gives the opportunity to a new treatment to show additional efficacy in the analysis it reflects more closely the scientific model as defined in the protocol (51). However, exclusions may affect the balance in the characteristics of the treatment groups, and hence, the PP may lead to biased estimates, especially, if rates and reasons for exclusions differ between the groups (14;36;71;73;74). CONSORT labeled PP analysis as a non-randomized, observational comparison, because any exclusion of the patients compromises the randomization and may introduce bias (36). The thread of bias due to exclusions carries over to meta-analysis of trials even if the included trials have a few exclusions; Tierney and Stewart survey used 14 meta-analyses of individual data on therapeutic questions in cancer and found an overestimation of the treatment effect when exclusions occurred, whereas a meta-epidemiological study that used data from 14 meta-analyses on osteoarthritis reported that the extent and direction of bias due to post-randomisation exclusions can be unpredictable (28;75). Therefore, it is best advised to perform PP as a secondary preplanned analysis (14).

Several studies in the medical literature have reported that post-randomization exclusions and missing outcome data may introduce bias in the estimates of the treatment effects. Therefore, the wide favor of the RCTs towards the ITT is warranted. PP is often contrasted with the ITT, because these two analyses represent different extremes so that, if they provide the same conclusion, then an overall conclusion can be safely drawn (76). In both analyses, if exclusions and losses to follow-up are associated with observed patient characteristics, then such data are considered MAR can be handled using the methods described in (77) for patient level data or in (1) for binary trial-level data.

As mentioned by the ICH, it is usually appropriate in confirmatory trials (i.e., adequately controlled trials in which the hypotheses are stated in advance and evaluated for at least one pre-specified primary outcome, such as, efficacy, safety, dose-response (78)) to plan the conduct of both ITT and PP analyses, so that any differences between them can be explicitly discussed and interpreted (51). However, the recommendations for the ITT and the PP analysis are different in superiority trials and in equivalence (or non-inferiority) trials; ITT is recommended in superiority trials because it does not yield over-optimistic results compared to PP, the non-adherent participants included in the ITT analysis decrease the estimated treatment effect (51). However, in equivalence trials ITT is cautiously recommended along with PP, because ITT is 'biased towards the null'-which is undesirable in equivalence trials-and the PP is less so (51;74).

Stewart et al studied the dropout rates for ITT and PP analysis in prospective, randomized, parallel trials on glaucoma or ocular hypertension (73). They found that dropout rates were decreased with the length of study for the ITT analysis. A possible explanation for this evidence may be the fact that more patients were randomized in trials with longer duration due to a greater anticipated dropout rate (73). In contrast, the dropout rates were increased with the length of study for the PP analysis which is plausible, because the longer the trial duration, the higher the risk of medical or life complication and hence, the higher the likelihood of exclusion. Understanding discontinuation rates for PP and ITT analyses may help in planning sample sizes for future clinical trials (73).

1.6.4 As-treated analysis

When participants fail to adhere to their randomized treatment (i.e., they stop taking the treatment or they switch to the compared treatment or to a treatment which is not part of the trial protocol) then it may seem irrational, at first look, to analyse the responses of those patients as if they were always compliant to their assigned treatment (79). In this case it may appear appealing to analyse the patients for the treatment they actually received. This approach is known as *as-treated analysis*. This approach has been mainly advocated for the analysis of treatment toxicity (1).

Despite being intuitive to analyse only those participants with treatment as actually received rather than assigned, as-treated analysis has been criticized of disturbing the balance of the prognostic factors, reducing the sample size as well as the power to

detect any difference between the compared treatments. Additionally, under the as-treated analysis, it is difficult to provide a reasonable definition of compliance in the assigned treatments; Lee et al. found that different definitions of compliance yielded different results (1). They further advised against the application of the as-treated analysis, even in the presence of balance in prognostic factors between the compliers and non-compliers, as there is an increased great risk of biased results (1).

1.7 Recommendations when dealing with missing outcome data

1.7.1 Recommendations to minimize missing outcome data in trial level

Premature discontinuation cannot be eliminated and there is not a maximum number of missing outcome data that could be acceptable. To avoid missing outcome data as much as possible and to prevent the repercussions that arise from their presence, it is extremely important to develop strategies that minimize this problem. There have been recommended such strategies that can be applied already at the study design stage (31;67). One strategy could be to estimate the predicted (and unavoidable) proportion of missing outcome data that is likely to be observed in the trial (67). This is highly recommended mainly for three reasons. First, missing outcome data may reduce considerable the sample size and hence, the power of the results. Second, proper planning might minimize the bias that is likely to be introduced by a strategy for handling missing outcome data. Last, missing outcome data tend to increase uncertainty in interpreting the results and hence, the performance of multiple sensitivity analyses is getting imperative. Another possible strategy is to carefully consider the eligibility criteria in order to exclude patients with characteristics that may prevent them from completing the study (31). A non-randomized study period could be also considered where all patients take the experimental intervention. The participants who remained free of side effects or are compliers will be randomly assigned to the experimental or the control intervention. The purpose of this design is to take into account the toxicity of the intervention which is a common reason that triggers premature discontinuation. Data collection on baseline covariates that are believed to affect the likelihood of a patient to dropout should be also implemented (31). These baseline covariates will help to identify the type of patients who dropout and assess the potential missing outcome data mechanism. Akl et al. presented a number steps that systematic reviewers must define in the protocol and undertake to minimize and assess the impact of both missing and excluded participants from the trials: (i) contact the trialists for the availability of unreported data, (ii) trials with both missing participants and excluded participants should handle these situations simultaneously by implementing appropriate strategies, (iii) conduct sensitivity analyses to assess the robustness of the results and (iv) discuss the implications of the missing outcome data on the primary analysis results by taking into account the amount of missing outcome data and the results of the sensitivity analysis (1).

1.7.2 The CONSORT guidelines

Consolidated Standards of Reporting Trials (CONSORT) and European Medicines Agency (EMA) encourage the inclusion of a table with the baseline characteristics of the randomized participants in order to check for possible imbalances across the treatment arms that may occur due to missing outcome data (3;12;36;67). It has been found that even small baseline imbalances on important prognostic factors can introduce bias in the meta-analysis results by pooling these trials (80). Hewitt et al. investigated the impact of attrition on baseline imbalance within individual trial and across multiple trials and they did not find any significant association mainly due to the small number of trials and the low level of attrition rate (3). However, they strongly support and recommend the CONSORT statement on the reporting of missing outcome data and their impact on the baseline characteristics across treatment arms in order to identify possible imbalances that could introduce bias in the meta-analysis results. Additionally, CONSORT guidelines require a clearly documented flow diagram of the randomized participants that includes the number of participants with missing outcome data for each treatment arm and the reasons these participants dropped out early in order to make judgments about the loss to follow-up (4;9;12;36;49;81;82). 'Carpenter and Kenward' and 'Wood et al.' underline the importance of reporting the number of patients with missing outcome data by treatment arm, because an imbalanced proportion of missingness between the treatment arms is likely to cause bias when the outcome of interest is related to early withdrawals (9;12). Effort should be put on the examination and the documentation of as much information as possible regarding the missing outcome data mechanisms in order to justify the analysis assumptions, since different reasons of missingness need to be handled differently in the meta-analysis (for instance, see, imputation case analysis in section 1.4) (9;36;67). To preserve ITT, CONSORT suggests the exclusion of any randomized participant from the analysis and encourages that researchers should continue to collect data after discontinuation (36;68). Kane et al. examined whether the CONSORT reporting guidelines improved clinical trial reporting and reported attrition in order to assure the credibility of the published results (83). They found that the CONSORT guidelines improved the reporting of clinical trials when they were implemented but did not substantially improve reported attrition rates.

1.7.3 Principles to select an appropriate primary analysis for a meta-analysis

There is no 'gold standard` applicable strategy to address missing outcome data and different strategies may yield different results (4;45;67). To avoid the temptation of data-driven selection of strategies, it is necessary to pre-specify in the study protocol the strategy to handle missing outcome data in the primary analysis as well as the set of sensitivity analyses. The selection of the primary analysis and the sensitivities analyses should be justified in detail (67). There are numerous articles in the literature that present recommendations on the reporting and handling of missing outcome data in an individual trial (for instance, (67;82;84)), but they are not on the

focus of the present thesis. Within a meta-analysis framework, the selection of the appropriate primary meta-analysis and sensitivity analyses should be guided by four principles (5). First, an ideal strategy should yield treatment effects with standard errors larger than those in the ACA. This principle is necessary in order to ensure that the uncertainty due to missing outcome data is carried into the meta-analysis. ICA fails to accomplish this, since it treats the imputed data as known and hence, increases the precision in the treatment effects. Second, an ideal strategy should use all available information (such as, reasons for missingness, evidence from related studies) in order to reduce bias. Bias is anticipated under ACA, since this approach ignores valuable information. Third, the strategy should be independent of the measure of treatment effect (e.g., odds ratio, mean difference). Last, the strategy should be simple to understand and easy to implement in widely used software.

In accordance with the four principles mentioned above, Higgins et al. proposed a combination of the ICA assumptions as a primary analysis, since it is plausible that the reasons for dropout may vary across the studies, along with a weighting scheme that incorporates the uncertainty due to missingness into the meta-analysis results (5). Despite the fact that the reasons of missingness include subjective judgments, this strategy approaches the reality in terms of the outcome that the missing participants could have provided if they had never left the trial. Similarly, an alternative strategy could be to use IMORs as a primary analysis, where uncertainty due to missingness could be 'expressed` using prior distributions (information outside the dataset) as suggested by subject experts. The proportion of missing outcome data and reasons for missingness may help the experts to determine this external information (5). Akl et. al. recommended ACA as a primary analysis or ICA using a specific assumption that is based on empirical evidence (1). They discouraged the use of ICA that assumes intervention-specific risks, because the increased total number of events leads to falsely increased precision, while the effect estimate remains the same (1).

1.7.4 Selection of an appropriate set of sensitivity analyses in meta-analysis level

In terms of sensitivity analysis, Higgins et al. suggested illustrating all possible imputation assumptions using the IMOR approach in a L'Abbé plot (5). This suggestion is based on the graphical sensitivity analysis of missing outcome data as proposed by Hollis (61). Another graphical implementation of sensitivity analysis is the contour plot described in section 1.4.2, which incorporates information both for the magnitude of the meta-analysis effect sizes and their significance under variant distributions of the proportion of the missing outcome data in each treatment group. Nevertheless, a difficulty in the use of the IMOR approach is to decide on the prior beliefs for the IMORs parameter. As mentioned by White et al., 'flat' priors for the logarithm of the IMORs are considered unrealistic, because they imply certainty that missing outcome data are either all event or all non-events (34). Therefore, it is recommended to use realistic priors based on subject matter experts (34). Higgins et al. advised against the implementation of the four extreme imputation assumptions (ICA-0, ICA-1, ICA-b and ICA-w) in the sensitivity analyses. This set of sensitivity

analyses provides the extreme bounds of the effect sizes, but it is considered implausible when there are many missing participants and hence, it should be avoided (5). As mentioned by Akl et al., to make the best use of the assumptions on the missing outcome, one could start with more plausible assumptions and if the results on the primary meta-analysis do not change substantially, then examine the impact of less plausible assumption (1). The rationale behind this stepwise approach is that risk of bias due to missingness may be low, when the meta-analysis results remain similar under less plausible assumptions. An alternative approach is to begin with the worst case scenario (ICA-w). If the results are robust then it is assured that the risk of bias due to missingness may be low. This approach might be very useful when the proportion of missing outcome data is low (1).

Chapter 2

Reporting and handling missing outcome data in mental health: A systematic review of Cochrane systematic reviews and meta-analyses

2.1 Introduction

Previous experience with systematic reviews in psychiatry demonstrated that despite the high dropout rate in the trials of this field, there is considerable inadequacy in the reporting information about the extent of missing outcome data, the analysis performed and the acknowledgment of the implications (7;29;85;86). This study aims to empirically inform on how aware the systematic reviewers are about the missing outcome data issue, to estimate the extent of inadequate description, to illuminate some common mistakes in the reporting and to investigate whether Cochrane systematic reviews in the mental health field have acknowledged the implications derived from missing outcome data.

2.2 Awareness on the missing outcome data issue

Missing outcome data are frequently encountered in trials of every clinical field and might have important implications on the results and conclusions. Bias and loss of power may result directly from the presence of missing outcome data. The degree of missing outcome data has been found to be considerable in psychiatric trials, as discussed extensively in the first chapter.

It has been already mentioned in the previous chapter that it is a mistake to associate the bias in trial results with the degree of dropout rate in the compared

intervention arms only. The cause of early dropout and the statistical methods applied to address premature discontinuation are the two key factors that also influence the validity of the analysis results (59). The evaluation of the risk associated with missing outcome data in a trial is well-described in the Cochrane Handbook of Systematic Reviews (87). Particularly, the Risk of Bias (RoB) table is the ideal tool for such evaluation as it contains an item that refers to attrition bias. This tool is part of the Cochrane Reviews since 2009. A detailed reporting on the RoB will assure a successful evaluation of the risk in studies, but it is rarely available in the published trial reports.

The reality is that many authors either ignore or mishandle missing outcome data despite the broad literature on statistical methods to handle missing outcome data in randomized controlled trials (RCTs) as well as the implications of missing outcome (9;58). As shown in a recent methodological review, missing outcome data in the primary outcome were present in 87% of the trials published in general medical journals of high impact. Of these trials, approximately one out of four completely ignored missing outcome data from the analysis (1).

There are many possible strategies to analyze a dataset with missing outcome data. It has been shown that the most favorite method to address missing outcome data is the intention-to-treat analysis followed by the per-protocol analysis and the as treated analysis (33). Under the intention-to-treat (ITT) analysis, all randomized participants in a trial are analyzed to the assigned intervention, irrespectively of whether they actually received it or not. As already mentioned in the previous chapter, the ITT analysis predominates because it preserves balance of both known and unknown prognostic factors and achieves high statistical power and precision. In the presence of missing outcome data, however, the ITT analysis is not feasible, because all randomized participants are required to be analyzed, in spite of whether they actually received the assigned intervention or completed the trial (33). As a result, modifications of the ITT principle are often implemented (71).

In an effort to perform ITT analysis in psychiatric studies, the last observation carried forward (LOCF) approach is often applied because it preserves most of the randomized sample and it is also easy to implement (40). Nonetheless, the use of LOCF has been repeatedly criticized due to its irrational assumptions (patients' responses after dropout remain constant) and the unrealistic results it often yields (40;43;50-52). As found by Hollis and Campbell, LOCF is the most popular imputation method in the primary analysis of trials, followed by extreme imputations (e.g. all missing participants experienced improvement or deterioration) (33). Additionally, Hollis and Campbell mentioned that the methods employed to ensure that the results pertain to ITT analyses were judged to be reported quite inadequately. The findings of Alshurafa et al. were in agreement with those of Hollis and Campbell (68).

The issue of missing outcome data is also present in the meta-analysis of trials. Missing outcome data can be handled in a meta-analysis, by employing an imputation scenario. In this case, the results will be unbiased but overprecise, provided that the

respective assumptions are plausible (5;32;63). However, there is rarely enough information about the missing outcome data in the trial reports to enable the implementation of an appropriate analysis model and hence, the systematic reviewer tends to evaluate the RoB tool associated with missing outcome data. Nevertheless, due to limitations in the description of trials, the assessment of the RoB tool is getting a real challenge. For instance, trialists often employ modifications to ITT analysis, but the exact method applied is rarely described (71). As reported by Wood et al., the majority of the trials that reported to have employed ITT analysis, did not actually include in the analysis all randomized participants (12). Overall, the evaluation of the RoB tool on the incomplete outcome data appears to be difficult due to important inadequacies in reporting and unclear use of the terminology.

2.3 Methodology

2.3.1 Search procedure

The literature examination included all reviews published in the Cochrane Database of Systematic Reviews between 1/1/2009 and 31/12/2012 (30). Since the new RoB tool was published in 2009, it was expected that reviewers would have searched for information on missing outcome data. The focus of the research was on reviews published by any of the three Cochrane Review Groups on mental health: the Depression, Anxiety and Neurosis Group, the Developmental, Psychosocial and Learning Problems Group and the Schizophrenia Review Group (30).

2.3.2 Detection of eligible systematic reviews and meta-analyses

The evaluation of reporting and strategies employed was based on eligible systematic reviews. Eligible meta-analyses were used for the evaluation of issues related to the analysis model used to handle missing outcome data (30).

Inclusion criteria for systematic reviews

A systematic review with at least one meta-analysis for a primary outcome that included more than three RCTs was considered for inclusion. Systematic reviews with RCTs of non-standard design, such as cross-over, cluster randomized controlled trials and factorial designs were excluded.

Inclusion criteria for meta-analyses

After having distinguished all eligible systematic reviews, the next step was to select the eligible meta-analyses. Specifically, only one meta-analysis per review was chosen on the primary outcome. In the presence of more than one eligible meta-analysis, priority was given to binary outcomes and then to statistical significance. In case all meta-analyses on binary outcomes were statistically significant (or non-

significant), the meta-analysis with higher dropout rate in the included studies was chosen. Meta-analyses of mortality were considered ineligible as well as meta-analyses without any information on the amount of missing data in the included studies.

Two reviewers (Loukia Spineli, LS and Nikolaos Pandis, NP) performed independently the selection of systematic reviews and the data extraction (30). In the presence of disagreement, a third reviewer (Georgia Salanti) was involved to reach a consensus. Prior to the initiation of full data extraction, calibration exercises were employed to ensure agreement across the two reviewers' results. To clarify any uncertainties, contact with the corresponding author of the systematic review took place.

2.3.4 Data extraction and management strategy

In the next step, LS and NP carefully screened all the selected systematic reviews and gathered information about missing outcome data on the primary outcomes only (30). Specifically, data on the characteristics of systematic review, such as publication year, any information regarding missing outcome data as well as the strategies relevant to the acknowledgment of the impact of missing outcome data on the findings were extracted. According to the guidance reported in the RoB tables, LS and NP attempted to replicate the reviewer's judgments about the presence of potential bias introduced by missing outcome data in all included trials of each review (87).

The data extraction from the selected meta-analyses included the type of outcome, treatment comparison and primary analysis (e.g., available cases analysis or ITT analysis) as well as the extent of missing outcome data and any methodological strategies to handle missing outcome data (30). Before finalization, data extraction forms were pilot-tested. The characteristics of systematic reviews and meta-analyses were summarized using percentages and medians.

2.3.5 Terminology relevant to missing outcome data

There is not a consensus on the terminology related to missing outcome data, some trialists and meta-analysts use terms like 'missing data' and 'intention-to-treat analysis' interchangeably, where others imply different things (68;71). A glossary of terms used in this chapter is presented in table 2.3.1. The definitions presented here may not be universally accepted (30).

Table 2.3.1
Glossary of terms related to missing outcome data

Terminology	Suggested definition
Attrition	This is a general term that describes the number of patient lost in a trial due to several reasons.
Missing outcome data	It refers to outcome data from participants who terminated (due to exclusions or withdrawal) the trial early for various reasons related or not to the structure of the trial (design, conduct and analysis). Some participants may have provided intermediate measurements. A term often used to describe these participants is *missing participants*. These participants may have provided a) at least one observed outcome (these participants are included in the analysis by using their observed outcome(s) under a specific statistical methodology, such as, LOFC, regression model, multiple imputations) or b) no observed outcome at all.
LOCF imputed outcomes	These are the outcomes obtained by participants who left the trial early, but they have provided at least one observed outcome. Then, these participants are usually included in the analysis by forwarding their last observed outcome to the remaining pre-specified time points (assessments) until the end of the trial.
Completely missing outcome data	These data constitute participants who terminated the trial early before the first post-baseline assessment and hence, they have not provided any measurement apart from that at baseline.
Imputation scenario	This is an assumption for the reason(s) the participants did not terminate the trial. Under this strategy, only *completely missing outcome* data are imputed. Common imputation scenarios are the following: all missing participants may have experienced the event of interest in both arms, all missing participants may have not experienced the event of interest in both arms and all missing participants in the experimental arm may have experienced the event of interest, while all missing participants in the control arm may have not experienced the event of interest. Apparently, LOCF imputed data cannot be used under this strategy as they need individual patient data.
Meta-analysis/data extraction according to the 'intention-to-treat' principal	The intention to treat principle requires the extraction and analysis of data from all participants in the groups they were randomized irrespectively of trial termination and compliance. To analyze a full data set, imputations are performed for the outcome of the missing participants.
Dropout rate	This term describes the proportion of participants who terminated the trial early due to any reason. The term *premature discontinuation rate* is used interchangeably.
Dropout threshold	This threshold reflects a dropout rate beyond which bias is likely to be introduced in the results of the trial. Such thresholds are used very often by meta-analysts (usually set arbitrarily) to exclude trials from the meta-analysis.

In this chapter, missing outcome data is distinguished into two categories: (i) the observed outcome data obtained from participants who left the trial early but they provided at least one outcome measurement before and (ii) the unobserved outcome data of those participants who left the trial early without providing any measurements apart from those at baseline (30). The first category is often included in the analysis of the original studies merely by manipulating the observed outcome(s) under a specific statistical methodology, such as LOCF. It is worth to mention that the fully observed outcomes are rarely distinguished from the imputed outcomes in the study

reports. As a result, the study outcome is treated as if estimated from observed data. The second category of missing outcome data constitutes the 'completely missing outcome data'. These data are often described in trial reports and hence, they can be easy analyzed at the meta-analysis level.

2.4 Results of the systematic review

Totally, 190 systematic reviews were eligible for this review and of those, 140 meta-analyses were considered eligible. The flow diagram in figure 2.4.1 illustrates the selection process (30). As reported in table 2.4.1, the majority of the systematic reviews derived from the Depression, Anxiety and Neurosis Group and the Schizophrenia Review Group and were published in 2009 (30). Complete absence of missing outcome data was clearly reported by 15 systematic reviews out of 190 (7.89%).

Table 2.4.1
Characteristics of 190 eligible systematic reviews

Characteristics	Categories		Count	%
Year of publication	2009		69	36.32
	2010		42	22.11
	2011		16	8.42
	2012		63	33.16
Cochrane Review Group	Depression, Anxiety and Neurosis Group		79	41.58
	Developmental, Psychosocial and Learning Problems Group		32	16.84
	Schizophrenia Review Group		79	41.58
The SR reported explicitly that the included trials of the primary outcomes had no missing outcome data.	Yes		17515	7.89
	No		175	92.11
The authors explicitly explained what they meant by the term used to describe missing outcome data	Yes		5	2.63
	No		183	96.32
	Unclear		2	1.05
The SR explicitly reported whether a LOCF strategy had been employed or not in the included trials	Yes		95	50.00
	The SR clearly distinguished the completely missing outcome data from the LOCF imputed data	*Yes*	*5*	*5.26*
		No	*90*	*94.74*
	No		93	48.95
	Unclear		2	1.05
Possible strategy applied to account for or address missing outcome data in their methods and results, as reported by the authors. * (Note that 175 SRs reported missing outcome data).	Dropout as an outcome		147	84.00
	Dropout as an outcome	*Primary*	*28*	19.05
		Secondary	*119*	80.95
	Imputation of missing outcome data in the primary analysis of (at least one) primary outcome		138	78.86
	Exclusion of trials from the meta-analysis using a specific dropout threshold		61	34.86
	Dropout threshold	*10%*	*1*	1.64
		20%	*8*	13.11
		30%	*3*	4.92
		40%	*14*	22.95
		50%	*34*	55.74
		60%	*1*	1.64
	Sensitivity analysis		28	16.00
The judgments for attrition bias in the RoB table were replicated	Yes, for all trials		36	46.75
	Yes, for some trials		36	46.75
	No, for all trials/No consensus		5	6.5

	No evaluation of attrition outcome because only allocation concealment was evaluated	113	59.47
Among the SRs with missing outcome data, the missing outcome data implications were reported (Note that 175 SRs reported missing outcome data).	Yes	61	34.86
	No	114	65.14
Missing outcome data implications were reported in * (Note that 61 SRs reported missing outcome data implications).	Abstract	10	16.39
	Results	10	16.39
	Conclusions	4	6.56
	Discussion	55	90.16
The reviewers discussed their findings in the context of missing outcome data by using information from * (Note that 61 SRs reported missing outcome data implications).	The amount of dropouts	47	77.05
	Judgments from the RoB table	0	0.00
	The meta-analysis of dropout	10	16.39
	The sensitivity analysis results	11	18.0

Note: SR: systematic review; LOCF: last observation carried forward; RoB: risk of bias table.

* A systematic review is possible to fit into at least one category.

In most of the systematic reviews it was totally ambiguous what was meant by missing outcome data, as there were no clear definitions of the terminology that related to missing outcome data; only 2.63% of the systematic reviews provided a clear definition on what was meant by missing outcome data (30). Half of the systematic reviews examined whether the trials implemented a LOCF strategy, but only 5.26% of the reviews clearly distinguished the LOCF imputed data from the completely missing outcome data (30). The majority of the systematic reviews with studies that have missing outcome data in their primary outcomes, analyzed missing outcome data as a secondary outcome (84%), 78.86% included missing outcome data in the primary analysis by employing a specific imputation strategy, 34.86% excluded from the meta-analysis those trials with percentage of missing data over a particular threshold (mainly for dropout threshold over 50%) before performing the primary analysis (30). A specified sensitivity analysis was applied only by 16% of the systematic reviews to investigate the impact of missing outcome data in the results (30).

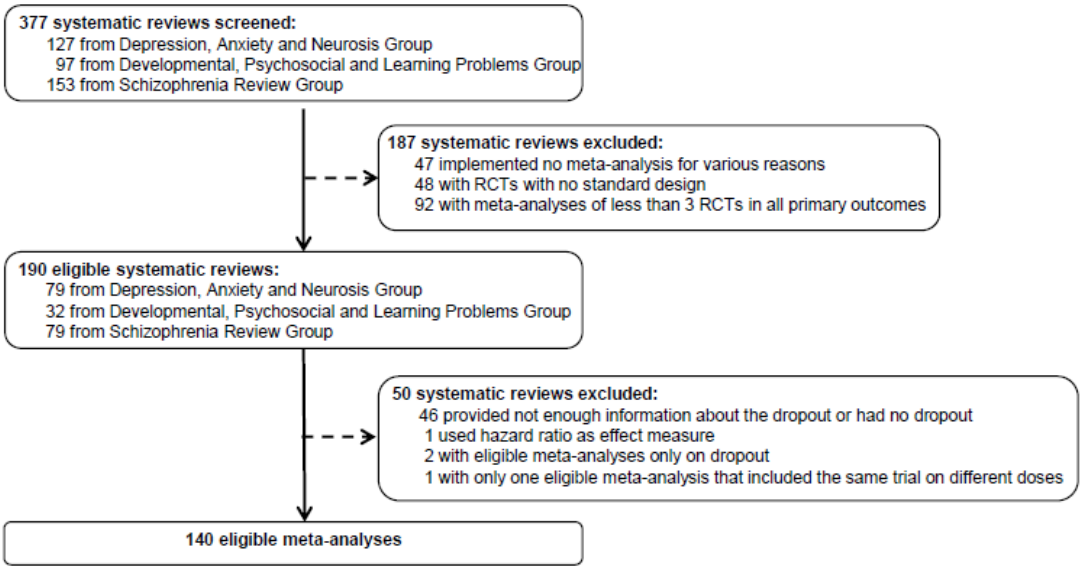

Figure 2.4.1 The flow diagram of the eligible systematic reviews and meta-analyses.

In general, author's judgments on attrition bias could be replicated satisfyingly. Specifically, in 36 out of 77 (46.75%) systematic reviews author's judgments were sufficiently justified for all included trials (30). Overall, based only on the verbatim text of the RoB table, the conclusions were the same with those drew by the authors in 1093 out of 1403 (77.90%) trials (30).

Only one in three systematic reviews discussed the possible implications of the missing outcome data and mostly in the discussion section (30). The reviewers referred mostly to the total amount of patients who dropped out in order to discuss their findings in the context of missing outcome data (30).

The characteristics of the 140 eligible meta-analyses (948 included trials with 101,021 participants in total) are summarized in table 2.4.2 (30). The median number of trials per meta-analysis was 4 (30). LOCF imputed data were included in of 3 trials per meta-analysis (median). The majority of meta-analyses (70.71%) assessed pharmacological interventions and 38.57% included placebo as control group. A dichotomous outcome measure was favored by most meta-analyses (78.57%) and mostly the risk ratio (57.14%).

Table 2.4.2
Characteristics of 140 meta-analyses for a primary outcome

Continuous characteristics				Median	Minimum	Maximum
Total number of trials				4	3	104
Total participants randomized (both arms)				404	78	7422
Total participants randomized in intervention				191	36	3775
Total participants randomized in control				192	38	3647
Total number of trials with LOCF data				3	1	14
Categorical characteristics	Levels				Count	%
Type of intervention	Combination/ augmentation				7	5.00
	Pharmacological				99	70.71
	Psychological				27	19.29
	Other (i.e., alternative, waiting list, no treatment)				7	5.00
Type of control	Placebo				54	38.57
	Combination/ augmentation				8	5.71
	Pharmacological				53	37.86
	Psychological				10	7.14
	Other (i.e., alternative, waiting list, no treatment)				15	10.71
Primary outcomes	Response to intervention				37	26.43
	Failure to respond to intervention				73	52.14
	Mean change of the symptoms based on a scale				23	16.43
	Other				7	5.00
Effect measure used	Odds ratio				32	22.86
	Risk ratio				80	57.14
	Risk differences				1	0.71
	Mean difference				16	11.43
	Standardized mean difference				11	7.86
Dropout rate range in the included trials	≤ 10%				24	17.14
	(10 – 30%]				79	56.43
	(30 – 50%]				29	20.71
	> 50%				8	5.71
Primary analysis is reported to be	Available case analysis				8	5.71
	Intention to treat	It was judged as	Imputation with or without LOCF		85	60.71
			LOCF only		18	12.86
			Unclear		6	4.28
	Modified intention to treat		Imputation with or without LOCF		2	1.43
			LOCF only		4	2.86

	Unclear	4	2.86
	Per protocol	2	1.43
	Other/Unclear	11	7.86
Primary analysis was judged to be[a]	Available cases analysis with or without LOCF imputed outcomes	30	21.43
	Imputation of missing outcome data as events or non-events	79	56.43
	Imputation of missing outcome data using intervention-specific event rate	9	6.43
	Imputation of missing outcome data using best case scenario for the experimental group	1	0.71
	Mixture of imputation methods across trials	5	3.57
	Unclear	16	11.43
The strategy implemented as sensitivity analysis was [b] (Note that 27 meta-analyses implemented sensitivity analysis)	Available cases analysis with or without LOCF imputed outcomes	11	40.74
	Imputation of missing outcome data as events or non-events	1	3.70
	Best or worst case scenario for the experimental group	13	48.15
	Subgroup analysis by dropout rate or by method of analysis of the individual trials	11	40.74
	No sensitivity analysis	113	80.71
Did the authors provide any justification to support their strategy of handling missing outcome data?	Yes, some justification was provided	9	6.43
	No justification was provided	98	70.00
	Description of the strategy but no justification	25	17.86
	Unclear	8	5.71
Did the authors comment on the inferences after sensitivity analysis? (Note that 27 meta-analyses implemented sensitivity analysis)	Yes, they reported that the inferences changed	2	7.41
	Yes, they reported that inferences were robust	24	88.89
	Unclear	1	3.70
	No sensitivity analysis was applied	113	-
The authors took into account the findings of the sensitivity analysis to interpret the results (Note that 27 meta-analyses implemented sensitivity analysis)	Yes	8	29.63
	No	18	66.67
	Unclear	1	3.70
	No sensitivity analysis was applied	113	-
The authors explicitly used a term to describe the categories of missing outcome data in the included trials	Yes	56	40.00
	No	80	57.14
	Unclear	4	2.86
The authors provided information on whether the included trials contained LOCF imputed outcomes?	Yes	79	56.43
	No	54	38.57
	Unclear	7	5.00

Note: LOCF: last observation carried forward.
[a] It was quite unclear in the systematic reviews whether the trials had already implemented LOCF strategy.
[b] Some meta-analyses had applied at least one sensitivity analysis.

At least half of the meta-analyses reported dropout rate in the range between 10% and 30%, while 26.42% of the meta-analyses reported dropout rate higher than 30% (30). Around 75% of the meta-analyses reported to have applied ITT analysis as a primary analysis by imputing missing outcome data (with LOCF strategy or other imputation scenarios) (30). The 'modified intention to treat' analysis was reported in 7.14% of the meta-analyses. After looking at the actual method of analysis

implemented rather than the method reported, the majority of trials seemed to have employed extreme scenarios by imputing completely missing outcome data either as events or non-events (30). Most of the meta-analyses (80.71%) did not perform any sensitivity analysis (30). All but 9 meta-analyses did not provide any explanation for the use of the assumed sensitivity analyses or the absence of it (30). Out of 27 meta-analyses that employed a strategy as sensitivity analyses, the majority (89%) reported that the inferences of the original analyses were robust (30). Table 2.4.3 presents some quotes from the systematic reviews regarding the explanation given by the authors to support their strategy of handling missing outcome data (30).

Table 2.4.3
Quotes given by the authors to justify their strategy on the missing outcome data

1.	"Where participants were withdrawn from the trial before the endpoint, it was assumed that their condition remained unchanged if they had stayed in the trial. This is conservative for outcomes related to response to treatment (because these participants will be considered to have not responded to treatment)"
2.	"For dichotomous data, analysis was undertaken assuming all those not completing the trial in the cognitive behavioral therapy group were treatment failures and all those not completing the trial in the control group were treatment successes, thereby allowing analysis of the most conservative treatment outcome"
3.	"Since little missing data were found across trials, we did not need to contact the authors of trials for further information or data. The meta-analysis of social competence used data from all original participants. Because little missing data were found, we did not conduct a sensitivity analysis to assess potential bias in the analysis or discuss the extent to which the results might be biased by missing data. Due to the heterogeneity shown by individuals with autism spectrum disorder, we did not impute missing data"

2.5 Discussion on the findings

The aim of this review was to collect empirical evidence about the stance of Cochrane systematic reviews in the mental health field toward the missing outcome data issue and the strategies employed to address missing outcome data. Several reviews have investigated the reporting of missing outcome data and the approaches used to handle missing outcome data at trial level (12;33;63;68). In these reviews, the percentage of missing outcome data was quite variable ranging from 0% to more than 20%. In terms of handling missing outcome data, LOCF and imputation under specific scenario(s) were the most frequently implemented approaches. However, none of these reviews discussed their findings in the context of missing outcome data.

Based on the results of this review, it can be concluded that there is great underreporting regarding any information about missing outcome data in the systematic reviews in mental health and despite the considerably large percentage of missing outcome data, the implications of missing outcome data are usually not addressed (30). As mentioned by Carpenter and Kenward, serious weaknesses exist in the description of the missing outcome data as well as in the statistical methodology employed to handle them (9). The present review also illuminated that most of the

analysts who considered missing outcome data in their primary analysis described the strategies they employed (imputation of missing outcome data as non-event in half of the meta-analyses) but they did not explicitly justify their selection (30). Gamble and Hollis collected 151 Cochrane systematic reviews and they investigated the strategies employed to handle missing outcome data on the binary outcome (63). They found that the majority of the systematic reviews imputed completely missing outcome data under extreme scenarios but without providing enough justification (63). For instance, missing outcome data in the experimental arm were assigned the beneficial outcome, while missing outcome data in the control arm were assigned the harmful outcome (63).

Despite the overall large number of missing outcome data, only 16% of the eligible meta-analyses performed a sensitivity analysis for missing outcome data (30). As reported by Gamble and Hollis, half the reviews had planned to undertake a sensitivity analysis; 39% of these presented their results graphically and only one review reported to have implemented a sensitivity analysis without providing further details, though (63). Interpretation of the sensitivity analysis results was often either unclear or not reported at all.

Finally, Gamble and Hollis found that the majority of the reviews who performed a sensitivity analysis and took into account the findings of the sensitivity analysis to interpret the results, used information from the point estimate, the changes in the statistical significance and the robustness of the inferences (63). The results of the present study are also in agreement (30). It needs to be clearly understood that both the results of the sensitivity analyses and the amount of missing outcome data should be always considered in order to understand the impact of missing outcome data on the findings.

Some limitations need to be considered when interpreting the findings of the present review. A major limitation is the restriction of the review to Cochrane reviews only (30). The findings may be more conservative, because all published reviews after 2009 are imposed to use the RoB tool and hence, more awareness and care about issues of attrition is expected. Reporting and methodology are expected to be even less adequate in non-Cochrane reviews. However, a comparison study needs to investigate this hypothesis before concluding. Another limitation of this review is that it investigated systematic reviews in mental health only and hence, the results cannot be generalized to all clinical fields (30). Last but not least, the focus of this study was on the primary outcomes. Note that the reporting and the use of sensitivity analysis will be even more inadequate for secondary outcomes (30).

Chapter 3

The impact of trial characteristics on premature discontinuation of antipsychotics in schizophrenia

3.1 Introduction

As mentioned already in the first chapter, patient dropout is very common in mental health trials and hence, it is important to understand why patients drop out from trials, so that measures can be taken to minimize its occurrence. The aim of this chapter is to investigate various trial characteristics that have an impact on premature discontinuation in antipsychotic trials for schizophrenia. To accomplice this, Poisson regression analysis will be applied to account for the impact of trial duration in discontinuation rate. Then, multinomial regression models will be employed to investigate the characteristics associated with the likelihood that a patient falls into different mutually exclusive categories: (i) of completing the study, (ii) being included in a LOCF analysis, or (iii) being completely missing from the analysis altogether.

3.2 A review on the trial characteristics and premature discontinuation in antipsychotic trials for schizophrenia

Dropout rate over 50% has been reported to be common in studies on pharmacological treatments for schizophrenia (10). Early dropouts have been criticized in the literature for materially affecting the credibility and clinical interpretation of findings from clinical trials. Therefore, premature discontinuation should be compelled to be investigated meticulously. Particularly, predictors of

dropout in a clinical trial may illuminate the design and conduct of the trial so that the quality of future trials is improved.

In a study of 31 trials comprising of 10,058 patients with a diagnosis of schizophrenia or schizoaffective disorder, Kemmler et al. found that placebo-controlled trials reported higher dropout rates compared to active-controlled trials (22). A plausible explanation of this finding may be that clinicians are more likely to withdraw patients who experienced no improvement earlier in a placebo-controlled study, when it is suspected that they have been assigned an ineffective treatment (placebo). In addition, when patients are aware that one of the treatments is placebo, they are more likely to terminate early a trial (88).

It has been observed that, within a trial, completers and dropouts tend to have different characteristics; for instance, it is more likely that younger patients will dropout early from a trial, while it is more likely that older patients will complete a trial (89). Severity of illness has been also found to predict which patients are more likely to dropout; the higher the severity of the disease, the more higher the likelihood of dropping out during an acute phase of the trial (10;89).

3.3 Methodology of the statistical analysis

3.3.1 Data extraction

In this study, 133 RCTs were included comparing 15 first- and second- generation antipsychotics either with each other or with placebo for the treatment of schizophrenia in the acute phase. Details on this published systematic review can been found in (86). In each trial, the total number of randomized patients was divided into three mutually exclusive groups: i) the completers (patients with complete data), ii) the LOCF group (patients who had provided at least one response before dropping out and hence, they were included in the analysis by forwarding their last observed outcomes into the subsequent pre-specified time-points); and iii) the completely missing outcome data group (patients who provided no outcome data apart from their baseline measurements and hence, they were excluded from the analysis) (29). The sum of the last two groups gives the total number of missing outcome data. Exclusions due to protocol violations after randomization were not the interest of this study.

The covariates of interest were the following: i) allocation concealment; ii) blinding status (outcome assessors or both patients and outcome assessors); iii) use of placebo as control; iv) standard error of standardized mean differences (SMDs) for efficacy (used as a measure of study precision); v) logarithm of study sample size; vi) number of treatment arms; vii) year of publication; viii) average patient age; ix) country of recruitment; x) number of recruiting centres; and xi) recruitment setting (29). Frequencies and percentages based on the number of trials and patients were used to summarize categorical trial-level characteristics, whereas means, standard deviations (SD) and ranges were used to summarize continuous trial-level characteristics.

3.3.2 Univariable and multivariable Poisson regression model

The dropout rate per patient-week in a trial was defined as the ratio of the total number of dropouts to the total number of patient-weeks of follow-up. The logarithm transformation of the dropout rate was considered because it assures normality that is required to implement the statistical tests. A univariable Poisson regression model was performed to model the relationship between the logarithm of the dropout rate and each trial characteristic (29). Particularly, a random-effects approach was employed because the impact of a trial characteristic on the dropout rate might differ across studies. A multivariable Poisson regression analysis was applied to study many characteristics jointly (29). Each regression coefficient was interpreted as a rate ratio (RR), where 1-RR indicates that a one unit increase in a studied continuous trial-level characteristic causes a relative change in the dropout rate.

The amount of heterogeneity in the dropout rate that could be explained by each characteristic, was assessed be comparing the heterogeneity variance before adjustment with the heterogeneity after adjustment (90). To make predictions on the dropout rate in a future included trial for a specific characteristic, the predicted dropout rate per patient-week was estimated for that characteristic (5).

3.3.3 Multinomial logistic model

As the possible classifications for a patient that pertain to the availability of his responses are three (completer, LOCF or completely missing), the multinomial distribution is ideal to describe the distribution of the randomized patients. The completers comprised the reference group. A univariable multinomial logistic regression analysis was performed to estimate the likelihood that a patient would be LOCF or completely missing rather than a completer while adjusting for trial characteristics (29). To examine simultaneously the association between all the trial-level characteristics, a multivariable logistic regression was applied.

The risk that a patient with a specific characteristic would provide at least one response or would be missing completely, compared with the risk that a patient with the same specific characteristic would complete the trial, was measured using a risk ratio (29). The interest pertained to whether this risk ratio would differ for different values of a specific characteristic. The comparison of two values of a specific characteristic produces a ratio of risk ratios, which is termed as relative risk ratio (RRR). Particularly, the exponentiation of the regression coefficient in a multinomial regression analysis gives a RRR. To investigate the amount of heterogeneity explained by a specific characteristic, the heterogeneity variance of risk ratios in the logarithm scale was calculated before and after adjusting for that characteristic (29).

3.3.4 Regression models and covariates with partial information

Partial information can be observed not only on the outcome but also on covariate data. In this case, some plausible assumptions about the missing covariate data need to be made in order to keep the corresponding trials in the analysis. Missing at random is

a common assumption also in the missing covariate data. Under this assumption, only the available data can be analyzed after ignoring the trials with missing covariate data. If the assumption is plausible, the reliability of the results will not be compromised (91).

A more sophisticated method to address missing covariate data has been previously proposed by Hemming et al (92). Hemming et al. described a meta-regression model that allows trials with partial information on covariates to be included in the analysis by jointly modeling the relationship between the outcome and the covariates with the inter-relationships between the covariates.

3.3.5 Model implementation

The models were implemented within a Bayesian framework using the software WinBUGS (93; 94). Vague priors were employed for all location parameters (e.g., means) and minimally informative priors for scale parameters (e.g., standard deviations) (29). Results are presented as posterior medians along with 95% credible intervals (CrIs). The regression models and the methods to account for partial information on covariates are presented in the Appendix B.

3.4 Results

3.4.1 Trial characteristics

In total, 33,073 patients were involved in the 133 antipsychotic trials (29). Of these patients, 64% completed the trials. Out of 11,922 patients who dropped out, 90% provided at least one measurement of the outcome and hence, they were included in the analysis based on the LOCF principle, whereas the remaining 10% did not provide any measurement at all apart from the baseline measurements. The total number of patient-weeks of follow-up was 215,890. The trial duration was 6.8 weeks on average, ranging from 4 to 12 weeks. Table 3.4.1 summarizes the patient and trial characteristics (29).

Table 3.4.1
Distribution of patients and trial characteristics

Categorical characteristics	**Levels**	**Studies (%)**	**Patients (%)**
		Total 133	**Total 33,073**
Allocation concealment	Adequate	47 (35.3%)	16,217 (49%)
	Unclear	86 (64.7%)	16,856 (51%)
Blinding status	Single	6 (4.5%)	863 (2.6%)
	Double	127 (95.6%)	32,210 (97.4%)
Placebo as control	PCT	61 (45.8%)	14,704 (44.5%)
	ACT	72 (54.1%)	18,369 (55.6%)
Number of arms	2	100 (75.2%)	22,059 (66.7%)
	>2	33 (24.8%)	11,014 (33.3%)
Recruitment country	International	62 (46.6%)	20,724 (62.7%)
	USA	38 (28.6%)	8,451 (25.6%)
	Europe	27 (20.3%)	2,615 (8.0%)
	Unknown	6 (4.5%)	1,283 (3.7%)

Recruiting centres	Single	22 (16.5%)	1,070 (3.2%)
	Multi	96 (72.2%)	30,028 (90.8%)
	Unknown	15 (11.3%)	1,975 (6%)
Recruitment setting	Inpatients	78 (58.6%)	1,891 (5.7%)
	Outpatients	7 (5.3%)	14,872 (45.0%)
	Mixed	28 (15.8%)	8,354 (25.3%)
	Unknown	27 (45.1)	7,956 (24.0%)
Continuous characteristics		**Mean (SD)**	**Min - Max**
Dropout rate		36% (17%)	0 – 71%
Dropout rate per patient-week		5.6% (3.3%)	0 – 16%
Standard error of SMDs		0.21 (0.12)	0.05 – 0.67
Study size		148.88[a] (2.87)[b]	16 – 1996
Publication year		1999 (9.2)	1970 - 2010
Average patient age		37.64 (5.84)	21 – 71.2
Duration (in weeks)		6.8 (2.3)	4 – 12

Note: PCT, Placebo-controlled trial; ACT, Active-controlled trial, SMD, Standardized mean difference.
[a] Geometric mean, [b] Geometric standard deviation.In total, 33,073 patients were involved in the 133 antipsychotic trials (29). Of these patients, 64% completed the trials. Out of 11,922 patients who dropped out, 90% provided at least one measurement of the outcome and hence, they were included in the analysis based on the LOCF principle, whereas the remaining 10% did not provide any measurement at all apart from the baseline measurements. The total number of patient-weeks of follow-up was 215,890. The trial duration was 6.8 weeks on average, ranging from 4 to 12 weeks. Table 3.4.1 summarizes the patient and trial characteristics

As shown in table 3.4.1, most of the trials had unclear allocation concealment, were double-blinded, included active treatment as control, studied two treatments, were conducted internationally, were multi-centre trials and treated inpatients (29). The mean proportion of dropouts across studies was 34% on average, with range from 0% to 71%. The dropout rate was on average 5.6% per patient-week with range from 0% to 16%.

3.4.2 Univariable and multiple Poisson regression analyses

The results from the univariable and multiple Poisson regression analyses are presented in table 3.4.2 (29). Trials associated with higher dropout rates were those with adequate allocation concealment, with double-blind design, that used placebo as control, with higher precision in estimates, with larger sample size, that compared at least three treatments, were more recently published, were conducted in USA and treated inpatients. In the unadjusted model, the between-study standard deviation for the dropout rate was 0.66 (95% CrI: 0.57 to 0.77). Most of the heterogeneity variance (40% relative drop) was explained by the use of placebo as control followed by the type of centre (20% relative drop).

table 3.4.2
Effects of the characteristics on the dropout rate per patient-week

Variables	**Contrast**	**Univariable analysis**			**Multiple analysis**		
		RR	**95% CrIs**		**RR**	**95% CrIs**	
			Lower	**Upper**		**Lower**	**Upper**
Allocation concealment	Adequate *vs.* Unclear	**1.31**	**1.01**	**1.63**	1.15	0.94	1.39
Blinding status	Double *vs.* Single	**2.25**	**1.22**	**3.72**	1.51	0.80	2.25
Placebo as control	PCTs *vs.* ACTs	**2.02**	**1.66**	**2.41**	**1.70**	**1.32**	**2.17**
Std. error of SMDs	Increase by 0.05	**0.92**	**0.87**	**0.97**	1.04	0.76	1.21
Logarithm of study size	Increase by 2	**1.56**	**1.22**	**2.02**	1.22	0.38	2.16

Number of arms	>2 *vs.* 2	**1.70**	**1.30**	**2.13**	1.02	0.77	1.36
Publication year	Increase by 10 years	**1.29**	**1.12**	**1.48**	**1.21**	**1.04**	**1.35**
Average patient age	Increase by 10 years	1.25	0.99	1.51	0.91	0.77	1.05
Recruitment country	USA *vs.* International	**1.58**	**1.20**	**2.02**	**1.41**	**1.12**	**1.69**
	Europe *vs.* International	0.87	0.62	1.17	1.19	0.88	1.55
	USA *vs.* Europe	**1.86**	**1.29**	**2.56**	1.21	0.87	1.63
Recruiting centres	Multi *vs.* Single	**2.27**	**1.73**	**2.94**	**1.83**	**1.36**	**2.34**
Recruitment setting	Mixed *vs.* In patient	**0.72**	**0.56**	**0.92**	0.83	0.65	1.15

Note: Statistically significant results are in bold. RR, risk ratio; CrI, credible interval; PCTs, Placebo-controlled trials; ACTs, Active-controlled trials; Std error, standard error; SMD, Standardized mean difference.

Table 3.4.3 presents predictions about the dropout rates per patient-week across all the characteristics in future clinical trials (29).

Table 3.4.3
Predicted dropout rates per-patient-week across all the characteristics

Variables	**Levels**	**Dropout rate per patient-week**	**95% Prediction Intervals**	
			Lower	**Upper**
Allocation concealment	Adequate	5.7%	1.6%	21.4%
	Unclear	3.7%	0.6%	21.2%
Blinding status	Single	2.3%	0.6%	9.2%
	Double	5.0%	1.4%	17.8%
Placebo as control	PCT	7.0%	2.4%	20.2%
	ACT	3.4%	1.2%	10.1%
Standard error of SMDs	Minimum std. error	6.1%	1.7%	22.2%
	Increase by 0.05	5.6%	1.5%	19.8%
Logarithm of study size	Minimum	3.0%	0.7%	10.7%
	Increase by 2	4.5%	1.2%	16.3%
Number of arms	2	4.2%	1.2%	14.4%
	>2	7.0%	2.1%	23.6%
Publication year	Minimum year	2.2%	0.6%	8.3%
	Increase by 10 years	2.8%	0.8%	10.2%
Average patient age	Minimum age	3.4%	0.8%	12.7%
	Increase by 10 years	4.0%	1.1%	14.0%
Recruitment country	International	4.2%	1.2%	14.0%
	USA	6.7%	1.8%	23.6%
	Europe	3.6%	1.1%	13.2%
Recruiting centres	Multi	5.5%	1.7%	17.6%
	Single	2.5%	0.8%	7.8%
Recruitment setting	Inpatients	5.3%	1.4%	18.5%
	Mixed	3.8%	1.0%	13.8%

Note: PCTs, Placebo-controlled trials; ACTs, Active-controlled trials; SMD, Standardized mean difference.

Figure 3.4.1 illustrates the relationship between the observed dropout rates and several continuous trial-level characteristics in antipsychotic trials (29). Dropout rates close to zero per patient-week seem to be associated with smaller trials and this might be an indication of reporting bias.

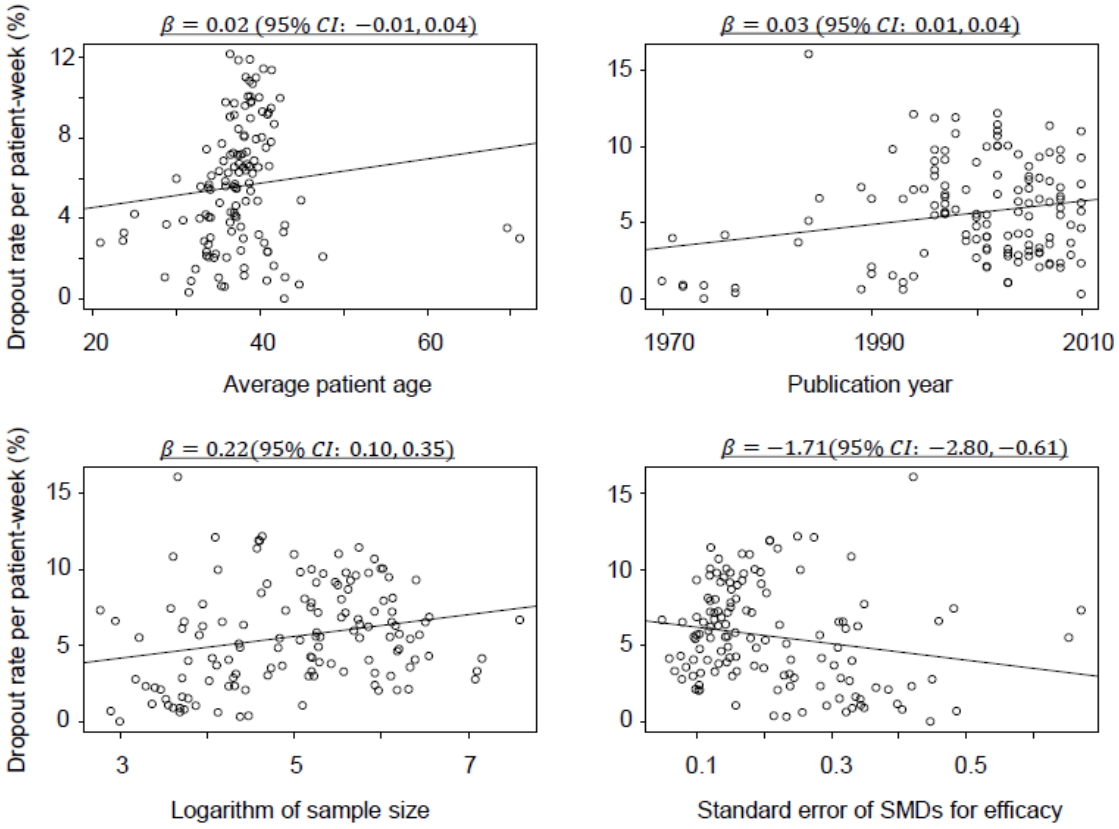

Figure 3.4.1. Scatterplots of the dropout rate per patient-week in continuous covariates. The line reflects the relationship between the dropout rate per patient-week and the trial characteristic. Parameter β is the regression coefficient that indicates the degree of change in the dropout rate per patient-week for a one unit increase in the trial characteristic.

3.4.3 Multivariable Poisson regression analysis

The number of trials with missing data was 6, 7, 15 and 27 in the location of trial conduction, average participant age, trial centre and patient type, respectively (29). A multivariable Poisson regression model was applied where trials with partial information on covariates could be included in the analysis after employing the methodology described in (91). The between-study standard deviation was reduced from 0.66 to 0.40 (95% CrI: 0.33 to 0.51), after adjusting for all characteristics. As shown in table 3.4.2, apart from the standard error of SMDs for efficacy, the comparison between European and international trials and the patient type, the rest of the covariates caused a slight decrease to the magnitude of the regression coefficients (29). The reduction in the coefficient of the variable 'Number of arms' in the

multivariable meta-regression may be explained by the correlation between the 'Placebo use' and the 'Number of arms' (87.8% of the multi-arm trials were placebo-controlled). The inclusion of placebo as a control arm had the most prominent impact on the dropout rate.

3.4.4 Univariable multinomial logistic regression analysis

To compare the LOCF and the completely missing patients with the completers, univariable multinomial logistic regression was implemented. Figure 3.4.2 depicts the proportion of completely missing patients across the characteristics in antipsychotic trials (29). Table 3.4.4 presents the results of the univariable multinomial logistic regression analysis (29).

As shown in table 3.4.4, in trials with adequate allocation concealment, that used placebo as control, had higher precision, larger size, compared at least three treatment arms, were more recently published, conducted in USA than in Europe or internationally, were multi-centred, treated inpatients and had shorter duration, the likelihood that a patient would be LOCF rather than a completer was significantly higher (29). On the other hand, in trials with double-blind design, that used placebo as control, had smaller sample size, randomized patients in USA, were single-centred and treated mixed patients (inpatients and outpatients), the likelihood that a patient would be completely missing rather than a completer was significantly higher (29).

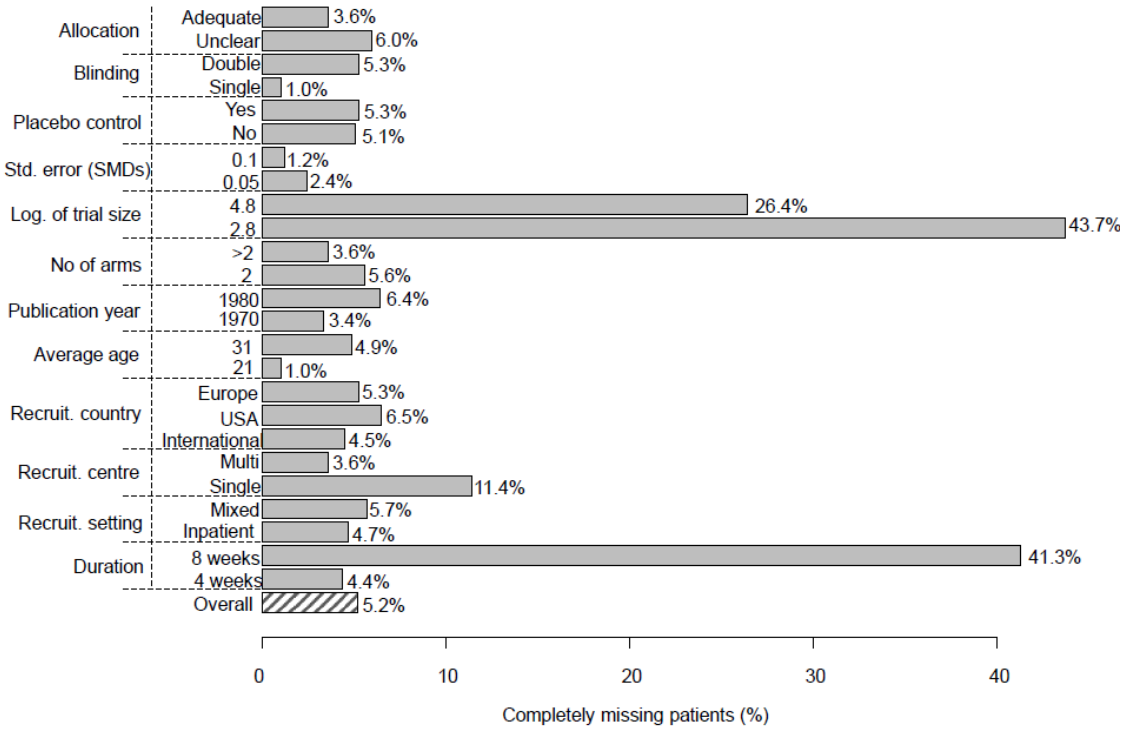

Figure 3.4.2. Bar plot of the proportion of completely missing patients in various characteristics

Table 3.4.4

Univariable multinomial logistic regression analysis: Trial characteristic effects on the compared groups

Variables	Contrast	RRR	95% CrIs		RRR	95% CrIs		σ	95% CrIs	
			Lower	Upper		Lower	Upper		Lower	Upper
		LOCF vs. Completers			*Missing vs. Completers*			***1.39***[a]	***1.23***	***1.57***
Allocation concealment	Adequate *vs.* Unclear	**1.70**	**1.08**	**2.77**	1.04	0.60	1.73	1.38	1.22	1.56
Blinding status	Double *vs.* Single	2.29	0.57	5.62	**7.37**	**1.22**	**28.43**	1.38	1.22	1.56
Placebo as control	PCTs *vs.* ACTs	**2.59**	**1.54**	**4.41**	**1.83**	**1.05**	**3.00**	1.33	1.17	1.50
Standard error of SMDs	Increase by 0.05	**0.75**	**0.67**	**0.83**	**1.21**	**1.09**	**1.33**	1.18	1.04	1.33
Logarithm of trial size	Increase by 2	**2.56**	**1.66**	**3.64**	**0.60**	**0.37**	**0.93**	1.25	1.11	1.42
Number of arms	>2 *vs.* 2	**2.76**	**1.51**	**4.70**	1.21	0.65	2.14	1.35	1.19	1.52
Publication year	Increase by 10 years	**1.05**	**1.03**	**1.08**	0.99	0.96	1.01	1.33	1.18	1.50
Average patient age	Increase by 10 years	1.32	0.91	1.89	0.70	0.43	1.09	1.37	1.21	1.55
Recruitment country	USA *vs.* International	**2.02**	**1.11**	**3.34**	**2.48**	**1.37**	**4.14**	1.35	1.19	1.53
	Europe *vs.* International	0.77	0.38	1.35	1.31	0.64	2.38			
	USA *vs.* Europe	**2.84**	**1.28**	**5.36**	2.07	0.87	4.12			
Recruiting centres	Multi *vs.* Single	**3.49**	**2.47**	**3.96**	**0.33**	**0.23**	**0.44**	1.15	1.01	1.31
Recruitment setting	Mixed *vs.* Inpatients	**0.64**	**0.52**	**0.77**	**1.24**	**1.11**	**1.51**	1.31	1.16	1.47
Duration	Increase by 4 weeks	**0.58**	**0.38**	**0.86**	1.22	0.77	1.85	1.36	1.20	1.53

Note: Statistically significant results are in bold. RRR, relative risk ratio; CrI, credible interval; PCTs, Placebo-controlled trials; ACTs, Active-controlled trials; SMD, Standardized mean difference.

[a]Between-study standard deviation of the logarithm of relative risk ratios in the unadjusted model is highlighted in bold and italic.

In the unadjusted model the between-study standard deviation was rather large (1.39, 95% CrI: 1.23 to 1.57) (29). The trial centre explained some of the heterogeneity (31% relative decrease in heterogeneity) followed by the standard error of SMDs for efficacy (28% relative decrease) and the logarithm of sample size (19% relative decrease).

3.4.5 Multivariable multinomial logistic regression analysis

Table 3.4.5 presents the results from the multivariable multinomial logistic regression analysis (29). Studying the likelihood that a patient will be LOCF than completer, the results were similar in the magnitude and direction to those from the univariable multinomial logistic regression analysis (29). However, in trials with single-blind design, patients were less likely to be LOCF rather than completers as shown by the univarible analysis, whereas multivariable analysis showed such a result for the most recently published trials (29). When studying the likelihood that a patient will be completely missing than completer, both analyses agreed only on the use of placebo as control and the number of recruiting centres. However, patients were significantly more likely to be completely missing rather than completers in trials with single-blind design and with younger or mixed patients, as opposed to the results from the univariable analysis. Heterogeneity decreased relatively by 42% after adjusting for all characteristics.

Table 3.4.5

Multivariable multinomial logistic regression analysis: Trial characteristic effects simultaneously on the compared groups

Variables	Contrast	RRR	95% CrIs		RRR	95% CrIs	
			Lower	Upper		Lower	Upper
		LOCF vs. Completers			*Missing vs. Completers*		
Allocation concealment	Adequate *vs.* Unclear	**1.37**	**1.14**	**1.54**	1.52	0.88	2.19
Blinding status	Double *vs.* Single	**0.28**	**0.26**	**0.30**	**0.47**	**0.34**	**0.71**
Placebo as control	PCTs *vs.* ACTs	**2.44**	**2.01**	**2.89**	**1.99**	**1.53**	**3.08**
Standard error of SMDs	Increase by 0.05	**0.75**	**0.72**	**0.79**	1.03	0.96	1.09
Logarithm of trial size	Increase by 2	**0.72**	**0.69**	**0.75**	0.95	0.75	1.11
Number of arms	>2 *vs.* 2	**1.31**	**1.18**	**1.48**	0.92	0.75	1.13
Publication year	Increase by 10 years	0.99	0.97	1.03	1.09	0.99	1.16
Average patient age	Increase by 10 years	1.06	0.98	1.16	**0.82**	**0.74**	**0.90**
Recruitment country	USA *vs.* International	**1.27**	**1.11**	**1.47**	1.60	0.94	2.22
	Europe *vs.* International	**2.07**	**1.39**	**2.64**	1.19	0.86	1.79
	USA *vs.* Europe	0.96	0.54	1.74	1.76	0.73	3.41
Recruiting centres	Multi *vs.* Single	**1.74**	**1.58**	**2.00**	**0.32**	**0.26**	**0.38**
Recruitment setting	Mixed *vs.* Inpatients	**0.90**	**0.78**	**0.99**	**1.37**	**1.16**	**1.60**
Duration	Increase by 4 weeks	**0.91**	**0.83**	**0.99**	0.98	0.73	1.33

Note: Statistically significant results are in bold. RRR. Relative risk ratio; CrI, credible interval; PCTs, Placebo-controlled trials; ACTs, Active-controlled trials; SMD, Standardized mean difference.

3.5 Discussion on the findings

The results from Poisson regression analysis indicated that the present study successfully confirmed previous findings regarding the association between higher dropout rates and placebo-controlled trials as well as suggestions that trials conducted in the USA might have higher dropout rates (22). However, this study showed that higher dropout rates were present in more recently published trials, or in double-blinded trials as opposed to Kemmler et al (29). In addition, this study investigated many more trials than Kemmler's. Kemmler et al investigated only trials conducted for drug registration purposes, and hence, the results of the present study might be generalized better (29).

Higher dropout rate was found to be associated with placebo-controlled trials (29). Theoretically, a patient is possible to be allocated to placebo, but it may increase the likelihood of premature discontinuation during the first weeks of treatment, especially if no benefit has emerged or clinical conditions become worse. Therefore, clinicians and patients tend to be particularly worried about risks related to the biological inefficacy of placebo in placebo-controlled trials (22).

This study revealed that trials conducted in USA were associated with higher dropout rate than trials conducted in Europe or elsewhere (Africa, Asia, Australia and South America) (29). Generally, antipsychotic placebo-controlled trials tend to be conducted in the USA and in turn placebo-controlled trials are found to be correlated with higher dropout rates (22). As reported in the present study, the great majority of trials conducted in USA were placebo-controlled (76%), compared to the European trials (7.4%) (29).

Dropout rate was found to be inversely associated with the trial sample size; the smaller the trial, the better the clinicians can motivate the patients to complete the study (29). In addition, multi-arm and multi-centre trials were found to be associated with higher dropout rates; larger number of arms and larger number of centres may affect indirectly the dropout rate through an overall higher sample size (29).

Moreover, recently published trials were found to be associated with higher dropout rates; these trials tended to provide more detailed information about their dataset, the methods and the results (29). They were also more likely to be conducted by many centres and hence, they had randomized more patients compared to older trials. In the present study, a few published trials before 1990 were multi-centred (27.3%), while the great majority of multi-centred trials were published after 1990 (87%) (29). Therefore, the presence of higher drop-out rates in recently published trials may be explained by correlation of publication year with the number of recruiting centres. Smaller teams tend to run single centres who closely supervise the trial conduct and who may achieve better motivation than doctors who participate in multi-centre trials sponsored by industry (29).

Both analyses showed an association between trials treating inpatients and higher dropout rates. However, this finding was not convincing in the multiple regression analysis (29). Patients with more severe schizophrenic symptoms tend to be enrolled as inpatients in antipsychotic trials (29). These patients are always more difficult to

treat and tend to drop out more frequently than less severely ill patients (29). In addition, many of the inpatient studies seem to be conducted for drug registration purposes, i.e. on treatments not yet on the market (29). These studies are particularly interested in the safety of these treatments due to possible unknown side-effects and hence patients may need to stay in the hospital in the early stages (29). Moreover, in studies with non-marketed treatments, doctors tend to be very careful and withdraw patients from the studies more frequently than if drugs were already released on the market (29).

As shown by the multivariable regression analysis, trials with adequate and unclear allocation concealment were found to have similar dropout rate (29). On the other hand, univariable analysis showed that adequate allocation concealment was strongly associated with higher dropout rate per patient-week (29). The confounding of adequate allocation concealment with the recently published trials may be a plausible explanation for this correlation in univariable analysis where the recently published trials were found to be associated with higher dropout rate (29). Elder trials (published before 1990) are believed to be less informative about the adequacy of allocation concealment as compared to newer trials; in the present study only a few trials (17%) published before 1990 had adequate allocation concealment, while the majority of the trials (38%) with adequate allocation concealment was published after 1990 (29).

There are some limitations in the present study that must be taken into account before generalizing the findings. First, the findings can be generalized only for RCTs recruiting participants with diagnosis of schizophrenia or related disorders only (29). Second, both the dropout rate per patient-week and the likelihood that a patient will be LOCF or completely missing rather than completer were studied regarding the trial-level characteristics and not the treatment group-level characteristics (e.g., the randomly assigned treatments, their side effects and anticipated efficacy) (29). Finally, the baseline severity of the condition (as measured at the beginning of the study based on a specific symptom's scale) was not included in this study because there were different rating systems of PANSS and BPRS (0-6 versus 1-7) and any transformation from the one scale to the other was not straightforward because it was frequently not indicated which specific rating system was used (95).

It is particularly important from clinical and methodological perspective that the trialists take into account the study characteristics that are more likely to trigger premature discontinuation (29). The risk of biased findings may be minimized by preventing high dropout rates (29). This is of greater importance during the design of comparative trials that aim to properly assess the efficacy and safety of new generation antipsychotic drug for schizophrenia (29).

Chapter 4

Evaluating the impact of imputations for missing participant outcome data in a network meta-analysis

4.1 Introduction

This chapter presents an extension of the idea of informatively missing odds ratio (IMOR) as proposed by White et al (34) into a network meta-analysis (NMA) model in order to explore the impact of completely missing outcome data on the inferences about the relative effectiveness of several competing treatments studied in two recently published NMAs in the field of psychopharmacology (85;96). To achieve this, two datasets comparing anti-manic treatments and antidepressants were used in the estimation of IMORs. The outcome was response to treatments. In the original meta-analyses, completely missing participants were assumed to have failed to respond irrespective of the treatment they were randomized. The robustness of this assumption in each dataset was evaluated by assessing a variety of plausible scenarios regarding completely missing outcome data and their uncertainty was studied by incorporating an IMOR parameter in the NMA model. The consistency of the conclusions was evaluated by comparing the odds ratios for efficacy under the initial analysis and under several assumptions about the missingness. To infer on the missing data mechanism, the prior IMOR distribution was compared with the posterior. The Markov chain Monte Carlo (MCMC) in WinBUGS was implemented to fit the models.

4.2 Addressing completely missing outcome data in conventional meta-analysis and network meta-analysis

Randomized controlled trials (RCTs) are frequently under the risk of missing outcome data, because participants either terminate the study early or are excluded due to post-randomization protocol violations. Missing outcome data may compromise the validity of inferences from RCTs since they reduce the precision of the estimated treatment effects. Missing outcome data may lead to biased estimates, if ignored and hence, it is particularly important that missing outcome data are appropriately addressed when RCTs are included in meta-analyses, because informative missing outcome data may lead to biased meta-analysis estimates and increased heterogeneity (28;75).

Within a systematic review, missing outcome data frequently are handled using non-statistical approaches, such as, by interpreting the summary treatment effect using the risk of bias assessments in each trial (87). Several statistical methods to address missing outcome data are also available (5;32;34;61;63;64). In general, it is important to consider a variety of different scenarios regarding the missingness mechanism (see, Chapter 1).

Missing outcome data occur very frequently in psychiatric trials (4). Treatment failure is a common assumption considered by most meta-analyses regarding the outcome of the missing participants. To explore the robustness of conclusions, a series of sensitivity analyses are employed under various scenarios. However, most of these scenarios are implemented by imputing the outcomes in missing participants, ignoring completely the uncertainty in these imputations.

White et al suggested the incorporation of an unknown parameter in meta-analysis model to quantify the deviance between MAR and informative missingness in each arm and trial (5;34). As a result, this model allows for uncertainty due to imputations and ‘adjusts’ the meta-analysis effect size according to specific assumptions on the outcome of missing participants. This parameter refers to binary outcomes and is called ‘informative missing odds ratio’ (IMOR). The IMOR describes the relationship between the occurrence of the outcome among missing participants and the occurrence outcome among the observed participants. So far only a few meta-analyses have applied this elegant approach.

Network meta-analysis (NMA) is an extension to meta-analysis that compare simultaneously many different treatments (97;98). The extension of the IMOR model into a NMA might provide particularly large evidence about the impact of various scenarios regarding the missing outcome data on the relative effectiveness of the interventions. Additionally, this extended model gives the opportunity to explore associations between treatments and informative missingness by implementing a series of models for the similarity of IMORs across arms and trials.

4.3 Methodology of statistical analysis

4.3.1 Datasets: anti-manic and antidepressant interventions

The data applied in this study were from two published NMAs; a dataset of 47 RCTs that compared 12 anti-manic treatments with another active treatment or placebo and a dataset of 109 RCTs that compared 12 second-generation antidepressants with each other (85;96). Placebo was the reference treatment in the first network, while fluoxetine was the reference treatment in the second. In the present analysis, both datasets included only studies that reported a binary efficacy outcome defined as at least a 50% reduction on the severity score between baseline and endpoint on a standardized rating scale. Figures 4.3.1 and 4.3.2 illustrate these two networks (85;96).

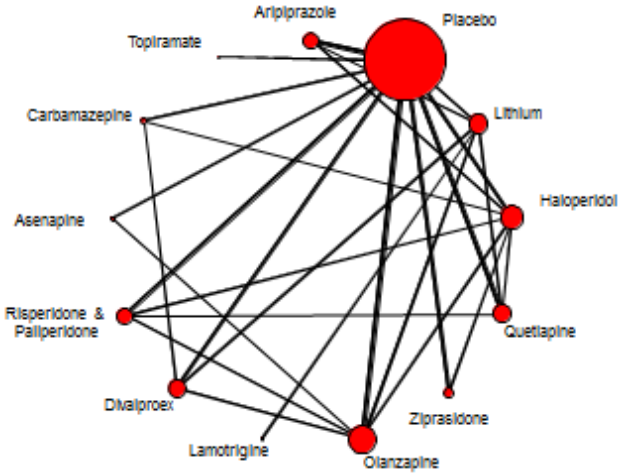

Figure 4.3.1. The network of antimanic treatments for the reduction of acute mania symptoms. The width of the lines reflects the number of trials that compared every pair of treatments, and the size of every node reflects the number of randomised participants.

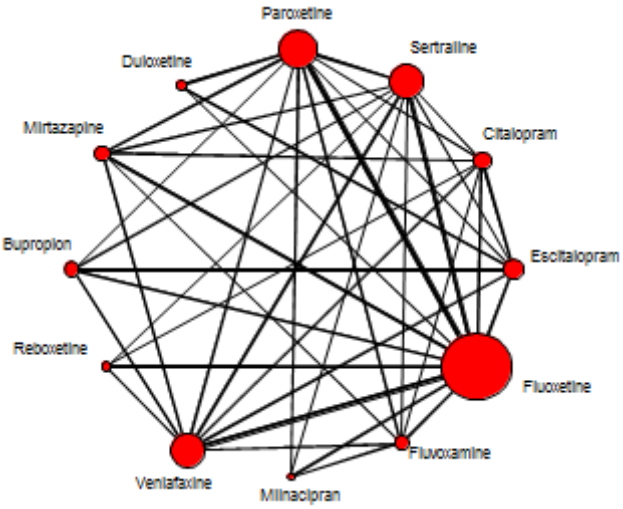

Figure 4.3.2. Network of antidepressant treatments for the reduction of major depression symptoms.

It is important to mention that the LOCF strategy was implemented in most trials in both datasets for patients who left the study early but provided at least one outcome. Participants who did not provide any outcome data were not included in the LOCF analyses. These participants are known as *completely missing participants* and they are the target of this study. Note that completely missing participants and LOCF imputed outcomes are usually labelled as missing outcome data. Initially, all completely missing participants in both datasets were assumed to have failed to experience improvement in their symptoms.

4.3.2 Incorporating completely missing outcome data as failures in the NMA model

Consider the total number of studies with *ns* and each study has na_i arms, where $i = 1, 2, \ldots, ns$. With $f_{i,k}$ denote the number of observed failures in arm k of study i, with $r_{i,k}$ the number of observed successes, with $m_{i,k}$ the number of completely missing outcome data and with $n_{i,k}$ the randomized sample size. Consider, first, that all completely missing participants failed. In each treatment and each trial, the number of successes is sampled from a binomial likelihood

$$r_{i,k} \sim Bin\left(\pi_{i,k}, n_{i,k}\right) \quad i = 1, 2, \ldots, ns \text{ and } k = 1, 2, \ldots, na_i$$

with $n_{i,k} = r_{i,k} + f_{i,k} + m_{i,k}$

The parameter $\pi_{i,k}$ is the probability of success in arm *k* of study *i*. The probabilities of success in two arms are indicated through the odds ratio

$$logit\left(\pi_{i,k}\right) = \begin{cases} u_i, & \text{if } k = 1 \left(\text{control treatment}\right) \\ u_i + \theta_{i,k}, & \text{if } k \geq 2 \left(\text{experimental treatment}\right) \end{cases} \qquad (4.3.1)$$

where $u_i = logit(\pi_{i,k})$ is the logarithm (log) of the odds of success in the control group. The parameter $\theta_{i,k}$ is the log odds ratio (OR) of success between experimental group *k* and the control group in study *i*. Under the random effects assumption, $\theta_{i,k}$ follows a normal distribution $N\left(\mu_{k,i}, \tau^2\right)$ with $\mu_{k,i}$ the mean log OR for treatment in arm *k* versus the treatment in the reference arm in study i and τ^2 the heterogeneity variance, which is assumed to be common for all pairwise comparisons.

For two random treatments *X* and *Y*, the parameters μ_{XY} (i.e. $\mu_{k,i} := \mu_{XY}$ where study *i* compares two treatments, *X* and *Y*) are linked through the *consistency equations*, that is, $\mu_{XY} = \mu_{RY} - \mu_{RY}$ with $\mu_{XX} = 0$ where *R* is an arbitrarily chosen reference treatment in the whole dataset (98;99). For each treatment, ranking probabilities can be estimated as well as cumulative ranking curves using the surface under the cumulative ranking (SUCRA) (97).

4.3.3 Including IMOR parameters in the NMA

The number of successes, the number of failures and the number of completely missing participants in a study *i* are mutually exclusive and can be described in a vector $y_{i,k} = (r_{i,k}, f_{i,k}, m_{i,k})$, which follows a multinomial likelihood

$$y_{i,k} = \left(r_{i,k}, f_{i,k}, m_{i,k}\right) \sim Multi\left(n_{i,k}, p_{1,i,k}, p_{2,i,k}, p_{3,i,k}\right) \quad i = 1,2,\ldots,ns \text{ and } k = 1,2,\ldots,na_i$$

where

$$p_{1,i,k} = \left(1 - a_{1,i,k}\right)\pi_{i,k}$$

$$p_{2,i,k} = \left(1 - a_{0,i,k}\right)\left(1 - \pi_{i,k}\right)$$

$$p_{3,i,k} = a_{0,i,k}\left(1 - \pi_{i,k}\right) + a_{1,i,k}\pi_{i,k}$$

is the probability of observed success, the probability of observed failure, and the probability of completely missing outcome data (either failures or successes), respectively. The parameter $\pi_{i,k}$ denotes the overall probability of success across observed and completely missing participants, whereas $a_{0,i,k}$ and $a_{1,i,k}$ imply the probability that a failure is completely missing and the probability that a success is completely missing, respectively (34). Then, the probabilities $\pi_{i,k}$ are parameterized as in equation (4.3.1) to obtain the log ORs.

Since the IMOR is the odds that a success is completely missing over the odds that a failure is completely missing, the log IMOR can be written as $\delta_{i,k} = logit(a_{1,i,k}) - logit(a_{0,i,k})$. Therefore, the probabilities of completely missing failure and success, $a_{0,i,k}$ and $a_{1,i,k}$, respectively, are related through the IMOR by definition. The probabilities $a_{0,i,k}$ and $a_{1,i,k}$ can be parameterized further with respect to $\delta_{i,k}$ and $\gamma_{i,k} = [logit(a_{0,i,k}) + logit(a_{1,i,k})]/2$ that represents the completely missingness on average across failures and successes

$$logit\left(a_{0,i,k}\right) = \gamma_{i,k} - 0.5 \times \delta_{i,k}$$

$$logit\left(a_{1,i,k}\right) = \gamma_{i,k} + 0.5 \times \delta_{i,k}$$

Various missing mechanisms across trials and treatments can be investigated with different models by modelling the parameters $\delta_{i,k}$. A MAR assumption is indicated by zero $\delta_{i,k}$ values, that is, the odds of missingness between successes and failures are equal. Negative $\delta_{i,k}$ values indicate that the odds of completely missing a success are less than the odds of completely missing a failure. By setting $\delta_{i,k} \neq 0$ or by determining some uncertainty in their common distribution, one can model any deviations from MAR. Extreme assumptions, such as all completely missing values are successes or failures, can be reflected by using very large positive or negative $\delta_{i,k}$ values, respectively.

4.3.4 Assumptions about IMORs

Specific fixed values can be assigned to the IMOR parameters, or some assumptions can be made for their similarity across trials and treatments (34). Two approaches might be implemented; the first assigns fixed $\delta_{i,k}$ values that reflect a specific data imputation process, while the second approach employs distributions where the mean indicates

specific assumption about the similarities of the parameters $\delta_{i,k}$ across trials and treatments and the variance reflects the uncertainty about the imputation assumption.

4.3.5 First approach: fixed IMOR models

Several fixed IMOR models can be employed that reflect different assumptions for the completely missing outcomes (7):

1. $\delta_{i,k} = 0$ indicates the MAR model, that is, the odds of unobserved success are the same to the odds of unobserved failure in any arm and study;
2. $\delta_{i,k} = -50$ indicates that all missing failures (AMF) model;
3. $\delta_{i,k} = 50$ indicates that all missing successes (AMS) model;
4. $\delta_{i,k} = -50$ for the placebo arm and $\delta_{i,k} = 50$ for the active treatments indicate the best-case (BC) model;
5. $\delta_{i,k} = 50$ for the placebo arm and $\delta_{i,k} = -50$ for the active treatments indicate the worst-case (WC) model.

Note that all these models above are equivalent to imputing the completely missing values $m_{i,k}$ as failures or successes (according to the scenario) before analysing the data.

4.3.6 Second approach: common IMOR and treatment-specific IMOR models

In the first approach, $\delta_{i,k}$ is set fixed to a specific value and hence, completely missing values are imputed without uncertainty. A better alternative would be to assign a prior distribution to $\delta_{i,k}$ that reflects both the assumption on the completely missing outcome data and the uncertainty in the imputations. For instance, one might believe that MAR holds 'on average' and fit a distribution $\delta_{i,k} \sim N(0, \sigma^2)$ where the mean reflects the MAR assumption and the variance σ^2 implies an uncertainty about the MAR. To simplify the IMOR model, the log IMORs can be assumed to be equal across arms and/or studies. For a more detailed description of the several possibilities to define the log IMORs, see White *et al*. (34). Focus of the present study is the common IMOR (on average MAR) model and several different treatment-specific IMOR models (7).

In order to apply a common IMOR (on average MAR) model, all IMORs are assumed to be equal, that is, $\delta_{i,k} = \delta^{all}$ and a prior normal distribution is placed on the parameter with mean $\Delta = 0$ and standard deviation $\sigma = 0.5$ (7). According to this model, there is considerable uncertainty in whether MAR is a plausible assumption, because the odds of completely missing a success range from 166% higher to 62% lower than the odds of completely missing a failure. The parameter $\delta_{i,k}$ actually measures the departure from MAR.

Within the treatment-specific IMOR model, the association between specific treatments and missingness mechanism can be also investigated. For instance, it is plausible to assume that participants assigned an active treatment may dropout

because of intolerable side-effects, whereas participant assigned to placebo or to an ineffective treatment may dropout because of lack of efficacy. In the treatment-specific IMOR model, the parameter $\delta_{i,k}$ is assumed to differ only across treatments, that is $\delta_{i,k} = \delta_{t_{i,k}}$, where $t_{i,k}$ is the treatment in the arm k in study i. Therefore, each treatment has a different parameter $\delta_{i,k}$, but a common normal prior distribution is assigned to them with mean Δ and standard deviation σ.

The mean Δ reflects the assumptions about the missingness mechanism, whereas the standard deviation σ reflects the uncertainty in these assumptions. A standard deviation equal to 0.5 may imply a rather large uncertainty in the imputations (7). Several assumptions can be considered to investigate the impact of completely missing outcome data on the relative treatment estimates:

1. $\Delta = 0$ implies that a success is as likely to be completely missing as a failure. This assumption coincides with the on average MAR assumption;
2. $\Delta = log(1/\lambda)$ with $\lambda > 1$ implies that it is more likely a failure to be completely missing than a success, that is, more completely missing failures (MMF);
3. $\Delta = log(\lambda)$ with $\lambda > 1$ implies that it is more likely a success to be completely missing than a failures, that is, more completely missing successes (MMS);
4. $\Delta = log(\lambda)$ for the active treatments and $\Delta = log(1/\lambda)$ for the placebo with $\lambda > 1$ imply that it is more likely a failure to be completely missing in the placebo arm, while it is more likely a success to be completely missing in the active treatment, that is, more failures in the placebo (MFP);
5. $\Delta = log(1/\lambda)$ for the active treatments and $\Delta = log(\lambda)$ for the placebo with $\lambda > 1$ imply that it is more likely a success to be completely missing in the placebo arm, while it is more likely a failure to be completely missing in the active treatment, that is, more successes in the placebo (MSP).

Clinical opinion is needed in order to specify the parameter λ. For the present study λ was set equal to 2 which indicates that the odds of completely missing a success are twice the odds of completely missing a failure under the MMS model (7).

4.3.7 Model estimation

All models were fit within the Bayesian framework using the WinBUGS software (93). For all the analyses, 100,000 updates were used and a burn-in of the 10,000 initial samples (7). For the parameters u_i and μ_{RX}, non-informative normal prior distributions were used centred at zero with a large variance equal to 10,000 (7). For the parameters $\gamma_{i,k}$ and $\delta_{i,k}$ normal distribution $N(0,1000)$ and $N(0,0.25)$ were chosen, respectively (7). For the square root of heterogeneity, half normal distribution $N(0,1)I(0,\infty)$ was used (7). Autocorrelation was reduced by thinning to every 40th value. The WinBUGS code is provided in the Appendix C.

4.4 Results

Tables 4.4.1 and 4.4.2 present the treatments, number of trials, the number of randomized patients, and the completely missing rate for the network of anti-manic and antidepressant treatments, respectively (7).

Table 4.4.1

Information on the number of trials, randomized sample and missingness summarized for antimanic treatment arms and placebo

Treatments	Total number of trials	Total number of patients	
		Completely missing (%)	Randomized
Aripiprazole	7	2	1284
Placebo	36	5	4102
Lithium	8	1	490
Haloperidol	8	3	1000
Quetiapine	7	3	832
Ziprasidone	5	5	1018
Olanzapine	13	7	1636
Lamotrigine	1	0	15
Divalproex	8	4	706
Risperidone & Paliperidone	7	7	1215
Asenapine	1	3	194
Carbamazepine	3	4	246
Topiramate	1	5	143

Table 4.4.2

Information on the number of trials, randomized sample and missingness summarized for antidepressant treatments arms

Treatments	Total number of trials	Total number of patients	
		Completely missing (%)	Randomized
Paroxetine	28	4	2989
Sertraline	25	5	5171
Citalopram	14	4	779
Escitalopram	16	2	569
Fluoxetine	53	6	2910
Fluvoxamine	11	12	689
Milnacipran	6	4	1623
Venlafaxine	27	3	1299
Reboxetine	8	3	1368
Bupropion	12	3	2289
Mirtazapine	12	5	1939
Duloxetine	8	3	2504

In the anti-manic network, the completely missing rate ranged from 0% to 7% and from 2% to 12% for the antidepressant network.

Figures 4.4.1 and 4.4.2 illustrate the relative effectiveness (measured in OR) of each anti-manic treatment versus placebo and each antidepressant versus fluoxetine under several different IMOR models (7). As shown in figure 4.4.1, only the two extreme scenarios of WC and BC seem to have a considerable impact on the relative effectiveness of the anti-manic treatments where their relative effectiveness drops and increases, respectively.

Figure 4.4.3 presents the heterogeneity under the various IMOR models (7). Large heterogeneity indicates increased variability between study effects and this is strong evidence against the employed scenario for the completely missing outcome data. In both datasets, WC and BC scenarios seem to increase the heterogeneity and hence, these assumptions might be too extreme to consider across all studies. Specifically, under the BC and WC scenarios, heterogeneity increases from 0.29 in MAR to 0.35 and 0.45, respectively, in anti-manic trials and from 0.08 in MAR to 0.22 and 0.17, respectively, in antidepressant trials.

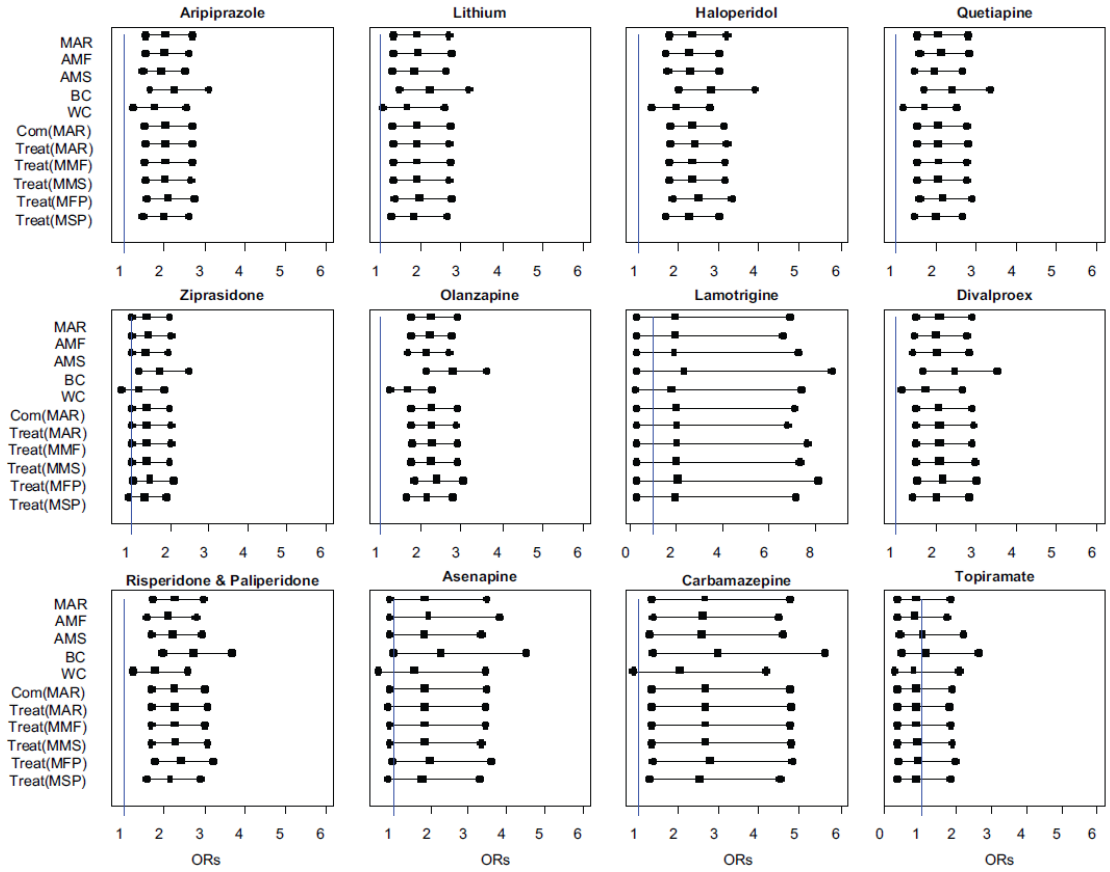

Figure 4.4.1 Efficacy of anti-manic treatments against placebo measured as OR with 95% posterior credible intervals for fixed IMOR (MAR, AMF, AMS, BC, and WC), common IMOR (on average MAR), treatment-specific IMOR (on average MAR), and several variations of the treatment-specific IMOR model (MMF, MMS, MFP, and MSP).

OR: odds ratio; IMOR: informative missing odds ratio; MAR: missing at random; AMF: all missing failures; AMS: all missing successes; BC: best case; WC: worst case; Com(MAR): common informative missing odds ratio (on average missing at random); Treat(MAR): treatment-specific

informative missing odds ratio (on average missing at random); Treat(MMF): treatment-specific informative missing odds ratio (more missing failures); Treat(MMS): treatment-specific informative missing odds ratio (more missing successes); Treat(MFP): treatment-specific informative missing odds ratio (more failures in placebo); Treat(MSP): treatment-specific informative missing odds ratio (more successes in placebo).

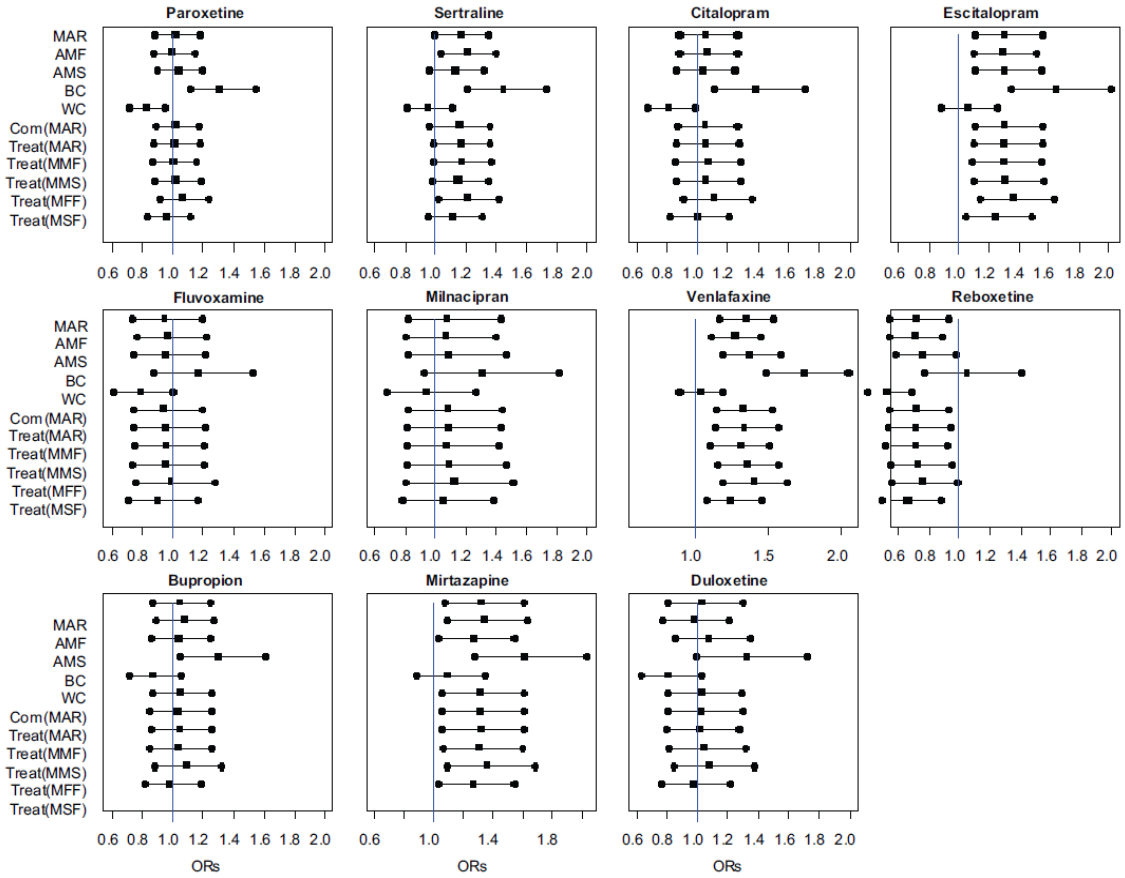

Figure 4.4.2 Efficacy of antidepressant treatments against fluoxetine measured as OR with 95% posterior credible intervals for fixed IMOR models (MAR, AMF, AMS, BC, and WC), common IMOR (on average MAR), treatment-specific IMOR (on average MAR), and several variations of the treatment-specific IMOR.

OR: odds ratio; IMOR: informative missing odds ratio; MAR: missing at random; AMF: all missing failures; AMS: all missing successes; BC: best case; WC: worst case; Com(MAR): common informative missing odds ratio (on average missing at random); Treat(MAR): treatment-specific informative missing odds ratio (on average missing at random); Treat(MMF): treatment-specific informative missing odds ratio (more missing failures); Treat(MMS): treatment-specific informative missing odds ratio (more missing successes); Treat(MFF): treatment-specific informative missing odds ratio (more failures in fluoxetine); Treat(MSF): treatment-specific informative missing odds ratio (more successes in fluoxetine); Treat(MFP): treatment-specific informative missing odds ratio (more failures in placebo); Treat(MSP): treatment-specific informative missing odds ratio (more successes in placebo)

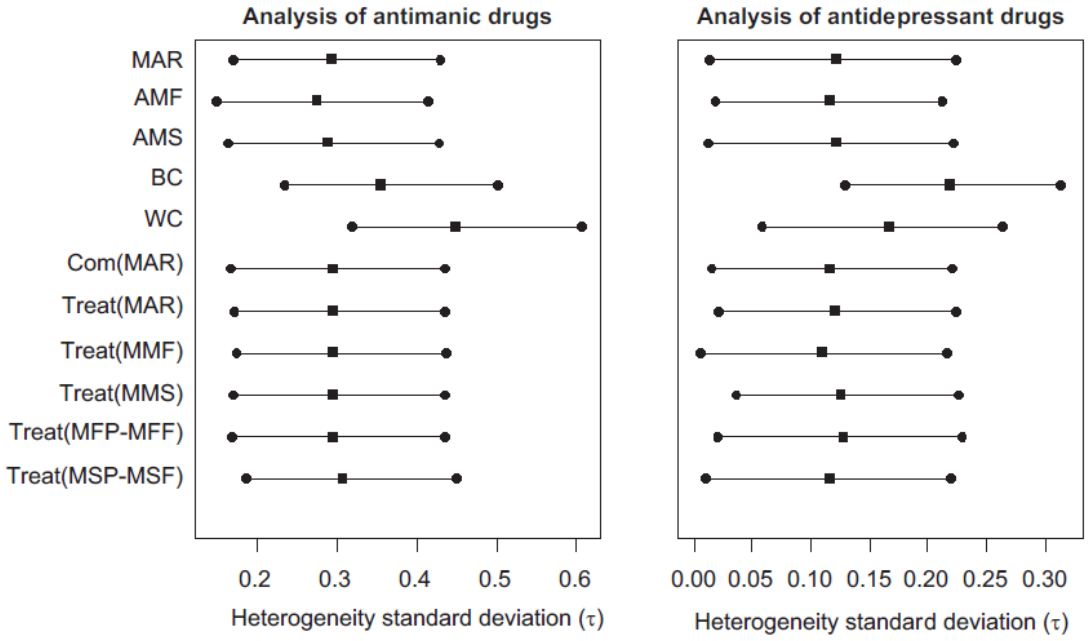

Figure 4.4.3 Common heterogeneity standard deviation (τ) of log OR with 95% posterior credible intervals for fixed IMOR models (MAR, AMF, AMS, BC, and WC), common IMOR (on average MAR), treatment-specific IMOR (on average MAR), and several variations of the treatment-specific IMOR model (MMF, MMS, MFP, MSP, MFF, and MSF) in both networks.

IMOR: informative missing odds ratio; MAR: missing at random; AMF: all missing failures; AMS: all missing successes; BC: best case; WC: worst case; Com(MAR): common informative missing odds ratio (on average missing at random); Treat(MAR): treatment-specific informative missing odds ratio (on average missing at random); Treat(MMF): treatment-specific informative missing odds ratio (more missing failures); Treat(MMS): treatment-specific informative missing odds ratio (more missing successes); Treat(MFP): treatment-specific informative missing odds ratio (more failures in placebo); Treat(MSP): treatment-specific informative missing odds ratio (more successes in placebo); Treat(MFF): treatment-specific informative missing odds ratio (more failures in fluoxetine); Treat(MSF): treatment-specific informative missing odds ratio (more successes in fluoxetine).

Overall, the results from the AMF scenario (the original analysis) seem to be robust to the several IMOR assumptions. As a result, the mean SUCRA values of the treatments as well as the rankings have small changes in both networks (see, table 4.4.3 and table 4.4.4) (7). For instance, the anti-manic treatment lithium ranks fifth under MAR and AMS models but seventh under AMF model, and the antidepressant treatment mirtazapine ranks second under MAR model but first and third under the AMF and AMS models, respectively.

Table 4.4.3

Mean SUCRA values for each antimanic treatment across various IMOR scenarios

Treatment[a]	Fixed IMORs					Common (on average MAR)	Treatment-specific IMORs				
	MAR	AMF	AMS	BC	WC		On average MAR	MMF	MMS	MFP	MSP
Carbamazepine	0.78	0.78	0.78	0.72	0.68	0.78	0.78	0.77	0.78	0.78	0.78
Haloperidol	0.77	0.73	0.78	0.77	0.74	0.77	0.77	0.76	0.77	0.77	0.76
Olanzapine	0.72	0.71	0.68	0.76	0.58	0.72	0.71	0.72	0.71	0.72	0.70
Risperidone & Paliperidone	0.72	0.65	0.74	0.73	0.66	0.72	0.73	0.71	0.73	0.73	0.70
Quetiapine	0.59	0.64	0.57	0.58	0.61	0.59	0.59	0.59	0.58	0.59	0.59
Aripiprazole	0.57	0.57	0.55	0.51	0.65	0.57	0.58	0.57	0.56	0.55	0.59
Divalproex	0.60	0.57	0.61	0.61	0.62	0.60	0.60	0.60	0.60	0.60	0.60
Lithium	0.48	0.51	0.48	0.47	0.55	0.48	0.48	0.48	0.48	0.47	0.49
Asenapine	0.47	0.52	0.47	0.49	0.46	0.47	0.46	0.46	0.47	0.47	0.46
Lamotrigine	0.39	0.40	0.37	0.40	0.41	0.38	0.38	0.40	0.39	0.40	0.38
Ziprasidone	0.25	0.26	0.25	0.28	0.27	0.25	0.25	0.25	0.25	0.25	0.25
Placebo	0.09	0.09	0.08	0.06	0.14	0.09	0.09	0.09	0.09	0.08	0.10
Topiramate	0.08	0.08	0.08	0.06	0.13	0.07	0.13	0.12	0.07	0.09	0.08

Note: SUCRA: surface under the curve ranking; IMOR: informative missing odds ratio; MAR: missing at random; AMF: all missing failures; AMS: all missing successes; BC: best case; WC: worst case; MMF: more missing failures; MMS: more missing successes; MFP: more failures in placebo; MSP: more successes in placebo.

[a]Treatments have been ordered according to their ranking in the All Missing Failures analysis.

Table 4.4.4

Mean SUCRA values for each antidepressant treatment across various IMOR scenarios

Treatment[a]	Fixed IMORs					Common (on average MAR)	Treatment-specific IMORs				
	MAR	AMF	AMS	BC	WC		On average MAR	MMF	MMS	MFF	MSF
Mirtazapine	0.87	0.90	0.82	0.81	0.88	0.87	0.87	0.88	0.86	0.86	0.88
Escitalopram	0.87	0.87	0.88	0.86	0.84	0.88	0.87	0.88	0.87	0.87	0.88
Venlafaxine	0.91	0.84	0.94	0.93	0.80	0.90	0.90	0.88	0.91	0.91	0.87
Sertraline	0.67	0.75	0.63	0.64	0.62	0.66	0.67	0.69	0.65	0.66	0.68
Milnacipran	0.48	0.48	0.51	0.45	0.57	0.51	0.51	0.49	0.51	0.50	0.54
Bupropion	0.42	0.48	0.40	0.42	0.41	0.42	0.39	0.42	0.39	0.43	0.39
Citalopram	0.45	0.46	0.39	0.54	0.30	0.43	0.45	0.48	0.45	0.47	0.43
Fluoxetine	0.32	0.33	0.30	0.06	0.74	0.32	0.33	0.32	0.32	0.24	0.43
Paroxetine	0.36	0.31	0.39	0.44	0.31	0.37	0.36	0.33	0.36	0.38	0.32
Duloxetine	0.39	0.29	0.48	0.45	0.29	0.39	0.39	0.37	0.42	0.41	0.35
Fluvoxamine	0.24	0.28	0.25	0.25	0.25	0.24	0.25	0.25	0.24	0.26	0.23
Reboxetine	0.01	0.01	0.02	0.13	0.01	0.01	0.01	0.01	0.02	0.02	0.01

Note: SUCRA: surface under the curve ranking; IMOR: informative missing odds ratio; MAR: missing at random; AMF: all missing failures; AMS: all missing successes; BC: best case; WC: worst case; MMF: more missing failures; MMS: more missing successes; MFF: more failures in fluoxetine; MSF: more successes in fluoxetine.

[a]Treatments have been ordered according to their ranking in the All Missing Failures analysis.

Figure 4.4.4 presents the estimated means and 95% credible intervals for log IMORs $\delta_{t_{i,k}}$ and δ^{all}(7). Mean log IMORs away from zero indicate deviation from the MAR assumption. The interval between the two vertical lines represents the prior information about the log IMORs. As shown in figure 4.4.4, most of the posterior intervals overlap with the prior, and this is a strong indication that the data provide little about whether the MAR assumption actually holds in both networks.

Specifically, in the network of anti-manic drugs, there is a very small indication that failures in placebo are more likely to be missing than successes (see, figure 4.4.4(a)) and this can be justified by its posterior interval which has moved slightly towards the negative values. The same conclusion might be drawn for paroxetine in the antidepressants network (see, figure 4.4.4(b)). Lack of efficacy in placebo and paroxetine may be a plausible reason why patients fail to respond to these treatments and tend to be missing. In contrast, successes in sertraline, fluoxetine, and fluvoxamine are more likely to be missing compared to failures because the corresponding posterior intervals are slightly moved towards the positive values.

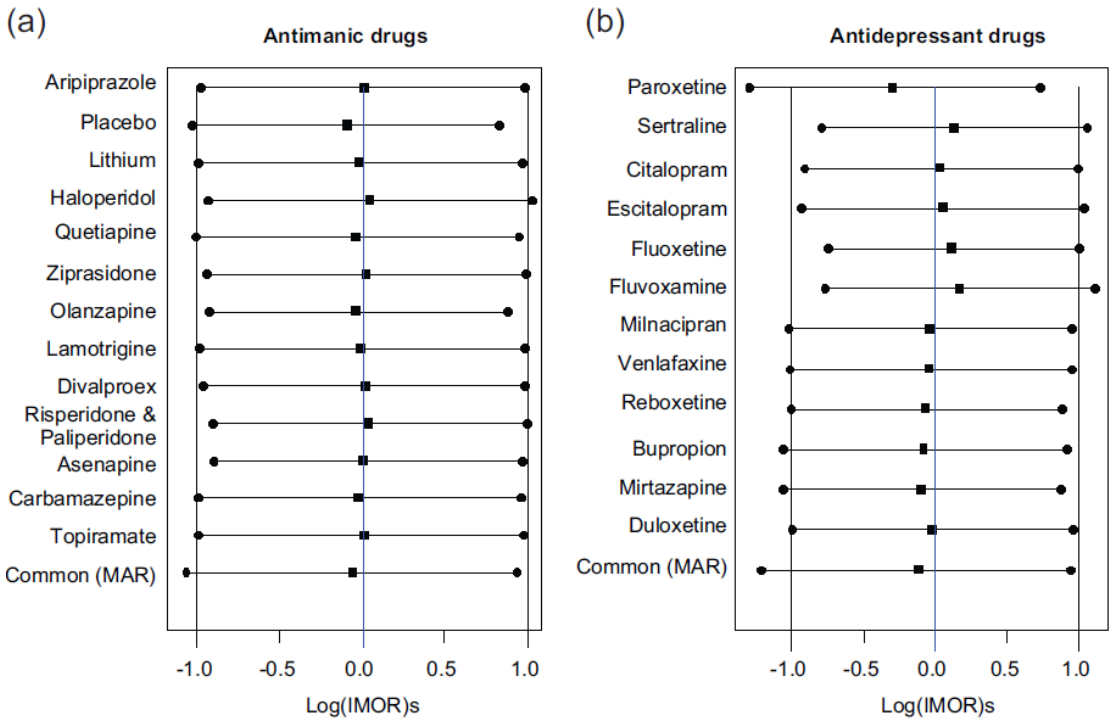

Figure 4.4.4 Mean log IMORs with 95% posterior intervals for treatment-specific (on average MAR) and common (on average MAR) IMOR models in anti-manic and antidepressant networks. The black vertical lines are the boundaries of the 95% prior interval.

IMOR: informative missing odds ratio; MAR: missing at random; Com(MAR): common informative missing odds ratio (on average missing at random).

As an ad hoc sensitivity analysis, the network of antimanic treatments was re-analysed assuming $\Delta = log(1/2)$ for paroxetine, $\Delta = log(2)$ for sertraline, fluoxetine, and fluvoxamine and $\Delta = 0$ for the rest of the treatments (7). Figure 4.4.5 presents the results from this ad hoc analysis through cumulative ranking curves along with the curves from the initial analysis (AMF scenario) (7). The SUCRA values and the ranking change for most treatments.

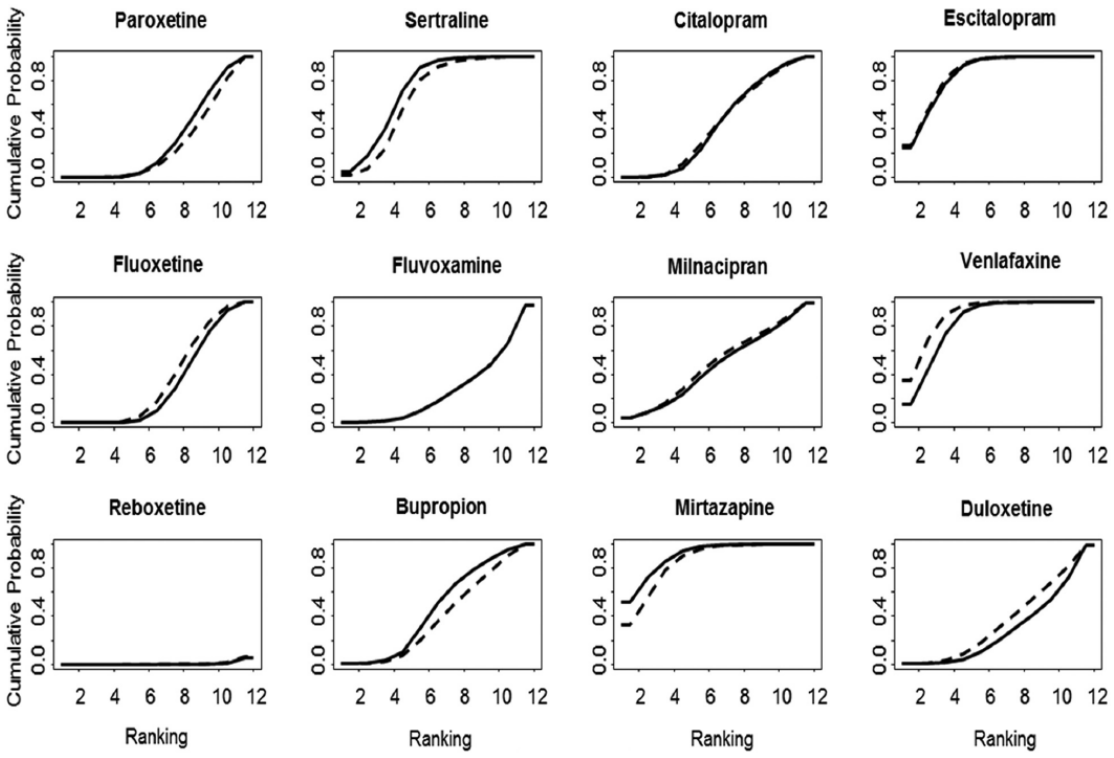

Figure 4.4.5 Cumulative ranking probability plots for the antidepressant treatments. Solid lines represent the AMF analysis, while dashed lines the ad hoc sensitivity analysis. The ranks are on the horizontal axis, and the cumulative probability of each treatment being the best option, among the best two options, among the best three options, and so on, is on the vertical axis. The surface under the cumulative ranking curve of each antimanic treatment reflects the overall ranking of the treatment.

AMF: all missing failures.

4.5 Discussion of the findings

Protocols for meta-analyses should include a pre-specified strategy to handle completely missing outcome data in the analyses. Different amount of and reasons for completely missing outcome data across arms may be a red alarm for possible significant changes in the relative effectiveness and ranking of treatments. In this case, transparency in the assumptions and the methodology of handling completely missing outcome data is necessary. Any assumption on the completely missing outcome data shall be supported by clinical judgement in order to employ the most plausible assumption in the primary analysis. A sensitivity analysis may include various other assumptions and the results may be compared through cumulative ranking probability plots. SUCRAs are a valuable tool in order to infer on the ranking of the treatments, but they cannot provide any information on whether differences between the ranking of the treatments are clinically important, and hence, they should always be presented along with the effect measures and their credible intervals.

This study presented a very flexible model to implement several different assumptions about the distribution of IMORs (i.e., how informative is the missingness across trials and arms) and to investigate their impact on the relative effectiveness of the compared treatments (7). The treatment-specific IMOR models received particular attention by the present study since they allow the degree of informative missingness to be different across the treatments and hence, they were examined in detail. For instance, it may be reasonable to assume that participants in aripiprazole leave the study due to intolerable adverse events despite its efficacy (and hence a prior IMOR more than 1 may be a plausible choice), whereas placebo participants leave the study due to lack of efficacy (and hence a prior IMOR less than 1 may be plausible). In both networks there was little information regarding the informative missingness across the treatments, possibly due to low missing outcome rate (7).

Since the completely missing outcome data are always present in psychiatric populations, the low completely missing outcome rates in both networks of the present study may be explained by two factors. First, the application of intention-to-treat (ITT) analysis (e.g., using LOCF) implies that participants with at least one post-randomization outcome measurement were included in the dataset and treated as observed. The ITT approach has been widely a preferred analysis strategy, because it preserves the benefits of randomization by analysing all participants assigned to the treatment they were randomized irrespectively of their compliance and discontinuation. However, the more recent guidelines have criticized the use of LOCF (50-52). Second, most of the antidepressant studies had a placebo run-in in order to select or exclude patients in a clinical trial before randomization. Even though the placebo run-in phase may increase internal validity and compliance rate, it may affect generalizability of the findings (100).

It is a really challenge to detect the missing data mechanism and any judgement is based only on assumptions. In meta-analysis level, researchers analyse the data based on untestable assumptions. Plausible assumptions require clinical opinion and some clinical research fields are more likely to deal with informative missingness than others. In addition, the reasons for missingness may be different between treatments, and hence, the risk of attrition bias may be considerably high (87). Meta-analyses should not ignore the risk of meta-confounding between the amount of missing outcome data in trials and the corresponding estimated treatment effect (5).

The present study has several limitations in the application of the findings. First, the differences in attrition were investigated only between treatments (7). Investigating the degree of informative missingness across trials and arms (i.e., trial-specific and arm-specific IMOR models) may supplement the results obtained by the treatment-specific IMOR models. Second, the present analysis has applied different prior assumptions to represent the beliefs about the completely missing outcome data (7). Although, non-informative priors could be implemented, they would not be useful because the data provide already very little information about the IMOR parameters. Third, no sensitivity analysis has been performed to justify the selection of prior distributions for the IMOR and particularly the choice of heterogeneity of the IMORs (7). Different prior distributions may provide different results, and hence, it is

necessary to perform sensitivity analysis to measure the robustness of the results regarding the selection of prior distributions of the IMORs. Consistency between the results of the primary analysis and those of the sensitivity analysis is a strong indication of robustness (101). Last, the IMOR approach refers to dichotomous outcome measures. Models to handle continuous missing outcomes need to be developed (7).

The most important limitation of the present study is that the IMOR model addresses only part of the completely missing outcome data problem in psychiatric trials. In both datasets, the completely missing outcome rate was very low in all treatments, because the LOCF imputed outcomes were treated as fully observed (7). However, it was impossible to treat the LOCF imputed outcomes as unobserved since the trials rarely reported the numbers of successes and failures for completers alone. Further extension of the IMOR model to account for the LOCF imputed outcomes is needed to fully study the robustness of the meta-analysis findings (7).

In conclusion, models that quantify deviation from MAR in order to address attrition in RCTs are available and can be applied not only in a conventional meta-analysis but also in a NMA (7). In the field of psychiatry, where dropout rates are high and missingness is informative, it is important to investigate the extent to which informative completely missing outcome data may have distorted the conclusions of RCTs and meta-analyses. The IMOR approach successfully allows the investigation of all possible assumptions for the outcome of the completely missing participants while taking into account the uncertainty of each assumption. From a clinical and meta-analysis perspective, it is important to use plausible prior distributions about these scenarios as suggested from experts of the field.

Chapter 5

Conclusions

The results of the systematic review of Cochrane systematic reviews and meta-analyses on the reporting and handling of missing outcome data in mental health (Chapter 2) showed that systematic reviews do not often provide a clear definition of terms that relate to attrition and they rarely distinguish between LOCF imputed data and completely missing outcome data. The majority of the systematic reviews tend to handle missing outcome data either as a dropout outcome or by applying an imputation strategy in the primary analysis, but without implementing a sensitivity analysis to investigate the appropriateness of their imputation strategy. Moreover, most of the systematic reviews do not report any possible implications of the missing outcome data in their results. These findings indicate that the description of missing outcome data, as well as the reporting of the techniques for handling missing outcome data, and the inclusion of the implications of missing outcome data in the findings of the systematic reviews is insufficient.

The empirical evaluation of attrition in antipsychotic trials (Chapter 3) revealed that trials using placebo as control, trials more recently published, trials conducted in the USA and multi-centre trials were significant 'prognostic factors' for a higher dropout rate. Furthermore, this empirical evaluation showed that a patient was more likely to drop out during the course of the study (and hence be included in an LOCF analysis), rather than complete the trial, when the trial had adequate allocation concealment, had a single-blind design, used a placebo as control, had higher precision, had at least three arms, was conducted in the USA or Europe, was multi-centred, treated inpatients and had a shorter duration. By contrast, the likelihood that a patient would drop out prior to any outcome data collection (and hence be excluded from all analyses), rather than complete the trial, was found to be higher in trials with double-blind design, using placebo as control, with lower precision, with smaller sample size, randomizing patients in the USA, with a single-centre and treating mixed patients. These results were in line with previous findings that reported differences between completers and LOCF or excluded patients, however our analyses explored and assessed a wider range of clinical and methodological variables, using a more sophisticated statistical approach (88).

In order to evaluate the impact of imputations for missing participant outcome data in a NMA (Chapter 4), the conventional meta-analysis model was extended to a NMA model that quantifies the relative effectiveness of multiple interventions and enables ranking of all treatments in the presence of attrition and reflects the uncertainty about imputing unobserved outcomes. Sensitivity analysis via assumptions about the IMOR parameter can contribute to the reliability of the inferences from meta-analysis and about the impact of attrition in meta-analysis. The performance of this model was evaluated on two published meta-analyses that compared antidepressants (96) and anti-manic drugs (85). In these examples involving 325 studies in total, the impact of missing outcome data was small and the relative effectiveness of the treatments was not affected by the different scenarios. There was only a small change in the ranking of most of the treatments. These can be partly explained by the low missing outcome data rate.

Summary in English

Systematic reviews in psychiatry have reported that the dropout rate is considerable high in the studies of this field. Despite the high dropout out, the reporting on the extent of missing outcome data, the analytical strategies undertaken and the acknowledgment of the implication of missing outcome data on the results have been particularly insufficient. In order to provide empirical evidence about the methodology reported to address and to describe the extend of the missing outcome data as well as the acknowledgement of their impact in the findings of the systematic reviews, 190 systematic reviews in the mental health field published in the Cochrane Database of Systematic Reviews after 1/1/2009 by three Cochrane Review Groups (the Depression, Anxiety and Neurosis Group, the Developmental, Psychosocial and Learning Problems Group and the Schizophrenia Review Group) were considered.

Of those, 175 systematic reviews included at least one study with missing outcome data. The majority of these 175 systematic reviews accounted for missing outcome data by considering a relevant primary or secondary outcome (e.g. acceptability). One in three systematic reviews only reported missing outcome data implications and primarily in the discussion section by commenting on the amount of the missing outcome data. One hundred and forty eligible meta-analyses with missing outcome data were examined. At least half of them had studies with total dropout rate between 10% and 30%. Intention-to-treat analysis was reported in three in four meta-analyses by including trials that had already implemented an imputation strategy. Less than 20% of the meta-analyses had employed any sensitivity analysis to investigate the impact of the missing outcome data on the primary analysis results.

The findings of this review indicated that systematic reviews often provide an unclear definition of terms that relate to attrition and rarely do they distinguish between data imputed under the ‘last observation carried forward’ (LOCF) approach and completely missing outcome data. The majority of the systematic reviews tend to handle missing outcome data either as a dropout outcome or by applying an imputation strategy in the primary analysis, but without implementing a sensitivity analysis to investigate the appropriateness of this strategy. Most of the systematic reviews do not report any possible implications of the missing outcome data in their results. Consequently, the description of the amount and the extend of missing outcome data, as well as the reporting of the techniques for handling missing outcome data, and the inclusion of the implications of missing outcome data in the findings of the systematic reviews seem to be particularly suboptimal.

Another importance consideration in the missing outcome data issue is the minimization of their occurrence. Patient dropout is common in mental health trials and the dropout rate has been found to be range from small to considerable high. In order to take the necessary measures to prevent as much as possible the missing outcome data, it is important to investigate and to understand the reasons that the

patients drop out early from trials. Therefore, an extensive research was initiated to identify trial characteristics that have an impact on premature discontinuation in antipsychotic trials for schizophrenia.

The results of this research showed that trials with adequate allocation concealment, double blinding, that used placebo as control, had higher precision, larger trial size, at least three treatment arms, were recently published, were conducted in the USA and enrolled inpatients were associated with higher dropout rates. Similar factors were associated with whether a patient was more likely to provide at least one or no post-baseline measurement before dropping out. In contrary, blinding status did not predict whether a patient was more likely to provide at least one outcome measurement before dropping out, and allocation concealment, larger sample size, number of arms, recent publication and recruiting inpatient did not predict where a patient was more likely to be completely missing from the trial.

The findings of this research revealed that high dropout rates in antipsychotic trials can be associated with various characteristics, particularly with the use of placebo as control and the size of the study. From clinical and methodological perspective, trialists should always take into account during the design and the conduct of a study the study characteristics that are more prone to trigger premature discontinuation. Preventing high dropout rates may reduce the risk of biased findings. This becomes even more important when designing confirmatory trials for new generation antipsychotic treatments for schizophrenia.

Although, accounting for trials characteristic is important to prevent and minimize missing outcome data to some extent, addressing the already existent missing outcome data using an appropriate analytic strategy is necessary to reduce the risk of obtaining biased and imprecise results. There are several statistical methods to account for the impact of missing outcome data in the trial level. The LOCF approach is frequently employed within the trials and has been criticized for its unrealistic assumptions and conservative results.

In meta-analysis level, missing outcome data are handled by making assumptions on the outcomes of the missing participants. The majority of the meta-analyses in psychiatry consider the assumption that all missing participants have failed to respond to the assigned treatment. These assumptions involve imputation of the outcomes in missing participants, ignoring usually the uncertainty in the imputations. Therefore, a conventional meta-analysis model has been suggested to better quantify uncertainty in the imputations by incorporating in the model an unknown parameter that quantifies the departure between uninformative and informative missingness in each arm and trial. This parameter refers to binary missing outcome data and it has been called the ‘informative missingness odds ratio’ (IMOR), as it describes the relationship between the unknown occurrence of the outcome among missing participants and the known outcome in observed participants.

A further step was to extend the idea of IMOR into a network meta-analysis (NMA) setting in order to explore the impact of missing outcome data on the inferences about the relative effectiveness of several competing treatments in psychiatric trials. Two dataset comparing anti-manic treatments and antidepressants

were used. The outcome was response to treatments. In the original meta-analyses, missing participants were assumed to have failed regardless the treatment they were randomized to. The robustness of this assumption in each dataset was evaluated by considering several IMOR models that reflected different assumption on the missing outcome data. By comparing the odds ratios for efficacy under the original analysis and under several assumptions about the missingness, the consistency of the conclusions was assessed. The missing data mechanism was studied by comparing the prior with the posterior IMOR distribution. Markov chain Monte Carlo (MCMC) was employed to fit the models in the WinBUGS software.

In both datasets, the relative effectiveness of the treatments was affected only by the worst- and best-case analyses. Moreover, heterogeneity increased in both datasets under these two extreme scenarios. Overall, there were small changes on the ranking of the anti-manic and antidepressant treatments. The posterior and prior IMOR distributions were very similar implying that there was little information about the true outcome in missing participants. There was a very weak indication that missing participants tended to fail in placebo and paroxetine, while the opposite occurred for sertraline, fluoxetine, and fluvoxamine.

In the field of psychiatry, where dropout rates are high and missing is informative, it is important to investigate the extent to which informative missing outcome data may have distorted the conclusions of the randomized controlled trials and meta-analyses. The IMOR approach allows the investigation of all possible assumptions for the outcome of the missing participants accounting simultaneously for the uncertainty of each assumption. Within a NMA framework, the IMOR approach also offers the opportunity to explore interactions between treatments and informative missingness by employing a range of assumptions on the IMOR parameter across arms and trials. From a clinical and meta-analysis perspective, it is desirable to use plausible prior distributions about these assumptions, elicited expert opinion.

Περίληψη στα Ελληνικά

Σύμφωνα με τις συστηματικές ανασκοπήσεις στην ψυχιατρική, το ποσοστό ελλιπουσών τιμών είναι ιδιαίτερα υψηλό στις μελέτες αυτού του πεδίου. Παρά το υψηλό ποσοστό ελλιπουσών τιμών, η αναφορά στην έκταση των ελλιπουσών τιμών αποτελέσματος, οι αναλυτικές στρατηγικές αντιμετώπισης αυτών και η αναγνώριση των συνεπειών των ελλιπουσών τιμών αποτελέσματος στα ευρήματα της μελέτης είναι ιδιαίτερα ανεπαρκή. 190 συστηματικές ανασκοπήσεις στο πεδίο της ψυχικής υγείας δημοσιευμένες στην βάση δεδομένων της Cochrane μετά το 1/1/2009 σε τρεις ομάδες ανασκόπησης (η ομάδα για την Κατάθλιψη, το Άγχος και τη Νεύρωση, η ομάδα για τα Αναπτυξιακά, Ψυχοκοινωνικά και Μαθησιακά Προβλήματα και η ομάδα ανασκόπησης στη Σχιζοφρένεια) συλλέχθηκαν με σκοπό να παρέχουν εμπειρική ένδειξη για την μεθοδολογία σχετικά με την αντιμετώπιση και την περιγραφή της έκτασης των ελλιπουσών τιμών αποτελέσματος, καθώς επίσης, και την αναγνώριση της επίδρασής τους στα ευρήματα των συστηματικών ανασκοπήσεων.

Από τις 190 συστηματικές ανασκοπήσεις, οι 175 περιείχαν τουλάχιστον μία μελέτη με ελλιπούσες τιμές αποτελέσματος. Η πλειοψηφία των 175 συστηματικών ανασκοπήσεων έλαβε υπόψη τις ελλιπούσες τιμές αποτελέσματος στο πρωταρχικό ή το δευτερεύον αποτέλεσμα (π.χ., αποδεκτικότητα της θεραπείας). Μόλις μία στις τρεις συστηματικές ανασκοπήσεις ανέφερε τις συνέπειες των ελλιπουσών τιμών αποτελέσματος κυρίως στα κεφάλαιο ‘Αποτελέσματα’ σχολιάζοντας το μέγεθος των ελλιπουσών τιμών αποτελέσματος. Εκατόν σαράντα κατάλληλες μετα-αναλύσεις με ελλιπούσες τιμές αποτελέσματος μελετήθηκαν εις βάθος. Τουλάχιστον οι μισές από αυτές είχαν μελέτες με συνολικό ποσοστό ελλιπουσών τιμών αποτελέσματος μεταξύ 10% και 30%. Η ‘με-σκοπο-την-θεραπεία’ ανάλυση αναφέρθηκε σε τρεις στις τέσσερις μετα-αναλύσεις οι οποίες περιείχαν μελέτες που είχαν ήδη εφαρμόσει μία μέθοδο αντικατάστασης (των ελλιπουσών τιμών αποτελέσματος). Λιγότερο από το 20% των μετα-αναλύσεων είχαν εφαρμόσει κάποια ανάλυση ευαισθησίας με σκοπό την εξέταση της επίδρασης των ελλιπουσών τιμών αποτελέσματος στα αποτελέσματα της πρωταρχικής ανάλυσης.

Τα αποτελέσματα της παρούσας μελέτης έδειξαν ότι οι συστηματικές ανασκοπήσεις συχνά παρέχουν ασαφή ορισμό της ορολογίας που αφορά τις ελλιπούσες τιμές αποτελέσματος και σπάνια διακρίνουν μεταξύ δεδομένων που έχουν υποστεί αντικατάσταση χρησιμοποιώνταν τη μέθοδο ‘προώθησης τελευταίας παρατήρησης’ και των εξ ολοκλήρου ελλιπουσών τιμών αποτελέσματος. Η πλειοψηφία των συστηματικών ανασκοπήσεων τείνουν να αντιμετωπίζουν τις ελλιπούσες τιμές αποτελέσματος είτε ως μεταβλητή αποτελέσματος είτε εφαρμόζοντας μία στρατηγική αντικατάστασης στην πρωταρχική ανάλυση, αλλά χωρίς να εφαρμόσουν ανάλυση ευαισθησίας ώστε να εξεταστεί η καταλληλότητα αυτής της στρατηγικής. Συνεπώς, η περιγραφή του μεγέθους και της έκτασης των

ελλιπουσών τιμών αποτελέσματος και η συμπερίληψη των συνεπειών αυτών στα ευρήματα των συστηματικών ανασκοπήσεων είναι ιδιαίτερα ανεπαρκής.

Η ελαχιστοποίηση της εμφάνισης των ελλιπουσών τιμών αποτελέσματος είναι επίσης ένα σημαντικό θέμα. Η πρόωρη απομάκρυνση των ασθενών από τις μελέτες συμβαίνει συστηματικά στις μελέτες ψυχικής υγείας και το ποσοστό των ελλιπουσών τιμών αποτελέσματος κυμαίνεται από χαμηλό προς ιδιαίτερα υψηλό. Για την πρόληψη όσο το δυνατόν περισσότερων ελλιπουσών τιμών αποτελέσματος, είναι σημαντική η εξερεύνηση και η κατανόηση των λόγων που οι ασθενείς εγκαταλείπουν νωρίς τις μελέτες. Επομένως, η παρούσα μελέτη ξεκίνησε τον εντοπισμό των χαρακτηριστικών μελετών που έχουν αντίκτυπο στην πρόωρη εγκατάλειψη των αντιψυχωτικών μελετών για τη σχιζοφρένεια.

Τα αποτελέσματα της παρούσας έρευνας έδειξαν ότι οι μελέτες με επαρκή απόκρυψη της τυχαιοποίησης, διπλά-τυφλές, που περιείχαν εικονική θεραπεία, με υψηλή ακρίβεια στις εκτιμήσεις, με μεγάλο δείγμα συμμετέχοντων, με τουλάχιστον τρεις θεραπείες, πρόσφατα δημοσιευμένες, που διεξάχθηκαν στις ΗΠΑ και επιστράτευσαν εσωτερικούς ασθενείς βρέθηκαν να σχετίζονται με το ενδεχόμενο ένας ασθενής να αξιολογηθεί τουλάχιστον μία φορά ή καθόλου πριν εγκαταλείψει τη μελέτη. Ωστόσο, η μέθοδος τυφλοποίησης δεν προέβλεψε το ενδεχόμενο ένας ασθενής να αξιολογηθεί τουλάχιστον μία φορά, ενώ η μέθοδος απόκρυψης της τυχαιοποίησης, η υψηλή ακρίβεια στις εκτιμήσεις, το μεγάλο δείγμα ασθενών, το πλήθος συγκρινόμενων θεραπειών, η πρόσφατη δημοσίευση και η επιστράτευση εσωτερικών ασθενών δεν προέβλεψαν το ενδεχόμενο ένας ασθενής να μην αξιολογηθεί καθόλου.

Τα ευρήματα της παρούσας μελέτης αποκάλυψαν ότι υψηλά ποσοστά των ελλιπουσών τιμών αποτελέσματος στις αντιψυχωτικές μελέτες μπορεί να σχετίζονται με διάφορα χαρακτηριστικά μελετών και κυρίως με την χρήση εικονικής θεραπείας ως θεραπεία ελέγχου και το μέγεθος της μελέτης. Από κλινική και μεθοδολογική πλευρά, οι ερευνητές θα πρέπει πάντα κατά τον σχεδιασμό και τη διεξαγωγή της κλινικής μελέτης να λαμβάνουν υπόψη εκείνα τα χαρακτηριστικά μελέτης που προκαλούν πρόωρη εγκατάλειψη της μελέτης. Ιδιαίτερα στο σχεδιασμό επιβεβαιωτικών μελετών για τη σύγκριση αντιψυχωτικών θεραπειών νέας γενιάς για την σχιζοφρένεια, ο εντοπισμός χαρακτηριστικών μελέτης που προκαλούν πρόωρη εγκατάλειψη της μελέτης είναι ιδιαίτερα σημαντικός.

Παρόλο που τα χαρακτηριστικά μελέτης είναι σημαντικά για την πρόληψη και την ελαχιστοποίηση των ελλιπουσών τιμών αποτελέσματος μέχρι κάποιο βαθμό, η αντιμετώπιση των παρόντων ελλιπουσών τιμών αποτελέσματος χρησιμοποιώντας μια κατάλληλη αναλυτική στρατηγική είναι απαραίτητη για την μείωση του κινδύνου μεροληπτικών και χαμηλής ακρίβειας αποτελεσμάτων. Υπάρχουν πολλές στατιστικές μέθοδοι που λαμβάνουν υπόψη την επίδραση των ελλιπουσών τιμών αποτελέσματος σε επίπεδο μελέτης. Η μέθοδος ‘προώθησης τελευταίας παρατήρησης’ εφαρμόζεται συχνά στις μελέτες και έχει λάβει αρνητικές κριτικές επειδή βασίζεται σε μη ρεαλιστικές υποθέσεις και δίνει συντηρητικά αποτελέσματα.

Σε επίπεδο μετα-ανάλυσης, οι ελλιπούσες τιμές αποτελέσματος αντιμετωπίζονται κάνοντας διάφορες υποθέσεις για το αποτέλεσμα των ελλιπόντων συμμετεχόντων. Η

πλεοψηφία των μετα-αναλύσεων στην ψυχιατρική υιοθετούν την υπόθεση ότι όλοι οι ελλιπόντες συμμετέχοντες απέτυχαν να αποκριθούν στο φάρμακό τους. Αυτές οι υποθέσεις συνοδεύονται από αντικατάσταση των αποτελεσμάτων των ελλιπόντων συμμετεχόντων αγνοώντας συνήθως την αβεβαιότητα των αντικαταστάσεων αυτών στα αποτελέσματα της ανάλυσης. Επομένως, ένα μοντέλο απλής μετα-ανάλυσης έχει προταθεί με σκοπό να ποσοτητκοποιήσει καλύτερα την αβεβαιότητα στις αντικαταστάσεις των ελλιπουσών τιμών αποτελέσματος εισάγοντας στο μοντέλο μία άγνωστη παράμετρο η οποία ποσοτικοποιεί την απόκλιση μεταξύ ελλιπουσών τιμών αποτελέσματος με και χωρίς πληροφορία σε κάθε θεραπεία και μελέτη. Αυτή η παράμετρος αφορά δίτιμα δεδομένα αποτελέσματος και ονομάζεται ‘αναλογία πιθανοτήτων ελλιπουσών τιμών αποτελέσματος με πληροφορία’ (*informative missingness odds ratio, IMOR*), όπου περιγράφει την σχέση μεταξύ του άγνωστου αποτελέσματος (επιτυχία ή αποτυχία) στους ελλιπόντες συμμετέχοντες και του γνωστού αποτελέσματος στους παρατηρούμενους συμμετέχοντες.

Η παρούσα μελέτη επέκτεινε την ιδέα για την IMOR παράμετρο στο δίκτυο μετα-ανάλυσης με σκοπό την εξερεύνηση της επίδρασης των ελλιπόντων τιμών αποτελέσματος στην συμπερασματολογία για την σχετική επίδραση διαφόρων ανταγωνιστικών θεραπειών στις ψυχιατρικές μελέτες. Ένα σετ δεδομένων για αντιμανιακές θεραπείες και ένα σετ δεδομένων για αντικαταθλιπτικές θεραπείες χρησιμοποιήθηκαν. Το αποτέλεσμα ήταν απόκριση στην τυχαιοποιημένη θεραπεία. Στις αρχικές μετα-αναλύσεις, οι ελλιπόντες συμμετέχοντες θεωρήθηκαν ότι απέτυχαν να αποκριθούν στη θεραπεία τους ασχέτως αν πήραν ή όχι την τυχαιοποιημένη θεραπεία. Η ακεραιότητα αυτής της υπόθεσης σε κάθε σετ δεδομένων αξιολογήθηκε εφαρμόζοντας διάφορα IMOR μοντέλα που αντανακλούν διαφορετική υπόθεση για τα ελλιπόντα δεδομένα αποτελέσματος. Η συμφωνεία των συμπερασμάτων αξιολογήθηκε συγκρίνοντας τις αναλογίες πιθανοτήτων για την αποτελεσματικότητα σύμφωνα με την αρχική ανάλυση με τις αναλογίες πιθανοτήτων για την αποτελεσματικότητα σύμφωνα με τις διάφορες υποθέσεις για τα ελλιπόντα δεδομένα αποτελέσματος. Ο μηχανισμός ελλιπόντων δεδομένων μελετήθηκε συγκρίνοντας την αρχική με την τελική κατανομή των IMOR παραμέτρων. Η Markov Monte Carlo αλυσίδα χρησιμοποιήθηκε για την εφαρμογή των μεντέλων στο πρόγραμμα WinBUGS.

Στα δύο σετ δεδομένων, η σχετική αποτελεσματικότητα των θεραπειών επηρεάστηκε μόνο από την χειρότερη και την καλύτερη υπόθεση για τα ελλιπόντα δεδομέναν αποτελέσματος. Επιπλέον, η ετερογένεια αυξήθηκε και στα δύο σετ δεδομένων σε αυτές τις δύο ακραίες υποθέσεις. Συνολικά, υπήρχαν μικρές διαφορές στην ιεράρχιση των αντιμανιακών και των αντικαταθλιπτικών θεραπειών. Η τελική κατανομή των IMOR παραμέτρων ήταν πολύ όμοια με την αρχική κατανομή αυτών και αυτό δείχνει ότι υπήρχε πολύ λίγη πληροφορία για το πραγματικό αποτέλεσμα (επιτυχία ή αποτυχία) των ελλιπόντων συμετέχοντων. Υπήρχε πολύ ελάχιστη ένδειξη ότι οι ελλιπόντες συμμετέχοντες έτειναν να μην αποκρίνονται στην εικονική θεραπεία και την παροξετίνη, ενώ το αντίθετο συνέβη για την σερτραλινη, την φλουοξετίνη και την φλουβοξαμίνη.

Στον τομέα της ψυχιατρικής όπου το ποσοστό ελλιπόντων δεδομένων αποτελέσματος είναι υψηλό και με πληροφορία, είναι σημαντικό να μελετηθεί κατά πόσο τα ελλιπόντα δεδομένα αποτελέσματος με πληροφορία μπορούν να διαστρεβώσουν τα συμπεράσματα των τυχαιοποιημένων κλινικών δοκιμών και μετα-αναλύσεων. Η IMOR μέθοδος επιτρέπει την μελέτη όλων των πιθανών υποθέσεων για το αποτέλεσμα των ελλιπόντων συμμετέχοντων λαμβάνοντας υπόψη ταυτόχρονα την αβεβαιότητα κάθε υπόθεσης. Εντός ενός δικτύου μετα-ανάλυσης, η IMOR προσέγγιση προσφέρει επίσης την δυνατότητα να μελετηθούν οι αλληλεπιδράσεις μεταξύ θεραπειών και ελλιπόντων δεδομένων αποτελέσματος με πληροφορία εφαρμόζοντας μεγάλου εύρους υποθέσεις για την IMOR παράμετρο κατά μήκους των θεραπειών και των μελετών. Από κλινική και μετα-αναλυτική πλευρά, είναι επιθυμητή η χρήση λογικών αρχικών κατανομών για αυτές τις υποθέσεις σύμφωνα με τη γνώμη ειδικού.

Reference List

1. Akl EA, Johnston BC, Alonso-Coello P, et al. Addressing dichotomous data for participants excluded from trial analysis: a guide for systematic reviewers. PLoS One. 2013 Feb; 8(2):e57132.

2. Dumville JC, Torgerson DJ, Hewitt CE. Reporting attrition in randomised controlled trials. BMJ. 2006 Apr; 332(7547): 969-971.

3. Hewitt CE, Kumaravel B, Dumville JC, et al. Trial attrition study group. Assessing the impact of attrition in randomized controlled trials. J Clin Epidemiol. 2010 Nov; 63(11):1264-1270.

4. Leon AC, Mallinckrodt CH, Chuang-Stein C, et al. Attrition in randomized controlled clinical trials: methodological issues in psychopharmacology. Biol Psychiatry. 2006 Jun; 59(11):1001-1005.

5. Higgins JP, White IR, Wood AM. Imputation methods for missing outcome data in meta-analysis of clinical trials. Clin Trials. 2008 Jun; 5(3):225-239.

6. Lane P. Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. Pharm Stat. 2008 Apr-Jun; 7(2):93-106.

7. Spineli LM, Higgins JP, Cipriani A, et al. Evaluating the impact of imputations for missing participant outcome data in network meta-analysis. Clin Trials. 2013 Jan; 10(3):378-388.

8. Ware JH, Harrington D, Hunter DJ, et al. Missing data. N Engl J Med. 2012 Oct; 367:1353-1354.

9. Carpenter JR, Kenward MG. Missing data in randomised controlled trials - a practical guide. Birmingham: National Institute for Health Research; 2008. www.pcpoh.bham.ac.uk/publichealth/methodology/projects/RM03_JH17_MK.shtml.

10. Rabinowitz J, Davidov O. The association of dropout and outcome in trials of antipsychotic medication and its implications for dealing with missing data. Schizophr Bull. 2008 Mar; 34(2):286-291.

11. Akl EA, Briel M, You JJ, et al. Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST-IT): systematic review. BMJ. 2012 May; 344:e2809.

12. Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. Clin Trials. 2004 Aug; 1(4):368-376.

13. Khan A, Schwartz K, Redding N, et al. Psychiatric diagnosis and clinical trial completion rates: analysis of the FDA SBA reports. Neuropsychopharmacology. 2007 Feb; 32(11):2422-2430.

14. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. Lancet. 2002 Mar; 359(9308):781-785.

15. Nemeroff CB. Improving antidepressant adherence. J Clin Psychiatry. 2003; 64(Suppl. 18):25-30.

16. Khan A, Warner HA, Brown WA. Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials: an analysis of the Food and Drug Administration database. Arch Gen Psychiatry. 2000 Apr; 57(4):311-317.

17. Hammad T. Relationship between psychotropic drugs and pediatric suicidality. 2004. pp. 1–131. [http://www.fda.gov/ohrms/dockets/ac/04/briefing/2004-4065b1-10-TAB08-Hammads-Review.pdf]. FDA Accessed June 2005.

18. Heo M, Papademetriou E, Meyers BS. Design characteristics that influence attrition in geriatric antidepressant trials: meta-analysis. Int J Geriatr Psychiatry. 2009 Sep; 24(9): 990-1001.

19. Liu-Seifert H, Adams DH, Kinon BJ. Discontinuation of treatment of schizophrenic patients is driven by poor symptom response: a pooled post-hoc analysis of four atypical antipsychotic drugs. BMC Med. 2005 Dec; 3:21-30.

20. Rabinowitz J, Davidov O. A composite approach that includes dropout rates when analyzing efficacy data in clinical trials of antipsychotic medications. Schizophr Bull. 2008 Nov; 34(6):1145-1150.

21. Wahlbeck K, Tuunainen A, Ahokas A, et al. Dropout rates in randomised antipsychotic drug trials. Psychopharmacology (Berl). 2001 May; 155(3):230-233.

22. Kemmler G, Hummer M, Widschwendter C, et al. Dropout rates in placebo-controlled and active-control clinical trials of antipsychotic drugs: a meta-analysis. Arch Gen Psychiatry. 2005 Dec; 62(12):1305-1312.

23. Prakash A, Risser RC, Mallinckrodt CH. The impact of analytic method on interpretation of outcomes in longitudinal clinical trials. Int J Clin Pract. 2008 Aug; 62(8):1147-1158.

24. Lieberman JA. Atypical antipsychotic drugs as a first-line treatment of schizophrenia: a rationale and hypothesis. J Clin Psychiatry. 1996; 57(Suppl. 11):68-71.

25. Sharif ZA. Common treatment goals of antipsychotics: acute treatment. J Clin Psychiatry. 1998; 59(Suppl. 19):5-8.

26. Martin JL, Pérez V, Sacristán M, et al. Meta-analysis of drop-out rates in randomised clinical trials, comparing typical and atypical antipsychotics in the treatment of schizophrenia. Eur Psychiatry. 2006 Jan; 21(1):11-20.

27. Liu-Seifert H, Zhang S, D'Souza D, et al. A closer look at the baseline-observation-carried-forward (BOCF). Patient Prefer Adherence. 2010 Jan; 4:11-16.

28. Tierney JF, Stewart LA. Investigating patient exclusion bias in meta-analysis. Int J Epidemiol. 2005 Feb; 34(1):79-87.

29. Spineli LM, Leucht S, Cipriani A, et al. The impact of trial characteristics on premature discontinuation of antipsychotics in schizophrenia. Eur Neuropsychopharmacol. 2013 Sep; 23(9):1010-1016.

30. Spineli LM, Pandis N, Salanti G. Reporting and handling missing outcome data in mental health: a systematic review of Cochrane systematic reviews and meta-analyses. Sumbitted research.

31. Myers WR. Handling missing data in clinical trials: an overview. Ther. Innov. Regul. Sci. 2000 Apr; 34(2):525-533.

32. White IR, Higgins JP, Wood AM. Allowing for uncertainty due to missing data in meta-analysis--part 1: two-stage methods. Stat Med. 2008 Feb; 27(5):711-727.

33. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. BMJ. 1999 Sep; 319(7211):670-674.

34. White IR, Welton NJ, Wood AM, et al. Allowing for uncertainty due to missing data in meta-analysis--part 2: hierarchical models. Stat Med. 2008 Feb; 27(5):728-745.

35. Leuchs AK, Zinserling J, Schlosser-Weber G, et al. Estimation of the treatment effect in the presence of non-compliance and missing data. Stat Med. 2014 Jun; 33(2):193-208.

36. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ. 2010 Feb; 340:c869. 2010.

37. Mulsant BH, Pollock BG, Nebes RD, et al. A double-blind randomized comparison of nortriptyline and paroxetine in the treatment of late-life depression: 6-week outcome. J Clin Psychiatry. 1996; 60(Suppl. 20):16-20.

38. Chen Q, Chen MH, Ohlssen D, et al. Bayesian modeling and inference for clinical trials with partial retrieved data following dropout. Stat Med. 2013 Oct; 32(4): 4180-4195.

39. Altman DG. Missing outcomes in randomized trials: addressing the dilemma. Open Med. 2009 May; 3(2):e51-3.

40. Hamer RM, Simpson PM. Last observation carried forward versus mixed models in the analysis of psychiatric clinical trials. Am J Psychiatry. 2009 Jun; 166(6):639-641.

41. Shih WJ. Problems in dealing with missing data and informative censoring in clinical trials. Curr Control Trials Cardiovasc Med. 2002 Jan; 3(1):4-10.

42. Rosenheck RA. Open forum: effectiveness versus efficacy of second-generation antipsychotics: haloperidol without anticholinergics as a comparator. Psychiatr Serv. 2005 Jan; 56(1):85-92.

43. Leucht S, Engel RR, Bäuml J, et al. Is the superior efficacy of new generation antipsychotics an artifact of LOCF? Schizophr Bull. 2007 Jan; 33(1):183-191.

44. Lieberman JA, Tollefson G, Tohen M, et al. Comparative efficacy and safety of atypical and conventional antipsychotic drugs in first-episode psychosis: a randomized, double-blind trial of olanzapine versus haloperidol. Am J Psychiatry. 2003 Aug; 160(8):1396-1404.

45. Mallinckrodt CH, Sanger TM, Dube S, et al. Assessing and interpreting treatment effects in longitudinal clinical trials with missing data. Biol Psychiatry. 2003 Apr; 53(8):754-760.

46. Mallinckrodt CH, Watkin JG, Molenberghs G, et al. Choice of the primary analysis in longitudinal clinical trials. Pharmaceut.Statist. 2004 Sep; 3(3):161-169.

47. Siddiqui O, Hung HM, O'Neill R. MMRM vs. LOCF: a comprehensive comparison based on simulation study and 25 NDA datasets. J Biopharm Stat. 2009 Feb; 19(2):227-246.

48. Mallinckrodt CH, Lane PW, Schnell D, et al. Recommendations for the Primary Analysis of Continuous Endpoints in Longitudinal Clinical Trials. Ther. Innov. Regul. Sci. 2008 Jul; 42(4):303-319.

49. Fielding S, Maclennan G, Cook JA, et al. A review of RCTs in four medical journals to assess the use of imputation to overcome missing data in quality of life outcomes. Trials. 2008 Aug; 9:51-56.

50. Helfand M, Berg A, Flum D, et al. Draft Methodology Report: 'Our questions, our decisions: Standards for patient-centered outcomes research'. Public comment draft report of the Patient-Centered Outcomes Research Institute (PCORI) Methodology Committee presented on 23 July 2012, and revised thereafter, 2012.

51. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. Statistical principles for clinical trials, Step 5: Note for guidance on statistical principles for clinical trials. In International Conference on Harmonization (ICH) Topic E9, 5 February 1998, Published in the Federal Register, p. 23.

52. Panel on Handling Missing Data in Clinical Trials, National Research Council. 'Front Matter'. The Prevention and Treatment of Missing Data in Clinical Trials. The National Academies Press, Washington, DC, 2010.

53. Streiner DL. The case of the missing data: methods of dealing with dropouts and other research vagaries. Can J Psychiatry. 2002 Feb; 47(1):68-75.

54. Shao J, Zhong B. On the treatment effect in clinical trials with dropout. J Biopharm Stat. 2006 Aug; 16(1):25-33.

55. Shao J, Jordan DC, Pritchett YL. Baseline observation carried forward: reasoning, properties, and practical issues. J Biopharm Stat. 2009 Jul; 19(4):672-684.

56. Rappaport BA. Overview of the August 19, 2010, ALSDAC Meeting to Discuss NDA 22-516 for Cymbalta for the Treatment of Chronic Pain. FDA. 2010.

57. Dr.Deng. On Biostatistics and Clinical Trials: LOCF, BOCF, WOCF, and MVTF [Internet]. 2010 Aug. Available from: http://onbiostatistics.blogspot.gr/2010/08/locf-bocf-wocf-and-mvtf.html.

58. Haukoos JS, Newgard CD. Advanced statistics: missing data in clinical research--part 1: an introduction and conceptual framework. Acad Emerg Med. 2007 Jul. 14(7):662-668.

59. Bell ML, Kenward MG, Fairclough DL, et al. Differential dropout and bias in randomised controlled trials: when it matters and when it may not. BMJ. 2013 Jan; 346:e8668.

60. Little RJA, Rubin DB. Statistical analysis with missing data. New York: Wiley; 1987.

61. Hollis S. A graphical sensitivity analysis for clinical trials with non-ignorable missing binary outcome. Stat Med. 2002 Dec; 21(24):3823-3834.

62. Little RJA. Modeling the drop-out mechanism in repeated-measures studies. J Am Statist Assoc. 1995 Sep; 90(431):1112-1121.

63. Gamble C, Hollis S. Uncertainty method improved on best-worst case analysis in a binary meta-analysis. J Clin Epidemiol. 2005 Jun; 58(6):579-588.

64. Magder LS. Simple approaches to assess the possible impact of missing outcome information on estimates of risk ratios, odds ratios, and risk differences. Control Clin Trials. 2003 Aug; 24(4):411-421.

65. Deeks JJ, Higgins JP, Altman DG. Analysing data and undertaking meta-analyses. In: Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.1 [updates September 2008] ed. The Cochrane Collaboration; 2008.

66. Borestein M, Hedges LV, Higgins JP, et al. Introduction to Meta-Analysis. John Wiley & Sons, Ltd; 2009.

67. European Medicines Evaluation Agency. Guideline on Missing Data in Confirmatory Clinical Trials. Committee for Medical Products for Human Use; 2009.

68. Alshurafa M, Briel M, Akl EA, et al. Inconsistent definitions for intention-to-treat in relation to missing outcome data: systematic review of the methods literature. PLoS One. 2012 Nov; 7(11):e49163.

69. White IR, Carpenter JR, Horton NJ. Including all individuals is not enough: lessons for intention-to-treat analysis. Clin Trials. 2012 Aug; 9(4):396-407.

70. Higgins JP, Deeks JJ, Altman DG, on behalf of the Cochrane Statistical Methods Group. Intention-to-treat issues. Cochrane handbook. 2009. Available from: http://www.cochrane-handbook.org.

71. Abraha I, Montedori A. Modified intention to treat reporting in randomised controlled trials: systematic review. BMJ. 2010 Jun; 340:c2697.

72. Montedori A, Bonacini MI, Casazza G, et al. Modified versus standard intention-to-treat reporting: are there differences in methodological quality, sponsorship, and findings in randomized trials? A cross-sectional study. Trials. 2011 Feb; 12:58-66.

73. Stewart WC, Jackson AL, Jenkins JN. Dropout rates for intent-to-treat and per protocol analyses. Am J Ophthalmol. 2004 Apr; 137(4):639-645.

74. McNamee R. Intention to treat, per protocol, as treated and instrumental variable estimators given non-compliance and effect heterogeneity. Stat Med. 2009 Sep; 28(21):2639-2652.

75. Nüesch E, Trelle S, Reichenbach S, et al. The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. BMJ. 2009 Sep; 339:b3244.

76. Lewis JA, Machin D. Intention to treat--who should use ITT? Br J Cancer. 1993 Oct; 68(4):647-650.

77. Groenwold RH, Donders AR, Roes KC, et al. Dealing with missing outcome data in randomized trials and observational studies. Am J Epidemiol. 2012 Feb; 175(3):210-217.

78. Donahue RMJ. Confirmatory trials. Wiley Encyclopedia of Clinical Trials; 2008.

79. Lee YJ, Ellenberg JH, Hirtz DG, et al. Analysis of clinical trials by treatment actually received: is it really an option. Stat Med. 1991 Oct; 10(10):1595-1605.

80. Trowman R, Dumville JC, Torgerson DJ, et al. The impact of trial baseline imbalances should be considered in systematic reviews: a methodological case study. J Clin Epidemiol. 2007 Dec; 60(12):1229-1233.

81. Akl EA, Briel M, You JJ, et al. LOST to follow-up Information in Trials (LOST-IT): a protocol on the potential impact. Trials. 2009 Jun; 10:40.

82. Joshi M, Royuela A, Zamora J. Proper analysis in clinical trials: how to report and adjust for missing outcome data. BJOG.2013 Jul; 120(8):915-919.

83. Kane RL, Wang J, Garrard J. Reporting in randomized clinical trials improved after adoption of the CONSORT statement. J Clin Epidemiol. 2007 Mar; 60(3):241-249.

84. Higgins JP, Deeks JJ, Altman DG, on behalf of the Cochrane Statistical Methods Group. General principles for dealing with missing data. Cochrane handbook. 2009. Available from: http://www.cochrane-handbook.org.

85. Cipriani A, Barbui C, Salanti G, et al. Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple-treatments meta-analysis. Lancet. 2011 Oct; 378(9799):1306-1315.

86. Leucht S, Cipriani A, Spineli LM, et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. Lancet. 2013 Sep; 382(9896):951-962.

87. Higgins JP, Altman DG, Sterne JAC. Assessing risk of bias in included studies. In: Higgins JP, Green S, editors. Cochrane Handbook for Sustematic Reviews of Interventions. The Cochrane Collaboration; 2011.

88. Hummer M, Holzmeister R, Kemmler G, et al. Attitudes of patients with schizophrenia toward placebo-controlled clinical trials. J Clin Psychiatry. 2003 Mar; 64(3):277-281.

89. Sonawalla SB, Farabaugh AH, Leslie VM, et al. Early drop-outs, late drop-outs and completers: differences in the continuation phase of a clinical trial. Prog Neuropsychopharmacol Biol Psychiatry. 2002 Dec. 26(7-8):1415-1419.

90. Aloe AM, Becker BJ, Pigott TD. An alternative to R^2 for assessing linear models of effect size. Res.Synth.Method. 2010 Dec; 1(3-4):272-283.

91. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A Stat Soc. 2009 Jan; 172(1):137-159.

92. Hemming K, Hutton JL, Maguire MG, et al. Meta-regression with partial information on summary trial or patient characteristics. Stat Med. 2010 May; 29(12):1312-1324.

93. Lunn DJ, Thomas A, Best N, et al. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. Stat Comput. 2000 Oct; 10(4):325-337.

94. Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. Stat Methods Med Res. 2001 Apr; 10(4):277-303.

95. Leucht S, Kane JM, Etschel E, et al. Linking the PANSS, BPRS, and CGI: clinical implications. Neuropsychopharmacology. 2006 Oct; 31(10):2318-2325.

96. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. Lancet. 2009 Feb; 373(9665):746-758.

97. Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. J Clin Epidemiol. 2011 Feb; 64(2):163-171.

98. Salanti G, Higgins JP, Ades AE, et al. Evaluation of networks of randomized trials. Stat Methods Med Res. 2008 Jun; 17(3):279-301.

99. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med. 2004 Oct; 23(20):3105-3124.

100. Cipriani A, Geddes JR. What is a run-in phase? Epidemiol Psychiatr Soc. 2010 Jan-Mar; 19(1):21-22.

101. Stojanovski J, Nur D, Prior Sensitivity Analysis for a Hierarchical Model. Proceeding of the Fourth Annual ASEARCH Conference. 2011 Feb; 64-67.

Appendices

Appendix A: Computational formulas for the ICA approach

Appendix B: The regression models and the methods to account for partial information on covariates

Appendix C: The WinBUGS code for the treatment-specific IMOR model

Appendix A

Computational formulas for the ICA approach

The following example aims to clarify the computational formulas for each ICA assumption. Consider a trial that compares two interventions. With r_E and r_C is denoted the number of events in the experimental and control intervention, respectively, with f_E and f_C the number of non-events, with m_E and m_C the number of missing outcome data and with n_E and n_C the total sample size. For each intervention it holds that

$$n_j = r_j + f_j + m_j \quad j = E, C$$

The observed risk of events is denoted by p_E and p_C for the experimental and control intervention, respectively, and it is defined as

$$p_E = r_E/(r_E + f_E) \text{ and } p_C = r_C/(r_C + f_C)$$

Table A.1 summarizes the computational formulas for each ICA assumption

Table A.1

Method	Experimental		Control	
	events	non-events	events	non-events
ICA-0	r_E	$f_E + m_E$	r_C	$f_C + m_C$
ICA-1	$r_E + m_E$	f_E	$r_C + m_C$	f_C
ICA-b	$r_E + m_E$	f_E	r_C	$f_C + m_C$
ICA-w	r_E	$f_E + m_E$	$r_C + m_C$	f_C
ICA-pE	$r_E + m_E \times p_E$	$f_E + m_E \times (1 - p_E)$	$r_C + m_C \times p_E$	$f_C + m_C \times (1 - p_E)$
ICA-pC	$r_E + m_E \times p_C$	$f_E + m_E \times (1 - p_C)$	$r_C + m_C \times p_C$	$f_C + m_C \times (1 - p_C)$
ICA-p	$r_E + m_E \times p_E$	$f_E + m_E \times (1 - p_E)$	$r_C + m_C \times p_C$	$f_C + m_C \times (1 - p_C)$

Note that the total sample size is preserved in both interventions under all ICA assumptions.

Appendix B

The regression models and the meta-regression with partial information on covariates

B.1 Definitions

We denote by ns the total number of studies. The total participants n_i randomized into study i are divided into completers (denoted by c_i), LOCF (denoted by l_i), or excluded (denoted by m_i) so that $n_i = c_i + l_i + m_i$. The total number of dropouts, d_i, is the sum of LOCF and excluded patients:

$$d_i = l_i + m_i.$$

B.2 Poisson regression model (univariable and multivariable)

In each trial the dropouts are assumed drawn from a Poisson distribution:

$$d_i \sim Poi(\lambda_i \cdot L_i),\ i = 1,2,\dots,ns$$

where L_i, the total number of patient-weeks, is the product of the study size n_i and its duration in weeks D_i. Parameter λ_i is the rate of dropouts per patient-week in study i. The logarithm of the dropout rate per patient-week can be modeled as a linear function of k covariates of interest in the following random-effects model

$$log(\lambda_i) = log(\lambda_i^*) + \boldsymbol{\beta}' \cdot \boldsymbol{T}_i \qquad (B.1)$$

$$log(\lambda_i^*) \sim N(\mu, \sigma^2)$$

where $log(\lambda_i^*)$ is the logarithm of the dropout rate per patient-week in study i, adjusted for covariates included in $\boldsymbol{T}_i$, a vector of k covariate values $(T_{i1}, T_{i2}, \dots, T_{ik})$ for trial i. Continuous covariates are centred at their minimum value to improve interpretability of the corresponding regression coefficients. The parameter $\boldsymbol{\beta}$ is a vector of regression coefficients $(\beta_1, \beta_2, \dots, \beta_k)$, each denoting the extent of change in the logarithm of dropout rate after adjusting for the corresponding covariate. The baseline drop-out rates depicted by $log(\lambda_i^*)$ are assumed normally distributed across trials with σ^2 representing the heterogeneity variance. For a dichotomous covariate l, the ratio of the dropout rate for covariate level $T_{il} = 1$ compared with level $T_{il} = 0$ is expressed as a rate ratio (RR), $RR_l = exp(\beta_l)$, while for a continuous covariate l, the

relative dropout rate for an increase by m units is $RR_l = exp(m \cdot \beta_l)$, where $l = 1,2,\dots,k$.

The predicted logarithm of the dropout rate per patient-week in a new study without the characteristic l $(T_{il} = 0)$ or at the minimum value of the continuous explanatory variable l is obtained as

$$log(\lambda^*_{new}) \sim N(\mu, \sigma^2)$$

and for a study with the characteristic l $(T_{il} = 1)$ or with continuous explanatory variable value increased by one unit the predicted logarithm of the dropout rate per patient-week is obtained as

$$log(\lambda^*_{new}|T_{il} = 1) \sim N(\mu + \beta_l, \sigma^2)$$

The exponentiation of $log(\lambda^*_{new}|T_{il})$ is the predicted dropout rate per patient-week for T_{il} in a new study. With predictions we can make inferences about the expected dropout rate per patient-week in future studies that might be included in the meta-analysis and are 'sufficiently similar' to those already included in analysis (1). The prediction interval incorporates both the heterogeneity of the dropout rate per patient-week and the sample variance of the summary dropout rate (2) .

B.3 Multinomial logistic model

In each trial, the three patient groups are assumed drawn from a multinomial distribution:

$$y_i = (c_i, l_i, m_i) \sim Multi(p_{i1}, p_{i2}, p_{i3}) \text{ with } i = 1,2,\dots,ns$$

where p_{i1}, p_{i2}, p_{i3} are the probabilities that a patient will be a completer, LOCF or excluded, respectively. To compare the probability that a patient will be either LOCF or excluded with the probability of being a completer, we set the completers as the baseline group and estimate the risk of being in one of the other groups compared with the baseline. Specifically, we define

$$\varphi_{ij} = p_{ij}/p_{i1} \text{ with } i = 1,2,\dots,ns,\ j = 2,3 \qquad (B.2)$$

where the parameter φ_{ij} is the risk ratio (RR) that a patient in the i^{th} study will be either LOCF $(j = 2)$ or excluded $(j = 3)$ rather than a completer. Starting from equation $(B.2)$ and noting that $\varphi_{i1} = 1$, for convenient implementation in WinBUGS we can write

$p_{i1} = 1/\sum_{j=1}^{3} \varphi_{ij}$ with $i = 1,2,\dots,ns,\ j = 2,3$

as the probability that a patient of the i^{th} study falls in the baseline group and

$p_{ij} = \varphi_{ij} / \sum_{j=1}^{3} \varphi_{ij}$ with $i = 1,2,\dots,ns,\ j = 2,3$

as the probability that a patient of the i^{th} study falls in the j^{th} group.

The log risk ratios (RR) $\theta_{ij} = log(\varphi_{ij})$ are assumed to be exchangeable across studies. They are also correlated because both incorporate the baseline category. Thus, they follow a bivariate normal distribution:

$$(\theta_{i2}, \theta_{i3})' \sim MVN\left((\mu_2, \mu_3)', \begin{pmatrix} \sigma_2^2 & C \\ C & \sigma_3^2 \end{pmatrix}\right)$$

where μ_2 and μ_3 are the summary log-RRs of being LOCF and excluded, respectively, rather than completers. Variances σ_2^2 and σ_3^2 are the comparison-specific heterogeneities of the log RRs across the studies that compare LOCF and excluded, respectively, with the completers. Assuming that heterogeneity variances are independent of the category, the covariance is $C = \sigma^2/2$.

We can investigate whether a vector of k covariates $\boldsymbol{T}_i = (T_{i1}, T_{i2}, \dots, T_{ik})$ can influence the risk of a patient being LOCF or excluded rather than a completer by fitting a model in which

$$\theta_{ij} = \theta_{ij}^* + \boldsymbol{\delta}_j' \cdot \boldsymbol{T}_i \text{ with } j = 2,3 \qquad (B.3)$$

where the adjusted $\left(\theta_{i2}^*, \theta_{i3}^*\right)'$ is drawn from a bivariate normal distribution $\left(\theta_{i2}^*, \theta_{i3}^*\right)' \sim MVN\left(\left(\mu_2^*, \mu_3^*\right)', \begin{pmatrix} \sigma_2^{*2} & C \\ C & \sigma_3^{*2} \end{pmatrix}\right)$ with $C^* = \sigma^{*2}/2$. The parameter $\boldsymbol{\delta}_j'$ is a vector of regression coefficients $(\delta_{j1}, \delta_{j2}, \dots, \delta_{jk})$ that are assumed identical across studies. The impact of $\boldsymbol{T}_i$ on the heterogeneity can be assessed by comparing σ^{*2} with σ^2 and calculating the percentage change as

$$(\sigma^2 - \sigma^{*2})/\sigma^2$$

The relative risk ratio (RRR) expresses how likely it is that a patient with a specific characteristic l will leave the study early or be excluded rather than be completer as opposed to another patient who does not have this characteristic. This relative measure of risk is derived as $RRR_{jl} = exp(\delta_{jl})$ for $j = 2,3$ and $l = 1,2,\dots,k$.

B.4 Model implementation

WinBUGS software was used to fit all models. For parameters that represent means, we used non-informative normal prior distributions centred at zero with large variance, such as 10,000. For standard deviations we used uniform prior distribution U(0,10). We used two chains with zero and positive starting values for the mean parameters, and 0.5 and 0.1 for the standard deviation parameters. The analyses used

100,000 updates with the first 10,001 samples discarded to allow for burn-in. To reduce autocorrelation we used thinning to every 40th or 80th value.

B.5 Meta-regression with partial information on covariates

Consider the univariable Poisson model $(B.1)$ with $\boldsymbol{T}_i = (T_{i1}, T_{i2})'$ a vector of two continuous covariates from trial *i*. Suppose that covariate T_2 has missing data for trial *i*. The idea behind the approach is to predict the missing data for the covariate T_{i2} through a linear regression on the covariate T_{i1} and then incorporate the "imputed" covariate T_{i2} in the Poison regression model $(B.1)$. We model covariate the true value, δ_{i2}, of T_{i2} as conditional on T_{i1}, being drawn from a conditional normal distribution $T_{i2}|T_{i1} \sim N(\delta_{i2}, \tau_2^2)$, such that:

$$\delta_{i2} = \gamma_0 + \gamma_1 \cdot T_{i1} \qquad (B.4)$$

γ_0 is the intercept and γ_1 is the regression coefficient that expresses the degree of change in the conditional mean δ_{i2} for each one-unit increase in T_{i1}. For the regression coefficients and the summary means we can apply informative normal distributions centred at zero with small precision, such as 0.0001 and for the standard deviation parameters we can apply uniform distribution, $U(0,10)$. In case the missing covariate is dichotomous, then the formula $(B.4)$ is written as $\log(p_{i2}) = \gamma_0 + \gamma_1 \cdot T_{i1}$ and covariate value T_{i2}, conditional on T_{i1}, is drawn from a binomial distribution $T_{i2}|T_{i1} \sim Bin(p_{i2}, n_i)$, where p_{i2} is the conditional probability that $T_{i2} = 1$ given the covariate T_{i1}.

The WinBUGS code below is for the univariable Poisson regression model $(B.1)$ adjusting for average age. Average age has missing data in seven trials.

```
model{
    for(i in 1:ns){
        lamda.new[i] <- lamda[i]*L[i]
        # As L_i we define the product of the randomized study size n_i and the duration in
        # weeks D_i of study i. This is the total number of patient-week of follow-up. With
        # lamda[i] we denote the rate of dropouts per patient-week in study i. Therefore,
        # lamda.new[i] is the total number of patients who dropped out of trial i.

        ## The Poisson likelihood

        d[i] ~ dpois(lamda.new[i])

        ## The Poisson meta-regression model

        log(lamda[i]) <- lamda.star[i] + b*(age[i] - 21)
        # The average age is centred at the minimum average age to improve interpretability
        # of the  regression coefficient.

        lamda.star[i] ~ dnorm(m,prec)
        # Parameter lamda.star[i] is the dropout rate per patient-week in study i adjusted for a
```

```
        # covariate matrix T_i and it is drawn from a normal distribution with mean m and
        # precision prec.
    }

    RR <- exp(b*10)
    # Risk ratio when age increases by 10
    rate.min <- exp(m)
    # Dropout rate by patient-week in the 21 years (minimum age)
    rate.incr <- exp(m + b*10)
    # Dropout rate by patient-week in the 31 years

    ## Predicted log(dropout rates per patient-week)

    predDR.min ~ dnorm(m,prec)
    # Predicted log dropout rate per patient-week in the 21 years

    x <- m + b*10
    predDR.incr ~ dnorm(x,prec)
    # Predicted log dropout rate per patient-week in the 31 years

    ## Specification of covariate densities - Covariate regression models

    for(i in 1:ns){

        age[i] ~ dnorm(m_age[i],tau_age)
        # Average age
        m_age[i] <- gamma20 + gamma2[1]*allocation[i]

        allocation[i] ~ dbin(p_alloc[i],n[i])
        # Allocation concealment
        logit(p_alloc[i]) <- gamma10

    # The covariate regression models have been organized in this hierarchy starting with the
    # covariate that has missing data and ending to the covariate without missing data. The
    # covariate of interest (here, the average age) is the first in this hierarchy. Average age is
    # conditional on the allocation concealment (the covariate of the next level).
    }

    ## Prior distributions

    b ~ dnorm(0,.0001)
    m ~ dnorm(0,0.0001)
    prec <- 1/pow(sd,2)
    sd  ~ dunif(0,10)
    tau_age <- 1/pow(sd_age,2)
    sd_age ~ dunif(0,10)

    gamma10 ~ dnorm(0,.0001)
    gamma20 ~ dnorm(0,.0001)
    # The intercepts of the corresponding covariate regression models

    for(k in 1:1){
        gamma2[k] ~ dnorm(0,.0001)
        # The regression coefficients of the corresponding covariate regression models
    }
}
```

The WinBUGS code for the univariable multinomial logistic regression model $(B.3)$ adjusted for the average age is given below

```
model{
    for(i in 1:ns){
        log(phi[i,1]) <- 0
        w[i,1] <- 0
        y[i,1:J] ~ dmulti(p[i,1:J],n[i])
        # The vector y[i,1:J] represents the three (J=3) patient groups that are drawn from a
        # multinomial distribution.

        p[i,1] <- 1/sum(phi[i,])
        # The likelihood that a patient will be completer

        for(j in 2:J){
            p[i,j] <- phi[i,j]/sum(phi[i,])
            # The likelihood that a patient will be LOCF or excluded.

            log(phi[i,j]) <- theta[i,j]
            # The unadjusted log-RR that a patient will be LOCF or excluded than completer.

            theta[i,j] <- theta.star[i,j] +  q[j]*(age[i] - 21)
            # The multinomial logistic regression adjusted for the average age. The
            # regression coefficient for each comparison is denoted by the parameter q[j].

            theta.star[i,j] ~ dnorm(md[i,j],taud[i,j])
            # The adjusted log-RR of being excluded than completer conditional on the
            # log-RR of being LOCF than completer is drawn from a normal distribution with
            # conditional mean md[i,3] and conditional precision taud[i,3].

            md[i,j] <-  md.star[j] + sw[i,j]
            # The unconditional mean is defined as  md.star [2] and md.star [3] for the
            # log-RR of being LOCF or excluded than completer, respectively

            taud[i,j] <- tau.star*2*(j-1)/j
            # The unconditional heterogeneities tau.star are assumed independent on the
            # comparison.

            w[i,j] <- theta.star[i,j] - md.star [j]
            # Adjustment for the log-RRs.

            sw[i,j] <- sum(w[i,1:j-1])/(j-1)
            # Cumulative adjustment for the log-RRs across the comparisons.
        }
    }

    for(j in 2:J){
        RRR[j,1] <- exp(q[j])
        # The RRR of the LOCF or excluded against the completers
    }

    ## Specification of covariate densities - Covariate regression models

    for(i in 1:ns){
        age[i] ~ dnorm(m_age[i],tau_age)
        # Average age
        m_age[i] <- gamma20 + gamma2[1]*allocation[i]


        allocation[i] ~ dbin(p_alloc[i],n[i])
        # Allocation concealment
        logit(p_alloc[i]) <- gamma10
```

```
    }

    ## Prior distributions

    for(j in 2:J){
        q[j] ~ dnorm(0,.0001)
        md.star[j] ~ dnorm(0,.0001)
    }

    tau.star <- 1/pow(sd.star,2)
    sd.star ~ dunif(0,10)
    tau_age <- 1/pow(sd_age,2)
    sd_age ~ dunif(0,10)
    gamma10 ~ dnorm(0,.0001)
    gamma20 ~ dnorm(0,.0001)
    # The intercepts of the corresponding covariate regression models

    for(k in 1:1){
        gamma2[k] ~ dnorm(0,.0001)}
        # The regression coefficients of the corresponding covariate regression models
    }
}
```

B.6 Reference List

1. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A Stat Soc. 2009 Jan; 172(1):137-159.

2. Borestein M, Hedges LV, Higgins JP, et al. Introduction to Meta-Analysis. John Wiley & Sons, Ltd; 2009.

Appendix C

The WinBUGS code for the treatment-specific IMOR model

The WinBUGS code below is for the treatment-specific IMOR model

```
model{
        for(k in 1:ns){
            w[k,1] <- 0
            beta[k,t[k,1]] <- 0
            # As index t[k,1] we define the treatment in the first arm (control arm) of trial k

            logit(pi[k,t[k,1]]) <- beta0[k]
            # The log odds of success in the control arm of trial k.
            beta0[k] ~ dnorm(0,.0001)
            # Non-informative priors for the logOR of success in the control arm.

            ## The multinomial likelihood

            for(i in 1:na[k]){
                y[k,i,1:J] ~ dmulti(prob[k,t[k,i],1:J],n[k,i])
                n[k,i] <- sum(y[k,i,])
                # With y[k,I,1:J] we denote the vector that comprises of the observed success,
                # the observed failure and the completely missing outcomes (in failure and
                # success) in arm I of trial k, with prob[k,t[k,i[,1:J] the vector with the
                # probability for observed success, observed failure and completely missing
                # outcomes (in failure or success), respectively, and with n[k,i] the sample
                # size.

               for(j in 1:J){
                    prob[k,t[k,i],j] <- 0.5*(3-j)*(2-j)*pi[k,t[k,i]]*(1-alpha1[k,t[k,i]]) +
                                        (j-1)*(3-j)*(1-pi[k,t[k,i]])*(1-alpha0[k,t[k,i]]) +
                                        0.5*(1-j)*(2-j)*(pi[k,t[k,i]]*alpha1[k,t[k,i]] +
                                        (1-pi[k,t[k,i]])*alpha0[k,t[k,i]])
                    # With pi[k,t[k,i]] we denote the overall probability of success (across
                    # observed and completely missing participants) in treatment t[k,i] of the
                    # arm I of trial k. The parameters alpha0[k,t[k,i]] and alpha1[k,t[k,i]] denote
                    # the probability that a failure is completely missing and the probability that
                    # a success is completely missing, respectively.
                }
            }

            for(i in 2:na[k]){
                logit(pi[k,t[k,i]]) <- beta0[k] + beta[k,t[k,i]]
                # The log odds of success in the experimental arm I of trial k.
```

```
        beta[k,t[k,i]] ~ dnorm(md[k,t[k,i]],taud[k,t[k,i]])
        # The trial-specific log IR distribution.
        md[k,t[k,i]] <- d[t[k,i]] - d[t[k,1]] + sw[k,i]
        # Mean of log OR distribution.
        taud[k,t[k,i]]<-tau*2*(i-1)/i
        # Precision of log OR distribution.
        w[k,i] <- (beta[k,t[k,i]] - d[t[k,i]] + d[t[k,1]])
        # Adjustment for multi-arm trials.
        sw[k,i]<-sum(w[k,1:i-1])/(i-1)
        # Cumulative adjustment for multi-arm trials.
    }
}

d[ref] <- 0
for(i in 1:(ref-1)){
    for(i in (ref+1):nt){
        d[i] ~ dnorm(0,.0001)
        # Non-informative priors for basic parameter of inconsistency equation.
     }
}

for(i in 1:nt){
    order[i]<- nt + 1 - rank(d[],i)
    # The order of treatment i when the outcome is positive. Omit 'nt + 1 -' when the
    # outcome is negative.
    most.effective[i] <- equals(order[i],1)
}

for(i in 1:nt){
    for(j in 1:nt){
        effectiveness[i,j] <- equals(order[i],j)
        # The ranking probability of treatment i to be at the j order.
        cumeffectiveness[i,j] <- sum(effectiveness[i,1:j])
        # The cumulative ranking probability of treatment i to be among the j best
        # treatment.
     }
 }

 for(i in 1:nt){
     SUCRA[i] <- sum(cumeffectiveness[i,1:(nt-1)])/(nt-1)
     # The surface under the cumulative rankings for treatment i.
 }

for(i in 1:nt){
    OR.ref[i] <- exp(d[i] - d[ref])
    # The OR of success between the treatment i and the reference treatment.
}

for(i in 1:nt){
    for(r in 1:nt){
```

```
            OR[i,r] <- exp(d[i] - d[r])
            # The OR of success between the treatment i and r.
        }
    }

    for(k in 1:ns){
        for(i in 1:na[k]){
            delta[k,t[k,i]] <- delta.treat[t[k,i]]

    ## Parameterization of alpha0 and alpha1

            logit(alpha1[k,t[k,i]]) <- gamma[k,t[k,i]] + 0.5*delta[k,t[k,i]]
            logit(alpha0[k,t[k,i]]) <- gamma[k,t[k,i]] - 0.5*delta[k,t[k,i]]
            gamma[k,t[k,i]] ~ dnorm(0,gamma.prec)
            # With gamma[k,t[k,i]] we denote the average completely missingness across
            # successes and failures.
        }
    }

    for(l in 1:nt){
        delta.treat[l] ~ dnorm(delta.mean,delta.prec)
        # Treatment-specific log IMOR distribution.
    }

    sd ~ dnorm(0,1)I(0,)
    # Non-informative prior for random-effects standard deviation.
    tau <- 1/pow(sd,2)
}
```