UNIVERSITÀ DEGLI STUDI DI MILANO

Scuola di dottorato in scienze biomediche, cliniche e sperimentali

Dipartimento di scienze cliniche e di comunità

Dottorato di ricerca in statistica biomedica

TESI DI DOTTORATO DI RICERCA

NETWORK META-ANALYSIS: A NOVEL APPROACH BASED ON A HIERARCHICAL DATA STRUCTURE

Ciclo XXVIII – Settore scientifico disciplinare MED/01

Dottoranda

Dottoressa Teresa Greco

Coordinatore del Dottorato e Tutor

Professor Adriano Decarli

Correlatori

Professor Giovanni Landoni

Dottoressa Valeria Edefonti

2015

SUMMARY

INTRODUCTION

Meta-analysis is a powerful tool to cumulate and summarize the knowledge in a research field through statistical instruments, and to identify the overall measure of a treatment's effect by combining several study-specific results. However, it is a controversial tool, because even small violations of certain rules can lead to misleading conclusions. Pooling data through meta-analysis can create problems, such as non linear correlations, multifactorial rather than unifactorial effects, limited coverage, or inhomogeneous data that fails to connect with the hypothesis. In this work we provided and discussed methods to overcome the limits of standard (univariate) meta-analysis, focusing on the ability to cope with multiple treatments and to deal with correlated data where correlation can derive from multiple endpoints, time-varying responses or from clustered observation.

In the first chapter we explore the principal steps (from writing a prospective protocol of analysis to results' interpretation) in order to minimize the risk of conducting a mediocre meta-analysis and to support researchers to accurately evaluate the published findings.
The second chapter represents an overview of conceptual and practical issues of a network meta-analysis. We start from general considerations on network meta-analysis to specifically appraise how to collect study data, structure the analytical network, and specify the requirements for different models and parameter interpretations. Specifically, we outline the key steps, from literature search to sensitivity analysis, necessary to perform a valid network meta-analysis on binomial data.
In the third party of this work, we focus our attention on data which can be analyzed with a binomial model applying the Bayesian hierarchical approach and using Markov Chain Monte Carlo approach. We also apply this analytical approach to a case study on the beneficial effects of anesthetic agents in order to further clarify the statistical details of the models, diagnostics, and computations.

In the fourth chapter we propose an alternative frequentist approach to estimate consistency and inconsistency models for a network meta-analysis. We discuss the *multilevel network meta-analysis* which include a three-level data structure: subjects within studies at the first level, studies within study designs at the second level and design configuration at the third level. We discuss multilevel modeling which may be carried out within widely available statistical programs such as SAS software, and we compare the results of a published Bayesian network meta-analysis on a binary endpoint which examines the effect on mortality of desflurane, isoflurane, sevoflurane, and total intravenous anaesthetics at the longest follow-up available.

In the final chapter we compare the Bayesian and the novel frequentist-multilevel approach in performing network meta-analysis on publicly available data and we investigate the descriptive characteristics that may contribute to decrease or increase the potential difference between the estimates derived from the two approaches. The two approaches were compared in terms of the difference between the pooled estimates or their standardized values, and of the Euclidean distance.

CHAPTER 1

LIMITS OF STANDARD META-ANALYSIS

1.1 CRITICAL ISSUES AROUND META-ANALYSES: AN INTRODUCTION

The statistical origin of meta-analysis reaches back to the 17th century when intuitions and experiences, in astronomy, suggest that combinations of data might be better than attempts to select amongst them. Karl Pearson [1] was probably the first medical researcher to report the use of formal techniques to combine data from different studies when examining the preventive effect of serum inoculations against enteric fever. He analyzed data comparing infection and mortality among soldiers who had volunteered for inoculation against typhoid fever in various places across the British Empire with that of other soldiers who had not volunteered. All individual estimates were presented for the first time in a table, together with the pooled estimate. Karl Pearson appears to have been the first to analyse clinical trial results using meta-analysis. He was especially thorough about questioning the consistency of individual trial results and equally keen to discover clues from this for better future research. However, a method for uncertainty estimation had not yet been identified. Although such techniques would be widely ignored in medicine for many years to come [2], social sciences, especially psychology and educational research, showed particular interest in them. Indeed, in 1976 the psychologist Gene Glass [3] coined the term “meta-analysis” in a paper entitled “Primary, Secondary and Meta-analysis of Research”, to help make sense of the growing amount of data in literature. A meta-analysis was defined as analysis of analysis. This is a powerful tool to summarizing several individual result into an overall measure of treatment effect.

Since the 80s, the amount of information generated by meta-analyses grew constantly, up to the point of becoming overwhelming. A PubMed search of the word “meta-analysis” in the title or in the abstract yielded 55,986 hints (update at August 11th, 2014). The 30% of them only in the years 2013 and 2014.

Meta-analysis is a powerful tool to cumulate and summarize the knowledge in a research field through statistical instruments, and to identify the overall measure of a treatment's effect by combining several individual results [4]. However, it is a controversial tool, because several conditions are critical and even small violations of these can lead to misleading conclusions. In fact, several decisions made when designing and performing a meta-analysis require personal judgment and expertise, thus creating personal biases or expectations that may influence the result [5,6].

As statistical means of reviewing primary studies, meta-analyses have inherent advantages as well as limitations [7]. Pooling data through meta-analysis can create problems, such as non linear correlations, multifactorial rather than unifactorial effects, limited coverage, or inhomogeneous data that fails to connect with the hypothesis. Despite these problems, the meta-analysis method is very useful: it establishes whether scientific findings are consistent and if they can be generalized across populations, it identifies patterns among studies, sources of disagreement among results, and other interesting relationships that may emerge in the context of multiple studies.

1.2 META-ANALYSIS' PROTOCOL REGISTRATION

It is important to write a prospective analysis' protocol, which specifies the objectives and methods of the meta-analysis. Having a protocol can help restrict the risk of biased post hoc decisions in methods, such as selective outcome reporting.

The PRISMA (Preferred Reporting Items Systematic Reviews and Meta-Analysis) guidelines [8] recommend the prior registration of the protocol of any systematic review and meta-analysis, requiring that this protocol should be made accessible before any hands-on work is done. The prior registration (i.e. through PROSPERO - International prospective register of systematic reviews) should prevent "the risk of multiple reviews addressing the same question, reduce publication bias, and provide greater transparency when updating systematic reviews". It is also true that meta-analyses are published only after passing through at least two steps: peer reviews and an editorial decision. These filters may be sufficient to decide whether a meta-analysis is good and novel enough to deserve

publication. Takkouche B et al. [9] stated that an additional committee or register does not increase the quality of what is published but it only increases bureaucracy.

In a recent letter on British Medical Journal, Krumholz H et al. [10] showed how different approaches on same patient level clinical trial data can bring to interpret the data differently, and emphasise different points. Krumholz H et al. assert the redundancy should be welcomed but it is important clarify similarities and contrasts with previous publications.

Rigorous meta-analyses undertaken according to standard principles (pre-specified protocol, comparable definitions of key outcomes, quality control of data, and inclusion of all information available) will ultimately lead to more reliable evidence on the efficacy and safety of interventions than either retrospective meta-analysis [11].

1.3 IDENTIFICATION AND SELECTION OF STUDIES

The first reason to criticize the meta-analytic method is that it provides evidence extracted and integrated from a number of primary studies, not from a random sampling; thus, results cannot lead to test relations such as causality [12]. However, meta-analysis may lead to support or rejection of the generalization of primary evidence, and may contribute to direct future research in a field. Moreover, meta-analysis results can improve understanding but sometimes they may not be very helpful in clinical practice. In this context, the definition of the scientific start-point (population and intervention) is crucial: the clinical question can either be broad or very narrow. Broad inclusion criteria could increase the heterogeneity between studies, making it difficult to apply the results to specific patients; narrow inclusion criteria make it hard to find pertinent studies and to generalize the results in clinical practice. Hence, the researcher should find the right compromise, focusing on the benefits for the patient.

One of the aims of meta-analysis is to take into account all the available evidence from multiple independent sources to evaluate a hypothesis [6]. However, meta-analysis usually includes only a small fraction of the published information, often derived from a small range of methodological designs (i.e. meta-analysis restricted to randomized clinical trials or to English languages). It is also true that with limited resources it is impossible to identify all the

evidence available in the literature. Systematic reviews, in contrast to traditional narrative reviews, require an objective and a reproducible search of a series of sources to identify as many relevant studies as possible [13]. The search strategy should be comprehensive and sensitive; searching more than one computerized database is strongly recommended. Commonly searched databases are: MEDLINE, including PubMed, The Cochrane Central Register of Controlled Trials (CENTRAL), and EMBASE. These databases are available to individuals free of charge, on a subscription or on a 'pay-as-you-go' basis. They can also be available free of charge through national provisions, professional organization or site-wide licenses at institutions such as universities or hospitals. There are also regional electronic bibliographic databases that include publications in local languages [13]. Additional studies can be identified employing the "backward snowballing" (i.e. scanning of references of retrieved articles and pertinent reviews) or investigating the "grey literature", namely the literature that is not formally published in sources such as books or journal articles (i.e. personal communications, conferences, abstracts, etc). Authors often provide supplementary data, not included in the original publications or relative to unpublished studies.

Decisions regarding what primary evidence to include in a meta-analysis depend on evidence availability. Practical problems, regarding access to primary data, include studies published in languages foreign to the researcher and evidence available only confidentially or in the "gray literature" of congress and dissertations. Similar issues are faced by analysts who want to perform a meta-analysis with individual patient data (which has several advantages over analysis on aggregate-level data [14]), since patient-level data is often confidential or protected by corporate interests.

Moreover, many other biases linked to study selection may influence the estimates and the interpretation of findings: citation bias, time-lag bias and multiple publications bias [13]. To overcome these biases, several tools are available. For example, the sensitivity analysis can spot bias by exploring the robustness of the findings under different assumptions.

1.4 QUALITY OF INCLUDED STUDIES

The conclusions of a meta-analysis depend strongly on the quality of the studies identified to estimate the pooled effect [15]. The internal validity may be affected by errors and incorrect evaluations during all the phases of a clinical trial (selection, performance, attrition, and detection bias [16]), so the assessment of the risk of study bias is a central step when one carries out a meta-analysis. The quality of randomized clinical trials should be evaluated with regard to randomization, adequate blinding and explanation for dropouts and withdrawals, which addresses the issues of both internal validity (minimization of bias) and external validity (ability to generalize results) [17]. The information gained from quality assessment is fundamental to determine the strength of inferences and to assign grades to recommendations generated within a review. The main problem during the quality assessment process is the inconsistent base for judgment: if the studies were re-examined, the same trained investigator might alter category assignments [6]. The investigator may also be influenced (consciously or unconsciously) by other unstated aspects of the studies, such as the prestige of the journal or the identity of the authors [6]. The published work can and should explain how the reviewers made these judgments, but the fact remains that these approaches can suffer from substantial subjectivity. Indeed, it is strongly recommended that reviewers use a set of specific rules to assign a quality category, aiming for transparency and reproducibility.

1.5 BIAS

1.5.1 PUBLICATION BIAS

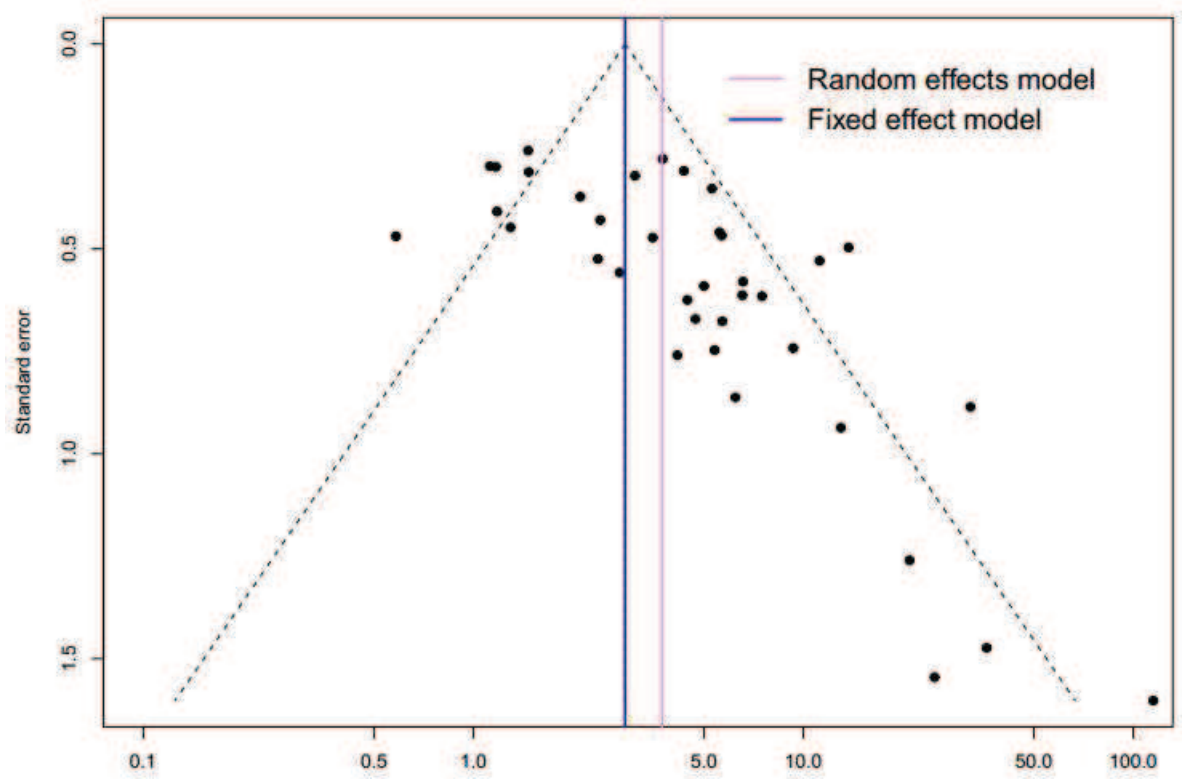
The biggest potential source of type I error (increase of false positive results) in meta-analysis is probably the publication bias [15]. This occurs when, in clinical literature, statistically significant “positive” results have either a better chance of being published, are published earlier or in journals with higher impact factors, and are more likely to be cited by others [18]. The graphical representation to evaluate the presence of publication bias is the funnel plot. In a funnel plot, effect size is plotted versus a measure of its precision, such as sample size. If no publication bias were present, we would expect that the effect size of each

included study to be symmetrically distributed around the underlying true effect size, with more random variation of this value in smaller studies. Asymmetry or gaps in the plot are suggestive of bias, most often due to studies which are smaller, non-significant or have an effect in the opposite direction from that expected, having a lower chance of being published [15]. Therefore, it is important to note that conclusions exclusively based on published studies can be misleading. Methods as Trim and Fill [19] allow estimation of an adjusted meta-analysis in the presence of publication bias.

1.5.2 SMALL-STUDY EFFECT

The small-study effect occurs when small studies have systematically different effects from the large ones. It has often been suggested that small trials tend to report larger treatment benefits than larger trials [20,21]. Such small-study effects can result from a combination of lower methodological quality of small trials or publication bias (small studies with negative effects are unpublished or less accessible than larger studies) or other reporting biases [15]. However, this effect could also reflect clinical heterogeneity, if small trials were more careful in selecting patients, so that a favorable outcome of the experimental treatment might be expected [22]. Researchers that are worried about the influence of small-study effects on the results of a meta-analysis in which there is evidence of between-study heterogeneity (I-square >0) should compare the fixed- and random-effects estimates of the treatment (figure 1.1). If the estimates are similar, then any small-study effect has little influence. If the random-effects estimate is more beneficial, researchers should consider whether it is reasonable to conclude that the treatment was more effective in smaller studies. This is because the weight given to each included study through the random effect model is less influenced by the sample size than that given by means of the fixed effects model. In the eventuality the small-study effect is present, the researcher should consider analyzing only large studies (if these tend to be conducted with more methodological stringency [23]). One must note that if there is no evidence of heterogeneity between studies, the fixed- and random-effects estimates will be identical, so there will be an actual difficulty in identifying the small-study effect. [13]

Figure 1.1: Example of small-study effect (Moore RA, BMJ 1998). Meta-analysis of 37 placebo-controlled randomized trials on the effectiveness and safety of topical non-steroidal anti-inflammatory drugs in acute pain.



1.6. DATA ANALYSIS

1.6.1 HETEROGENEITY

The degree of heterogeneity is another important limitation, and the random effects model should be used during the data analysis phase to incorporate in the treatment effect the identifiable or non-variability between-studies [24]. It is fundamental to observe that exploring heterogeneity in a meta-analysis should start at the stage of protocol writing, by identifying a priori which factors are likely to influence the treatment effect. Visual inspection of the meta-analysis plots may show whether the results of a subgroup of studies have the same overall direction of the treatment effect. One should pay attention to meta-analysis in which results have a discordant treatment effect for groups of studies and no explanation of variance has been done. Sources of variation should be identified and their impact on effect size should be quantified using statistical tests and methods, such as

analysis of variance (ANOVA) or weighted meta-regression [15]. Actually, when high heterogeneity is evident, individual data should be not pooled and definitive conclusions should be drawn when more studies become available.

1.6.2 RARE EVENTS

Meta-analysis makes it possible to look at events that were too rare in the original studies to show a statistically significant difference. However, analyzing rare events represents a problem because small changes in data can determine important changes in the results and this instability can be exaggerated by the use of relative measures of effect instead of absolute ones. To overcome this problem several methods have been proposed [13,25,26].

1.6.3 AGGREGATE DATA ANALYSIS

Another problem that affects meta-analysis carried out with aggregate data, is the ecological fallacy that arises when the averages of the patient's features fail to properly reflect the individual-level association [27]. The best scenario is when data at an individual-level is available, but it is equally true that there is resistance from authors to allow ready access to their own dataset containing individual patient data. Very often aggregate data is the only information offered.

1.6.4 MULTIPLE SUBGROUP ANALYSIS

Finally, it is essential to spend a few moments discussing the common problem that occurs when one wants to perform multiple subgroup analysis, according to multiple baseline characteristics, and then examine the significance of effects not set a priori into the protocol. Testing effects suggested by data and not planned a priori considerably increase the risk of false-positive results [28]. To minimize this error it is important to identify the effects to test before data collection and analysis [5]; otherwise, one may adjust the p-value according to the number of analysis performed. In general, post hoc analysis should be deemed exploratory and not conclusive.

1.7 DISCUSSION

Important decisions in a systematic review are often based on understanding the medical domain and not the underlying methodology. The clinical question must be relevant to clinicians and the outcomes must be important for patients. Efforts are made to avoid bias by including relevant research, using adequate statistical methodology and interpreting results based on the context and available evidence. Published reports should include quality criteria and should describe the selected tools and their reliability and validity. The synthesis of the evidence should reflect the a priori analytic plan including quality criteria, regardless of statistical significance or the direction of the effect. Published reviews should also include justifications of all post hoc decisions to synthesize evidence. Organizing and carrying out a meta-analysis is hard work, but the findings can be significant. In the best-case scenario, by revealing the magnitude of effect sizes associated with prior research, meta-analysis can suggest how future studies might be best designed to maximize their individual power. On the other hand, low-powered analysis based on a small number of studies can still provide useful insights by revealing publication bias through a funnel plot or highlighting a deficiency in a particular topic that deserves further attention.

Meta-analysis represents a powerful way to summarize data and effectively increase sample size to provide a more valid pooled estimate. However, the results of a meta-analysis should be interpreted in the light of the various checks previously discussed in this work, which can inform the readers of the likely reliability of the conclusions.

CHAPTER 2

THE NETWORK META-ANALYSIS

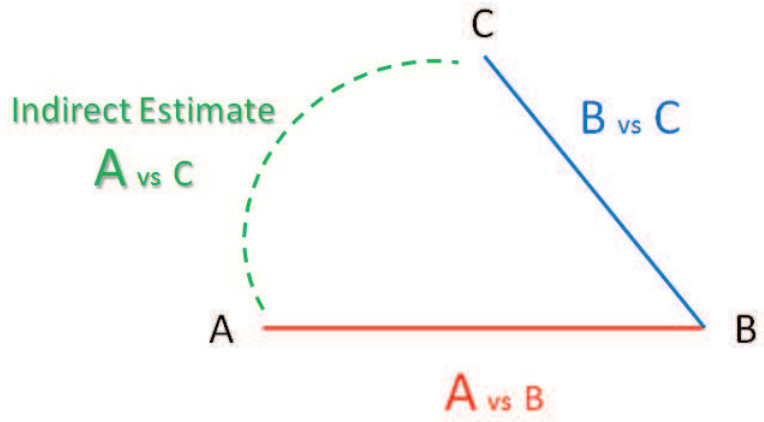
2.1 SCOPE AND BACKGROUND

The search for accurate and reliable sources of evidence represents a ongoing challenge in medicine, as only a comprehensive yet synthetic analytical effort can accurately guide clinical decision making. Any single empirical observation on the apparent relationship between events and exposures or between events and interventions may provide useful information [29], but systematic reviews and meta-analyses of large high quality randomised controlled trials (RCT) with low heterogeneity represent the highest degree of evidence, as they offer increased precision and external validity [30]. Moreover, meta-analyses offer a quick and cost-effective method to gather information for clinical decision making. When head-to-head treatment comparisons are not available or conclusive, the limitations of standard (i.e. pairwise) meta-analyses can be overcome by network meta-analyses (NMA, i.e. mixed treatment comparisons [MTC]), which can provide estimates of treatment efficacy or safety of multiple treatment regimens. Different treatment strategies are analyzed by statistical inference methods rather than simply summing up trials that evaluated the same intervention compared to another intervention, standard care, or placebo. If a first trial compares drug A to drug B, showing that drug A is significantly superior to drug B, and a second trial investigates the same or a similar patient population comparing drug B versus drug C (demonstrating that drug B is equivalent to drug C), NMA may allow to infer that drug A is also potentially superior to drug C for this given patient population, even though there was no direct test of drug A against drug C (figure 2.1).

Specifically, if two particular treatments have never been compared against each other but have been compared to a common comparator, then an adjusted indirect treatment comparison (ITC) can exploit the direct effects of the two treatments versus the common comparator to estimate the indirect treatment effect [31-33]. In such perspective, ITC

represents the simplest type of NMA or MTC. In addition to this, it is important to understand that both direct and indirect information provide data for evidence synthesis, and thus any NMA is inherently more efficient and accurate.

Figure 2.1: Example of indirect effect estimation.



As pairwise meta-analysis, NMA is accurate and clinically useful only when it combines studies that are similar enough to be grouped, with the aim to explore and limit as much as possible the sources of variability while concomitantly maximizing the statistical precision. The results obtained from the combination of direct and indirect estimates may also strengthen the validity within comparison [34]. Even when the results of the direct evidence are conclusive, merging them with the results of indirect estimates in MTC may give a more accurate estimate optimizing the existing information of the network. [35,36].

The pioneering work by Thomas Lumley [35] presented the first methodology to perform a meta-analysis for direct and indirect comparisons and proposed the term “network meta-analysis”. In this work, Lumley detailed the approach to using potentially very complex networks of treatment comparisons to detect inconsistency (or incoherence) between randomized trials, to estimate treatment differences and to assess the uncertainty in these estimates. Moreover, Lumley suggested the application of Bayesian approaches, to model both heterogeneity between treatments and the underlying inconsistency, due to their flexibility.

The extension to handle multi-armed trials in the classical field was faced from authors such as Salanti et al. [37], Jackson et al. [38], White et al. [39,40] and Higgins et al. [41]. Salanti and collaborators [37] detailed the general set-up for NMA with both arm-based, where the effect measures are reported for each arm (i.e. odds, absolute risk, hazard or mean), and contrast-based models, where results are presented as the difference in effect between arms (i.e. odds ratio, risk ratio, hazard ratios or mean difference). However, this paper left open some important topics such as the quantification of inconsistency, the evaluation of bias and the development of a user-friendly software to NMA models. White and colleagues [40] have updated a STATA (College Station, TX, USA) command, *mvmeta,* to perform a multivariate meta-regression and obtain suitable difference effect estimates. Jackson, Riley and White [38] explored the potential of the multivariate model for fitting a network data structure adopting a two-stage approach to analysis. The trial-specific parameter of interest and the variance-covariance matrix are obtained at the first stage and then these estimates are combined at the second stage. In this case, the aggregate input data are managed as contrast-level summaries, namely as the relative difference in effect between arms (i.e. odds ratio, risk ratio, hazard ratio, or mean difference). White et al. [39] and Higgins et al. [41] review the meaning of inconsistency, best modeled by a design-by-treatment interaction, and the method to fit both consistency and inconsistency models.

On the other hand, Lu and Ades [36] proposed an alternative Bayesian approach to make NMA for multi-arm studies by including both direct and indirect comparisons. Moreover, they explored results from Markov Chain Monte Carlo (MCMC) algorithm to set up a strategy for selecting the best treatment regimen.

2.2 NETWORK EVIDENCE

The evidence of the network must include all randomised clinical trials of relevant treatments (interventions, drugs, or procedures) that have been compared directly in a reasonably similar patient and diagnostic setting. The inclusion of all relevant evidence in systematic reviews is crucial to avoid bias and maximize precision [16]. The literature search for a NMA applies the same basic standards exploited for pairwise comparisons. Indeed, a

standard meta-analysis involves a single search for any trial that compares the treatment of interest with any other therapy, that may be theoretically exploited also in a NMA focusing on differences in treatment effects.

However, the choice of treatments to include in a NMA is more challenging. Since the literature search is time consuming and requires resources, one may decide that it is not worthwhile to search for all the possible indirect evidences. Hawkins et al. [42] suggest an efficient search strategy to identify clinical trials that may provide indirect evidence when comparing different treatment comparators: a series of iterative searches where the set of comparators included in each search is dependent on the results of the previous one. This iterative process continues or stops considering the marginal cost of searching for higher order indirect data and the marginal benefit of progressively less informative data. If the search is stopped before finding all the entire evidence, the missed treatments are assumed as missing at random [43], but it is important to pay attention to the applicability of results.

2.3 CONFIGURATION OF NETWORK

Before starting a NMA, it is important to have a complete view of the distribution of included studies. The network diagram allows an intuitive approach to symbolically represent all the direct comparisons among treatments. This graph consists of a set of nodes representing the interventions linked by lines that depict how many RCT have been included. Two important properties of network configuration are geometry and asymmetry [37,44]. Geometry refers to the overall structure of treatment contrasts, while asymmetry summarizes the amount of data for a specific comparison.

The network structure must be carefully built and examined so that each pattern of data may be used to reveal particular characteristics that may assist in the choice of the analytical method [33]. For example, the diagram in figure 2.2.a (star-shaped) allows an ITC analysis of treatments B, C and D all linked to the common comparator A. The graph in figure 2.2.b comprises three nodes representing three interventions (A, B, C) and three edges (arrows). An important property of this network is that each contrast has both direct and indirect evidence (closed loops). For example, the BC comparison obtains direct information from

trials that compare BC and indirect evidence from trials that analyse AB and AC treatment differences. Moreover, the structure of the network can become extremely complex as in figure 2.2.c (multiple loops). A network where some pairwise contrasts have both direct and indirect evidence can be analyzed performing an MTC analysis [34,36]. Actually, all connected networks can be examined using NMA, as even pairwise meta-analysis is only a special case of NMA, and can be analyzed using the exact same model.

Figure 2.2: Examples of network configurations.

Figure 2.2.a

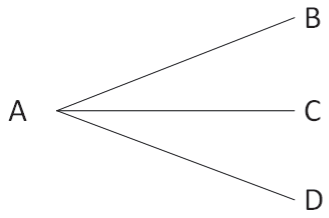


Figure 2.2.c

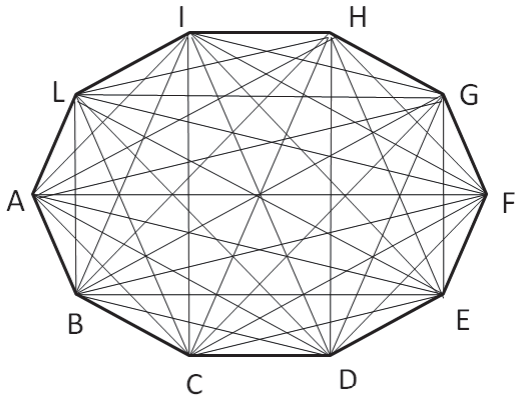


Figure 2.2.b

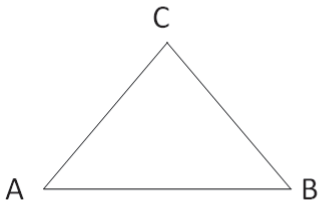
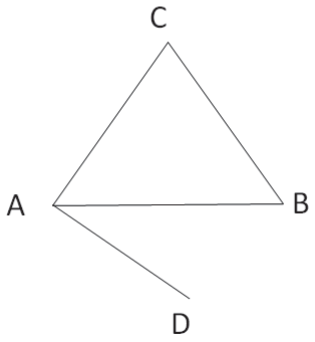


Figure 2.2.d



Network configuration can help to establish which treatment can be defined as reference (usually the one that appears more frequently than others, or the one which is most commonly used in clinical practice) or to show the head-to head comparative relations. In the network structure, the effect estimates may be included, with the corresponding 95%

confidence, credibility or credible intervals, the number of studies and the study reference for each pairwise comparisons.

Focusing on diagram asymmetry, it is crucial to understand the extent to which different nodes or links are present in the diagram, weighting for the number of trials. Salanti et al. [37,44] propose metrics and tests employed in the ecological literature [45]. Specifically, they suggest to investigate the diversity and the co-occurrence inherent in the network [46,47]. Diversity refers to the number of nodes contained in a network and to different frequencies in treatments. Conversely, co-occurrence is measured with the C-score [48] or other widely cited similarity scores (e.g. Jaccard, Dice, or Cosine coefficients) [49], and represents the tendency of a particular comparison to occur more frequently than expected by chance. In other words it tests for the presence of favorite couples of treatment.

2.4 ASSUMPTIONS

2.4.1 PRESERVATION OF THE RANDOMISATION PROCESS

When there is no or insufficient evidence from direct comparison trials, it may be possible to use results of different studies to obtain the pooled estimates of relative treatment effect. One fundamental assumption is the preservation of the randomisation process within each trial, comparing the estimates of relative effect among treatments. Let us suppose that we have three treatments (A, B, C) compared head-to-head in *N* trials. Treatment estimates will not be accurate if the researcher only calculates the indirect effect estimate of B versus C, by AB and AC trials, by balancing the observed fraction of respondents on treatment B from AB trials to the observed fraction of respondents on treatment C from AC trials. In fact, in this way, the analysis fails to separate the treatment effects from the other sources of variability [50,51]. However, one can compare the (log) odds ratio of A versus B from the AB trials to the (log) odds ratio for A versus C from AC trials [33]. This indirect comparison, adjusted according to the results of their direct comparison with a common comparator, largely preserves the force and validity of the randomised trials [54].

2.4.2 HOMOGENEITY AND CONSISTENCY

Both homogeneity (no variation in treatment effect between trials within pairwise contrasts) and consistency (no variation in treatment effect between pairwise contrasts) ensure the validity of the analysis. Evaluation of heterogeneity represents another milestone and should be based on the use of a random-effect meta-analysis approach assuming each individual estimate is different at random and generated from a common distribution. For a large sample, classical inference based on the classic DerSimonian-Laird methods is unbiased [55]; but this is not true for small samples. For a large sample, classical inference based on the standard DerSimonian-Laid method is unbiased [55], even when the distribution of the effects is extremely non-normal [56,57]. However, the performance of the method deteriorates rapidly as the number of studies decreases, especially for meta-analyses of 5 studies or fewer [55-57]. The advantage of Bayesian methods in comparison to such frequentist approaches is that inference is exact for any sample size, assuming the prior assumptions are valid.

The between-trial variability can be attributed to specific characteristics (i.e. inclusion criteria, choice of outcomes, differences in follow-up, or methods of randomisation) that can be sources of confounding and bias. The true treatment effect would be similar across all trials of network even if these did not include one or both of these two comparative elements. If the included trials have differences that are modifiers of the relative treatment effect, the similarity will be violated and the pooled estimated will be biased [58-61]. The classic measure of heterogeneity is Cochran Q, calculated as the weighted sum of squared differences between individual study effects and the pooled effect across studies. However, the Q test statistic has generally low power, even more so when the data are sparse [55,62]. Another commonly used statistic is I-square [13,54,63], that describes the percentage of variation across studies that is due to heterogeneity rather than chance. However, use of I-square can be challenging since it is not independent from the meta-analysis size [64].

Nevertheless, Bayesian meta-analysis allows the incorporation, into the random effect model, of between-study heterogeneity, including a prior distribution for it as well. Heterogeneity should also be taken into account by performing adjusted analysis, planning

appropriate subgroup analyses or using meta-regression techniques to adjust for differences in study-level characteristics.

2.4.3 THE CONSISTENCY EQUATION

The exchangeability assumption [65] justifies the fact that the treatment effects may be non-identical but their magnitudes cannot be differentiated a priori. Within the context of NMA, it is important that the indirect estimate is not biased and that there is no divergence between the direct and indirect comparisons. If the AB and AC trials are comparable in effect modifiers (and are thus similar), an indirect estimate ($\tilde{\theta}_{BC}$) for the true difference effect between B versus C can be obtained from the direct estimates of A versus B ($\hat{\theta}_{AB}$) and direct estimates of A versus C ($\hat{\theta}_{AC}$). To perform NMA, it is indispensable that the following *consistency equation* is satisfied

$$\tilde{\theta}_{BC} = \hat{\theta}_{AC} - \hat{\theta}_{AB} \quad (2.1)$$

where the effectiveness of each treatment is measured on a scale symmetric to zero such as log odds ratio, log-hazard ratio or difference in mean [33,36,37,66,67,68].

Consistency regards a loop (closed network) rather than individual comparisons. Indeed, to verify presence of inconsistency, the treatments involved must belong to a loop in the network configuration. Lu and Ades [43] propose a general method for assessing evidence inconsistency in the framework of Bayesian hierarchical models. They suggest to represent evidence consistency as a set of linear relations among basic parameters on the log odds scale. Then, these relations will be complicated by introducing some random terms, called inconsistency factors (ICF), and finally this model which incorporates ICF will be compared with the standard one without ICF. Dias et al. [69] also propose an extension of Bucher methods [31] to carry out tests for inconsistency in a network with multiple loops and with only two-arm trials. In this work we compared the goodness of fit of the consistency model (that obtains the indirect treatment effects by means of the consistency equation) with the inconsistency model (which estimates all relative effects for all treatment contrasts).

Health decisions should be based on models that are internally coherent and if the data cannot be fitted by a consistent model some adjustment must be made to correct for possible causes of discrepancy. More details are provided in the following sections, but

careful reading of the pioneering work of Lu and Ades [43], and Dias et al. [69] is also recommended.

2.5 STATISTICAL DETAILS

2.5.1 CONTRAST-LEVEL AND ARM-LEVEL SUMMARY DATA

The input data in NMA are usually the summary statistics extracted from the published literature (aggregate data or study-level data), rather than the original data directly collected from trial authors (individual patient data or patient-level data). Besides, the aggregate input data are available in two formats: as arm-level summaries, where effect measures are reported for each arm (i.e. odds, absolute risk, hazard, or mean), or as contrast-level summaries (i.e. odds ratio, risk ratio, hazard ratios, or mean difference), where results are presented as the difference in effect between arms. One advantage of the arm-level approach is that it is possible to adopt the exact likelihood for the data (i.e. binomial for binary data) rather than its normal approximation, as for the contrast-level summary. Both frequentist and Bayesian approaches can be used to specify models based on either two format [70]. Hereafter, we discuss the analysis of data by means of arm-level summaries, which enable more flexible and precise analyses.

2.5.2 FIXED AND RANDOM EFFECT MODELS FOR A BINARY OUTCOME

Suppose that *N* RCTs make mixed comparisons among *K* treatments. The number of events on treatment *k* in the trial *j* is denoted with r_{jk} and the number of total observations with n_{jk}. Then let p_{jk} be the probability of event occurrence, then the number of events, r_{jk}, leads a Binomial distribution:

$$r_{jk} \sim Bi(p_{jk}, n_{jk}), j = 1,2, \dots N;\ k = 1,2, \dots K \quad (2.2)$$

The probability of event occurrence p_{ik} is modeled on the logit sale as:

$$logit(p_{jb}) = \log\left(\frac{p_{jb}}{1-p_{jb}}\right) = \mu_j, j = 1,2, \dots N;\ k = b = 1,2, \dots K \quad (2.3)$$

$$logit(p_{jk}) = \log\left(\frac{p_{jk}}{1-p_{jk}}\right) = \mu_j + \delta_{j,bk}\ , j = 1,2, \dots N;\ k = 2,3, \dots K;\ b < k \quad (2.4)$$

where μ_i are the trial-specific baselines and represent the log odds of event in the referent treatment (*k=b*), while $\delta_{j,bk}$ are the trial-specific log odds ratio of event occurrence of the treatment group *k* compared with referent treatment.

The nature of effect $\delta_{j,bk}$ depends of assumption underlying the fitted model: fixed or random effect model. The difference consists in the way variability of the between-trial results is treated [16]. The fixed effect model considers this variability as exclusively due to random variation (assume between-trial variance equal to zero) and individual studies are simply weighted by their precision. Therefore, if all the studies were infinitely large they would give identical results. For fixed effect model the equation (2.4) will be replaced as follow:

$$logit(p_{jk}) = \mu_j + d_{j,bk}, j = 1,2, \dots N;\ b = 1,2, \dots K;\ k = 2,3, \dots K;\ b < k \quad (2.5)$$

where μ_j are the trial-specific baselines and $d_{j,bk}$ are the fixed ($\sigma^2_{j,bk} = \sigma^2_j = 0$), trial-specific log odds ratio of event occurrence of the treatment group k compares with referent treatment.

The random effect model, instead, assumes a different underlying effect for each study and takes this into consideration as an additional source of variation. This model has been advocated if there is heterogeneity in between-trial results. For a random effect model the trial-specific log odd ratio $\delta_{j,bk}$ is commonly generated from a Normal distribution

$$\delta_{j,bk} \sim N(d_{bk}, \sigma^2). \quad (2.6)$$

We assumed equal within-trial variance between relative treatment effect ($\sigma^2_{j,bk} = \sigma^2_j$). For more details, Lu and Ades [36] explain the heterogeneous within-trial variance models. The Bayesian structure requires the prior specification for unknown parameter μ_j, $\delta_{j,bk}$ and σ. Dias et al. [71] recommend to give independent weakly priors such as $\mu_j, \delta_{j,bk} \sim N(0, 100^2)$ and $\sigma \sim Uniform(0,2)$.

From the consistency assumption, the indirect estimate δ_{st} is:

$$\delta_{st} = \delta_{bt} - \delta_{bs}, b = 1,2, \dots K;\ s = 2,3, \dots K;\ \ t = 3,4 \dots K; s < t \quad (2.7)$$

The *K-1* direct treatment effects δ_{bk} (between k and baseline treatment groups) represent the *basic parameters* of the model on which priors distributions of Bayesian approach are placed [71], while the *functional parameters* δ_{st} are all the remaining contrasts that are function of basic parameters.

2.5.3 NETWORK META-REGRESSION

Network meta-regression represents a useful tool to explain the heterogeneity between the different treatment effects in the studies by regression of aggregate (study-level) covariates or on individual patient data, if available, exactly like head to head comparisons [35,43,72,73]. Nixon et al. [74] develop methods to simultaneously compare several treatments and to adjust for study-level covariates by combining ideas from MTC and meta-regression. In general, the meta-regression model is fitted specifying fixed or random effect models and adjusting the log odds ratio for study-level prognostic factors. The meta-regression procedure can reduce bias and inconsistency when covariates are distributed uniformly [75,76].

The meta-regression model with fixed treatment effect is:

$$logit(p_{jb}) = \mu_j + \beta x_i, j = 1,2, \dots N;\ b = 1,2, \dots K \quad (2.8)$$

$$logit(p_{jk}) = \mu_j + d_{j,1k} + \beta x_j\ , j = 1,2, \dots N;\ k = 2,3, \dots K;\ b < k \quad (2.9)$$

where x_j is the trial-level covariate for trial j, which can represent a subgroup or a continue variable. In the meta-regression with random treatment effect, the equation (2.9) is replaced with

$$logit(p_{jk}) = \mu_j + \delta_{j,bk} + \beta x_j$$

$$j = 1,2, \dots N;\ b = 1,2, \dots K;\ k = 2,3, \dots K;\ b < k \quad (2.10)$$

where the trials-specific log odds ratios are generated from a common distribution $\delta_{j,bk} \sim N(d_{bk}, \sigma^2)$. In the Bayesian framework, the parameters μ_j, d_{bk}, β and σ will be given independent weakly priors such as $\mu_j, d_{bk}, \beta \sim N(0, 100^2)$ and $\sigma \sim Uniform(0, 5)$ [73].

If the number of studies in a network is limited, the validity of incorporating study-level covariates with meta-regression model may be questionable, given the limited statistical power and risk of overfitting [51,77]. Besides, aggregate covariates adjustment might be prone to ecological bias [27], that represents the failure of study-level associations to properly reflect individual-level associations. Network meta-analyses of IPD are considered the gold standard, as they provide the opportunity to explore differences in effects between subgroups. When individual patient data are available, meta-regression usually have sufficient power to evaluate heterogeneity and to identify effect-modifying factors [72,76].

2.5.4 MULTI-ARM TRIALS

Let us suppose we include in a network, based on contrast-level summaries, one or more multi-arm trials where the number of comparators is 3 or greater. A single multi-arm trial j which compares a_i treatments produces a vector δ_j of $a_j - 1$ random treatment effect, $\delta_j = \left(\delta_{j,12}, \dots, \delta_{j,ba_j}\right)^T$ that are correlated. Over the between-trial variance, it needs to include the random effect covariance [36,37,71]. The specification of the variance-covariance matrix for the random effects vary from constant and equal structure to totally unrestricted positive-definite matrix [37,43]. The assumption of homogeneous between-trial variance means that all σ_{bk}^2 are the same and equal to σ^2, and this implies that the covariance between two contrasts in a multi-arm trial is $\sigma^2/2$ [58]. The univariate Normal distribution (2.6) for multi-arm trial j which compares a_i treatments will be a multivariate Normal distribution

$$\delta_j = \begin{pmatrix} \delta_{j,12} \\ \vdots \\ \delta_{j,ba_j} \end{pmatrix} \sim N_{a_i-1}\left(\begin{pmatrix} \delta_{j,12} \\ \vdots \\ \delta_{j,ba_j} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \cdots & \sigma^2/2 \\ \vdots & \ddots & \vdots \\ \sigma^2/2 & \cdots & \sigma^2 \end{pmatrix}\right). \tag{2.11}$$

Let us imagine we have $K = 4$ trials included in the network and $\binom{K}{2} = \frac{K!}{2!(K-2)!} = 6$ contrasts between A, B, C and D (figure 2.2.d). One study is a multi-arm trial that compares A, B and C treatment and 2 studies are two-arm trials that produce AB and AD comparisons. This network will estimate 3 basic parameters (d_{AB}, d_{AC}, d_{AD}) and 3 functional parameters (d_{BC}, d_{BD}, d_{CD}) obtained from the consistence equation, in formulation (2.1). One can specified the following random effect model:

$$logit\begin{pmatrix} p_{1B} \\ p_{1C} \\ p_{2B} \\ p_{3D} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} + \begin{pmatrix} 1\,0\,0\,0 \\ 0\,1\,0\,0 \\ 0\,0\,1\,0 \\ 0\,0\,0\,1 \end{pmatrix} \begin{pmatrix} \delta_{1,AB} \\ \delta_{1,AC} \\ \delta_{2,AB} \\ \delta_{3,AD} \end{pmatrix} \tag{2.12}$$

with

$$\begin{pmatrix} \delta_{1,AB} \\ \delta_{1,AC} \\ \delta_{2,AB} \\ \delta_{3,AD} \end{pmatrix} \sim N_4\left(\begin{pmatrix} \delta_{1,AB} \\ \delta_{1,AC} \\ \delta_{2,AB} \\ \delta_{3,AD} \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & \sigma^2/2 \\ 0 & 0 & \sigma^2/2 & \sigma^2 \end{pmatrix}\right) \tag{2.13}$$

2.6 DISCUSSION

We have provided a comprehensive and detailed overview of the conceptual and practical issues involved in performing a and interpreting NMA on binomial data while applying a Bayesian hierarchical model. We have discussed the general topics related to NMA, including how to collect study data, structure the network, and set assumptions about the network that lead to different models and interpretations of model parameters. Many papers have been published on these topics and other will follow suite, describing methods for NMA with binary data in a concise way and in quite some detail [33-37,78,41-44,50,51,60,61,63,66,69-76]. We have strived to put together the most important topics (making available the major references) and we offer, for the first time, a thorough yet manageable guideline to conduct (from literature search to results interpretation) a rigorous NMA on binomial data, applying the Bayesian hierarchical model.

CHAPTER 3

BAYESIAN NETWOK META-ANALYSIS

3.1 BAYESIAN FRAMEWORK AND WINBUGS

Here we focus our attention on data which can be analyzed with a binomial model applying the Bayesian hierarchical approach proposed by Smith et al. [79] and using Markov Chain Monte Carlo (MCMC) approaches.

There is a large literature base on Bayesian analysis, hierarchical modeling and implementation of Markov Chain Monte Carlo (MCMC) methods to perform statistical inference [58,65,80]. In contrast with the classical statistical theory, the Bayesian approach can incorporate any available information about the parameters before we observe the data, hence the parameters are considered as random variables that are characterized by a prior distribution. Priors may be divided into four categories according to the analytical goal and use: a) informative, based on existing evidence; b) weakly informative, that provide enough information to avoid results that contradict the previous knowledge; c) least informative, determined solely by the model and observed data to minimize the amount of subjective acquaintance; d) non-informative [81]. Van Dongen et al. [82] stated that non-informative priors, especially when derived from small sample sizes, lead to different results from reference non-Bayesian models, and weak priors generate information closer to the referral model. Consequently the priors should be looked as a distribution that should reflect a biologically plausible parameter space [82]. With a large number of comparisons in a well-defined network configuration and with a large number of trials included, a reasonable choice of prior distributions will have minor effects on posterior inferences. If the data are sparse or there are no events in one or more arms of contrast, the prior distribution becomes more important. In general, if the information is strong the inference is based primarily on prior beliefs; if it is weak the numerical estimation can be unstable. Various weakly informative prior distributions have been suggested for scale parameters in

hierarchical models. It has been suggested [71,83] to use vague priors for μ_i and d_{bk} parameters, such as $N(0, \sigma^2)$ with variance equal to 0.001 or 0.0001. A Uniform distribution, $\sigma \sim Uniform(0, A)$, can be used as prior for the standard deviation of a Binomial distributions and logit link function. The upper limit of distribution, *A*, represents a huge range of trial-specific treatment effect [71]. For a finite but sufficient large A, inferences are not sensitive to the choice of *A* [83]. The approach to set a Gamma prior on precision, $1/\sigma \sim Gamma(\varepsilon, \varepsilon)$, produces a sharply peaked near zero distribution and further distorts posterior inferences, because of the marginal likelihood that σ^2 remains close to zero. Where σ is estimated to be near zero, the resulting inferences will be sensitive to ε. On the other hand, the use of an Inverse-Gamma distribution is suitable when data are sparse, improving stability and convergence. Usually the hyperparameter ε is set to a low value such as 0.001. As priors are part of the model specification, initial values are part of the computing process. Initial values can be derived from the current dataset or may be generated from prior distributions. The evaluation of posterior distributions is dependent on the MCMC chains convergence. Most convergence checking, such as the Gelman and Rubin approach, are graphical, and either compare the results from different chains or divide one chain into sections and compare these sections. If the simulation has not yet converged, the chains or part-chains will look different when plotted [80]. Finally, the Monte Carlo error (an index that reflect the number of simulation and the autocorrelation degree) should be no more than 5% of the posterior standard deviation of the parameters of interest to minimize the bias inherent to the resampling method [71].

3.2 GOODNESS OF FIT

Statistical models, in addition to drive the inference process to provide prediction results, allow to describe how well the model itself fits a set of observations and to discriminate between alternative models. The *likelihood ratio test* (LRT) represents one of the classic ways to compare two nested models [84,85]. Alternatives include the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The AIC [86] for a given model is a function of its maximized log-likelihood and the number of estimable parameters p:

$$AIC = -2\log L\left(\hat{\theta}\,|y\right) + 2p \quad (3.1)$$

For a non-hierarchical model with *p* parameters and *n* observations, the Bayes (or Schwarz) Information Criterion [87] is given by

$$BIC = -2\log L\left(\hat{\theta}\,|y\right) + 2p \cdot \log(n) \quad (3.2)$$

The advantage of the AIC and BIC statistics is that these can also be used for non-nested models. To compare two competitive models, as the comparison between the fixed and the random effect models, smaller values of these model assessment statistics are better, efficiency remains paramount. Subsequently, Spiegelhalter et al. [88] developed a model comparison criterion called the Deviance Information Criterion (DIC), that is a generalisation and Bayesian version of AIC and is also related to the BIC, following the original suggestion of Dempster [89] for model choice in the Bayesian framework. Indeed, the frequentist approach to model assessment is based on *deviance*, which the difference in the log-likelihoods between the fitted and the saturated model (the model with as many parameters as observations, with perfect t to the data). Similarly, Dempster suggested to examine the posterior distribution of the classical deviance defined by

$$D(\theta) = -2\log f(y|\theta) + 2\log f(y) \quad (3.3)$$

for observations y and parameter vector θ. The DIC is thus based on the posterior distribution of $D(\theta)$ and it is defined as the sum of two components. The first component measures the goodness of fit of a model by the posterior expectation of the overall residual deviance:

$$E_{\theta|y}[D] = \overline{D}. \quad (3.4)$$

The second measures the complexity of the model by the effective number of parameters, p_D, defined as the difference between the posterior mean of the overall residual deviance and the deviance evaluated at the posterior mean of the parameter of interest:

$$p_D = E_{\theta|y}[D] - D\left(E_{\theta|y}[\theta]\right) = \overline{D} - D(\hat{\theta}). \quad (3.5)$$

Ultimately, models may be compared using a DIC [71,88,90-92] defined sum of expression (3.4) and (3.5)

$$DIC = \overline{D} + p_D = 2\overline{D} - D\left(\hat{\theta}\right) = D\left(\hat{\theta}\right) + 2p_D. \quad (3.6)$$

The model with the smallest DIC is estimated to be the model that would best and most efficiently predict the observed data. It is difficult to say what would constitutes an

important difference in DIC, as both subjectivity and experience must be applied. As a rule of thumb, a difference of more than 10 might definitely rule out the model with the higher DIC, differences between 5 and 10 are considerable, but if the difference in DIC is less than 5 and the models provide very different inferences, care should be taken when referring the model with the lowest DIC [78]. The above mentioned statistics (AIC, BIC and DIC) are easily calculated during an MCMC run by monitoring both θ and $D(\theta)$. The DIC tool of WinBUGS system directly provides the posterior mean of the overall residual deviance (Dbar), the deviance of the posterior means of interested parameter (Dhat), the p_D and the DIC value.

Another promising method for comparing different models, nested or not, is to use only the posterior distribution of the sum of residual deviance $\overline{D}$ of each competing model [43,81]. The sum of residual deviance for a binomial likelihood function is provided by:

$$\overline{D} = \sum_{i=1}^{N} Dev_i = \sum_{i=1}^{N} \sum_{k=1}^{K} 2 \left[log\left(\frac{r_{jk}}{n_{jk}p_{jk}}\right) + \left(n_{jk} - r_{jk}\right) log\left(\frac{n_{jk}-r_{jk}}{n_{jk}-n_{jk}p_{jk}}\right)\right] \quad (3.7)$$

where, as mentioned above, r_{jk} denotes the number of events on treatment *k* in the trial *j*, n_{jk} represents the number of total observations and p_{jk} is the probability of event occurrence. The posterior distribution of the model deviance difference can be obtained as $\overline{D}_{1,2} = \overline{D}_1 - \overline{D}_2$ and it may be used to calculate the posterior probability

$$P\left[\overline{D}_{1,2} > \beta(\overline{D})\right] \quad (3.8)$$

as an analytic method for model selection. The choice of the value of β can vary for different purposes.

In order to make an association with the frequentist approach, the difference between the deviances of two nested models is approximately a chi-squared distribution with *df* degrees of freedom, where $df = p_2 - p_1$ is the difference between the number of parameters estimated. In this case one can choose $\beta = \chi^2_{1-\alpha;df}$. The higher this probability the stronger is the evidence in favour of model 2 against model 1. In addition it is possible to calculate the value of β that gives:

$$P\left[\overline{D}_{1,2} > \beta(\overline{D})\right] = 0.5 \quad (3.9)$$

and use, for example, the table of Kass and Raftery [93] (table 3.1) to quantify the evidence against model 1. It is worth noting that these numbers are driven more from intuition, rather than a scientific justification [81]. The posterior probability check is performed in WinBUGS using the step function.

Table 3.1: Scale of evidence proposed by Kass and Raftery (1995)

β	**Evidence in favour of model 2**
0-2	Not worth more than a bare mention
2-6	Positive
6-10	Strong
>10	Very strong

3.3 RANK PROBABILITY ESTIMATE

An advantage of Bayesian approach is that the posterior distribution of estimate, with its credible interval, can be interpreted in terms of probability which allows an intuitive and direct interpretation of which treatment is the best or the subsequent. In each MCMC run, every treatment is ranked according to its estimated magnitude. Then, the proportion of MCMC cycles in which the treatment *k* ranks first gives the probability that such specific treatment *is the best* among all *K* treatments. Other probabilities are calculated for being the second best, the third best and so on for each treatment. Salanti et al. [94] propose some graphical and numerical summaries of rank probabilities (rankograms). These authors also suggest a simple method to show the cumulative rank probabilities for each treatment estimating the surface under the cumulative rank curve (SUCRA). For each treatment *k* and for each rank *w* (*k,w=1,2,…K*), it is possible to calculate the vector of cumulative probabilities $cum_{k,w}$ and the SUCRA index will be:

$$SUCRA_k = \frac{\sum_{w=1}^{K-1} cum_{k,w}}{K-1} \tag{3.10}$$

The SUCRA index simplifies the entire information about treatment ranking into a single number. SUCRA is equal to 1 if the treatment is surely the best, and equal to 0 if the treatment is surely the worst.

3.4 SENSITIVITY ANALYSIS

Various techniques may be used to check whether the assumptions of the model are valid and whether the fit of the model is adequate. In the Bayesian setting, it is important to pay attention to the robustness of the posterior distribution. One can assess how posterior distribution changes over different prior distributions [80]. When prior information is available, sensitivity analysis focuses on the structure of the prior distribution. When weakly priors are used, it focuses on how different choices of prior parameters may influence the posterior inference. Besides, sensitivity analysis can be performed discussing the different findings from competing models (fixed or random effect models, consistency or inconsistency model) or executing the NMA on a subgroup of RCTs (high quality RCTs only, or specific stratification by other baseline covariates).

3.5 CASE STUDY ON ANAESTHETIC DRUGS

3.5.1 INTRODUCTION

To clarify the statistical features of the network meta-analysis on a binary endpoint we start from a published Bayesian network meta-analysis [95] that compared the effect on mortality of three different volatile agents (desflurane, isoflurane, sevoflurane), and total intravenous anesthetics (TIVA).

Anesthetics have pharmacological properties that go beyond their effects on blood pressure and heart rate and they might induce cardiac protection. These effects influence perioperative [96-98] and long term clinically relevant outcomes [99,100]. An international, web based consensus conference [101]recently included volatile anesthetics among the few drugs that might reduce mortality in patients undergoing cardiac surgery. The scientific community agrees that there is initial evidence suggesting that different anaesthetic drugs could lead to apparent differences in survival rate in patients undergoing cardiac surgery [101], with volatile agents having beneficial effects (or TIVA having detrimental effects). At the same time there are few (if any) direct comparisons between different anaesthetic agents to define which treatment is the best.

We analyzed the data applying the Bayesian hierarchical approach proposed by Smith et al. [79] and using Markov Chain Monte Carlo (MCMC) approach. The WinBUGS code to analyse data on anaesthetic drugs is performed using the indications of Ades et al. [102] and Dias et al. [69,71,73].

3.5.2 SEARCH STRATEGY AND NETWORK CONFIGURATION

Several databases (BioMedCentral, MEDLINE/PubMed, Embase, and the Cochrane Central Register of clinical trials) were searched to identify articles comparing a TIVA or an anesthesia plan including administration of isoflurane, desflurane or sevoflurane with no restriction in dose and time of administration. Duplicate publications, nonhuman experimental studies and studies with no mortality data were excluded. No language restriction was enforced, and non-English articles were translated and included in the analyses. The primary treatment strategies of interest in this Bayesian network meta-analysis were 1) TIVA, 2) isoflurane, 3) desflurane and 4) sevoflurane.

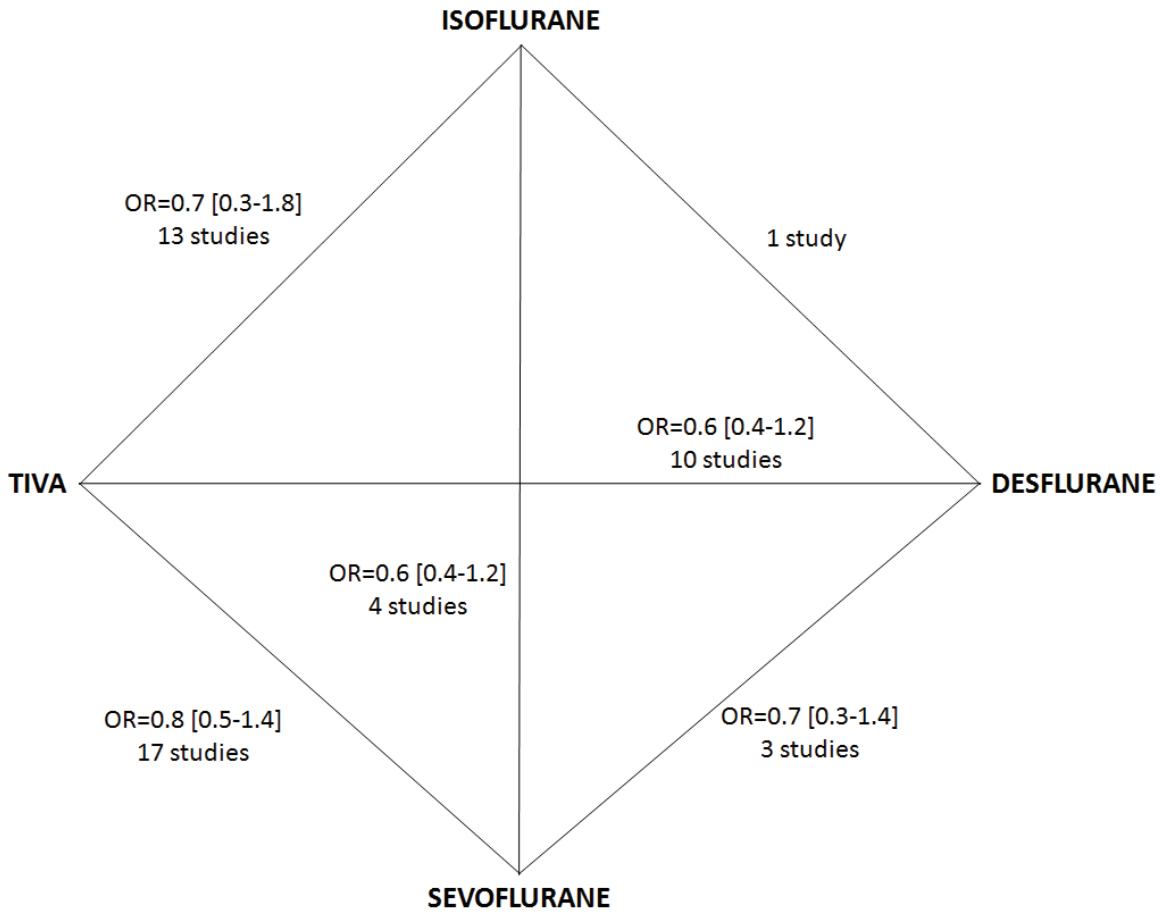
The systematic and reproducible search together with additional hints identified 2,630 full manuscripts that were screened by several authors at the title/abstract level. The 112 remaining papers were studied in details, excluding articles with no mortality data, those which were not randomized or were nor performed in the cardiac surgery setting.

Thirty-eight randomized trials [103-140] were included in the final analyses and data (year of publication, setting, number of patients in each group, comparators, length of follow up) extracted by several authors.

To have a complete view of the distribution of the included studies, we build the network configuration of the anesthetic agent (figure 3.1). The network diagram graphically represents all the direct comparisons among drugs. This graph consists of a set of nodes representing the interventions linked by lines that depict how many RCT have been included. It is worthy to cautiously examine the network structure because the data pattern reveals particular characteristics that may assist in the choice of the analytical method [141]. In this case, figure 3.1 includes four nodes representing the four interventions and six edges (arrows). This network has four closed loops and each contrast has both direct and indirect

evidence. We have chosen to show in the picture the odds ratio estimates with corresponding 95% confidence intervals and the number of studies for each pairwise comparison. The most frequently treatment, TIVA was chosen as reference.

Figure 3.1: Network configuration [95]. This diagram represent all the direct comparisons among drugs. This graph consists of a set of nodes representing the interventions linked by lines that depict how many RCT have been included.



3.5.3 DATA PROCESSING

The 38 included trials [103-140] were published between 1991 and 2012 and randomized 3,996 patients. The median of randomized patients per trial was 60 (range 20-414). Volatile agents were administered to 2,348 (59%) patients while TIVA was given to 1,648 (41%) patients. Specifically 1,086 (27%) patients were randomized to propofol (the most commonly used TIVA agent), 622 (16%) to isoflurane, 701 (17%) to desflurane and 1,025 (26%) to

sevoflurane. Most studies (24/38 [63%]) were performed in coronary artery bypass graft (CABG) surgery.

The data structure of mortality outcome for the four anesthetic agents is presented in the appendix 1. The list command specifies *nt* treatments and *ns* studies. The number of arms in each trial is reported into the vector *na*. The matrix *t(38X3)*, *ns* X maximum number of arms in a trial, identifies the code of treatment, *r(38X3)* the number of events, and *n(38X3)* the total number of patients, for each included study. The hash symbol (#) permits to write a comment text that will be ignored by WinBUGS. The NA code is required when the data is not available.

3.5.4 FIXED OR RANDOM EFFECT MODELS

The treatment effect evaluation depends of hypothesis underlying the variability between the included trials. The fixed effect model accounts this variability as completely due to chance (assuming between-trial variance equal to zero). In this case the weight of each individual study coincides with his precision. The random effect model has been advocated if there is evidence of between-trial heterogeneity.

The forest plot of overall standard meta-analysis (figure 3.2) showed that the use of all volatile agents (isoflurane, desflurane, or sevoflurane) was associated with a reduction in mortality when compared to TIVA at the longest follow-up available (OR=0.51, 95% IC 0.33 to 0.81, p for effect =0.004) and the visual inspection of funnel plot (figure 3.3) did not identify an important skewed or asymmetrical shape.

Figure 3.2: Forest plot of volatile agents (isoflurane, desflurane, or sevoflurane) versus total intravenous anaesthesia (TIVA) for the risk of mortality at the longest follow-up available. CI=confidence interval; OR=odds ratio. [95]

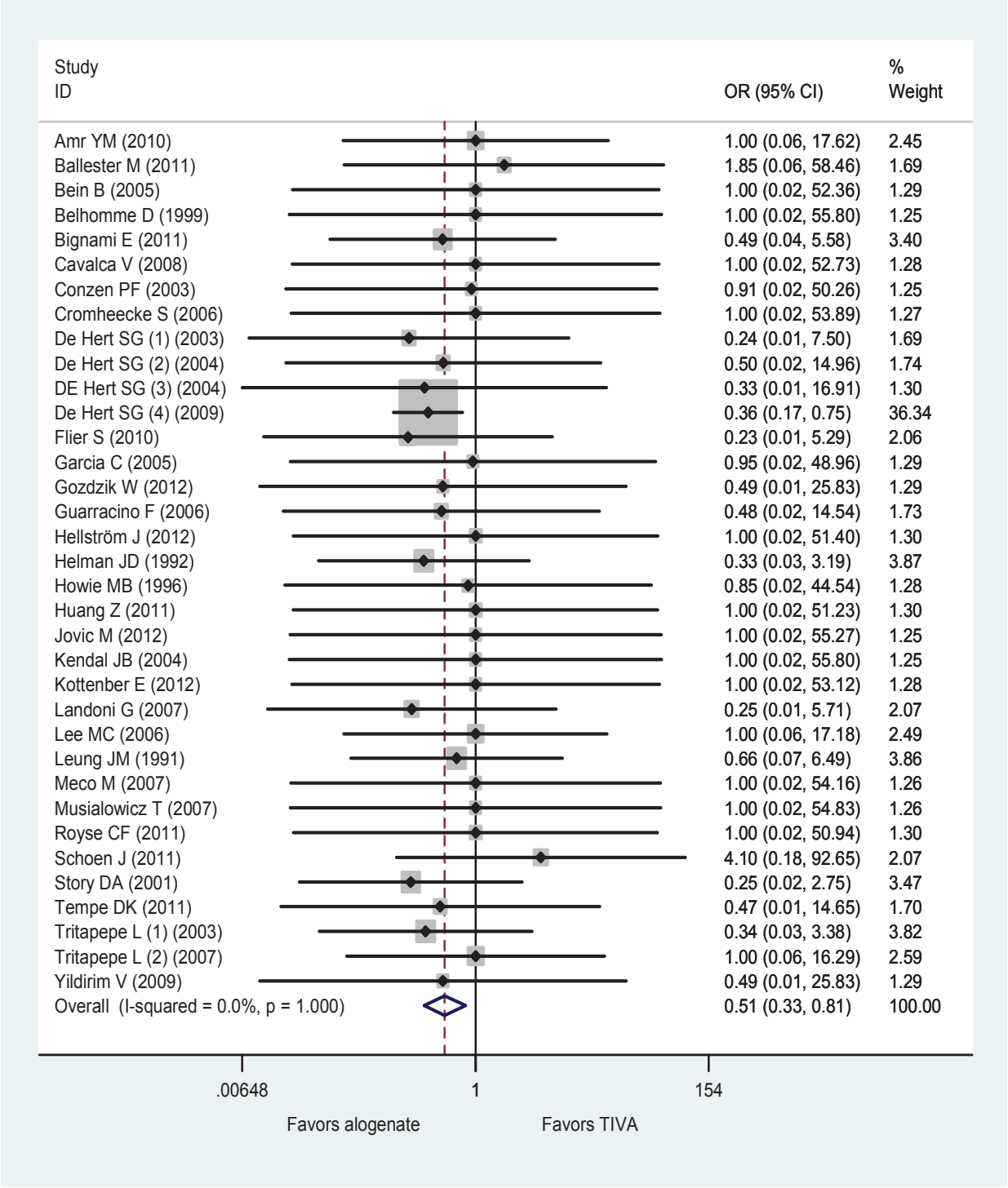
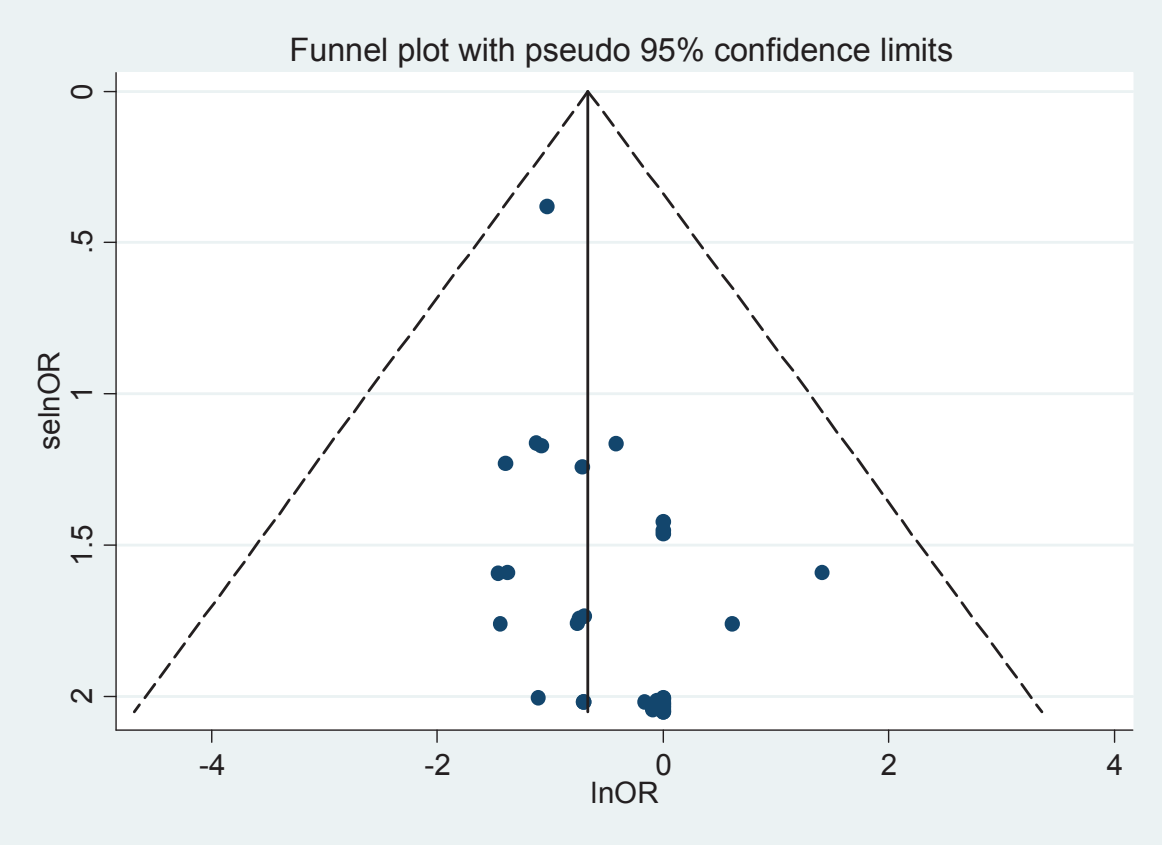


Figure 3.3: Funnel plot of volatile agents (isoflurane, desflurane, or sevoflurane) versus total intravenous anesthesia (TIVA) for the risk of mortality at the longest follow-up available. SE=confidence interval; OR=odds ratio. [95]



The WinBUGS code used to analyse data on anesthetic drugs by means of Bayesian network meta-analysis is the following:

```
#MCT fixed effect model
model{
 for(i in 1:ns){
 mu[i]~dnorm(0, 0.0001)
  for(k in 1:na[i]){
  r[i,k]~dbin(p[i,k], n[i,k])                    #binomial likelihood
logit(p[i,k])<-mu[i] + d[t[i,k]] - d[t[i,1]]     #model
}
 }
d[1]<-0
for(k in 2:nt){d[k]~dnorm(0,0.0001)}
}
```

```
# MCT random effect model
model{
 for(i in 1:ns){
 w[i,1]<-0
delta[i,1]<-0
 mu[i]~dnorm(0, 0.0001)
for(k in 1:na[i]){
  r[i,k]~dbin(p[i,k], n[i,k])                    #binomial likelihood
logit(p[i,k])<-mu[i]+delta[i,k]                  #model
}
  for(k in 2:na[i]){
  delta[i,k]~dnorm(md[i,k],taud[i,k])            #trial-specific LOR distribution
  md[i,k]<-d[t[i,k]] - d[t[i,1]] + sw[i,k]
taud[i,k]<-tau *2*(k-1)/k                        #adjustament for multi-arm trial
w[i,k]<-(delta[i,k] - d[t[i,k]] + d[t[i,1]])
sw[i,k]<-sum(w[i,1:k-1])/(k-1)
  }
 }
d[1]<-0
for(k in 2:nt) {d[k]~dnorm(0,0.0001)}
tau~dgamma(0.001,0.001)
sd<-pow(tau,-0.5)
}
```

We carried out the two models running 3 chains (20,000 iterations after a burn-in of 20,000 and 100,000 interactions after a burn-in of 100,000 for the fixed and random effect model, respectively), and monitoring the unknown parameter δ and τ (treatment difference effects and variance in the random effect model only). We used vague priors to produce the posterior distributions for the treatment effects: Normal distribution with mean equal to 0 and variance equal to 0.0001. Besides, we ran the random effect model with a more informative prior (Inverse-Gamma distribution) on the variance parameter [80-83] to overcome the zero-cell count problem.

The fixed or random effect model can be selected calculating the posterior mean of residual deviance (Dres) and the DIC statistics, directly by WinBUGS software. The model with smaller DIC is estimated to be the model that best and most parsimoniously predicts the data observed. The DIC tool, directly implemented in WinBUGS, provides the posterior mean of the overall residual deviance (Dbar), the deviance of the posterior means of interested parameter (Dhat), the$_D$ p_D, which is a measure of the complexity of the model by the

effective number of parameters, and the DIC value. The DIC statistic is given by $D_{bar} + p_D$ or directly from the WinBUGS tool: DIC_{FE}=149.4 and DIC_{RE}=150.1. The residual deviance is done by $D_{bar} = DIC - p_D$ and is equal to $Dres_{FE}$=127.5 and $Dres_{RE}$=126.5. We can conclude that the two models have the same goodness of fit. However, we choose the fixed model because it leads to more precise estimates than random effect model (that is more conservative), and because it is more parsimonious (table 3.1).

The consistency assumption (no discrepancy between direct and indirect comparisons) was verified by the posterior distribution of residual deviance difference in order to compare the consistency model (which estimates only the basic parameters [72]) with the inconsistency model (which estimates both basic and functional parameters). The posterior probability check is performed in WinBUGS using the step function [141]:

```
#Post probability check
diff<-sumresdev1 - sumresdev2
 p<-1-step(beta-diff)
for(b in 1:10){
 probability[b] <- 1-step(b-diff)
 }
```

In the case of halogenated agents study, the probability in favors of inconsistency model was equal to 0.03; hence we calculated the indirect estimate as difference from the appropriate direct estimates and the indirect 95% CrI by normal approximation.

3.5.5 MODEL RESULTS

The pooled odds ratio (OR) estimates were calculated exponentiating the corresponding posterior mean of log odds ratio (lnOR) obtained from Bayesian software (table 3.1). The indirect estimates were calculated from the consistency equation (2.1), by taking the difference between the corresponding direct estimates, while the 95% credible intervals were calculated from the actual posterior distribution by means of the normal approximation. For example, for the isoflurane-desflurane (2-3) indirect estimate we have: $lnOR_{23} = lnOR_{13} - lnOR_{12} = -0.8505 + 0.8776 = 0.0271$ where the treatment 1, TIVA, is there reference. The standard error of the logarithm of indirect estimate $SE(lnOR_{23})$ is

obtained by $\sqrt{SE(lnOR_{13})^2 + SE(lnOR_{12})^2} = \sqrt{(0.3423)^2 + (0.5018)^2} = 0.6074$ and the 95% credible interval is equal to $0.0271 \pm 0.6074 \cdot 1{,}96$. The corresponding odds ratio OR_{23} is calculated as $e^{0.0271}$ and it 95% credible interval as $e^{0.0271 \pm 0.6074 \cdot 1{,}96}$. Table 3.2 reports the posterior distribution of means and 95% credible intervals, for the anesthetic agents difference effects, derived by Bayesian hierarchical model with Markov Chain Monte Carlo algorithm.

We found that the use of sevoflurane (posterior mean of OR =0.31, 95% CrI 0.14 to 0.64) and desflurane (posterior mean of OR =0.43, 95% CrI 0.21 to 0.82) was associated with a reduction in mortality when compared to TIVA at the longest follow-up. A sensitivity analysis showed that when the largest study was removed only the use of desflurane resulted associated with a significant reduction in mortality with respect to TIVA (posterior mean of OR =0.30, 95% CrI 0.09 to 0.88). A network meta-regression was performed to evaluate the association between log-risk of mortality and both the length of study follow-up and the year of publication. Heterogeneity should also be taken into account by adjusted analysis or planning appropriate subgroup analyses. Bayesian meta-regressions showed no significant effect of average follow-up (regression coefficient =-0.0008, CrI -0.004 to 0.002)and of average of publication's years (regression coefficient =-0.058, CrI -0.048 to 0.185) against log-risk of mortality.

Table 3.1: WinBUGS output carring out both fixed- and random- effect model. Posterior distribution of mean, standard deviation (SD), median and 95% confident interval limits for the estimated treatment difference effects, variability estimate and goodness of fit indices.

	Fixed effect model					**Random effect model**				
Contrast	**Mean**	**SD**	**2.5%**	**Median**	**97.5%**	**Mean**	**SD**	**2.5%**	**Median**	**97.5%**
Isoflurane vs TIVA	-0.8776	0.5018	-1.892	-0.8697	0.08594	-0.9229	0.5729	-2.102	-0.9056	0.1644
Desflurane vs TIVA	-0.8505	0.3423	-1.546	-0.8415	-0.1999	-0.903	0.4295	-1.814	-0.8814	-0.1397
Sevoflurane vs TIVA	-1.158	0.3814	-1.936	-1.146	-0.4388	-1.106	0.4541	2.008	-1.107	-0.191
Tau (τ)						167.3	383.0	0.5407	26.33	1286.0
Dbar	**127.473**					**126.526**				
Dhat	**105.546**					**102.967**				
pD	**21.928**					**23.559**				
DIC	**149.401**					**150.085**				
Dres	**127.473**					**126.5**				

TIVA: Total intravenous anesthesia

Table 3.2: Posterior distribution of means and 95% credible intervals [95].

	Fixedeffect model	
Contrast	**OR**	**95% credibleinterval**
Isoflurane vs TIVA	0.42	0.15-1.09
Desflurane vs TIVA	0.43	0.21-0.82
Sevoflurane vs TIVA	0.31	0.14-0.64
Desflurane vs isoflurane*	1.03	0.31-3.38
Sevoflurane vs isoflurane*	0.76	0.22-2.60
Sevoflurane vs desflurane*	0.74	0.27-2.01

* Indirect treatment difference effect calculated from consistency equation (2.1) - TIVA: Total intravenous anesthesia

3.5.6 POSTERIOR RANK PROBABILITIES

Table 3.3 reports, for each anesthetic agent, the posterior distribution of the probability to be the best, the second, the third and the worst, showing a trend of TIVA to be the worst in terms of long term survival after cardiac surgery. The code to ranking the treatment calculating the posterior mean of the probability that each treatment is the best, the second, the third and the fourth, is the following:

```
# Rank probabilities to be the best treatment, or the subsequent, to prevent an adverse events
for (k in 1:nt) {
 for (m in 1:nt) { best[k,m]<- equals(rank(d[],k), m) }
}
```

Table 3.3: Posterior distribution of mean and 95% credible interval, for the anesthetic agent difference effects, derived by Bayesian hierarchical model with Markov Chain Monte Carlo algorithm. [95]

Anesthetic agents	Probability to be the best	Probability to be the second	Probability to be the third	Probability to be the worst
TIVA	<0.001	<0.001	0.04	0.96
Isoflurane	0.26	0.31	0.40	0.04
Desflurane	0.18	0.38	0.44	0.005
Sevoflurane	0.57	0.32	0.12	<0.001

TIVA: Total intravenous anesthesia

3.5.7 SENSITIVITY ANALISIS

Sub-analyses were performed including studies which reported 30-day mortality and using propofol as TIVA.

When we repeating the Bayesian network meta-analyses using short term mortality (less or equal to 30-days after surgery) as an endpoint, we found only a trend towards a reduction in mortality when comparing desflurane versus TIVA (posterior mean of OR =0.41, 95% CrI 0.15-1.04). For what concerns the analysis including all studies using propofol as TIVA, we found a significant difference in the treatment effects between sevoflurane and propofol (posterior mean of OR =0.37, 95% CrI 0.13 to 0.98). Moreover, the Bayesian meta-regressions of average of publication's years against mortality log-risk showed a significant association when analyzing only those studies using propofol (regression coefficient =0.259,

CrI 0.007 to 0.545). Adjusting the analysis for the effect of year of publication, we observed a more intense difference effect between sevoflurane and propofol (posterior mean of OR =0.30, 95% CrI 0.10 to 0.86).

3.6 DISCUSSION

We have provided a comprehensive and detailed overview of the conceptual and practical issues involved in performing a and interpreting NMA on binomial data while applying a Bayesian hierarchical model. We have discussed the general topics related to NMA, including how to collect study data, structure the network, and set assumptions about the network that lead to different models and interpretations of model parameters. The presented case study on the beneficial effects of anaesthetic agents and the practical guide with the actual WinBUGS codes will allow transparency and ease of replication of all steps that are required when carrying out such quantitative syntheses. Additionally, we propose and applied the posterior probability check method [81] to compare the posterior distribution of the sum of residual deviance of consistency and inconsistency models.

This Bayesian network meta-analysis had confirmed that isoflurane, desflurane and sevoflurane reduce mortality after cardiac surgery when compared to TIVA. Unfortunately, even the Bayesian network meta-analysis with direct and indirect comparisons was unable to identify if one of the volatile agent was better or worse that the other ones in terms of improved survival. Traditional limitations of meta-analyses due to variations in the treatment regimens, in populations or major subgroups within trials, and in the conduct of the trials also apply to this Bayesian network meta-analysis. Bayesian network meta-analysis incorporates both the direct and indirect comparisons between treatments. However indirect evidence is susceptible to confounding [32] and thus should be interpreted with caution since it does not always agree with the corresponding direct estimates [4]. Although the consistency hypothesis was not rejected in this Bayesian network meta-analyses, additional methodological and empirical work needs to be done to evaluate the direct and indirect comparisons across a number of types of interventions.

CHAPTER 4

MULTILEVEL NETWORK META-ANALYSIS

4.1 BACKGROUND

The easiest way to compare two treatment arms is to look at the relative difference in the effect size estimate (i.e. weighted mean difference, relative risk, odds ratio) between the group of interest and the reference group, treating such effect size estimates as independent. However this ability to cope with multiple treatments implies that NMA provides naturally a more general framework to deal with correlated data [37,39,41,70,142,143] where correlation can derive from multiple endpoints, time-varying responses or from clustered observation. Multilevel modeling approaches [144-149] offer a valuable framework for carrying out NMA taking advantage of an existing hierarchical data structure.

In this chapter we propose an alternative frequentist approach to estimate the consistency and inconsistency models in the context of NMA following Higgins at al. [41] definition in using *design* to refer to the set of treatments compared in a trial. We discuss multilevel modeling which also provides a unified analysis method to meta-analysis, and which may be carried out within widely available statistical programs such as the SAS software (SAS Institute Inc. Cary, NC, USA). Therefore, we present the *multilevel network meta-analysis* which includes a three-level data structure: subject within studies at first level, studies within study designs at second level and the design configuration at the third level. This approach differs from an alternative two-stage modeling because it is a one-stage strategy which works directly on an arm-based data structure, where the effect are measured for each arm (i.e. odds, absolute risk, hazard, or mean), instead of on the contrast-level summaries.

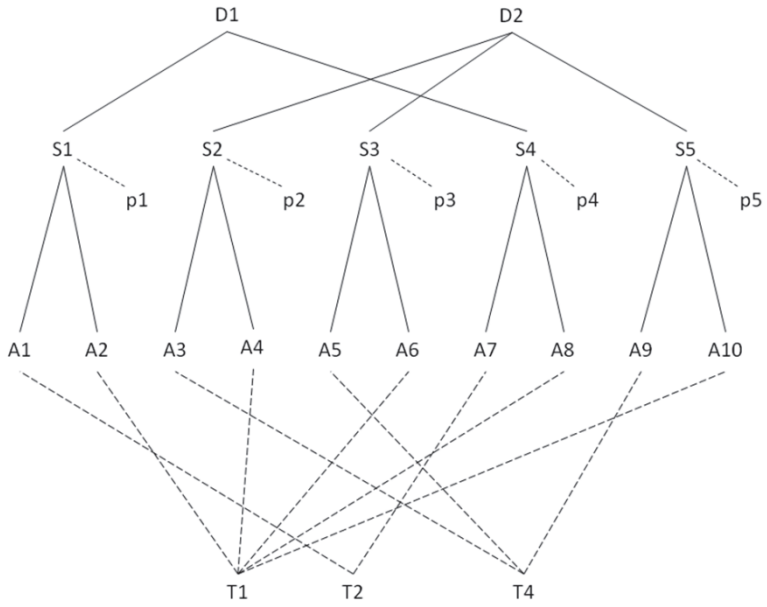
4.2 MULTILEVEL NETWORK META-ANALYSIS

In the last 15-20 years, multilevel methodology has evolved from a specialty area of statistical research into a standard analytical tool used by many applied researchers. Multilevel modeling is now an accepted statistical tool to analyze nested sources of heterogeneity derived from hierarchical data, taking into account the variability associated with each level of the hierarchy. Discussions of methodological and statistical issues including performing a meta-analysis using multilevel model are available from works authored by Goldstein [146], van Houwelingen [150] and Hox [144,148]. These researchers posed existing methods for meta-analysis of two-arm clinical trials into the general framework of multilevel modeling. Flexibility is the major advantages of using these models instead of classical meta-analysis approaches [144,148]: it is easier to include study characteristics as explanatory variables in an attempt to explain existing heterogeneity, and to add additional levels into the model to accommodate multiple treatment comparisons.

In general, multilevel analysis assumes a linear regression model on, both, individual level that relates the outcome to the treatment-group variable and on a second level for each study included in the analysis. Even without having the original data, it is often possible to carry out a multilevel meta-analysis on the summary statistics [148]. Mean and variance of the regression coefficients across the studies are properly estimated. In case of heterogeneous results, if the variance of the regression slopes of the treatment-group variable is large and significant, researchers can refer to the study characteristics as explanatory variables at the second (study) level to predict the differences in the regression coefficients [144].

In the present work, we have a three-level data structure: including the subjects nested within studies, studies nested within designs and study designs. Consider a sample of the first 10 hypothetical treatment arms (level 1) of the data frame (appendix 1) of the case study on anesthetic agents, taken from 5 consecutive studies (level 2) which show two different designs (level 3): isoflurane vs TIVA and sevoflurane vs TIVA. It is clearly observed that the treatment and the sample size represent the variables at first and second level, respectively. The unit diagram in figure 4.1 and the classification diagram in figure 4.2 highlight the relations underlying the data. The unit diagram conveys the three-level structure of the hierarchical data in terms of actual units and exact relationships between levels.

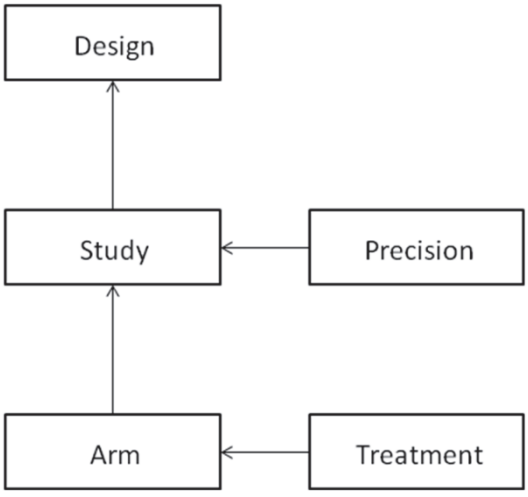
Figure 4.1: Unit diagram for the three-level data structure of the case study on anesthetic drugs.



Nodes A_i identify each of 10 arms, nodes S_i each study and D_i the two design configuration. The data are strictly hierarchical, since each unit belongs to one and only one higher level unit. The precision p_i of each study is a second level covariate while each arm is linked at one of four treatments at the first level.

The classification diagram in figure 4.2 provides a simple summary of the entire subset of data: it has only one node for each classification in the model. The arrows indicates the nested relationship.

Figure 4.2: Classification diagram for the three-level data structure of illustrative example on anesthetic drugs.



4.3 CLUSTER-SPECIFIC AND POPULATION-AVERAGED EFFECTS

Cluster-specific (also known as subject-specific or conditional) models and population-averaged (also known as marginal or unconditional) models are two different approaches to model covariate effects on outcomes in the presence of a clustering structure. Random effect model, such as the multilevel ones, is typically used to estimate the cluster-specific effects while the generalized estimating equations (GEE) model by Liang and Zeger (1986) [151] specifies the marginal or population-averaged distribution of the treatments. The major difference between them is whether the observations are analyzed and interpreted within the same cluster, or across clusters.

In cluster-specific models, the regression coefficients have a cluster-specific interpretation and significance of the between-cluster variance is crucial to assess the clinical relevance of the corresponding hierarchical level. The regression coefficient of an explanatory variable *X*, often referred to as the subject-specific effect, is a measure of the difference in the mean response (depending on the nature of the link function) in the same cluster for a 1-unit change in *X*, holding constant all the other covariates and the combination of unobserved individual features represented by the random effect. The interpretation of results as a population-average are frequently not of interest [152], indeed the cluster-specific model may be the more used approach. However interpretation of the parameters is difficult as it assumes the existence of latent (unobserved group-level) variables which are included directly in the model through the random effect term.

A marginal model consists of (i) a generalized linear model explaining the response and predictors association and (ii) a specification of the structure of the correlations among individuals (in this case: arms) on the same group (in this case: study) [151,153]. An advantage of the marginal model is that coefficient estimates and their standard errors are robust to misspecification of the correlation structure. In marginal models, the coefficients have a population-averaged interpretation and the coefficients of *X* describe differences in the mean responses across all observations (and hence across all clusters). The clustering is an individual characteristic that needs to be taken into account, but which is not the main focus of the analysis. The model specifies no parameter representing the between-cluster variance and the effect of the cluster cannot be obtained. Population-averaged comparisons make no specific use of within-cluster comparisons varying covariates and substantially underestimate the within-cluster effect.

In the case of the present work, the choice between multilevel and marginal approaches depends on the magnitude of the heterogeneity between-studies. If the treatment effect is essentially the same in each study, the within-group dependencies will be treated as nuisance and the model provides predictions for the entire population. If it is important to allow for clustering structure, one can fit a multilevel model with the group-level random effect.

4.4 THREE-LEVEL RANDOM INTERCEPT MODEL

Meta-analysis in medicine is meant to evaluate the significant effect size among treatments, taking into account both the within-study (among patients evaluated in a same trial) and the between-studies (which identify the magnitude of heterogeneity into the meta-analysis) correlation structure.

We began by defining the structure of the two-level random intercept model [41,144-149]. Consider *N* arms (at level 1) nested within *J* studies (at level 2), with n_j arms in study *j* and K treatments. We indicate with y_{ij} the response for arm *i* in study *j* and with τ_{ijk} the usual series of dummy variables used to parameterize the treatment effect (there are k-1 factor levels plus 1 representing the reference treatment). The response y_{ij} is usually the mean of a continuous outcome but in the presence of binary data, as an approximation, may also be the log odds of an event frequency.

The random intercept model for a response y_{ij} is as follows:

$$y_{ij} = \beta_0 + \beta_{1k}\tau_{ijk} + u_j + e_{ij} \qquad for\ i = 1,2,\dots,N;\ j = 1,2,\dots,J;\ k = 2,\dots,K \qquad (4.1)$$

where β_0 is the mean of y_{ij} (across all studies) in the reference treatment 1, β_{1k} indicates the difference in the effect between treatment *k* and the reference treatment 1, u_j and e_{ij} are the first and second residual terms, respectively. The overall relationship between y_{ij} and τ_{ijk} is represented by a straight line with intercept β_0 and slope β_{1k}. The residual component terms are assumed to be independent and identically distributed (iid) as s normal distribution with zero mean: $u_j \sim N(0, \sigma_u^2)$ iid and $e_{ij} \sim N(0, \sigma_e^2)$ iid *for* $i = 1,2,\dots,N;\ j = 1,2,\dots,J$.

The multilevel model consists of two parts: (i) a fixed part $\beta_0 + \beta_{1k}\tau_{ijk}$ (with fixed parameters β_0 and β_{1k}) which formalizes the link betweeny_{ij} and τ_{ijk} and (ii) a random part

$u_j + e_{ij}$ (with random parameters σ_u^2 and σ_e^2) which contains the first- and second- level residuals. The total residual variance is the sum of the two residual components, $\sigma_u^2 + \sigma_e^2$.
The model in the equation (4.1) can be rewritten in terms of two equations to highlight the random nature of the intercept:

$$y_{ij} = \beta_{0j} + \beta_{1k}\tau_{ijk} + e_{ij} \qquad for\ i = 1,2,\dots,N;\ j = 1,2,\dots,J; k = 2,\dots,K$$

$$\beta_{0j} = \beta_0 + u_j \qquad for\ j = 1,2,\dots,J \tag{4.2}$$

The intercept for a given study *j* is $\beta_0 + u_j$ and the study effect u_j is given by the difference between the study *j* 's mean ($\bar{y}_j$) and the overall mean β_0. The slopes β_{1k} are assumed to be the same for each study so that the predicted study-specific regression lines are parallels.
To define the second level of the hierarchical structure (the study classification),we assume that the trial effect is random, implicitly admitting that some existing studies or treatments may not be included in the meta-analysis for any reason. However, the missed studies or treatments are still assumed to be missing at random.
The model (1) can be made more complex by allowing for the presence of heterogeneity of treatment effects between studies. Ignoring the heterogeneity of the treatment effect may grossly underestimate the uncertainty. The model will include the cross-level term, namely a random treatment effect at the study level, and becomes:

$$y_{ij} = \beta_0 + \beta_{1k}\tau_{ijk} + u_{0j} + u_{1j}\tau_{ijk} + e_{ij}$$

$$for\ i = 1,2,\dots,N;\ j = 1,2,\dots,J;\ k = 2,\dots,K \tag{3}$$

with the second level residual term assumed to follow a bivariate normal distribution with zero mean: $\boldsymbol{u} = \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N_2(0, \Omega_u)$ iid where $\Omega_u = \begin{bmatrix} \sigma_{u_0}^2 & \sigma_{u_{01}} \\ \sigma_{u_{01}} & \sigma_{u_1}^2 \end{bmatrix}$ is the covariance matrix of the random effects.
In the frequentist multilevel approach, it is easy to include and assess the effect of the study design by adding this extra level to the hierarchy of the data structure. Furthermore, if the outcome is continuous (therefore the response is reported as mean together with the sample standard deviation) the model may adjust for its variability by incorporating the arm-specific standard deviation as a first level covariate.
The three-level random intercept model including design is:

$$y_{ijl} = \beta_0 + \beta_{1k}\tau_{ijlk} + v_l + u_{0jl} + u_{1jl}\tau_{ijlk} + e_{ijl}$$

with $v_l \sim N(0, \sigma_v^2)$ iid, $\boldsymbol{u} \sim N(0, \Omega_u)$ iid and $e_{ijl} \sim N(0, \sigma_e^2)$ iid

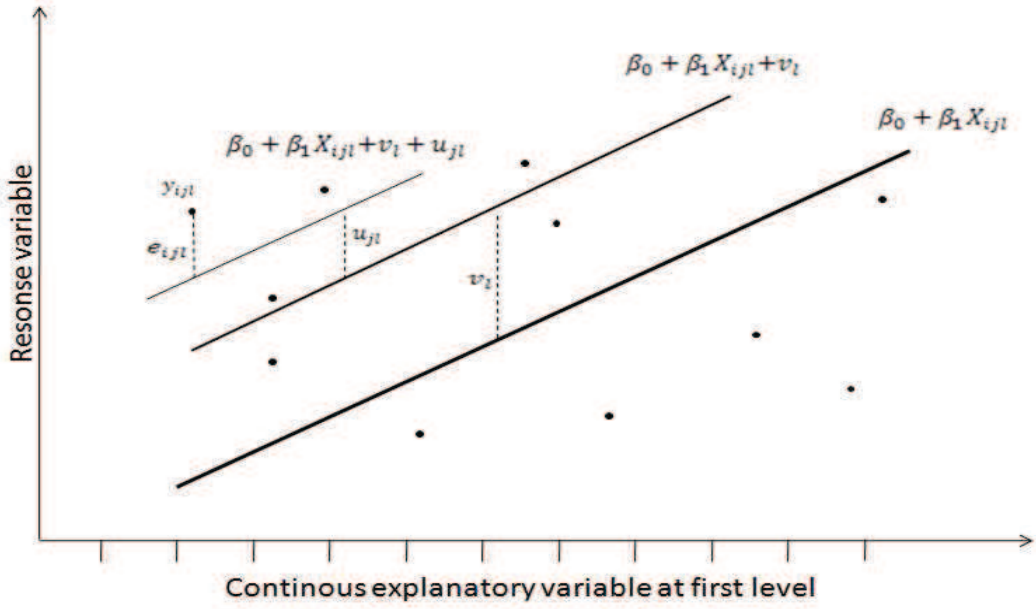
$$for\ i = 1,2,\dots,N;\ j = 1,2,\dots,J; k = 2,\dots,K;\ l = 1,2\dots,L \tag{4}$$

where the overall intercept β_0 measures the mean of y_{ijl} in the reference treatment across all designs and all studies, v_l represents the residual component of the third level *l*, $\beta_0 + v_l$ is the mean of y_{ijl} for design *l*, and $\beta_0 + v_l + u_{0jl}$ is the mean of y_{ijl} for the study *j*. Now, the random part of the model is given by $v_l + u_{0jl} + u_{1jl}\tau_{ijlk} + e_{ijl}$ and the random parameters are σ_v^2, Ω_u and σ_e^2, respectively.

For the purpose of the data analysis, it is a good practice to start from the most complicated model (i.e. with a three-level structure and cross-level terms) and test the importance of the variance components to choose whether or not to use an extra level in the hierarchical structure.

Figure 4.3 provides a graphical interpretation of the intercept and random effects of a hypothetical random intercept model based on a three-level data structure, by fixing a given study *j* and a design *l,* with a continuous explanatory variable at first level.

Figure 4.3: Graphical illustation of the variance component of a random slope model of a three-level data structure.



4.5 PUBLICATION BIAS

Publication bias, the main cause of small study effects, is one of the major sources of type I error (which increases the probability of false positive results). A meta-analysis is affected by

publication bias when studies with statistically significant and positive results have a better chance of being published, are published earlier or in journals with higher impact factors, and/or are more likely to be cited by others. A graphical evaluation of this bias is provided by the funnel plot, where the individual effect size is plotted versus a measure of its precision. Asymmetry or gaps in the plot are suggestive of such bias. Conversely, if the effect size of each included study is symmetrically distributed around the underlying true effect size, no publication bias is present.

To assess publication bias we start from the random intercept model (4.3) and add to this model a second level explanatory variable representing a measure of precision and the treatment-by-precision interaction term to evaluate if the treatment effect varies with the study precision.

Let P_{ij} indicate the precision of each study included in the meta-analysis. The random intercept model for a response y_{ij} becomes:

$$y_{ij} = \beta_0 + \beta_{1k}\tau_{ijk} + \beta_2 P_{ij} + \beta_3 \tau_{ijk} P_{ij} + u_{0j} + u_{1j}\tau_{ijk} + e_{ij}$$

$$for\ i = 1,2,\dots,N;\ j = 1,2,\dots,J;\ k = 2,\dots K \qquad (4.5)$$

where, as before, we assume that the effect of publication bias on the response y_{ij} is the same for all studies.

The significance of the treatment-by-precision interaction is suggestive of publication bias, although the non significance of the precision covariates (main and interaction terms) cannot exclude the presence of this bias.

Other authors [154-156] suggest the use of different measures of study precision, including the inverse of the sample size or the sampling variance, instead of the inverse of the squared standard error. These are all, highly correlated, valid alternatives [148]. Publication bias may be assessed as previously described for those measures too.

4.6 TESTING CONSISTENCY

According to its widest meaning [157] the presence of inconsistency, in a network structure, implies that the treatment effects vary among different designs [39,157]. For example, there may be an inconsistency network when the treatment effect difference between arms A and

B is different in studies comparing A and B only (AB design) and in studies which evaluated together the arms A, B and C (ABC design).

In the frequentist approach, the consistency assumption may be tested by looking at the interaction treatment-by-design term. The corresponding inconsistency model may be fitted by adding two fixed effects to the two-level random intercept model: (i) a fixed main effect for the *d*th design factor, δ_{ijd}, and (ii) a fixed effect for the interaction between the *k*th treatment and the *d*th design factors, $\tau_{ijk}\delta_{ijd}$. The random intercept model in equation (4.3) may be rewritten therefore as:

$$y_{ij} = \beta_0 + \beta_{1k}\tau_{ijk} + \beta_{2k}\delta_{ijd} + \beta_{3kd}\tau_{ijk}\delta_{ijd} + u_{0j} + u_{1j}\tau_{ijk} + e_{ij}$$

$$for\ i = 1,2,\dots,N;\ j = 1,2,\dots,J; k = 2,\dots K; d = 2,\dots,D \quad (4.6)$$

Testing consistency implies testing for the significance of the interaction treatment-by-design term. However, this may be onerous, with *(K-1)x(D-1)*factor levels, and the convergence of the estimation algorithm may not be achieved. In this case, it is reasonable to test for the significance of the main effects before and then to consider the interaction terms.

The proposed three-level random intercept model (4.4) includes the design factor as a random parameter to capture the heterogeneity of treatment effect due to the different comparisons.

In the Bayesian inconsistency model the consistency assumption is tested implementing the posterior probability check method [141], which allows comparison of the difference in the residual deviance between the consistency and the inconsistency model (which assumes a prior distributions for all independent treatment difference).

4.7 ESTIMATION PROCEDURE

In meta-analysis, the question of whether all studies report or not the same outcome is an essential issue. Therefore, it is very important to have models which estimate the contribution of each random effect to the variance of the dependent variable. Indeed, the standard errors of the coefficients of higher level predictors may be underestimated when a single-level model is used. The fact that multilevel models, also known as variance

component models, estimate the variability accounted for each level of the hierarchy and obtain correct standard error estimates is just one reason for using multilevel modeling.

Multilevel models for continuous responses are usually fitted using maximum likelihood procedures. However, for binary responses these procedures are highly computer-intensive. The main difficulty to fit a generalized linear mixed model is that the likelihood does not have a closed form. Assuming that $f(y|u)$ is the conditional distribution of the data and $p(u)$ is the distribution of the random effect, one would maximize the marginal likelihood to obtain the maximum likelihood estimates:

$$L = \int f(y|u)p(u)du \quad (4.6)$$

When the random effect enters in the model in a nonlinear form the integral cannot be solved in a closed form. Two available solution methods to proceed are: methods based on linearization and methods based on integral (or numerical) approximation [38,146,149,158].

Linearization methods use expansions to approximate the model with another one based on pseudo-data with fewer nonlinear components. Pseudo-likelihood methods for generalized linear mixed models involve Taylor series expansions (linearizations) creating pseudo-data for each optimization. Those data are then transformed to have zero mean in a restricted likelihood method. In the restricted (or residual) maximum likelihood estimate, the covariance parameter estimates are the maximum likelihood estimates for the transformed data but the fixed effect estimates are generalized least squares estimates. In other words, the restricted method is a function of the variance component only. In non-restricted procedures both the covariance parameters and the fixed effect estimates are maximum likelihood estimates, but the former have greater bias.

Direct maximum likelihood via integral approximation essentially replaces the integration in (4.6) with a summation. This occurs approximating the normal distribution for the random effects by a discrete distribution with q points. Integral approximation procedures allow to compare nested models with the true likelihood ratio tests. The most commonly used method of numerical integration is Gauss-Hermite with numerical or adaptive quadrature. The Gauss-Hermite quadrature approximates the original integral multiplying the integrand by a function having a normal density distribution. This results in a finite weighted sum that assesses the function at certain points, q. The approximation improves as q increases. It is a good practice to sequentially increase q until changes in both estimates and standard errors are negligible. The adaptive version of the Gauss-Hermite quadrature improves the

efficiency reducing the number of quadrature points needed because it centers the q points with respect to the mode of the function being integrated [159]. Numerical integration is computationally intensive for models with more complex population structures, in large datasets or with multiple random effects. Maximum likelihood simulation is an alternative to numerical quadrature which is more efficient when there are a large number of random parameters, although it is still highly computationally intensive [160].

Quasi-likelihood methods, including marginal and penalized quasi-likelihood, are other types of numerical integration procedures to direct maximum likelihood using an exponential family representation of each component of the joint distribution in equation (4.6). Hence, the integrand of (4.6) is an exponential function of the random effect u. For example, the Laplace approximation is done using the second-order of Taylor series expansion of this exponent around the point $\tilde{u}$ [159].

4.8 CASE STUDY ON ANESTHETIC AGENTS

In the following we present an application of multilevel NMA on the effect on mortality of anesthetics drugs. We derive results of NMA from both fixed and random effect models. We compare the obtained results with those of a previously published Bayesian NMA on a binary endpoint which in detail examined the effect on mortality of desflurane, isoflurane, sevoflurane, and total intravenous anaesthetics (TIVA) at the longest clinical follow-up available [95,161]. Anaesthetic drugs have pharmacological properties which go beyond their effects on blood pressure and heart rate and they might induce cardiac protection. An international, web based consensus conference recently included volatile anaesthetics among the few drugs that might reduce mortality in patients undergoing cardiac surgery [101].

The data structure of included 38 randomized controlled trials published between 1991 and 2012, with data on 3,996 patients [95], is presented in the appendix 1. The analyses were carried out in SAS 9.2.

4.8.1 FIXED EFFECT MODEL

Firstly, we implemented a fixed effect model following an approach based on the population-averaged interpretation of the coefficients. We modeled data using GEE to

consider correlating features within study. For simplicity, we assumed the same correlation between any two elements of a cluster (exchangeable correlation matrix). The GENMOD procedure in SAS extends the traditional linear model theory to generalized linear models by allowing the mean of a population to depend on a linear predictor through a nonlinear link function. The GEE implementation in the GENMOD procedure is a marginal method that does not incorporate any random effects. We model the mean of the average response (i.e. treatment effect differences in term of odds ratios) over the sub-populations (across all studies).

The following is the code to perform the fixed effect model taking into account within-study dependences:

```
proc genmod data=q.alog;
 class treat study;
 model M/N=treat /dist=binomial link=logit type3;
 repeated subject=study /type=exch;
 lsmeans treat /diff cl or;
run;
quit;
```

GENMOD procedure refers to the method of moments for the effect estimations.

The CLASS statement allows to establish which are the categorical variables included and to determine which variables in the model will define the classification levels. In this case, the variable TREAT defines the treatment (TIVA, isoflurane, desflurane, or sevoflurane) administrated in each arm and STUDY defines the identifier of each trial included in the meta-analysis.

The MODEL statement names the response and explanatory variables, including main effects of interest, covariates, interactions, and nested effects. The procedure allows the input of binary response data that are grouped: *M* represents the number of events (death) and *N* represents the sample size in each arm. A TYPE3 analysis is similar to the Type III sums of squares calculated for the general linear model, except that likelihood ratios are used instead of sums of squares. The TYPE3 command produces a table that contains the likelihood ratio statistics, degrees of freedom, and p-values based on the limiting chi-square distributions for each effect in the model. LINK=LOGIT and DIST=BINOMIAL identify the appropriate transformation and distribution for binary data.

The REPEATED statement specifies the within-groups variance and covariance structure of multivariate responses for the GEE model and controls the iterative fitting algorithm. SUBJECT=STUDY states which are the subjects in the input dataset which corresponded to the studies in our case. SUBJECT option in GENMOD procedure does not allow to specify more than one classification level and therefore is not possible to account for both study and design (namely the type of treatment comparison used in each study) effects. However one can test the significance of the design effect as a covariate including this variable in the model. Moreover, fixed effect assumption in meta-analyses implies that responses from different subjects are assumed to be statistically independent, and responses within subjects are assumed to be correlated according to a working correlation matrix identified by the TYPE command. In this case, the command EXCH refers to an exchangeable within-group correlation which assumes non-zero but equal correlations between each pair of individuals in the same group. This is equivalent to the within-group correlation structure assumed in the random intercept multilevel model (TYPE=CS - compound symmetry).

The LSMEANS statement computes least-squares means corresponding to the basic treatment effects (i.e. those involving the reference treatment). The DIFF, CL and OR options allow to compute all the treatment effect differences (in this case study, expresses in term of the natural logarithm of the odds ratio), the odds ratios and their 95% confident interval.

Results from GEE models are valid under the assumption that the distribution of missingness (DOM) is completely at random (MCAR), which means that the probability of dropout is unrelated to any characteristics of the included observations [162-164]. Accordingly, PROC GENMOD ignores any observation with a missing value for any variable involved in the model. Alternatively, PROC GLIMMIX can fit marginal models but the covariance parameters are estimated by likelihood-based techniques and not by the method of moments as with PROC GENMOD. For likelihood or Bayesian estimation procedures, we may generally ignore the DOM when the missing data are missing at random (MAR), which means that the probability of dropout may be related to covariates and to pre-dropout responses [162-164].

The following is the alternative code, using the GLIMMIX procedure, to perform the fixed effect model using the generalized linear mixed model:

```
proc glimmix data=q.alog empirical=classical;
 class treat study;
```

```
 model m/N=treat /dist=binomial link=logit ddfm=bw solution;
 random _residual_ /type=cs subject=study;
 lsmeans treat /diff or cl;
run;
```

The option EMPIRICAL=CLASSICAL requests that the covariance matrix of the fixed-effects parameter estimates is computed by using one of the asymptotically consistent estimators, known as sandwich or empirical estimators [159]. As before, the CLASS statement defines the classification levels of the model and the MODEL specifies the association of interest. The DDFM=BW (abbreviation of between) option in the MODEL statement defines the degrees-of-freedom method and requires that the data are processed by subjects. The SOLUTION option requests to print out the fixed effect parameter estimates.

The RANDOM statement is better explained into the next section (see the "4.8.2 Random effect model" paragraph). As before, the LSMEANS statement produces the estimates of the average logits of the treatment groups. Since the indirect estimate can be calculated as the difference of the corresponding direct estimates, the consistency equation (2.1) is satisfied.

Table 4.1 shows the results of fitting the different fixed effect consistency models to estimate the anaesthetic agent difference effects. The left part of the table reports the mean of the posterior distribution of the odds ratios, and the corresponding 95% credible intervals, estimated from the Bayesian hierarchical model with a MCMC simulation. The right part of the table reports the estimates of the odds ratio, the corresponding 95% confidence intervals and p-values, from the multilevel NMA fitted by both GENMOD and GLIMMIX procedures.

Table 4.1: Fixed effect consistency models to estimate the anaesthetic agent difference effects. Comparison between the mean of the posterior distribution of the odds ratios and the corresponding 95% credible intervals, derived by Bayesian hierarchical model with Markov Chain Monte Carlo algorithm, and the estimate odds ratios, the corresponding 95% confidence intervals and p-values, derived from multilevel network meta-analysis.

	Bayesian approach		**Multilevel network meta-analysis (PROC GENMOD)**		
Contrast	**OR**	**95% credible interval**	**OR**	**95% confidence interval**	**P-value**
Sevoflurane vs TIVA	0.31‡	0.14-0.64	0.34‡	0.15-0.78	0.0109
Desflurane vs TIVA	0.43‡	0.21-0.82	0.49‡	0.31-0.78	0.0028
Isoflurane vs TIVA	0.42	0.15-1.09	0.51	0.26-1.00	0.0504
Sevoflurane vs desflurane *	0.74	0.27-2.01	0.68	0.23-1.99	0.4807
Sevoflurane vs isoflurane *	0.76	0.22-2.60	0.66	0.22-1.93	0.4476
Desflurane vs isoflurane *	1.03	0.31-3.38	0.98	0.42-2.38	0.9408
			Multilevel network meta-analysis (PROC GLIMMIX)		
Contrast			**OR**	**95% confidence interval**	**P-value**
Sevoflurane vs TIVA			0.34‡	0.15-0.78	0.0122
Desflurane vs TIVA			0.54‡	0.35-0.85	0.0091
Isoflurane vs TIVA			0.50	0.24-1.06	0.0687
Sevoflurane vs desflurane *			0.63	0.23-1.72	0.3627
Sevoflurane vs isoflurane *			0.68	0.25-1.86	0.4471
Desflurane vs isoflurane *			1.08	0.42-2.78	0.8703

* Indirect treatment difference effect calculated from consistency equation (2.1)

‡ Significant treatment difference effect

TIVA: Total intravenous anesthesia

Results are not materially different, suggesting that the proposed frequentist multilevel modeling approach is suited to NMA data.

Adding the design variable as fixed effect, the code begins:

```
proc genmod data=q.alog;
 class treat study design;
 model m/N=treat design /dist=binomial link=logit type3covb;
 repeated subject=study /type=exch;
 lsmeans treat /diff cl or;
run;
quit;
```

The fixed effect of the design variable was not associated at the response variable (p=0.3) indicating the plausibility of the consistency model. Accordingly, the magnitude of treatment estimates and the corresponding width of the confidence interval was not materially changed with the introduction of the information of the design.

When we refitted the model adding the interaction treatment-by-design term, the SAS software returns a message indicating convergence problems. This is probably due to the scarce information available overall which does not allow to estimate a model including also a interaction term.

4.8.2 RANDOM EFFECT MODEL

The GLIMMIX procedure in SAS software defines two types of random effects. The program distinguishes between *G*-side and *R*-side random effects depending on whether the parameters of the covariance structure, for random components, are contained in the *G* or in *R* matrix. The GLIMMIX procedure can fit models that have none, one, or more of each type of effects.

Suppose to define the general matrix structure of the generalized linear mixed model [148,159] as $Y = X\beta + Zu + e$where the random effect *u* is normally distributed with mean 0 and variance *G* and the residual effects e is normally distributed with mean 0 and variance *R* [36,37]. Models with only *R*-side random effects, $Y = X\beta + e,$ are also known as marginal (or population-averaged) models.

The variance-covariance matrix in a model with only a *R*-side random component is given by: $var(Y) = A^{1/2}RA^{1/2}$ where *A* is a diagonal matrix containing the variance function.

The command to specify the *R*-side covariance structure in the GLIMMIX procedure is:

random _residual_ /type=cs subject=study(design);

which is equivalent to the covariance structure implied by the REPEATED command in the GENMOD procedure:

repeated subject=study(design)/type=exch;

However, the GEE estimation in the GENMOD procedure allows for the estimation of only *R*-side variance-covariance matrices. On the other hand, the GLIMMIX procedure allows for the specification of both *G*-side and *R*-side variance-covariance matrix using the RANDOM statement. The following code fits the random effect consistency model using the GLIMMIX procedure:

```
proc glimmix data=q.alog;
 class treat study design;
 model m/N=treat /dist=binomial link=logit ddfm=bw solution;
 random intercept treat /type=cs subject=study(design);
 random intercept /type=cs subject=design;
 lsmeans treat /diff or cl;
run;
```

We added the design level in the class statement to accommodate for the three-level data structure.

In SAS GLIMMIX procedure, the default estimation method for generalized linear mixed models is the restricted pseudo-likelihood with a subject-specific expansion [158]. One can choose another estimation methods specifying the option METHOD in the DATA statement.

The RANDOM statements define the hierarchical multilevel structure. In this case, we specified an intercept term that randomly varies at the level of the design effect and at the level of study effect (within design). The TYPE option defines the covariance structure of *G*. In this case, we specified a compound symmetry structure (TYPE=CS) because it is equivalent to the exchangeable structure specified in the GENMOD procedure.

Table 4.2 shows the results of fitting different random effect consistency models to estimate the treatment difference effects in the case study.

Table 4.2. Random effect consistency models to estimate the anaesthetic agent difference effects. Comparison between the mean of the posterior distribution of the odds ratios and the corresponding 95% credible intervals, derived by Bayesian hierarchical model with Markov Chain Monte Carlo algorithm, and the estimate odds ratios, the corresponding 95% confidence intervals and p-values, derived from multilevel network meta-analysis.

	Bayesian approach		**Multilevel network meta-analysis (residual pseudo-likelihood)**	
Contrast	**OR**	**95% credible interval**	**OR**	**95% confidence interval**
Sevoflurane vs TIVA	0.33[b]	0.13-0.83	0.35[b]	0.14-0.83
Desflurane vs TIVA	0.41[b]	0.16-0.87	0.53	0.24-1.17
Isoflurane vs TIVA	0.40	0.12-1.18	0.55	0.19-1.63
Sevoflurane vs desflurane [a]	0.82	0.24-2.78	0.66	0.23-1.88
Sevoflurane vs isoflurane [a]	0.83	0.20-3.49	0.62	0.19-2.15
Desflurane vs isoflurane [a]	1.02	0.25-4.15	0.95	0.27-3.36
Variability	**Estimate**	**95% credible interval**	**Estimate**	**Standard error**
Between trials standard deviation	0.34	0.03-1.37		
- Study intercept variance			0.6412	0.3574
- Treatment slope variance			2,78E-17	.
- Design intercept variance			0.0046	0.2575

[a] Indirect treatment difference effect calculated from consistency equation

[b] Significant treatment difference effect

TIVA: Total intravenous anesthesia

As before, the left part of the table 4.2 reports the mean of posterior distribution of odds ratios and the corresponding 95% credible intervals estimated with the Bayesian hierarchical model. The right part of the table shows the estimates of odds ratio, with the corresponding 95% confidence intervals and p-values, fitting the multilevel NMA. Results are comparable, although the multilevel model found only a trend towards a reduction in mortality when comparing desflurane versus TIVA.

The multilevel models produce more conservative estimates than the Bayesian one. Moreover, the variance estimate for the design classification variable is close to zero indicating that the design does not influence the treatment effect on the response and this may point to the consistency of model.

To formally test for this hypothesis, we specified the model as in (4.5) adding the design and the interaction treatment-by-design terms. The SAS code used is the following:

```
proc glimmix data=q.alog method=quad(qpoints=2);
 class treat study design;
 model m/N=treat design treat*design /dist=binomial link=logit ddfm=bw solution;
 random intercept treat /type=cs subject=study;
run;
```

Given the complexity of the model, we specified the maximum likelihood estimation method with the adaptive Gauss-Hermite quadrature (METHOD=QUAD) and we chose to impute two quadrature points (QPOINTS=2). The type III test on fixed effects for both the design and the interaction term were not significant (p=0.4 and p=0.3 respectively) suggesting a homogeneous influence of the design type on the response.

The publication bias is examined adding the precision term as a fixed effect:

```
proc glimmix data=q.alog;
 class treat study design;
 model m/N=treat precision /dist=binomial link=logit ddfm=bw solution;
 random intercept treat/ subject=study(design);
 random intercept / subject=design;
 lsmeans treat /diff or cl;
run;
```

The non significant in the type III test on the precision term suggested the absence of publication bias (p=0.3).

4.9 DISCUSSION

We propose multilevel modeling as an alternative approach to carry out a NMA generalizing the multilevel analysis approach described by Hox [144,148], Raudenbush [145], Goldstein [165] and Gage [166].

We suggest to consider the arm-based data, instead of contrast-based ones, as the first level of the hierarchy. The use of arm-level summaries provides several important advantages in terms of precision and flexibility when multi-arm trials are included in the NMA. Compared to the contrast level approach, the arm-level one adopts the exact likelihood of the data (i.e. the binomial distribution for binary data) rather than its normal approximation [70]. Indeed, the inference for likelihood-based meta-analysis is the same, for an arm-level or contrast-level data structures, only when two-arm studies are included. Moreover, multi-arm trials force to take into account the within-study correlation structure as well. Indeed, when using contrast-level data the researcher has to specify the variance-covariance matrix for each multi-arm trial to reflect the data correlation structure. This implies an adjustment of the likelihood. Furthermore, as discussed in Zhang [167] the arm-based method is more robust to presence of missing data and is more accurate compared to the contrast-based one.

Model flexibility is another advantage of multilevel analysis, as compared to standard meta-analysis. Indeed, a multilevel framework naturally allows to add extra levels to the model, to include covariates at each level of the hierarchy, and to accommodate for multiple outcomes. It is therefore easier to generalize the multilevel methodology to a network with multiple treatments, defining the statistical implications and model parameterization to perform a multilevel NMA.

Moreover, a NMA can be viewed as the analysis of experimental data when incomplete block treatments are used [168]. Complete and balanced data are not required for a multilevel analysis and the estimates are not affected if individuals may vary in their number of measurements.

Multilevel NMA is a relatively new approach with several issues that are still open to discussion. These include an assessment of the minimum number of trials per design to ensure an adequate statistical power, the definition and evaluation of the bias related to network asymmetry, a more effective specification of the multivariate inconsistency model, and the implementation of a bootstrap procedure to provide a ranking of the treatments.

CHAPTER 5

COMPARISON BETWEEN BAYESIAN AND FREQUENTIST MULTILEVEL NETWORK META-ANALYSES

5.1 INTRODUCTION

In the chapter 3 we highlighted the key steps to perform a valid network meta-analysis (NMA), from literature search to sensitivity analysis, using the Bayesian approach [141], and in chapter 4 we proposed an alternative frequentist method [169], which we called multilevel NMA, for a three-level data structure (arms within studies, studies within study designs and design configuration) that models directly the arm-level information [167].

In the present chapter, we compare the Bayesian and our frequentist-multilevel approach, in performing NMA on publicly available data, and we investigate the descriptive characteristics on either individual studies or NMAs that may contribute to decrease/increase the potential difference between the estimates derived from the two approaches. To do this, we selected a set of published NMAs on any outcomes from a published systematic and narrative review [170] we collected the raw data used in the original analyses and we re-fitted the Bayesian and multilevel models for NMA with both fixed and random effects. The two approaches were compared in terms of the raw or standardized (divided by its standard error (SE)) differences between the derived pooled estimates, and of the Euclidean distance between the standardized estimates.

5.2 METHODS

5.2.1 SEARCH STRATEGY

We searched MEDLINE/PubMed for papers in which any possible approach to NMA was applied, without any restriction on type of included studies (updated on April 15th, 2014).

We excluded papers with the following characteristics: (i) not presenting an indirect comparisons, (ii) methodological or descriptive reports, (iii) commentaries, letters or editorial style reviews, or (iv) protocols of NMAs. Furthermore, we selected NMAs with: (i) at least four treatments, (ii) at least two closed loops, and (iii) at least one dichotomous primary outcome. Authors of the original papers were contacted by email in case of missing raw data.

5.2.2 DATA EXTRACTION

For each included paper (NMA level), the following data were extracted and collected in a spreadsheet file: first author, year and journal of publication, type of intervention or procedure, medical condition, outcome, number of treatments, number of studies included in the analysis, number of multi-arm studies, number of pairwise comparisons, minimum and maximum number of studies for pairwise comparison, total number of patients, total number of events, statistical software and model used in the original paper for data analysis, method to assess the publication bias and method to assess the consistency hypothesis, respectively (table 5.1). Moreover, for each pairwise comparison (arm level) we extracted the information on the number of patients, the number of events, and the number of individual studies involved (table 5.2).

5.2.3 STATISTICAL ANALYSIS

We applied both the Bayesian and our frequentist-multilevel (fixed- and random- effect) models to estimate the treatment effect of each comparison of each NMA, in terms of the pooled logarithm (log) of the odd ratio (OR). We fitted the Bayesian hierarchical models for NMA using the WinBUGS software (freely available in the BUGS project website [171]) and the frequentist-multilevel NMA using SAS 9.4 (SAS Institute Inc., Cary, NC, USA). When we re-fitted our NMA models to the published data with the Bayesian and frequentist-multilevel approaches, SAS and/or WinBUGS software (table 5.2) indicated an estimation issue (i.e. large SEs of the pooled estimates) in the 2.3% of cases (5/216) and a convergence issue in the 5.1% of cases (11/216). In 9 cases, we achieved convergence at the expense of a different model specification. In detail, in case of “trap 66” error message from WinBUGS

software [171], we reduced the variance from 0.0001 to 0.001 in all the prior distributions (2 cases). In case of non-convergence of models fitted by SAS software, we removed the assumption of constant intra-study correlation (7 cases) in favours of a corresponding unstructured correlation matrix. In the remaining 2 cases of unsolved convergence issues [172,173] large SEs of the pooled estimates were present too.

We evaluated the difference between the two approaches (separately for fixed- and random- effect models) using the Generalized Estimating Equation (GEE) models [151] taking into account the intracorrelation within NMA by assuming the same correlation between any two comparisons within each NMA. We considered as model outcomes: (i) the raw pooled log(OR), (ii) the standardized pooled log(OR), calculated as the ratio between each pooled estimate and its SE and (iii) the Euclidean distance calculated as $\sqrt{[(x_0-x_1)^2]}$, where x_0 is the standardized log(OR) derived from the Bayesian NMA and x_1 is the standardized log(OR) derived from the multilevel NMA. Models were fitted either on the 27 available NMAs and on the 20 NMAs with no estimation or convergence issues.

Furthermore, we evaluated the possible effect of NMA-level covariates (one at time) on the presence of estimation or convergence issues, separately for each approach, using the fixed-effect logistic regression model.

Statistical significance was set at the two-tailed 0.05 level. For all levels of analysis, we accounted for the problem of multiple testing by proposing adjusted p-values calculated referring to the Benjamini-Hochberg method [174]. The corresponding unadjusted p-values were reported in the Supplemental Material.

Table 5.1: Descriptive characteristics of the 27 published network meta-analyses providing raw data for our analysis.

Author (year)	Journal	Treatments	Condition	Outcome
Baker WL (2009)	Pharmacotherapy	LABA (fornoterol, salmenterol), ICS (budesonide, futicasone, triamcinolone), LABA + ICS (futicasone plus salmenterol, budesonide plus fornoterol), tiotropium and placebo.	COPD	COPD exacerbation
Chatterjee S (2013)	BMJ	Beta blockers (atenolol, bisoprolol, bucindolol, carvedilol, metoprolol, enalapril, and nebivolol) and placebo.	Chronic heart failure	All cause mortality
Cipriani A (2011)	Lancet	Antimanic drugs (aripiprazole, asenapine, carbamazepine, divalproex, gabapentin, haloperidol, lamotrigine, lithium, olanzapine, quetiapine, risperidone + paliperidone, topiramate, ziprasidone) and placebo.	Acute mania	Dropout rate of the allocated treatment at 3 weeks
Cipriani A (2009)	Lancet	Antidepressant drugs (bupropion, citalopram, duloxetine, escitalopram, fluoxetine, fluvoxamine, milnacipran, mirtazapine, paroxetine, reboxetine, sertraline, and venlafaxine).	Unipolar major depression	Response rate (efficacy analysis)
Dias S (2010)	Stat Med	Thrombolytic drugs and angioplasty (streptokinase, alteplase, accelerated alteplase, streptokinase+ alteplase, reteplase, tenocteplase, urokinase, and anistreptilase, and percutaneous transluminal coronary angioplasty).	Acute myiocardial infarction	Death in 30 or 35 days
Dogliotti A (2014)	Heart	Oral antithrombotics (aspirin, aspirin + clopidogrel, vitamin K antagonists, dabigatran 110 mg, dabigatran 150 mg, rivaroxaban, apixaban) and placebo/control.	Non-valvular atrial fibrillation	Stroke
Dong YH (2013)	Thorax	Inhaled medications (tiotropium Soft Mist Inhaler, tiotropium HandiHaler, LABA, ICS, LABA-ICS combination) and placebo.	COPD	Overall death
Dunkley AJ (2012)	Diabetes Obes Metab	Interventions (lifestyle, pharmacological, lifestyle + pharmacological treatments), no treatment.	Metabolic syndrom	Metabolic syndrome reversed
Elliott WJ (2007)	Lancet	Antihypertensive drugs (ACE inhibitors, ARB, CCB, thiazide diuretic, β blocker) and placebo.	Hypertension (impaired glucose tolerance, insulin resistance, and obesity)	Diabetes incidence
Filippini G (2003)	Cochrane Database Syst Rev	Immunomodulators and immunosuppressants (IFNß-1b Betaseron, IFNß-1a Rebif, IFNß-1a Avonex, glatiramer acetate, natalizumab, mitoxantrone, methotrexate, azathioprine, immunoglobulins, and long-term corticosteroids)	Multiple sclerosis	Clinical relapse over 24 months

Author (year)	Journal	Treatments	Condition	Outcome
		and placebo.		
Fretheim A (2012)	BMC Med	Antihypertensive treatments (ACE inhibitor, ARB, α blocker, β blocker, diuretic, CCB, diuretic plus β blocker), placebo and conventional drug.	Primary prevention of cardiovascular disease	All cause mortality
Lam SK (2007)	British Medical Journal	Combined resynchronisation and implantable cardioverter defibrillator therapy, cardiac resynchronisation therapy, implantable cardioverter defibrillator therapy and medical therapy.	Left ventricular impairment and symptomatic heart failure	All cause mortality
Landoni G (2013)	Br J Anaesth	Anaesthetic drugs (desflurane, isoflurane, sevoflurane, total intravenous anaesthesia).	Coronary artery bypass grafting patients with standard cardiopulmonary bypass	All cause mortality
Li LT (2014)	Colorectal Dis	Skin closure techniques (primary closure, primary closure with a drain, secondary closure, delayed primary closure, loose primary closure, circular closure).	Enterostomy (ileostomy or colostomy) reversal in adult patients	Surgical site infection
Owen A (2010)	Int J Cardiol	Antithrombotic treatments (warfarin, aspirin low dose <300mg daily, aspirin high dose aspirin >300mg daily) and control.	Non valvular atrial fibrillation	Stroke
Psaty BM (2003)	JAMA	Six most commonly used antihypertensive classes (diuretic, β-blocker, ACE inhibitors, ARB, CCB and α-blocker) and placebo.	Cardiovascular desease	Coronary heart desease
Ramsberg J (2012)	PLoS One	Antidepressants (amitriptyline, citalopram, dothiepin, duloxetine, escitalopram, fluoxetine, fluvoxamine, imipramine, lofepramine, maprotiline, milnacipran, mirtazapine, nortriptyline, paroxetine, reboxetine, sertraline, venlafaxine).	Unipolar major depression	Remission drug
Reich K (2012)	Br J Dermatol	Monoclonal antibodies (efalizumab, etanercept 25 mg, etanercept 50 mg, adalimumab, ustekinumab 45 mg, ustekinumab 90 mg, infliximab) and placebo.	Moderate to severe psoriasis	PASI 75 response rate
Reichenpfader U (2014)	Drug Saf	Second-generation antidepressants (bupropion, citalopram, desvenlafaxine, duloxetine, escitalopram, fluoxetine, fluvoxamine, mirtazapine, nefazodone, paroxetine, sertraline, trazodone, venlafaxine) and placebo.	Major depressive disorder	Sexual dysfunction
Ribeiro RA (2013)	Int J Cardiol	Statins intensity regimes (high, intermediate and low) and placebo/no treatment.	Major cardiovascular events	All cause mortality
Sciarretta S (2011)	Arch Intern Med	Antihypertensive treatments (ACE inhibitor, ARB, diuretis, CCB, β-blocker,	Hypertension	Incidence of heart

Author (year)	Journal	Treatments	Condition	Outcome
		conventional treatment, and α-blocker) and placebo.		failure
Tadrous M (2014)	Osteoporos Int	Bisphosphanates (alendronate, risedronate, etidronate, zoledronic-acid) and placebo.	Primary osteoporosis	Gastrointestinal related adverse event
Thijs V (2008)	Eur Heart J	Antiplatelet agents (aspirin, thienopyridines (ticlopidin or clopidogrel), aspirin and dipyridamole, combination of thienopyridines and aspirin, and placebo.	Transient ischaemic attack or stroke	Vascular event
van Valkenhoef G (2012)	J Clin Epidemiol	Second-generation antidepressants (fluoxetine, paroxetine, sertraline, venlafaxine) and placebo.	Depressive disorder	Response rate
Wang H (2010)	J Hosp Infect	Central venous catheteres (standard, heparin-bonded, silver alloy-coated, silver-impregnated, silver iontophoretic, chlorhexidine and silver sulfadiazine, chlorhexidine and silver sulfadiazine catheter blue plus, minocyclinee-rifampicin, benzalkonium chloride, miconazolee-rifampicin).	Nosocomial infection	Catheter-related bloodstream infection on number of central venous catheteres studied
Wu HY (2013)	BMJ	Antihypertensive treatments (ACE inhibitor, ARB, α blockers, β blockers, CCB, diuretics, ACE inhibitor + CCB, ACE inhibitor + diuretics, ARB + CCB, ARB + diuretic, ACE inhibitor + ARB) and placebo.	Diabetes	All cause mortality
Yang B (2014)	PLoS One	Sodium ozagrel, sodium ozagrel + edaravone, ozagrel + edaravone, edaravone, edaravone + kininogenase and placebo.	Cerebral hemorrhage	Acute cerebral stroke

LABA: long-acting β2 agonists; ICS: inhaled corticosteroids; COPD: chronic obstructive pulmonary disease; ACE: angiotensin converting enzyme; ARB angiotensin receptor blocker; CCB: calcium channel blockers; IFN: interferon; PASI: Psoriasis Area and Severity Index.

Table 5.2: Numeric characteristics extracted from the 27 published network meta-analyses providing raw data for our analysis. Information on potential estimation or convergence issues emerged during our re-analysis of the corresponding network meta-analysis. Some descriptive statistics are provided at the bottom of the table to summarize available information.

Author (year)	N.of treatments	N.of studies	N.of multi-arm studies	N.of pairwise comparisons	Mini/max n.of studies per comparison	N.of events	N.of patients	Approach and software used in the original paper	Evaluation of publication bias	Evaluation of inconsistency	Presence of problems fitting the Bayesian model	Presence of problems fitting the multilevel model
Baker WL (2009)	5	39	10	10	1/19	11864	28235	Bayesian approach, WinBUGS	None	None	Convergence problems. “trap 66” error message solved reducing the variance of all prior distributions	None
Chatterjee S (2013)	8	21	0	9	1/8	3871	23122	Bayesian approach, WinBUGS	None	None	None	Convergence problems solved with an unstructured correlation matrix
Cipriani A (2011)	14	64	18	30	1/7	5712	15858	Bayesian approach, R and WinBUGS	None	Difference between indirect and direct estimates and evaluation of cosistency/	Large standard error of the estimates	None

Author (year)	N.of treatments	N.of studies	N.of multi-arm studies	N.of pairwise comparisons	Mini/max n.of studies per comparison	N.of events	N.of patients	Approach and software used in the original paper	Evaluation of publication bias	Evaluation of inconsistency	Presence of problems fitting the Bayesian model	Presence of problems fitting the multilevel model
										inconsistency models fit and parsimony		
Cipriani A (2009)	12	111	2	42	1/11	13951	24595	Bayesian approach, WinBUGS	None	Incoherence defined as the disagreement between direct and indirect evidence with a 95% CI excluding 1	Large standard error of the estimates	None
Dias S (2010)	9	50	2	16	2/9	12484	154201	Bayesian approach, WinBUGS	None	Back-calculation and node-splitting methods	None	None
Dogliotti A (2014)	8	20	4	28	1/9	3004	79808	Bayesian approch, R-GeMTC package	None	Node-splitting method	None	None
Dong YH (2013)	6	41	10	10	1/16	2408	52516	Bayesian approach, WinBUGS	None	None	Convergence problems. “trap 66” error message solved	

Author (year)	N.of treatments	N.of studies	N.of multi-arm studies	N.of pairwise comparisons	Mini/max n.of studies per comparison	N.of events	N.of patients	Approach and software used in the original paper	Evaluation of publication bias	Evaluation of inconsistency	Presence of problems fitting the Bayesian model	Presence of problems fitting the multilevel model
											reducing the variance of all prior distributions	
Dunkley AJ (2012)	4	12	3	6	1/8	1483	3907	Bayesian approach, WinBUGS	None	None	None	None
Elliott WJ (2007)	6	22	3	31	1/5	10865	152216	Frequentist approach, R - Lumley program [35]	None	Degree of incoherence using the Lumley's definition	None	None
Filippini G (2003)	11	25	1	16	1/5	5267	9152	Bayesian approach, WinBUGS	Reporting bias evaluated by means of an adaptation of the funnel plot for pairwise MA (comparison-adjusted funnel plot [200])	Bucher method (difference between direct and indirect estimates in each closed loop) and comparison between model's DIC	Large standard error of the estimates	None
Fretheim A (2012)	9	25	4	15	1/7	14218	164971	Bayesian approach, WinBUGS	None	Node-splitting method and calculation	None	Convergence problems solved with an unstructured

Author (year)	N.of treatments	N.of studies	N.of multi-arm studies	N.of pairwise comparisons	Mini/max n.of studies per comparison	N.of events	N.of patients	Approach and software used in the original paper	Evaluation of publication bias	Evaluation of inconsistency	Presence of problems fitting the Bayesian model	Presence of problems fitting the multilevel model
										of Bayesian p-values for effect to check for inconsistency between direct and indirect evidence		correlation matrix
Lam SK (2007)	4	11	1	5	1/4	1363	7359	Bayesian approach, WinBUGS	None	None	None	None
Landoni G (2013)	4	38	5	6	1/17	76	3996	Bayesian approach, WinBUGS	None	Residual deviance difference between consistency and incosistency models - Post probability check	None	None
Li LT (2014)	6	15	6	11	1/5	235	2929	Bayesian approach, R - R2OpenBUGS package	None	Comparison between direct and indirect evidence	None	Convergence problems solved with an unstructured correlation

Author (year)	N.of treatments	N.of studies	N.of multi-arm studies	N.of pairwise comparisons	Mini/max n.of studies per comparison	N.of events	N.of patients	Approach and software used in the original paper	Evaluation of publication bias	Evaluation of inconsistency	Presence of problems fitting the Bayesian model	Presence of problems fitting the multilevel model
												matrix
Owen A (2010)	4	14	2	5	2/5	411	8250	Bayesian approach, WinBUGS	None	None	None	None
Psaty BM (2003)	7	36	3	12	1/17	8286	180291	Frequentist approach, R - Lumley program [35]	None	Bucher method (difference between direct and indirect estimates in each closed loop)	None	Convergence problems solved with an unstructured correlation matrix
Ramsberg J (2012)	17	87	0	35	1/19	8370	19878	Bayesian approach, WinBUGS	None	Node-splitting method	Large standard error of the estimates	None
Reich K (2012)	8	20	5	11	1/5	3903	10108	Bayesian approach, WinBUGS	None	None	None	None
Reichenpfader U (2014)	12	37	12	29	1/5	2217	10417	Bayesian approach, WinBUGS	None	None	Large standard error of the estimates	Convergence problems solved with an unstructured correlation matrix
Ribeiro RA (2013)	4	44	2	5	3/19	12904	173503	Bayesian approach, WinBUGS	None	Inconsistency test proposed	None	None

Author (year)	N.of treatments	N.of studies	N.of multi-arm studies	N.of pairwise comparisons	Mini/max n.of studies per comparison	N.of events	N.of patients	Approach and software used in the original paper	Evaluation of publication bias	Evaluation of inconsistency	Presence of problems fitting the Bayesian model	Presence of problems fitting the multilevel model
										by [63]		
Sciarretta S (2011)	8	26	2	15	1/5	8554	223313	Bayesian approach, WinBUGS	None	Inconsistency model.	None	Convergence problems solved with an unstructured correlation matrix
Tadrous M (2014)	5	46	1	7	2/27	8523	33471	Bayesian approach, WinBUGS	None	Test of incosistency.	None	None
Thijs V (2008)	5	25	3	7	2/11	7413	50886	Frequentist approach - SAS	None	"lme" function in the package R [35]	None	None
van Valkenhoef G (2012)	5	24	4	10	1/6	2924	5110	Bayesian approach, R-GeMTC package	None	Node-splitting method and performing incosistency models	None	None
Wang H (2010)	10	45	1	14	1/18	467	12085	Bayesian approach, WinBUGS	None	None	Large standard error of the estimates	Convergence problems not solved
Wu HY (2013)	11	62	12	24	1/28	2400	36810	Bayesian approach, WinBUGS	None	Node-splitting method	Large standard error of the estimates	Convergence problems not solved
Yang B (2014)	5	145	0	6	3/57	10333	12983	Bayesian approach,	None	Loop incosistency	None	Convergence problems

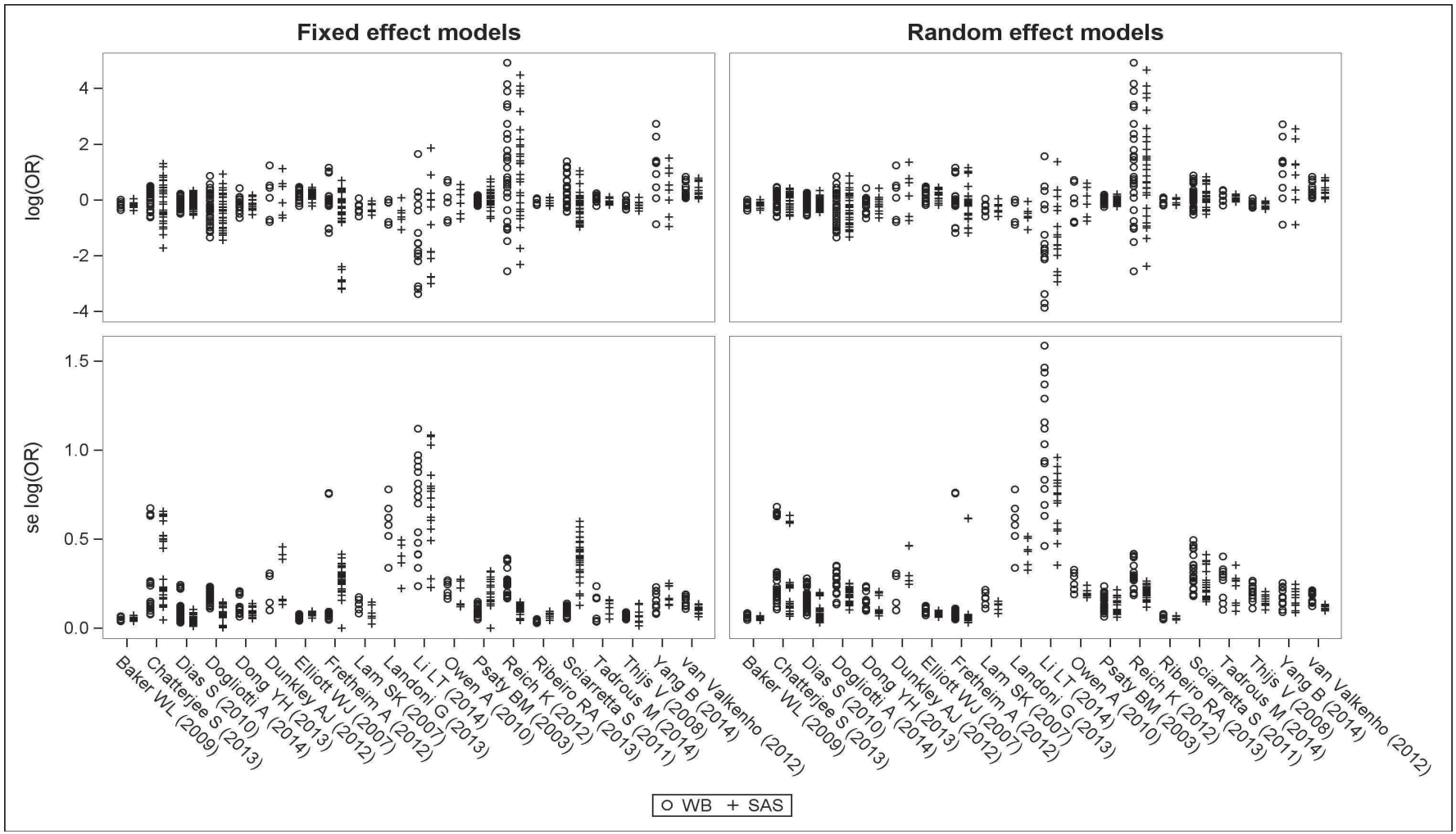
Author (year)	N.of treatments	N.of studies	N.of multi-arm studies	N.of pairwise comparisons	Mini/max n.of studies per comparison	N.of events	N.of patients	Approach and software used in the original paper	Evaluation of publication bias	Evaluation of inconsistency	Presence of problems fitting the Bayesian model	Presence of problems fitting the multilevel model
								WinBUGS and R softwares		and Bayesian model by using the of incosistence factor [36]		solved with an unstructured correlation matrix
MEDIAN	7	36	3	11	1/9	5267	23122	-	-	-	-	-
1st QUARTILE	5	21.5	1.5	7	1/5	2308.5	9630	-	-	-	-	-
3rd QUARTILE	9.5	45.5	5	20	1/17.5	9443.5	66162	-	-	-	-	-
MINIMUM	4	11	0	5	1/4	76	2929	-	-	-	-	-
MAXIMUM	17	145	18	42	3/57	14218	223313	-	-	-	-	-

5.3 RESULTS

Twenty-seven out of the 71 NMAs potentially satisfying the inclusion criteria provided the raw data for the current analysis [95,172,173,175-199]. Tables 5.1 and 5.2 show the individual characteristics of the NMAs included in the analysis. The NMAs under consideration were published between 2003 and 2014. As reported at the bottom of table 5.2, the median number of investigated treatments was 7 (iqr: 5-9.5; range: 4-17), whereas the median number of individual studies included was 36 (iqr: 21.5-45.5; range: 11-145). The median number of multi-arm trials was 3 (iqr: 1.5-5; range: 0-18). Only 6 [172,189,194,196,197,199] out of the 27 NMAs (22%) reported more than one study in each comparison. Most of the included NMAs (89%) were carried out with a Bayesian approach in the original analysis. Only the NMA of Filippini *et al.* [185] reported the use of a method to investigate on the presence of publication bias, while there was heterogeneity in the methods used to test the consistency hypothesis. In 7 NMAs [172,173,178,179,185,191,193] the fitted Bayesian models (both fixed- and random- effect) estimated a very large SE for the treatment effect, and this led to 95% credible intervals between <0.001 and >100. Two NMAs [172,173] showed a convergence problem that remained even when we removed the assumption of correlation matrix within each NMA.

Figures 1 shows the distribution of the pooled log(OR)s (top panel) and corresponding SEs (lower panel) derived from the fixed- (left side) and the random- (right side) effect models, respectively, fitted for each of the 20 NMAs not affected by estimation or convergence issues. Each point of the graphs represents a study-specific estimate of each treatment effect and corresponding SE (on logarithm scale) derived from the Bayesian or frequentist-multilevel approach.

Figure 5.1: Distributions of the pooled odd ratio and corresponding standard error (on a logarithmic scale) for each study-specific effect difference, derived from the fitting of the fixed- and the random- effect models to each of the 20 network meta-analyses not affected by estimation or convergence issues. Abbreviations. Log: logarithm; OR: odds ratio; s.e.: standard error; WB: WinBUGS software; SAS: SAS software.



No matter of the estimation method used, the distributions of the pooled log(OR) were fairly comparable between the frequentist and Bayesian approaches. Three partial exceptions were represented by the Chatterjee [176], Fretheim [186] and Psaty [190] NMAs. However, these differences were attenuated in the random effect model.

Table 5.3 shows the results derived from the univariate analysis evaluating the presence of a potential significant difference in the effects between the Bayesian and the frequentist-multilevel approaches using several available GEE models specified in the table. No matter of the estimation method used, the "approach" effect was not significantly different from zero, when the raw or standardized log(OR)s were considered in the analysis. However, we observed a significant effect of the intercepts included in the random effect regressions that considered the Euclidean distance as the outcome variable (intercept=0.77 and p-value <0.001; intercept=0.72 and p-value <0.001 for models including or excluding the NMAs affected by estimation or convergence issues, respectively). Table A1 of Appendix 2 reports the corresponding unadjusted p-values with results in agreement with the previous ones.

Table 5.4 shows the corresponding results from the multiple analysis including one of the available 14 predictive covariates at a time. Among the NMA-level covariates, the most relevant ones were the percentage of events in each NMA and, marginally, the number of studies included in the NMA. After including in the models any potentially relevant covariate (defined as having a univariate p-value<0.25) at either levels, the differences in the raw or standardized log(OR)s derived according to the Bayesian and frequentist-multilevel approaches were still non-significant.

Table A2 of Appendix 2 reports the corresponding unadjusted p-values, showing consistent results with the previous ones, except for the significant effect of the percentage of events included in each arm, in 3 models.

Table 5.3: Effect estimates, corresponding 95% confidence intervals, and adjusted p-values derived from the univariate analysis* to assess the presence of a difference between the Bayesian and the frequentist multilevel approaches to network meta-analysis (NMA).

Model ID	Number of NMAs included	Number of pooled estimates analyzed	Effects derived from a fixed/random model	Model**	QIC	Parameter	Estimate	Standard Error	95% confidence interval	Adjusted p-value
A.1	27	844	Fixed	Log(OR) = β_0 + approach	1646	Intercept	0.075	0.116	-0.152 to 0301	0.864
						approach	-0.660	0.413	-1.470 to 0.150	0.551
A.2	27	844	Random	Log(OR) = β_0 + approach	1627	Intercept	-0.024	0.086	-0.193 to 0.145	0.869
						approach***	0.038	0.046	-0.052 to 0.128	0.749
B.1	20	330	Fixed	Log(OR) = β_0 + approach	679	Intercept	0.035	0.113	-0.186 to 0.255	0.869
						approach***	-0.145	0.087	-0.316 to 0.025	0.551
B.2	20	330	Random	Log(OR) = β_0 + approach	680	Intercept	0.003	0.115	-0.222 to 0.229	0.976
						approach***	-0.005	0.017	-0.039 to 0.029	0.869
C.1	27	844	Fixed	Standardized log (OR) = β_0 + approach	1624	Intercept	0.041	0.285	-0.519 to 0.601	0.932
						approach***	-30.265	29.550	-88.183 to 27.652	0.682
C.2	27	844	Random	Standardized log (OR) = β_0 + approach	1634	Intercept	0.121	0.358	-0.581 to 0.823	0.869
						approach***	0.110	0.122	-0.130 to 0.349	0.738
D.1	20	330	Fixed	Standardized log (OR) = β_0 + approach	662	Intercept	-3.561	3.487	-10.395 to 3.273	0.682
						approach***	-77.288	70.083	-214.648 to 60.072	0.682
D.2	20	330	Random	Standardized log (OR) = β_0 + approach	678	Intercept	0.239	0.504	-0.748 to 1.226	0.869
						approach***	0.070	0.156	-0.235 to 0.374	0.869
E.1	27	844	Fixed	Euclidean distance = β_0	779	Intercept	36.133	32.001	-26.590 to 98.856	0.682
E.2	27	844	Random	Euclidean distance = β_0	761	Intercept	0.771	0.122	0.532 to 1.009	**<0.001**
F.1	20	330	Fixed	Euclidean distance = β_0	332	Intercept	91.173	77.177	-60.091 to 242.437	0.682
F.2	20	330	Random	Euclidean distance = β_0	337	Intercept	0.724	0.147	0.436 to 1.013	**<0.001**

* This analysis included the main "approach" effect only.

** Generalized Estimating Equations models taking into account the within-NMA correlation

*** Variable "approach" was equal to 0 for the Bayesian approach and to 1 for the frequentist-multilevel one.

Bold style typeface: significant p-value.

Statistical significance was set at the two-tailed 0.05 level. Adjustment for multiple testing was provided by the Benjamini-Hochberg method [174]. Standardized log (OR)=log(OR)/SE[log(OR)]; Euclidean distance= $\sqrt{[(x_0-x_1)^2]}$, where x_0= $\log(OR)_{Bayesian}$ and x_1= $\log(OR)_{multilevel}$

QIC: Quasi-likelihood under the independence model criterion.

Table 4: Adjusted p-values from tests of significance of potential covariates of interest, as derived from the multiple analysis* to assess the presence of a difference between the Bayesian and the frequentist multilevel approaches to network meta-analysis (NMA).

		Model **											
Parameter	**Level**	**A.1**	**A.2**	**B.1**	**B.2**	**C.1**	**C.2**	**D.1**	**D.2**	**E.1**	**E.2**	**F.1**	**F.2**
n_treatment_nma	NMA	0.523	0.970	0.892	0.757	0.523	0.776	0.523	0.882	0.523	0.533	0.523	0.523
n_studies_nma	NMA	0.812	0.523	0.086	**0.027**	0.523	0.523	0.523	0.086	0.523	0.731	0.523	0.523
n_multiarm_nma	NMA	0.591	0.523	0.523	0.523	0.676	0.523	0.640	0.523	0.623	0.121	0.644	0.715
n_pairwise_nma	NMA	0.810	0.691	0.764	0.776	0.523	0.776	0.776	0.916	0.523	0.523	0.731	0.810
min_n_studies_nma	NMA	0.470	0.523	0.523	0.523	0.523	0.559	0.523	0.591	0.523	0.125	0.523	0.344
max_n_studies_nma	NMA	0.747	0.523	0.523	0.433	0.523	0.523	0.523	0.523	0.523	0.523	0.523	0.523
n_events_nma	NMA	0.505	0.523	0.597	0.523	0.523	0.776	0.523	0.812	0.523	0.850	0.523	0.776
n_patients_nma	NMA	0.594	0.946	0.849	0.967	0.523	0.523	0.523	0.523	0.523	0.523	0.523	0.523
perc_nma	NMA	0.086	**0.027**	**0.006**	**0.006**	0.523	0.194	0.523	**0.006**	0.523	0.523	0.523	0.523
n_effect_estimated	NMA	0.596	0.916	0.916	0.757	0.523	0.757	0.523	0.934	0.523	0.704	0.523	0.523
n_arms_arm	Arm	0.157	0.523	0.523	0.523	0.523	0.523	0.523	0.523	0.523	0.523	0.523	0.757
n_events_arm	Arm	0.523	0.776	0.574	0.965	0.523	0.892	0.523	0.970	0.523	0.626	0.523	0.523
n_patients_arm	Arm	0.523	0.591	0.523	0.523	0.776	0.776	0.554	0.523	0.791	0.523	0.554	0.757
perc_arm	Arm	0.554	0.975	0.344	0.289	0.523	0.626	0.523	0.099	0.523	0.523	0.523	0.523

* This multiple analysis included one extra covariate of interest, together with the "approach" effect (when applicable).

** Generalized Estimating Equations models taking into account the within-NMA correlation.

Bold style typeface: significant p-value. Statistical significance was set at the two-tailed 0.05 level. Adjustment for multiple testing was provided by the Benjamini-Hochberg method [174].

NMA: network meta-analysis; **n_treatment_nma**: number of treatments included in the nma; **n_studies_nma**: number of studies included in the NMA; **n_multiarm_nma**: number of multi-arm studies included in the NMA; **n_pairwise_nma**: number of pairwise comparisons taken into account in the NMA; **min_n_studies_nma**: minimum number of studies included in each comparisons; **max_n_studies_nma**: maximum number of studies included in each comparison; **n_events_nma**: total number of events analyzed in the NMA; **n_patients_nma**: total number of patients analyzed in the NMA; **perc_nma**: percentage of events in the NMA; **n_effect_estimated**: number of effects estimated in the NMA; **n_arms_arm**: number of arms taken into account for each comparisons; **n_events_arm**: total number of events analyzed in each arm; **n_patients_arm**: total number of patients analyzed in each arm; **perc_arm**: percentage of events in each arm.

Table 5.5 shows results from the univariate analysis assessing the potential effect of NMA characteristics on the presence of estimation or convergence issues for each approach. The presence of WinBUGS fitting problems is potentially associated with the number of pairwise comparisons taken into account in the NMA (slope=0.178, 95% CI 0.051 to 0.304, p-value=0.059) No significant effect was found to explain possible convergence issues in SAS random-effect models. In agreement with previous results, table A3 of Appendix 2 reports the corresponding unadjusted p-values.

Table 5.5: Adjusted p-values derived from univariate analyses to evaluate the predictors of convergence problems using Bayesian or multivariate approach.

Parameter	Problem with winBUGS	Problem with SAS
n_treatment_nma	0.599	0.679
n_studies_nma	0.390	0.679
n_multiarm_nma	0.407	0.679
n_pairwise_nma	0.059	0.679
min_n_studies_n	0.948	0.957
max_n_studies_n	0.948	0.679
n_events_nma	0.872	0.679
n_patients_nma	0.407	0.679
perc_nma	0.407	0.679
n_effect_estima	0.599	0.679

Bold style typeface: significant p-value. Statistical significance was set at the two-tailed 0.05 level. Adjustment for multiple testing was provided by the Benjamini-Hochberg method [174].
nma: network meta-analysis; **n_treatment_nma**: number of treatments included in the nma; **n_studies_nma**: number of studies included in the nma; **n_multiarm_nma**: number of multi-arm studies included in the nma; **n_pairwise_nma**: number of pairwise comparisons taken into account in the nma; **min_n_studies_nma**: minimum number of studies included in each comparisons; **max_n_studies_nma**: maximum number of studies included in each comparisons; **n_events_nma**: total number of events analyzed in the nma; **n_patients_nma**: total number of patients analyzed in the nma; **perc_nma**: percentage of events in the nma; **n_effect_estimated**: number of effects estimated in the nma.

5.4 DISCUSSION

With no intention to add arguments to any controversy between Bayesian and frequentist approaches [201,202], the purpose of this chapter is to provide a comparison between the frequentist (multilevel) and the Bayesian hierarchical approaches in the estimation of the treatment effect differences in data on 27 previously published NMAs selected by a systematic review on the attractiveness of NMA [167]. Our analysis revealed that there is no

material difference in the pooled estimates obtained with the two approaches when the raw or standardized differences are modelled. However, the Euclidean distance between the standardized pooled estimates is significant in the random effect model.

The absence of a significant difference between the pooled estimates from the two approaches in most of the examined scenarios may be explained in part by our decision to re-fit the Bayesian hierarchical models for each NMA referring to non-informative priors. This choice is motivated by the need to have a uniform criterion in the re-analysis of the NMAs and by the absence of substantive *a priori* knowledge in such different fields of applications of the included NMAs. Our decision is also in agreement with general suggestions in the use of Bayesian approaches to NMA [36,71,141].

We identified a significant p-value for the difference between the two approaches only modelling the Euclidean distance with the random-effect model. Between the two models satisfying these criteria, the one with the best goodness of fit (QIC=337) showed an intercept estimate of 0.72 (95% confidence interval 0.44-1.01). This may suggest that the estimates derived from the Bayesian approach may be systematically greater than those derived from the frequentist-multilevel approach, without knowing whether this affects the significance of the corresponding estimate. It is equally true that the corresponding SE is not negligible (0.15) and this may point to some caution in the interpretation of this result.

Furthermore, it is not surprising to us that this significant difference was found only in the random effect model.

We were able to carry out our analysis on 27 NMAs out of the 71 potentially satisfying the inclusion criteria. We derived the raw data for most of these cases from the original papers. In a few cases, the authors replied to a general request of data we sent via email to the corresponding author of the papers with missing information. Although we acknowledge that the raw data provided by the authors in the original papers may be more reliable, we do not believe that the few cases of information sent to us via email were different from the previous ones in this respect. So, if a bias exists, it affects the entire sample of available NMAs in the same direction. As to the direction of this bias, we may speculate that the included studies have a higher quality, as compared to those which did not provide their original data in any form. We also recognize that the number of NMAs included in our re-

analysis is limited. However, the analysis was actually based on more than 800 comparisons available, and not just on the 27 original NMAs. This reassures against any power issue in this case.

In addition, a main limit of our analysis is that we included all the available NMAs satisfying sensible inclusion criteria, no matter of the specific field of research and outcome considered. The immediate implication is that we cannot give any numeric interpretation to a potential existing difference between the approaches. This is just a preliminary attempt to assess if there are relevant differences between available NMA-approaches or not. Future work may consider NMAs showing a similar outcome of interest to overcome this issue.

We based our comparison between the approaches on the explicit modelling of the difference between the effect estimates and referred to the SE of these estimates to standardize them. Future work may benefit from a parallel direct modelling of the SE of the estimates, to provide an extra insight on the potential existing differences in the estimates' precision.

With the fitting of a large amount of univariate and multiple models, adjustment of the p-values for multiple testing problem was worth a try. An adjustment method like the one from Benjamini and Hochberg provided a less stringent control of Type I Error, compared to family-wise error rate controlling procedures, such as the Bonferroni correction, and this represents a valid alternative to be used in this set-up. In our application, unadjusted and adjusted p-values were in agreement for most of the analyses this reassuring against the effect of chance, especially for the results presented in table 4.4. In this case, it is evident that the percentage of events in the NMA is a significant explanatory variable for the pooled estimate after adjusting for the "approach" variable. This is true for the raw and for the standardized log(OR), although in a less consistent way. Finally, we acknowledge that the proposed re-analysis is based on our frequentist version of NMA modelling. Future work may consider to extend the comparison to include alternative frequentist approaches to NMA [35].

CONCLUSIONS

Meta-analysis is a powerful tool to cumulate and summarize the knowledge in a research field. Nevertheless, the results of a meta-analysis should be interpreted in the light of the various checks which can inform the readers of the likely reliability of the conclusions.
In this work we provided and discussed methods to cope with multiple treatments and to deal with correlated data where correlation can derive from multiple endpoints, time-varying responses or from clustered observation.

We have provided a comprehensive and detailed overview of the conceptual and practical issues involved in performing a and interpreting network meta-analysis on binomial data. We have discussed the general topics related to network meta-analysis, including how to collect study data, structure the network, and set assumptions about the network that lead to different models and interpretations of model parameters. We have strived to put together the most important topics (making available the major references) and we offer, for the first time, a thorough yet manageable guideline to conduct (from literature search to results interpretation) a rigorous network meta-analysis on binomial data, applying both the Bayesian and frequentist approaches. We presented a case study on the beneficial effects of anaesthetic agents and the practical guide with the actual WinBUGS and SAS codes to allow transparency and ease of replication of all steps that are required when carrying out such quantitative syntheses.

We suggest to consider the arm-based data, instead of contrast-based ones, as input data structure. In fact, the use of arm-level summaries allows to adopt the exact likelihood of the data rather than its normal approximation [70] and to not specify the variance-covariance matrix for each multi-arm trial to reflect the data correlation structure.

In the network meta-analysis framework, Bayesian and frequentist approaches are expected to give approximately the same results because it is a common practice to use a non-informative priors in the Bayesian strategy [36]. Indeed, our analyses revealed that there is no material difference in the pooled estimates obtained with the Bayesian and frequentist-multilevel approaches.

A drawback of the Bayesian approach is the complexity in model specification, which requires familiarity with the WinBUGS software and the MCMC methods. On the other hand, multilevel models essentially are suited for simpler regression structures where a single outcome variable depends on a few covariates, and therefore they do not allow to inspect the full range of relationships between variables. The multilevel approach, taking into account the clustering structure, provides correct estimates for standard errors, confidence intervals and tests [166] which are generally more conservative than those stemming from the Bayesian approach and the traditional ones obtained by ignoring the presence of groups.

REFERENCES

1. Pearson K. Report on Certain Enteric Fever Inoculation Statistics. Br Med J 1904; 2:1243-6.

2. Egger M, Ebrahim S, Smith GD. Where now for meta-analysis? Int J Epidemiol 2002; 31:1-5.

3. Glass GV. Primary, secondary and meta-analysis of research. Educ Res 1976; 5:3-8.

4. Biondi-Zoccai G, Landoni G, Modena MG. A journey into clinical evidence: from case reports to mixed treatment comparisons. HSR Proceedings in Intensive Care and Cardiovascular Anesthesia 2011; 3:93-96.

5. Walker E, Hernandez AV, Kattan MW. Meta-analysis: Its strengths and limitations. Cleve Clin J Med 2008; 75:431-9.

6. Stegenga J. Is meta-analysis the platinum standard of evidence? Stud Hist Philos Biol Biomed Sci 2011; 42: 497-507.

7. Biondi-Zoccai G, Lotrionte M, Landoni G, Modena MG. The rough guide to systematic reviews and meta-analysis. HSR Proceedings in Intensive Care and Cardiovascular Anesthesia 2011; 3:161-173.

8. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009; 339:b2700.

9. Takkouche B, Norman G. Meta-analysis protocol registration: sed quis custodiet ipsos custodes? [but who will guard the guardians?].Epidemiology 2010; 21:614-5.

10. Krumholz H. The case for duplication of meta-analyses and systematic reviews. BMJ 2013; 347:f5506.

11. Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomised controlled trials. Lancet 1998; 351:47-52.

12. Timothy MD, Ryan WT. Do we really understand a research topic? funding answers thought meta-analysis. University of Technology, Sydney, Faculty of Business. Available on http://www.academia.edu/2385330/Do_We_Really_Understand_a_Research_Topic_Finding_Answers_Through_Meta-Analysis (Accessed July, 2015).

13. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. Available on http://www.cochrane.org/handbook/chapter-6-searching-studies (Accessed July, 2014).

14. Lyman GH, Kuderer NM. The strengths and limitations of meta-analysis based on aggregate data. BMC Med Res Methodol 2005; 25:14.

15. Harrison F. Getting started with meta-analysis. Methods in Ecology and Evolution 2011, 2:1-10.

16. Egger M, Smith GD, Altman DG. Systematic Reviews in Health Care: Meta-Analysis in Context. London: BMJ Publishing Group 2001:69-86.

17. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analysis? Lancet 1998; 352:609-13.

18. Dubben HH, Beck-Bornholdt HP. Systematic review of publication bias in studies on publication bias. BMJ 2005; 331:433-4.

19. Rothstein HR, Sutton AJ, Dr. Michael Borenstein Director Associate Professor Lecturer PI4Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments. Chapter 8. The Trim and Fill Method. New York John Wiley & Sons, Ltd 2005.

20. Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. J Clin Epidemiol 2000; 53:1119-29.

21. Sterne JA, Egger M, Smith GD. Review Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. BMJ 2001; 323:101-5.

22. Nüesch E, Trelle S, Reichenbach S, et al. Small study effects in meta-analysis of osteoarthritis trials: meta-epidemiological study. BMJ 2010; 341:c3515.

23. Bellomo R, Warrillow SJ, Reade MC. Why we should be wary of single-center trials. Crit Care Med 2009; 37:3114-9.

24. Ng TT, McGory ML, Ko CY, Maggard MA.Meta-analysis in surgery: methods and limitations. Arch Surg 2006; 141:1125-30; discussion 1131.

25. Cai T, Parast L, Ryan L. Meta-analysis for rare events. Stat Med 2010; 29:2078-89.

26. Lane PW. Meta-analysis of incidence of rare events. Stat Methods Med Res 2013; 22:117-32.

27. Greenland S, Morgenstern H. Ecological bias, confounding, and effect modification. Int J Epidemiol 1989; 18:269-74.

28. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine--reporting of subgroup analysis in clinical trials. N Engl J Med 2007; 357:2189-94.

29. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. Users' guides to the medical literature. IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. The Journal of the American Medical Association 1995; 274:1800-4.

30. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996; 312:71-2.

31. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol 1997; 50:683-91.

32. Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. BMJ 2003; 326:472.

33. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health 2011; 14:417-28.

34. Caldwell DM, Ades AE, Higgins JPT. Simultaneous Comparison of Multiple Treatments: Combining Direct and Indirect Evidence; BMJ 2005; 331:897-900.

35. Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med 2002; 21:2313-24.

36. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004; 23:3105-24.

37. Salanti G, Higgins JP, Ades AE, Ioannidis JP. Evaluation of networks of randomized trials. Stat Methods Med Res 2008; 17:279-301.

38. Jackson D, Riley R, White IR. Multivariate meta-analysis: Potential and promise. Stat Med 2011;30:2481-2498.

39. White IR, Barrett JK, Jackson J, Higgins JPT. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. Res Synth Methods 2012; 3:111-125.

40. White IR. Multivariate meta-analysis. The Stata Journal 2009; 9:40-56.

41. Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. Res Synth Methods 2012; 3: 98-110.

42. Hawkins N, Scott DA, Woods B. How far do you go? Efficient searching for indirect evidence. Medical Decision Making 2009; 29:273-81.

43. Lu G, Ades AE. Assessing Evidence Inconsistency in Mixed Treatment Comparisons. J Am Stat Assoc 2006; 101:447-459.

44. Salanti G, Kavvoura FK, Ioannidis JP. Exploring the geometry of treatment networks. Ann Intern Med 2008; 148544-53.

45. Magurran AE. Ecological Diversity and Its Measurement. New Jersey: Princeton University Press 1998.

46. Hamilton AJ. Species diversity or biodiversity? J Environ Manage 2005; 75:89-92.

47. Tiho S, Josensb G. Co-occurrence of earthworms in urban surroundings: A null model analysis of community structure. Eur J Soil Biol 2007; 43:84-90.

48. Stone L, Robert A. The checkerboard score and species distributions. Oncologia 1990; 85:74-79.

49. Salton G. Introduction to Modern Information Retrieval. New York: McGraw-Hill 1983.

50. Sutton A, Ades AE, Cooper N, Abrams K. Use of indirect and mixed treatment comparisons for technology assessment. Pharmacoeconomics 2008; 26:753-67.

51. Jansen JP, Crawford B, Bergman G, Stam W. Bayesian meta-analysis of multiple treatment comparisons: an introduction to mixed treatment comparisons. Value Health 2008; 11:956-64.

52. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002; 21:1539-1558.

55. Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. Stat Med 2001; 20:825-40.

56. Kontopantelis E, Reeves D. Performance of statistical methods for meta-analysis when true study effects are non-normally distributed: A simulation study. Stat Methods Med Res 2012; 21:409-26.

57. Kontopantelis E, Reeves D. Performance of statistical methods for meta-analysis when true study effects are non-normally distributed: a comparison between DerSimonian-Laird and restricted maximum likelihood. Stat Methods Med Res 2012; 21:657-9.

58. Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. Stat Med 1996; 15:2733-49.

59. Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. Int J Epidemiol 2012; 41:818-27.

60. Song F, Loke YK, Walsh T, Glenny AM, Eastwood AJ, Altman DG. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. BMJ 2009; 3(338): b1147.

61. Donegan S, Williamson P, Gamble C, Tudur-Smith C. Indirect comparisons: a review of reporting and methodological quality. PLoS One 2010; 5:e11054.

62. Hardy RJ, Thompson SG. Detecting and describing heterogeneity in meta-analysis. Stat Med 1998; 17:841-56.

63. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003; 327:557-560.

64. Mittlböck M, Heinzl H. A simulation study comparing properties of heterogeneity measures in meta-analyses. Stat Med 2006; 25:4321-33.

65. Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation, New York: Wiley 2004.

66. Cooper NJ, Sutton AJ, Morris D, Ades AE, Welton NJ. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. Stat Med 2009; 28: 1861-81.

67. Caldwell DM, Welton NJ, Ades AE. Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency. J Clin Epidemiol 2010; 63: 875-82.

68. Donegan S, Williamson P, D'Alessandro U, Tudur Smith C. Assessing the consistency assumption by exploring treatment by covariate interactions in mixed treatment comparison meta-analysis: individual patient-level covariates versus aggregate trial-level covariates. Stat Med 2012; 31:3840-57.

69. Dias S, Welton WJ, Sutton AJ, Caldwell MD, Lu G, Ades AE. NICE DSU Technical Support Document 4: inconsistency in networks of evidence based on randomised controlled trial. Available on http://www.nicedsu.org.uk. (Accessed July, 2014).

70. Franchini AJ, Dias S, Ades AE, Jansen JP, Welton NJ. Accounting for correlation in network meta-analysis with multi-arm trials. Research Synthesis Methods 2012; 3:142–160.

71. Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU Technical Support Document 2: A generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. Available on http://www.nicedsu.org.uk (Accessed July, 2014).

72. Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. Value Health 2011; 14:429-37.

73. Dias S, Sutton AJ, Welton WJ, Ades AE. NICE DSU Technical Support Document 3: heterogeneity: subgroups, meta-regression, bias and bias-adjustment. Available on http://www.nicedsu.org.uk (Accessed July, 2014).

74. Nixon RM, Bansback N, Brennan A. Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. Stat Med 2007; 26:1237-54.

75. Lu G, Ades A. Modeling between-trial variance structure in mixed treatment comparisons. Biostatistics 2009; 10:792-805.

76. Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. J Clin Epidemiol 2002; 55:86-94.

77. Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI; Anti-Lymphocyte Antibody Induction Therapy Study Group. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. Stat Med 2002; 21:371-87.

78. Spiegelhalter D, Thomas A, Best N, Lunn D. WinBUGS User Manual. Version 1.4, 2003. Available on http://www.mrc-bsu.cam.ac.uk/bugs (Accessed July, 2014).

79. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. Stat Med 1995; 14:2685-99.

80. Ntzoufras I. Bayesian Modeling Using WinBUGS. New York: Wiley 2009.

81. Adamakis S, Raftery CL, Walsh RW, Gallagher PT. A Bayesian approach to comparing theoretic models to observational data: A case study from solar flare physics. astro-ph.SR 2012; arXiv:1102.0242v3 Available at: http://arxiv.org/abs/1102.0242 (Accessed July, 2014).

82. Van Dongen S. Prior specification in Bayesian statistics: three cautionary tales. J Theor Biol 2006; 242:90-100.

83. Gelman A. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 2006;1:515-533.

84. Crowder MJ. Maximum likelihood estimation for dependent observations. Journal of the Royal Statistical Society 1976; B38:45-53.

85. Gouriéroux C, Holly A, Monfort A. Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in linear models with inequality constraints on the regression parameters. Econometrica 1982; 50:63-80.

86. Akaike H. Information Theory as an Extension of the Maximum Likelihood Principle. In Second International Symposium on Information. Budapest: Akademiai Kiado 1973; 267-81.

87. Schwarz G. Estimating the dimension of a model. Ann Stat 1978; 6:461-464.

88. Spiegelhalter DJ, Best NG, Carlin BP. Bayesian deviance, the effective number of parameters, and the comparison of arbitrarily complex models. Technical Report, MRC Biostatistics Unit, Cambridge, UK 1998. Available on http://yaroslavvb.com/papers/spiegelhalter-bayesian.pdf (Accessed July, 2014).

89. Dempster AP. The direct use of likelihood for significance testing. Stat Comput 1997; 7:247-252.

90. Berg A, Meyer R, Yu J. Deviance information criterion for comparing stochastic volatility models. Journal of Business & Economic Statistics 2004; 22:107-120.

91. Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A. Bayesian Measures of Model Complexity and Fit. Journal of the Royal Statistical Society Series B 2002;64(4):583-639.

92. Yuan Y, Johnson VE. Goodness-of-Fit Diagnostics for Bayesian Hierarchical Models. Biometrics 2012; 68:156-64.

93. Kass RE, Raftery AE. Bayes Factors. J Am Stat Assoc 1995; 90:773-795.

94. 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; 64:163-71.

95. Landoni G, Greco T, Biondi-Zoccai G, et al. Anaesthetic drugs and survival: a Bayesian network meta-analysis of randomized trials in cardiac surgery. Br J Anaesth 2013; 111:886-96.

96. Jakobsen CJ, Berg H, Hindsholm KB, Faddy N, Sloth E. The influence of propofol versus sevoflurane anesthesia on outcome in 10,535 cardiac surgical procedures. J Cardiothorac Vasc Anesth 2007; 21:664-71.

97. Landoni G, Biondi-Zoccai GG, Zangrillo A, et al. Desflurane and sevoflurane in cardiac surgery: A meta-analysis of randomized clinical trials. J Cardiothorac Vasc Anesth 2007; 21:502-11.

98. Bignami E, Biondi-Zoccai G, Landoni G, et al. Volatile anesthetics reduce mortality in cardiac surgery. J Cardiothorac Vasc Anesth 2009; 23:594-9.

99. Garcia C, Julier K, Bestmann L, et al. Preconditioning with sevoflurane decreases PECAM-1 expression and improves one-year cardiovascular outcome in coronary artery bypass graft surgery. Br J Anaesth 2005; 94:159-65

100. De Hert S, Vlasselaers D, Barbé R, et al. A comparison of volatile and non volatile agents for cardioprotection during on-pump coronary surgery. Anaesthesia 2009; 64:953-60.

101. Landoni G, Rodseth R, Santini F, et al. Randomized Evidence for Reduction in Perioperative Mortality. J Cardiothorac Vasc Anesth 2012;26:764-72.

102. Ades AE, Sculpher M, Sutton A, et al. Bayesian methods for evidence synthesis in cost-effectiveness analysis. Pharmacoeconomics 2006; 24:1-19.

103. Amr YM, Yassin IM. Cardiac protection during on-pump coronary artery bypass grafting: ischemic versus isoflurane preconditioning. Semin Cardiothorac Vasc Anesth 2010; 14:205-11.

104. Ballester M, Llorens J, Garcia-de-la-Asuncion J, et al. Myocardial oxidative stress protection by sevoflurane vs. propofol: a randomised controlled study in patients undergoing off-pump coronary artery bypass graft surgery. Eur J Anaesthesiol 2011; 28:874-81.

105. Bein B, Renner J, Caliebe D, et al. Sevoflurane but not propofol preserves myocardial function during minimally invasive direct coronary artery bypass surgery. Anesth Analg 2005; 100:610-6.

106. Belhomme D, Peynet J, Louzy M, Launay JM, Kitakaze M, Menasché P. Evidence for preconditioning by isoflurane in coronary artery bypass graft surgery. Circulation 1999; 100:II340-4.

107. Bignami E, Landoni G, Gerli C, et al. Sevoflurane vs. propofol in patients with coronary disease undergoing mitral surgery: a randomised study. Acta Anaesthesiol Scand 2012; 56:482-90.

108. Cavalca V, Colli S, Veglia F, et al. Anesthetic propofol enhances plasma gamma-tocopherol levels in patients undergoing cardiac surgery. Anesthesiology 2008; 108:988-97.

109. Conzen PF, Fischer S, Detter C, Peter K. Sevoflurane provides greater protection of the myocardium than propofol in patients undergoing off-pump coronary artery bypass surgery. Anesthesiology 2003; 99:826-33.

110. Cromheecke S, Pepermans V, Hendrickx E, Let al. Cardioprotective properties of sevoflurane in patients undergoing aortic valve replacement with cardiopulmonary bypass. Anesth Analg 2006; 10:289-96.

111. De Hert SG, Cromheecke S, ten Broecke PW, et al. Effects of propofol, desflurane, and sevoflurane on recovery of myocardial function after coronary surgery in elderly high-risk patients. Anesthesiology 2003; 99:314-23.

112. De Hert SG, Van der Linden PJ, Cromheecke S, et al. Choice of primary anesthetic regimen can influence intensive care unit length of stay after coronary surgery with cardiopulmonary bypass. Anesthesiology 2004; 101:9-20.

113. De Hert SG, Van der Linden PJ, Cromheecke S, et al. Cardioprotective properties of sevoflurane in patients undergoing coronary surgery with cardiopulmonary bypass are related to the modalities of its administration. Anesthesiology 2004; 101:299-310.

114. De Hert S, Vlasselaers D, Barbé R, et al. A comparison of volatile and non volatile agents for cardioprotection during on-pump coronary surgery. Anaesthesia 2009; 64:953-60.

115. Flier S, Post J, Concepcion AN, Kappen TH, Kalkman CJ, Buhre WF. Influence of propofol-opioid vs isoflurane-opioid anaesthesia on postoperative troponin release in patients undergoing coronary artery bypass grafting. Br J Anaesth 2010; 105:122-30.

116. Garcia C, Julier K, Bestmann L, Zollinger A, von Segesser LK, Pasch T, et al. Preconditioning with sevoflurane decreases PECAM-1 expression and improves one-year cardiovascular outcome in coronary artery bypass graft surgery. Br J Anaesth 2005; 94:159-65.

117. Goździk W, Adamik B, Gomulkiewicz A, Gozdzik A, Dziegiel P. Upregulation of HMBG1 and TLR -4 cardiac right atrium mRNA expression during CABG surgery – the effects of sevoflurane conditioning and postconditioning. Abstracts presented at WCA 2012. Br J Anaesth 2012; 108:ii37–ii44.

118. Guarracino F, Landoni G, Tritapepe L, et al. Myocardial damage prevented by volatile anesthetics: a multicenter randomized controlled study. J Cardiothorac Vasc Anesth 2006; 20:477-83.

119. Hellström J, Owall A, Sackey PV. Wake-up times following sedation with sevoflurane versus propofol after cardiac surgery. Scand Cardiovasc J 2012; 46:262-8.

120. Helman JD, Leung JM, Bellows WH, et alThe risk of myocardial ischemia in patients receiving desflurane versus sufentanil anesthesia for coronary artery bypass graft surgery. The S.P.I. Research Group. Anesthesiology 1992; 77:47-62.

121. Hemmerling T, Olivier JF, Le N, Prieto I, Bracco D. Myocardial protection by isoflurane vs. sevoflurane in ultra-fast-track anaesthesia for off-pump aortocoronary bypass grafting. Eur J Anaesthesiol 2008; 25:230-6.

122. Howie MB, Black HA, Romanelli VA, Zvara DA, Myerowitz PD, McSweeney TD. A comparison of isoflurane versus fentanyl as primary anesthetics for mitral valve surgery. Anesth Analg 1996; 83:941-8.

123. Huang Z, Zhong X, Irwin MG, Ji S, et al. Synergy of isoflurane preconditioning and propofol postconditioning reduces myocardial reperfusion injury in patients. Clin Sci (Lond) 2011; 121:57-69.

124. Kendall JB, Russell GN, Scawn ND, Akrofi M, Cowan CM, Fox MA. A prospective, randomised, single-blind pilot study to determine the effect of anaesthetic technique on troponin T release after off-pump coronary artery surgery. Anaesthesia 2004; 59:545-9.

125. Jovic M, Stancic A, Nenadic D, et al. Mitochondrial molecular basis of sevoflurane and propofol cardioprotection in patients undergoing aortic valve replacement with cardiopulmonary bypass. Cell Physiol Biochem 2012; 29:131-42.

126. Kottenberg E, Thielmann M, Bergmann L, et al. Protection by remote ischemic preconditioning during coronary artery bypass graft surgery with isoflurane but not propofol - a clinical trial. Acta Anaesthesiol Scand 2012; 56:30-8.

127. Landoni G, Calabrò MG, Marchetti C, et al. Desflurane versus propofol in patients undergoing mitral valve surgery. J Cardiothorac Vasc Anesth 2007; 21:672-7.

128. Lee MC, Chen CH, Kuo MC, Kang PL, Lo A, Liu K. Isoflurane preconditioning-induced cardio-protection in patients undergoing coronary artery bypass grafting. Eur J Anaesthesiol 2006; 23:841-7.

129. Leung JM, Goehner P, O'Kelly BF, et al Isoflurane anesthesia and myocardial ischemia: comparative risk versus sufentanil anesthesia in patients undergoing coronary artery bypass graft surgery. The SPI (Study of Perioperative Ischemia) Research Group. Anesthesiology 1991; 74:838-47.

130. Meco M, Cirri S, Gallazzi C, Magnani G, Cosseta D. Desflurane preconditioning in coronary artery bypass graft surgery: a double-blinded, randomised and placebo-controlled study. Eur J Cardiothorac Surg 2007; 32:319-25.

131. Musialowicz T, Niskanen M, Yppärilä-Wolters H, Pöyhönen M, Pitkänen O, Hynynen M. Auditory-evoked potentials in bispectral index-guided anaesthesia for cardiac surgery. Eur J Anaesthesiol 2007; 24:571-9.

132. Royse CF, Andrews DT, Newman SN, et al. The influence of propofol or desflurane on postoperative cognitive dysfunction in patients undergoing coronary artery bypass surgery. Anaesthesia 2011; 66:455-64.

133. Schoen J, Husemann L, Tiemeyer C, et al. Cognitive function after sevoflurane- vs propofol-based anaesthesia for on-pump cardiac surgery: a randomized controlled trial. Br J Anaesth 2011; 106:840-50.

134. Searle NR, Martineau RJ, Conzen P, et al. Comparison of sevoflurane/fentanyl and isoflurane/fentanyl during elective coronary artery bypass surgery. Sevoflurane Venture Group. Can J Anaesth 1996; 43:890-9.

135. Story DA, Poustie S, Liu G, McNicol PL. Changes in plasma creatinine concentration after cardiac anesthesia with isoflurane, propofol, or sevoflurane: a randomized clinical trial. Anesthesiology 2001; 95:842-8.

136. Tempe DK, Dutta D, Garg M, Minhas H, Tomar A, Virmani S. Myocardial protection with isoflurane during off-pump coronary artery bypass grafting: a randomized trial. J Cardiothorac Vasc Anesth 2011; 25:59-65.

137. Thomson IR, Bowering JB, Hudson RJ, Frais MA, Rosenbloom M. A comparison of desflurane and isoflurane in patients undergoing coronary artery surgery. Anesthesiology 1991; 75:776-8.

138. Tritapepe L, Giorni C, Di Giovanni C, Pompei, F. Cuscianna, E, Pietropaoli, P. Desflurane-sufentanil reduce troponin-I production after CABG. Eur J Anaesthesiol 2003; 20:A16.

139. Tritapepe L, Landoni G, Guarracino F, et al. Cardiac protection by volatile anaesthetics: a multicentre randomized controlled study in patients undergoing coronary artery bypass grafting with cardiopulmonary bypass. Eur J Anaesthesiol 2007; 24:323-31.

140. Yildirim V, Doganci S, Aydin A, Bolcal C, Demirkilic U, Cosar A. Cardioprotective effects of sevoflurane, isoflurane, and propofol in coronary surgery patients: a randomized controlled study. Heart Surg Forum 2009; 12:E1-9.

141. Greco T, Landoni G, Biondi-Zoccai G, D'Ascenzo F, Zangrillo A. A Bayesian network meta-analysis for binary outcome: how to do it. Stat Methods Med Res 2013. [Epub ahead of print]

142. Kalaian HA, Raudenbush SW. A Multivariate Mixed Linear Model for Meta-Analysis. Psychological Methods 1996; 1:227-235

143. Gleser LJ, Olkin I. Stochastically dependent effect sizes. In: H. Cooper & L. V. Hedges (Editor), The handbook of research synthesis. New York: Russell Sage Foundation 1994; 339-355.

144. Hox JJ. Applied multilevel analysis. Amsterdam: TT-publikaties, 1995.

145. Raudenbush SW. Hierarchical linear models and experimental design. Applied Analysis of Variance in Behavioral Science (Statistics: A Series of Textbooks and Monographs) Lynne Edwards (Editor) Chapman and Hall/CRC. 1993. Avaliable from http://www.unt.edu/rss/class/Jon/MiscDocs/Raudenbush_1993.pdf (Accessed July, 2014).

146. Goldstein H. Multilevel statistical models. London: Edward Arnold. New York, Halstead Press. 2003.

147. Turner RM, Omar RZ, Yang M, Goldstein H, Thompson SG. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. Statist Med 2000; 19:3417-3432.

148. Hox JJ. Multilevel analysis: Techniques and applications. Psychology Press, 2002.

149. LEMMA (Learning environment for multilevel methodology and applications). University of Bristol - Centre for Multilevel Modelling. 2013.

Available from http://www.bristol.ac.uk/cmm/learning/online-course/index.html (Accessed July, 2014).

150. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. Stat Med 2002; 21:589-624.

151. Liang KY, Zeger SI. Longitudinal data analysis using generalized linear models. Biometrika 1986; 73:13-22.

152. Lee Y, Nelder JA. Conditional and Marginal Models: Another View. Statistical Science. 2004; 19:219-238.

153. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. Biometrics 1988; 44:1049-1060.

154. Macaskill P, Walter SD, Irwig L. A comparison of methods to detect publication bias in meta-analysis. Stat Med 2001; 20:641-54.

155. Sterne JAC, Egger M. Regression methods to detect publication and other bias in meta-analysis. In: Rothstein, Hannah R., Alexander J. Sutton, and Michael Borenstein (Editors). Publication bias in meta-analysis: Prevention, assessment and adjustments John Wiley & Sons 2005; 99-110.

156. Sterne JAC, Becker BJ, Egger M. The funnel plot. In: Rothstein, Hannah R., Alexander J. Sutton, and Michael Borenstein (Editors). Publication bias in meta-analysis: Prevention, assessment and adjustments 2005; 75-98.

157. Piepho HP, Williams ER, Madden LV. The use of two-way linear mixed models in multitreatment meta-analysis. Biometrics 2012; 68:1269-77.

158. Schabenberger O. Introducing the GLIMMIX Procedure for Generalized Linear Mixed Models. Statistics and Data Analysis. Statistics and Data Analysis. SUGI 30. 2008; Paper 196-30. Available from: http://www2.sas.com/proceedings/sugi30/196-30.pdf (Accessed July, 2014).

159. Agresti A. Categorical Data Analysis. Wiley Series in Probability and Statistics Book Series. John Wiley & Sons, Inc 2002.

160. Ng ES, Carpenter JR, Goldstein H, Rasbash J. Estimation in generalised linear mixed models with binary outcomes by simulated maximum likelihood. Statistical Modelling 2006; 6:23-42.

161. Greco T, Landoni G, Saleh O, Zangrillo A. Case study in anesthesia and intensive care. In: Biondi-Zoccai G (Editor). Network Meta-Analysis: Evidence Synthesis with Mixed Treatment Comparison. Hauppauge, NY: Nova Science Publishers 2014; 263-281.

162. Rubin DB. Inference and missing data. Biometrika 1976; 63:581-592.

163. Little RJ. Modeling the dropout mechanism in repeated-measures studies. J Am Stat Assoc 1995; 90:1112-1121.

164. Little RJA, Rubin DB. Statistical Analysis with Missing Data, Second edition. New York: Wiley 2002.

165. Goldstein H, Yang M, Omar R, Turner R, Thompson S. Meta-analysis using multilevel models with an application to the study of class size effects. Journal of the Royal Statistical Society: Series C (Applied Statistics) 2000; 49:399-412.

166. Gage NA. Hierarchical Linear Modeling Meta-Analysis of Single-Subject Design Research. J Spec Educ 2012.

167. Zhang J, Lin L. Choosing the appropriate statistics. In: Biondi-Zoccai G (Editor). Network Meta-Analysis: Evidence Synthesis with Mixed Treatment Comparison. Hauppauge, NY: Nova Science Publishers 2014; 139-151.

168. Jones B, Roger J, Lane PW, Lawton A, Fletcher C, Cappelleri JC, et al. Statistical approaches for conducting network meta-analysis in drug development. Pharm Stat 2011; 10:523-31.

169. Greco T, Edefonti V, Biondi-Zoccai G, Decarli A, Gasparini M, Zangrillo A, Landoni G. A multilevel approach to network meta-analysis within a frequentist framework. Contemp Clin Trials 2015; 42:51-9.

170. Greco T, Biondi-Zoccai G, Saleh O, Pasin L, Cabrini L, Zangrillo A, Landoni G. The attractiveness of network meta-analysis: a comprehensive systematic and narrative review. Heart Lung Vessel 2015; 7(2):133-42.

171. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 2000; 10:325-337.

172. Wang H, Huang T, Jing J, Jin J, Wang P, Yang M, Cui W, Zheng Y, Shen H. Effectiveness of different central venous catheters for catheter-related infections: a network meta-analysis. J Hosp Infect 2010; 76(1):1-11.

173. Wu HY, Huang JW, Lin HJ, Liao WC, Peng YS, Hung KY, Wu KD, Tu YK, Chien KL. Comparative effectiveness of renin-angiotensin system blockers and other antihypertensive drugs in patients with diabetes: systematic review and bayesian network meta-analysis. BMJ 2013;347:f6008.

174. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B 57 1995; 289-300.

175. Baker WL, Baker EL, Coleman CI. Pharmacologic treatments for chronic obstructive pulmonary disease: a mixed-treatment comparison meta-analysis. Pharmacotherapy 2009; 29(8):891-905.

176. Chatterjee S, Biondi-Zoccai G, Abbate A, D'Ascenzo F, Castagno D, Van Tassell B, Mukherjee D, Lichstein E. Benefits of β blockers in patients with heart failure and reduced ejection fraction: network meta-analysis. BMJ 2013; 346:f55.

178. Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins JP, Churchill R, Watanabe N, Nakagawa A, Omori IM, McGuire H, Tansella M, Barbui C. Comparative efficacy and

acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. Lancet 2009; 373(9665):746-58.

179. Cipriani A, Barbui C, Salanti G, Rendell J, Brown R, Stockton S, Purgato M, Spineli LM, Goodwin GM, Geddes JR. Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple-treatments meta-analysis. Lancet 2011; 378(9799):1306-15.

180. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. Stat Med 2010; 29:932-44.

181. Dogliotti A, Paolasso E, Giugliano RP. Current and new oral antithrombotics in non-valvular atrial fibrillation: a network meta-analysis of 79 808 patients. Heart 2014; 100(5):396-405.

182. Dong YH, Lin HH, Shau WY, Wu YC, Chang CH, Lai MS. Comparative safety of inhaled medications in patients with chronic obstructive pulmonary disease: systematic review and mixed treatment comparison meta-analysis of randomised controlled trials. Thorax 2013; 68(1):48-56.

183. Dunkley AJ, Charles K, Gray LJ, Camosso-Stefinovic J, Davies MJ, Khunti K. Effectiveness of interventions for reducing diabetes and cardiovascular disease risk in people with metabolic syndrome: systematic review and mixed treatment comparison meta-analysis. Diabetes Obes Metab 2012; 14(7):616-25.

184. Elliott WJ, Meyer PM. Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis. Lancet 2007;369(9557):201-7.

185. Filippini G, Del Giovane C, Vacchi L, D'Amico R, Di Pietrantonj C, Beecher D, Salanti G. Immunomodulators and immunosuppressants for multiple sclerosis: a network meta-analysis. Cochrane Database Syst Rev 2013; 6:CD008933.

186. Fretheim A, Odgaard-Jensen J, Brørs O, Madsen S, Njølstad I, Norheim OF, Svilaas A, Kristiansen IS, Thürmer H, Flottorp S. Comparative effectiveness of antihypertensive medication for primary prevention of cardiovascular disease: systematic review and multiple treatments meta-analysis. BMC Med 2012; 10:33.

187. Lam SK, Owen A. Combined resynchronisation and implantable defibrillator therapy in left ventricular dysfunction: Bayesian network meta-analysis of randomised controlled trials. BMJ 2007; 335(7626):925.

188. Li LT, Hicks SC, Davila JA, Kao LS, Berger RL, Arita NA, Liang MK. Circular closure is associated with the lowest rate of surgical site infection following stoma reversal: a systematic review and multiple treatment meta-analysis. Colorectal Dis 2014; 16(6):406-16.
189. Owen A. Antithrombotic treatment for the primary prevention of stroke in patients with non valvular atrial fibrillation: a reappraisal of the evidence and network meta analysis. Int J Cardiol 2010; 142(3):218-23.
190. Psaty BM, Lumley T, Furberg CD, Schellenbaum G, Pahor M, Alderman MH, Weiss NS. Health outcomes associated with various antihypertensive therapies used as first-line agents: a network meta-analysis. JAMA 2003; 289(19):2534-44.
191. Ramsberg J, Asseburg C, Henriksson M. Effectiveness and cost-effectiveness of antidepressants in primary care: a multiple treatment comparison meta-analysis and cost-effectiveness model. PLoS One 2012; 7(8):e42003.
192. Reich K, Burden AD, Eaton JN, Hawkins NS. Efficacy of biologics in the treatment of moderate to severe psoriasis: a network meta-analysis of randomized controlled trials. Br J Dermatol 2012; 166(1):179-88.
193. Reichenpfader U, Gartlehner G, Morgan LC, Greenblatt A, Nussbaumer B, Hansen RA, Van Noord M, Lux L, Gaynes BN. Sexual dysfunction associated with second-generation antidepressants in patients with major depressive disorder: results from a systematic review with network meta-analysis. Drug Saf 2014; 37(1):19-31.
194. Ribeiro RA, Ziegelmann PK, Duncan BB, Stella SF, da Costa Vieira JL, Restelatto LM, Moriguchi EH, Polanczyk CA. Impact of statin dose on major cardiovascular events: a mixed treatment comparison meta-analysis involving more than 175,000 patients. Int J Cardiol 2013; 166(2):431-9.
195. Sciarretta S, Palano F, Tocci G, Baldini R, Volpe M. Antihypertensive treatment and development of heart failure in hypertension: a Bayesian network meta-analysis of studies in patients with hypertension and high cardiovascular risk. Arch Intern Med 2011; 171(5):384-94.
196. Tadrous M, Wong L, Mamdani MM, Juurlink DN, Krahn MD, Lévesque LE, Cadarette SM. Comparative gastrointestinal safety of bisphosphonates in primary osteoporosis: a network meta-analysis. Osteoporos Int 2014; 25(4):1225-35.

197. Thijs V, Lemmens R, Fieuws S. Network meta-analysis: simultaneous meta-analysis of common antiplatelet regimens after transient ischaemic attack or stroke. Eur Heart J 2008; 29(9):1086-92.

198. van Valkenhoef G, Tervonen T, Zhao J, de Brock B, Hillege HL, Postmus D. Multicriteria benefit-risk assessment using network meta-analysis. J Clin Epidemiol 2012; 65(4):394-403.

199. Yang B, Shi J, Chen X, Ma B, Sun H. Efficacy and safety of therapies for acute ischemic stroke in China: a network meta-analysis of 13289 patients from 145 randomized controlled trials. PLoS One 2014; 9(2):e88440.

200. Chaimani A, Salanti G. Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. Res Synth Method 2012;3:161-176.

201. Bayarri MJ, Berger JO. The Interplay of Bayesian and Frequentist Analysis. Statistical Science 2004, 19(1):58-80.

202. Hong H, Carlin BP, Shamliyan TA, Wyman JF, Ramakrishnan R, Sainfort F, Kane RL. Comparing Bayesian and frequentist approaches for multiple outcome mixed treatment comparisons. Med Decis Making 2013; 33(5):702-14.

Appendix 1

Data frame: Data structure of the 38 randomized controlled trials included in the preciously published Bayesian network meta-analysis [95].

Study	Group	Author	Year	Design	Treatment	N	M	Follow_up	Sample Size
1	1	Amr YM	2010	Isoflurane vs TIVA	Isoflurane	15	1	Hospital stay	30
1	2	Amr YM	2010	Isoflurane vs TIVA	TIVA	15	1	Hospital stay	30
2	3	Ballester M	2011	Sevoflurane vs TIVA	Sevoflurane	21	1	1 year	40
2	4	Ballester M	2011	Sevoflurane vs TIVA	TIVA	19	0	1 year	40
3	5	Bein B	2005	Sevoflurane vs TIVA	Sevoflurane	26	0	Hospital stay	52
3	6	Bein B	2005	Sevoflurane vs TIVA	TIVA	26	0	Hospital stay	52
4	7	Belhomme D	1999	Isoflurane vs TIVA	Isoflurane	10	0	3 days	20
4	8	Belhomme D	1999	Isoflurane vs TIVA	TIVA	10	0	3 days	20
5	9	Bignami E	2011	Sevoflurane vs TIVA	Sevoflurane	50	1	1 year	100
5	10	Bignami E	2011	Sevoflurane vs TIVA	TIVA	50	2	1 year	100
6	11	Cavalca V	2008	Sevoflurane vs TIVA	Sevoflurane	22	0	24 hours	44
6	12	Cavalca V	2008	Sevoflurane vs TIVA	TIVA	22	0	24 hours	44
7	13	Conzen PF	2003	Sevoflurane vs TIVA	Sevoflurane	12	0	Hospital stay	23
7	14	Conzen PF	2003	Sevoflurane vs TIVA	TIVA	11	0	Hospital stay	23
8	15	Cromheecke S	2006	Sevoflurane vs TIVA	Sevoflurane	15	0	Hospital stay	30
8	16	Cromheecke S	2006	Sevoflurane vs TIVA	TIVA	15	0	Hospital stay	30
9	17	De Hert SG (1)	2003	Sevoflurane vs desflurane vs TIVA	Sevoflurane	15	0	36 hours	45
9	18	De Hert SG (1)	2003	Sevoflurane vs desflurane vs TIVA	Desflurane	15	0	36 hours	45
9	19	De Hert SG (1)	2003	Sevoflurane vs desflurane vs TIVA	TIVA	15	1	36 hours	45
10	20	De Hert SG (2)	2004	Sevoflurane vs desflurane vs TIVA	Sevoflurane	80	0	Hospital stay	320
10	21	De Hert SG (2)	2004	Sevoflurane vs desflurane vs TIVA	Desflurane	80	0	Hospital stay	320
10	22	De Hert SG (2)	2004	Sevoflurane vs desflurane vs TIVA	TIVA - TIVA	160	2	Hospital stay	320
11	23	DE Hert SG (3)	2004	Sevoflurane vs TIVA	Sevoflurane	150	0	30 days	200
11	24	DE Hert SG (3)	2004	Sevoflurane vs TIVA	TIVA	50	0	30 days	200
12	25	De Hert SG (4)	2009	Sevoflurane vs desflurane vs TIVA	Sevoflurane	132	4	1 year	414

Study	Group	Author	Year	Design	Treatment	N	M	Follow_up	Sample Size
12	26	De Hert SG (4)	2009	Sevoflurane vs desflurane vs TIVA	Desflurane	137	9	1 year	414
12	27	De Hert SG (4)	2009	Sevoflurane vs desflurane vs TIVA	TIVA	145	18	1 year	414
13	28	Flier S	2010	Isoflurane vs TIVA	Isoflurane	51	0	1 year	100
13	29	Flier S	2010	Isoflurane vs TIVA	TIVA	49	2	1 year	100
14	30	Garcia C	2005	Sevoflurane vs TIVA	Sevoflurane	37	0	1 year	72
14	31	Garcia C	2005	Sevoflurane vs TIVA	TIVA	35	0	1 year	72
15	32	Goździk W	2012	Sevoflurane vs TIVA	Sevoflurane	40	0	24 hours	60
15	33	Goździk W	2012	Sevoflurane vs TIVA	TIVA	20	0	24 hours	60
16	34	Guarracino F	2006	Desflurane vs TIVA	Desflurane	57	0	30 days	112
16	35	Guarracino F	2006	Desflurane vs TIVA	TIVA	55	1	30 days	112
17	36	Hellström J	2012	Sevoflurane vs TIVA	Sevoflurane	50	1	30 days	100
17	37	Hellström J	2012	Sevoflurane vs TIVA	TIVA	50	0	30 days	100
18	38	Helman JD	1992	Desflurane vs TIVA	Desflurane	100	1	3 postoperative days	200
18	39	Helman JD	1992	Desflurane vs TIVA	TIVA	100	3	3 postoperative days	200
19	40	Hemmerling T	2008	Sevoflurane vs isoflurane	Sevoflurane	20	0	Hospital stay	40
19	41	Hemmerling T	2008	Sevoflurane vs isoflurane	Isoflurane	20	0	Hospital stay	40
20	42	Howie MB	1996	Isoflurane vs TIVA	Isoflurane	27	0	4 hours after surgical ICU	50
20	43	Howie MB	1996	Isoflurane vs TIVA	TIVA	23	0	4 hours after surgical ICU	50
21	44	Huang Z	2011	Isoflurane vs TIVA	Isoflurane	60	0	Hospital stay	120
21	45	Huang Z	2011	Isoflurane vs TIVA	TIVA - TIVA	60	0	Hospital stay	120
22	46	Jovic M	2012	Sevoflurane vs TIVA	Sevoflurane	11	0	Hospital stay	22
22	47	Jovic M	2012	Sevoflurane vs TIVA	TIVA	11	0	Hospital stay	22
23	48	Kendal JB	2004	Isoflurane vs TIVA	Isoflurane	10	0	48 hours	20
23	49	Kendal JB	2004	Isoflurane vs TIVA	TIVA	10	0	48 hours	20
24	50	Kottenber E	2012	Isoflurane vs TIVA	Isoflurane	19	0	72 hours	38
24	51	Kottenber E	2012	Isoflurane vs TIVA	TIVA	19	0	72 hours	38
25	52	Landoni G	2007	Desflurane vs TIVA	Desflurane	59	0	30 days	120
25	53	Landoni G	2007	Desflurane vs TIVA	TIVA	61	2	30 days	120
26	54	Lee MC	2006	Isoflurane vs TIVA	Isoflurane	20	1	Hospital stay	40

Study	Group	Author	Year	Design	Treatment	N	M	Follow_up	Sample Size
26	55	Lee MC	2006	Isoflurane vs TIVA	TIVA	20	1	Hospital stay	40
27	56	Leung JM	1991	Isoflurane vs TIVA	Isoflurane	62	1	Surgical time	186
27	57	Leung JM	1991	Isoflurane vs TIVA	TIVA	124	3	Surgical time	186
28	58	Meco M	2007	Desflurane vs TIVA	Desflurane	14	0	72 hours	28
28	59	Meco M	2007	Desflurane vs TIVA	TIVA	14	0	72 hours	28
29	60	Musialowicz T	2007	Isoflurane vs TIVA	Isoflurane	12	0	Surgical time	24
29	61	Musialowicz T	2007	Isoflurane vs TIVA	TIVA	12	0	Surgical time	24
30	62	Royse CF	2011	Desflurane vs TIVA	Desflurane	91	0	1 year	182
30	63	Royse CF	2011	Desflurane vs TIVA	TIVA	91	0	1 year	182
31	64	Schoen J	2011	Sevoflurane vs TIVA	Sevoflurane	64	2	Hospital stay	128
31	65	Schoen J	2011	Sevoflurane vs TIVA	TIVA	64	0	Hospital stay	128
32	66	Searle NR	1996	Sevoflurane vs isoflurane	Sevoflurane	140	1	Hospital stay	273
32	67	Searle NR	1996	Sevoflurane vs isoflurane	Isoflurane	133	4	Hospital stay	273
33	68	Story DA	2001	Sevoflurane vs isoflurane vs TIVA	Sevoflurane	120	1	Hospital stay	360
33	69	Story DA	2001	Sevoflurane vs isoflurane vs TIVA	Isoflurane	120	0	Hospital stay	360
33	70	Story DA	2001	Sevoflurane vs isoflurane vs TIVA	TIVA	120	2	Hospital stay	360
34	71	Tempe DK	2011	Isoflurane vs TIVA	Isoflurane	23	0	72 hours	45
34	72	Tempe DK	2011	Isoflurane vs TIVA	TIVA	22	1	72 hours	45
35	73	Thomson IR	1991	Deslurane vs isoflurane	Desflurane	21	2	Hospital stay	41
35	74	Thomson IR	1991	Deslurane vs isoflurane	Isoflurane	20	1	Hospital stay	41
36	75	Tritapepe L (1)	2003	Desflurane vs TIVA	Desflurane	52	1	30 days	107
36	76	Tritapepe L (1)	2003	Desflurane vs TIVA	TIVA	55	3	30 days	107
37	77	Tritapepe L (2)	2007	Desflurane vs TIVA	Desflurane	75	1	ICU stay	150
37	78	Tritapepe L (2)	2007	Desflurane vs TIVA	TIVA	75	1	ICU stay	150
38	79	Yildirim V	2009	Sevoflurane vs Isoflurane vs TIVA	Sevoflurane	20	0	30 days	60
38	80	Yildirim V	2009	Sevoflurane vs Isoflurane vs TIVA	Isoflurane	20	0	30 days	60
38	81	Yildirim V	2009	Sevoflurane vs Isoflurane vs TIVA	TIVA	20	0	30 days	60

M: number of deaths; N: number of total patients per arm; Sample size: number of total patients per study; TIVA: Total intravenous anaesthesia

Appendix 2

Table A1: Effect estimates, corresponding 95% confidence intervals, and unadjusted p-values derived from the univariate analysis* to assess the presence of a difference between the Bayesian and the frequentist multilevel approaches to network meta-analysis (NMA).

Model ID	Number of NMAs included	Number of pooled estimates analyzed	Effects derived from a fixed/random model	Model**	QIC	Parameter	Estimate	Standard Error	95% confidence interval	Unadjusted p-values
A.1	27	844	Fixed	Log(OR) = β_0 + approach	1646	Intercept	0.075	0.116	-0.152 to 0301	0.518
						approach***	-0.660	0.413	-1.470 to 0.150	0.110
A.2	27	844	Random	Log(OR) = β_0 + approach	1627	Intercept	-0.024	0.086	-0.193 to 0.145	0.782
						approach***	0.038	0.046	-0.052 to 0.128	0.412
B.1	20	330	Fixed	Log(OR) = β_0 + approach	679	Intercept	0.035	0.113	-0.186 to 0.255	0.759
						approach***	-0.145	0.087	-0.316 to 0.025	0.095
B.2	20	330	Random	Log(OR) = β_0 + approach	680	Intercept	0.003	0.115	-0.222 to 0.229	0.976
						approach***	-0.005	0.017	-0.039 to 0.029	0.763
C.1	27	844	Fixed	Standardized log (OR) = β_0 + approach	1624	Intercept	0.041	0.285	-0.519 to 0.601	0.886
						approach***	-30.265	29.550	-88.183 to 27.652	0.306
C.2	27	844	Random	Standardized log (OR) = β_0 + approach	1634	Intercept	0.121	0.358	-0.581 to 0.823	0.736
						approach***	0.110	0.122	-0.130 to 0.349	0.369
D.1	20	330	Fixed	Standardized log (OR) = β_0 + approach	662	Intercept	-3.561	3.487	-10.395 to 3.273	0.307
						approach***	-77.288	70.083	-214.648 to 60.072	0.270
D.2	20	330	Random	Standardized log (OR) = β_0 + approach	678	Intercept	0.239	0.504	-0.748 to 1.226	0.635
						approach***	0.070	0.156	-0.235 to 0.374	0.655
E.1	27	844	Fixed	Euclidean distance = β_0	779	Intercept	36.133	32.001	-26.590 to 98.856	0.259
E.2	27	844	Random	Euclidean distance = β_0	761	Intercept	0.771	0.122	0.532 to 1.009	**<0.001**
F.1	20	330	Fixed	Euclidean distance = β_0	332	Intercept	91.173	77.177	-60.091 to 242.437	0.238
F.2	20	330	Random	Euclidean distance = β_0	337	Intercept	0.724	0.147	0.436 to 1.013	**<0.001**

* This analysis included the main "approach" effect only.

** Generalized Estimating Equations models taking into account the within-NMA correlation

*** Variable "approach" is equal to 0 for the Bayesian approach and to 1 for the frequentist-multilevel one.

Bold style typeface: significant p-value. Statistical significance was set at the two-tailed 0.05 level. Adjustment for multiple comparison was provided by the Benjamini-Hochberg method. Standardized log (OR)=log(OR)/SE[log(OR)]; Euclidean distance= $\sqrt{[(x_0-x_1)^2]}$, where x_0= $\log(OR)_{Bayesian}$ and x_1= $\log(OR)_{multilevel}$

QIC: Quasi-likelihood under the independence model criterion.

Table A2: Unadjusted p-values from tests of significance of potential covariates of interest, as derived from the multiple analysis* to assess the presence of a difference between the Bayesian and the frequentist multilevel approaches to network meta-analysis (NMA).

		Model **											
Parameter	**Level**	**A.1**	**A.2**	**B.1**	**B.2**	**C.1**	**C.2**	**D.1**	**D.2**	**E.1**	**E.2**	**F.1**	**F.2**
n_treatment_nma	NMA	0.309	0.959	0.835	0.613	0.302	0.679	0.206	0.819	0.278	0.349	0.194	0.300
n_studies_nma	NMA	0.739	0.185	**0.0037**	**0.001**	0.289	0.186	0.319	**0.003**	0.275	0.575	0.316	0.183
n_multiarm_nma	NMA	0.415	0.064	0.166	0.131	0.511	0.251	0.476	0.149	0.452	**0.007**	0.483	0.553
n_pairwise_nma	NMA	0.728	0.527	0.6318	0.680	0.288	0.678	0.653	0.874	0.284	0.338	0.574	0.724
min_n_studies_nma	NMA	0.050	0.231	0.241	0.238	0.300	0.379	0.260	0.4095	0.300	**0.008**	0.241	**0.031**
max_n_studies_nma	NMA	0.592	0.259	0.0839	0.044	0.334	0.206	0.312	0.102	0.330	0.096	0.312	0.152
n_events_nma	NMA	0.057	0.194	0.425	0.207	0.234	0.676	0.086	0.738	0.209	0.784	0.095	0.660
n_patients_nma	NMA	0.421	0.918	0.7808	0.964	0.245	0.335	0.246	0.339	0.241	0.281	0.255	0.337
perc_nma	NMA	**0.004**	**0.001**	**<0.001**	**<.0001**	0.271	**0.015**	0.240	**<0.001**	0.280	0.207	0.269	0.268
n_effect_estimated	NMA	0.429	0.878	0.868	0.622	0.294	0.607	0.192	0.901	0.267	0.541	0.184	0.225
n_arms_arm	Arm	**0.011**	0.223	0.108	0.141	0.314	0.199	0.325	0.152	0.321	0.232	0.330	0.622
n_events_arm	Arm	0.091	0.664	0.3929	0.942	0.296	0.839	0.298	0.962	0.294	0.462	0.286	0.243
n_patients_arm	Arm	0.086	0.414	0.2479	0.211	0.683	0.654	0.371	0.231	0.701	0.238	0.372	0.611
perc_arm	Arm	0.369	0.975	**0.033**	**0.024**	0.281	0.459	0.252	**0.005**	0.284	0.330	0.268	0.111

* This multiple analysis included one extra covariate of interest, together wirh the “approach” effect.

** Generalized Estimating Equations models taking into account the within-NMA correlation.

Bold style typeface: significant p-value. Statistical significance was set at the two-tailed 0.05 level.

NMA: network meta-analysis; **n_treatment_nma**: number of treatments included in the nma; **n_studies_nma**: number of studies included in the NMA; **n_multiarm_nma**: number of multi-arm studies included in the NMA; **n_pairwise_nma**: number of pairwise comparisons taken into account in the NMA; **min_n_studies_nma**: minimum number of studies included in each comparisons; **max_n_studies_nma**: maximum number of studies included in each comparison; **n_events_nma**: total number of events analyzed in the NMA; **n_patients_nma**: total number of patients analyzed in the NMA; **perc_nma**: percentage of events in the NMA; **n_effect_estimated**: number of effects estimated in the NMA; **n_arms_arm**: number of arms token into account for each comparisons; **n_events_arm**: total number of events analyzed in each arm; **n_patients_arm**: total number of patients analyzed in each arm; **perc_arm**: percentage of events in each arm.

Table A3: Unadjusted p -values derived from univariate analyses to evaluate the predictors of convergence problems using Bayesian or multivariate approach.

Parameter	Problem with winBUGS	Problem with SAS
n_treatment_nma	0.403	0.251
n_studies_nma	0.078	0.555
n_multiarm_nma	0.135	0.471
n_pairwise_nma	**0.006**	0.611
min_n_studies_n	0.948	0.957
max_n_studies_n	0.945	0.241
n_events_nma	0.698	0.260
n_patients_nma	0.199	0.540
perc_nma	0.204	0.400
n_effect_estima	0.420	0.386

Bold style typeface: significant p-value. Statistical significance was set at the two-tailed 0.05 level.
nma: network meta-analysis; **n_treatment_nma**: number of treatments included in the nma; **n_studies_nma**: number of studies included in the nma; **n_multiarm_nma**: number of multi-arm studies included in the nma; **n_pairwise_nma**: number of pairwise comparisons taken into account in the nma; **min_n_studies_nma**: minimum number of studies included in each comparisons; **max_n_studies_nma**: maximum number of studies included in each comparisons; **n_events_nma**: total number of events analyzed in the nma; **n_patients_nma**: total number of patients analyzed in the nma; **perc_nma**: percentage of events in the nma; **n_effect_estimated**: number of effects estimated in the nma.