USING VARIABLE SELECTION METHODS IN NETWORK META-ANALYSIS

University of Ioannina
School of Education
Department of Primary Education

A Thesis
by
GEORGIOS S. SEITIDIS

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

Supervisor

Dimitris Mavridis   Department of Primary Education, University of Ioannina

Thesis committee

Ioannis Ntzoufras   Department of Statistics, Athens University of Economics & Business

Anna Chaimani   Centre of Research in Epidemiology and Statistics (CRESS-U1153), Université Paris Cité and Université Sorbonne Paris Nord

External examination committee

Stavros Nikolakopoulos   Department of Psychology, University of Ioannina
Department of Biostatistics, University Medical Center Utrecht

Elena Dragioti   Department of Nursing, University of Ioannina
Department of Medical and Health Sciences, Linköping University

Konstantinos Tsilidis   Department of Hygiene and Epidemiology, University of Ioannina
Department of Epidemiology and Biostatistics, Imperial College London

Anna-Bettina Haidich   Department of Hygiene, Aristotle University of Thessaloniki

ACKNOWLEDGMENTS

I want to begin by expressing my deepest appreciation and gratitude to my advisor, Prof. Dimitris Mavridis, for his unwavering commitment and guidance throughout the entire duration of my PhD. His tireless efforts, constant encouragement, and invaluable input have been instrumental in bringing this thesis to fruition. Without his support, this project would have never materialised, and I am truly grateful for his dedication.

In addition to my advisor, I would also like to extend my heartfelt thanks to my co-advisors, Prof. Ioannis Ntzoufras and Dr. Anna Chaimani, for their invaluable support and feedback over the past four years. Their guidance has helped me to become a better researcher, and I am honored to have had the opportunity to work with them. Dr. Chaimani has generously allowed me to visit her at her university and collaborate with her on my thesis. During my time there as an Erasmus$^+$ student, I had the opportunity to collaborate with a multitude of researchers hailing from diverse research fields. Through this experience, I engaged in several research projects that were both stimulating and enriching. Also, throughout these years, Prof. Ntzoufras has been a constant source of support for me, always willing to answer any questions and address any concerns I had.

I would also like to express my gratitude to the members of the external examination committee: Prof. Nikolakopoulos, Prof. Dragioti, Prof. Tsilidis and Prof. Haidich. Their thorough review and insightful suggestions have greatly improved the quality of my work, and I am grateful for their contributions. Special recognition to Prof. Nikolakopoulos, who supported me by providing invaluable help, guidance and advice whenever I needed it.

Lastly, I want to thank my parents, Stylianos and Spiridoula, my brother, Stavros, and my wife Dimitra for their unending love and support. Their encouragement and assistance have been instrumental in helping me achieve my goals, and I am truly grateful for everything they have done for me.

DEDICATION

To Georgios G. Kaponis

ABSTRACT

The current thesis aims to address two common challenges encountered in network meta-analysis (NMA), the evaluation of consistency and the handling of multi-component interventions. Consistency assessment is vital as the validity of the NMA findings is primary affected by the satisfaction of this assumption. Significant discrepancies between direct and indirect evidence may lead to biased NMA estimates. In particular, a brief overview of (N)MA and the standard method used to evaluate consistency is presented. Additionally, a novel method for identifying inconsistencies is proposed that evaluates network consistency both globally and locally. This was accomplished by integrating the Stochastic Search Variable Selection method into the NMA framework and treating inconsistency factors as variables in a generalized linear model. Historical evidence and differences between direct and indirect evidence that are considered of practical significant, can be also incorporated into the inconsistency detection process. The performance of the proposed method was evaluated in two published networks under various scenarios. Furthermore, to facilitate the implementation of the method, the user-friendly R-package *ssifs* was developed and published on CRAN. Concerning the handling of multi-component interventions, novel ways of visualizing NMA results are proposed that allow for the easy identification of the most efficacious components, exploration of the network's geometry and examination of the components' behaviour. To implement these tools, the user-friendly R-package *viscomp* was developed and published on CRAN. Lastly, some concluding remarks and suggestions for future research are provided.

ΠΕΡΙΛΗΨΗ

Η συγκεκριμένη διατριβή στοχεύει στην αντιμετώπιση δύο συχνών προβλημάτων που συναντώνται στην μετα-ανάλυση δικτύου, την αξιολόγηση της υπόθεσης της συνέπειας και τον χειρισμό παρεμβάσεων πολλών συνιστωσών. Η αξιολόγηση της υπόθεσης της συνέπειας είναι κομβικής σημασίας, καθώς η αξιοπιστία των αποτελεσμάτων της μετα-ανάλυσης δικτύου εξαρτάται σε μεγάλο βαθμό από αυτή την υπόθεση. Μεγάλες διαφορές μεταξύ άμεσης και έμμεσης πληροφορίας μπορεί να οδηγήσει σε μεροληπτικά αποτελέσματα. Στην διατριβή αρχικά γίνεται μια ανασκόπηση της μετα-ανάλυσης και της μετα-ανάλυσης δικτύου, καθώς και των συχνών μεθόδων που χρησιμοποιούνται για την αξιολόγηση της υπόθεσης της συνέπειας. Έπειτα παρουσιάζεται μια καινοτόμα μέθοδος για την εύρεση ασυνεπειών στο δίκτυο, η οποία έχει την δυνατότητα να ανιχνεύει τοπικά που υπάρχουν ασυνέπειες και να αξιολογεί γενικά το δίκτυο αν πληρεί την υπόθεση της συνέπειας. Αυτό επιτυγχάνεται αντιμετωπίζοντας το μοντέλο της μετα-ανάλυσης δικτύου ως ένα μοντέλο παλινδρόμησης και χρησιμοποιώντας την μέθοδο στοχαστικής αναζήτησης επιλογής μεταβλητών. Σημαντικά πλεονεκτήματα που προσφέρει η προτεινόμενη μέθοδος είναι η δυνατότητα εύρεσης ασυνεπειών χρησιμοποιώντας εξωτερική πληροφορία, και η εύρεση ασυνεπειών βάσει διαφορών που θεωρούνται πρακτικά σημαντικές. Η προτεινόμενη μέθοδος αξιολογήθηκε σε δύο δημοσιευμένα δίκτυα μετα-ανάλυσης υπό διάφορα σενάρια. Επιπρόσθετα, για την εύκολη χρήση της μεθόδου, κατασκευάστηκε το στατιστικό πακέτο *ssifs*, το οποίο είναι δημοσιευμένο στο CRAN. Αναφορικά με τον χειρισμό παρεμβάσεων πολλών συνιστωσών, προτείνονται νέοι τρόποι οπτικοποίησης των αποτελεσμάτων της μετα-ανάλυσης δικτύου, οι οποίοι επιτρέπουν την εύκολη εύρεση των πιο αποτελεσματικών συνιστωσών, την διερεύνηση του δικτύου και της συμπεριφοράς των συνιστωσών. Για την εύκολη χρήση των προτεινόμενων εργαλείων οπτικοποίησης των αποτελεσμάτων της μετα-ανάλυσης δικτύου, κατασκευάστηκε το στατιστικό πακέτο *viscomp*, το οποίο είναι δημοσιευμένο στο CRAN. Τέλος παρατίθενται ορισμένες καταληκτικές παρατηρήσεις και προτάσεις για μελλοντική έρευνα.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Network Meta-Analysis

In systematic reviews, meta-analyses are typically conducted when only two interventions for a given indication are compared in the included studies (Nikolakopoulou et al., 2014b). For example, if the study aims at comparing the efficacy of two paracetamol drugs for treating head-pain symptoms, then meta-analysis should be conducted using the studies that 1) compare these drugs head-to-head and 2) include the pain level as an outcome. Meta-analysis generally provides an improved level of precision for the summary effect estimates. Also, it may provide insight into issues that cannot be addressed by single trials, since heterogeneity can provide valuable insight into the population and conditions under which an intervention is effective. However, it is not uncommon for a given indication to have several interventions available. In this case, pair-wise meta-analysis cannot provide adequate answers to crucial research questions, such as which interventions are the "best", under which conditions, and which interventions have an optimal balance of benefits and harms. If for example, the research question concerns the comparative effectiveness of the various paracetamol drugs in reducing pain symptoms, pair-wise meta-analysis cannot be used to determine which drug is the more effective.

The extension of pair-wise meta-analysis to multiple comparisons is termed Network Meta-Analysis (NMA) (Seitidis et al., 2022a; Mavridis et al., 2015; Salanti et al., 2008; Lu and Ades, 2004; Caldwell et al., 2005). It was firstly introduced by Lumley (2002) and it is also referred in the literature as multiple-treatment meta-analysis or mixed-treatment comparison (Druyts et al., 2013; Lu and Ades, 2006; Mills et al., 2011; Salanti, 2012). The rationale behind NMA is that we synthesize jointly both sources of evidence, direct and indirect. Suppose that we have a set of studies comparing treatments A versus B and A versus C. It should be noted that that we do not have direct evidence for the comparison B versus C since there are no studies comparing these two treatments (Figure 1.1a). However, the comparisons AB and AC use treatment A as a common comparator. Hence, by subtracting

the difference between AB and AC, we can indirectly infer about the comparison BC.

(a) Studies that compare B versus C are absent. Only indirect evidence for the comparison BC is available.

(b) Studies that compare B versus C are present. Both direct and indirect evidence are available for the comparison BC.
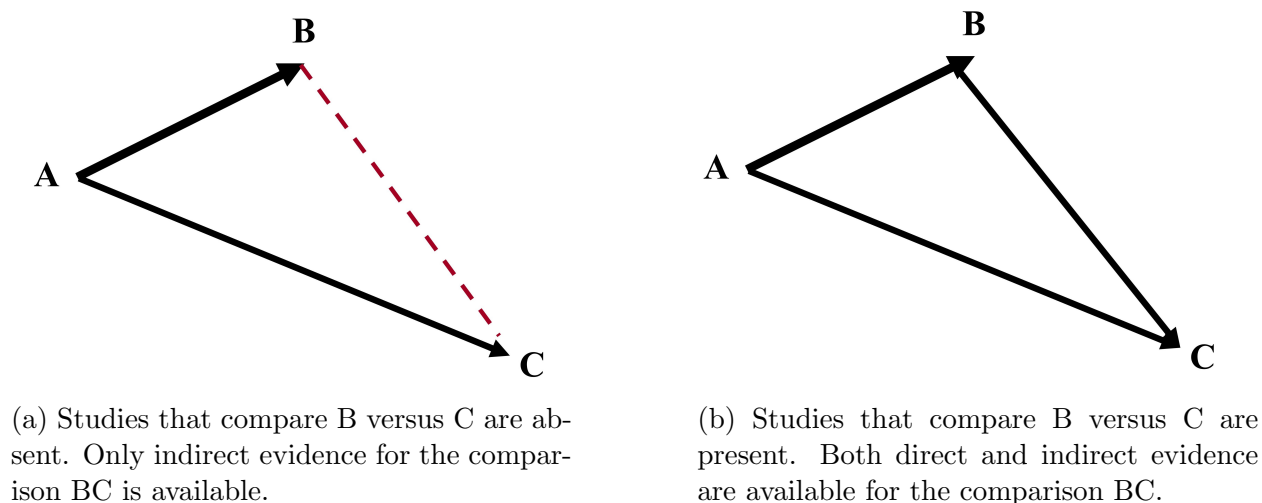
Black solid lines denote the direct evidence, while red dashed line denotes the indirect evidence.

Figure 1.1: Example of a ABC network.

Due to its efficiency, NMA has been used extensively for estimating the relative efficacy of multiple interventions (Nikolakopoulou et al., 2014a). A systematic review showed an increasing rate of published NMAs, with 456 NMAs of RCTs including at least four different interventions published between 1999 and 2015 (Petropoulou et al., 2017). Moreover, international organizations like the World Health Organization (WHO), the National Institute for Health and Care Excellence (NICE), the National Health Service (NHS) in the United Kingdom, have established NMA as a powerful tool for evidence synthesis in order to provide useful insight for decision-making (Salanti, 2012). Advantages of NMA include more precise estimates since all the relevant evidence, direct and indirect, is harvested jointly. As a consequence, estimates of comparative efficacy between interventions that have never been compared head-to-head can be obtained. Another important advantage is that it can provide information with respect to the relative ranking of interventions for a specific outcome (e.g., efficacy or safety). To emphasize further the benefits of NMA compared to pairwise meta-analysis, an empirical study showed that network estimates are more precise than their pairwise counterparts and in 20% of the cases (10 out of 49) the network estimates showed statistical significance that we wouldn't have attained based on simple meta-analysis alone

(Nikolakopoulou et al., 2018).

The implementation of NMA can be performed using a variety of models in a frequentist or a Bayesian setting. These include treating the NMA model as a weighted meta-regression model (Lumley, 2002), a hierarchical model (Lu and Ades, 2004), a multivariate meta-analysis (White, 2015), or an application of graph theory to electrical networks (Rücker, 2012), among other approaches.

1.1   Presentation of Network Meta-Analysis Results

The first step for the presentation of the NMA results is to visualize the geometry of the network. The network plot is used to display how treatments are "connected" with each other. Interventions/treatments are denoted by nodes, where the size of the node is usually proportional to the number of participants randomized in the corresponding node. Nodes are connected with an edge, if there is at least one study that compares them directly. The thickness of the edge could be used to demonstrate the number of studies that compares the corresponding nodes or to reflect the standard error in the corresponding comparison. A real example of a network plot is presented in Figure 1.2, which compares the relative effects of no contact (reference), self-help, individual counseling and group counseling on smoking cessation, reported by the Agency for Health Care Policy and Research smoking cessation guideline panel (Lu and Ades, 2006). The network plot indicates that most of the studies evaluates the effectiveness of individual counseling compared to no contact, and most participants have been also randomized in these two nodes.
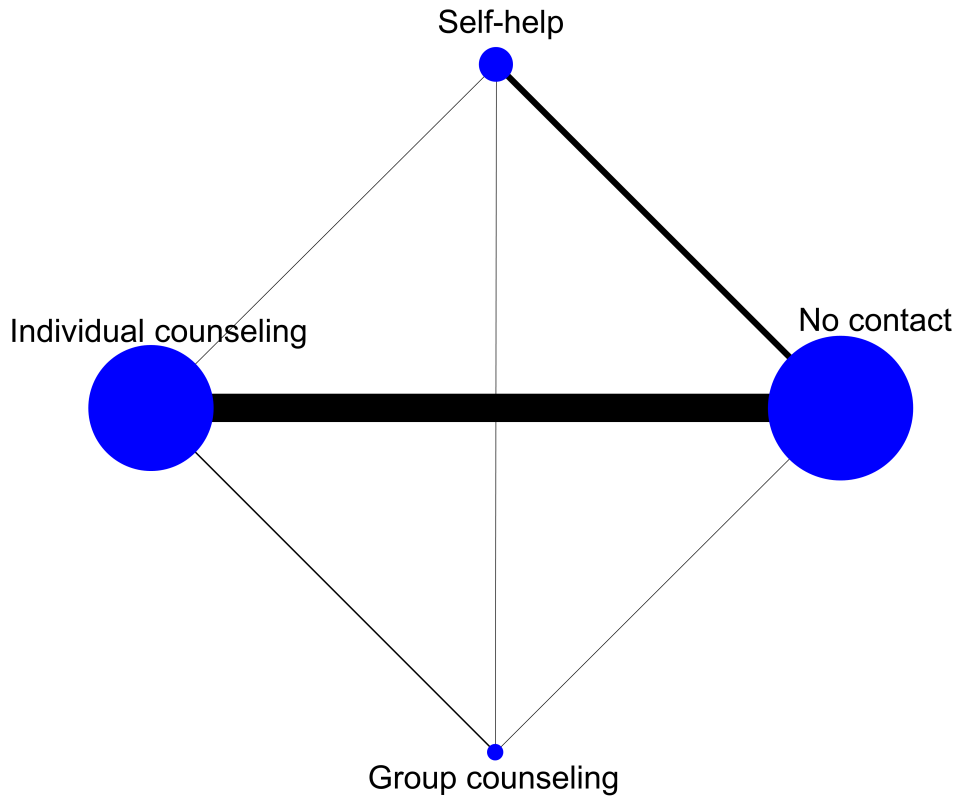
Edge's thickness and node's size is analogous to the number of direct treatment comparisons and participants, respectively.

Figure 1.2: Network plot for the smoking cessation example.

NMA results are typically presented through the forest plot. Forest plot displays the treatments' effect estimate with the corresponding confidence intervals, compared to a pre-specified reference treatment. Usually, placebo or standard (usual) care is used as a reference. Figure 1.3 presents a forest plot for the smoking cessation example. The plot indicates that the interventions individual counseling and group counseling significantly increases the probability of smoking cessation, since the corresponding confidence intervals do not cross the line of no-effect (red dashed line). Note also, that the intervention no-contact is used as a reference category.
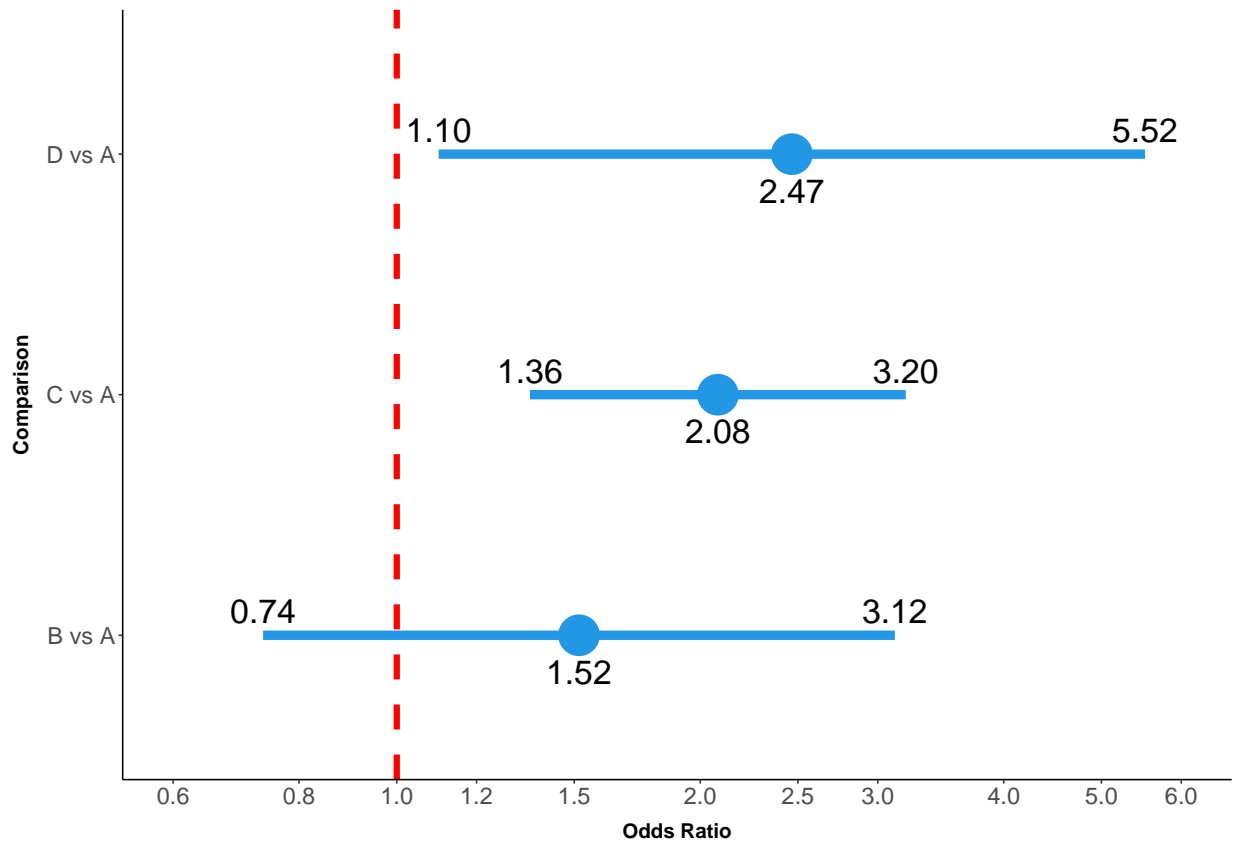
Abbreviations: A = no contact, B = self-help, C = individual counseling, D = Group counseling

The effectiveness of the interventions is measured by using the odds ratio as summary measure. Axis x is displayed on a logarithmic scale for a better representation of odds ratios below 1.

Figure 1.3: Forest plot for the smoking cessation example.

A drawback of the forest plot is that in large networks with many nodes and comparisons, the presentation of the NMA results may be challenging. For this reason, the league table could be used instead, which displays all the available treatments' effect estimates. Consider a network with T nodes. The league table will be a $T \times T$ matrix which presents the available $\binom{T}{2} = \frac{T(T-1)}{2}$ treatment effect estimates. Diagonal elements refer to the names of the nodes, while off-diagonal elements refer to the comparison between the nodes of the underlying column and row. The lower triangle of the league table is equivalent with the upper triangle, with the only difference that the upper triangle is in the opposite direction. For example, if

the element of the 1st row and 2nd row refer to the comparison A versus B, the corresponding element of the 2nd row and 1st column will refer to the comparison B versus A. For that reason, to simplify the league table we usually exclude the first upper triangle.

Table 1.1: League table displaying the odds ratio estimates with the corresponding 95% confidence intervals, for the smoking cessation example.

No contact	1.52 [0.74, 3.12]	2.08 [1.36, 3.20]	2.47 [1.10, 5.52]
0.66 [0.32, 1.36]	**Self-help**	1.37 [0.65, 2.89]	1.63 [0.67, 3.93]
0.48 [0.31, 0.74]	0.73 [0.35, 1.53]	**Individual counseling**	1.18 [0.55, 2.55]
0.41 [0.18, 0.91]	0.62 [0.25, 1.49]	0.84 [0.39, 1.82]	**Group counseling**

NMA provides a hierarchy of the interventions based on the effectiveness or the safety, by using either the P-scores in the frequentist framework (Rücker and Schwarzer, 2015), or the surface under the cumulative ranking (SUCRA) in the Bayesian framework (Salanti et al., 2011); P-scores and SUCRAs range between 0-1. A value equal to 1 indicates that the certain intervention is the best, while a value equal to zero indicates that is the worst (Rücker and Schwarzer, 2015; Salanti et al., 2011). These two metrics can be also visualized through the rankogram (Salanti et al., 2011; Rücker et al., 2022) . Mavridis et al. (2020) have extended P-scores to multiple outcomes and to account for minimum clinical important differences. Table 1.2 and Figure 1.4 indicate that the best intervention is the group counseling followed by the individual counseling, since they associated with the largest P-scores and probabilities to be ranked in the first places.

Table 1.2: P-scores for the smoking cessation example.

Intervention	P-score
Group counseling	0.8376
Individual-counseling	0.7103
Self-help	0.4042
No contact	0.0479
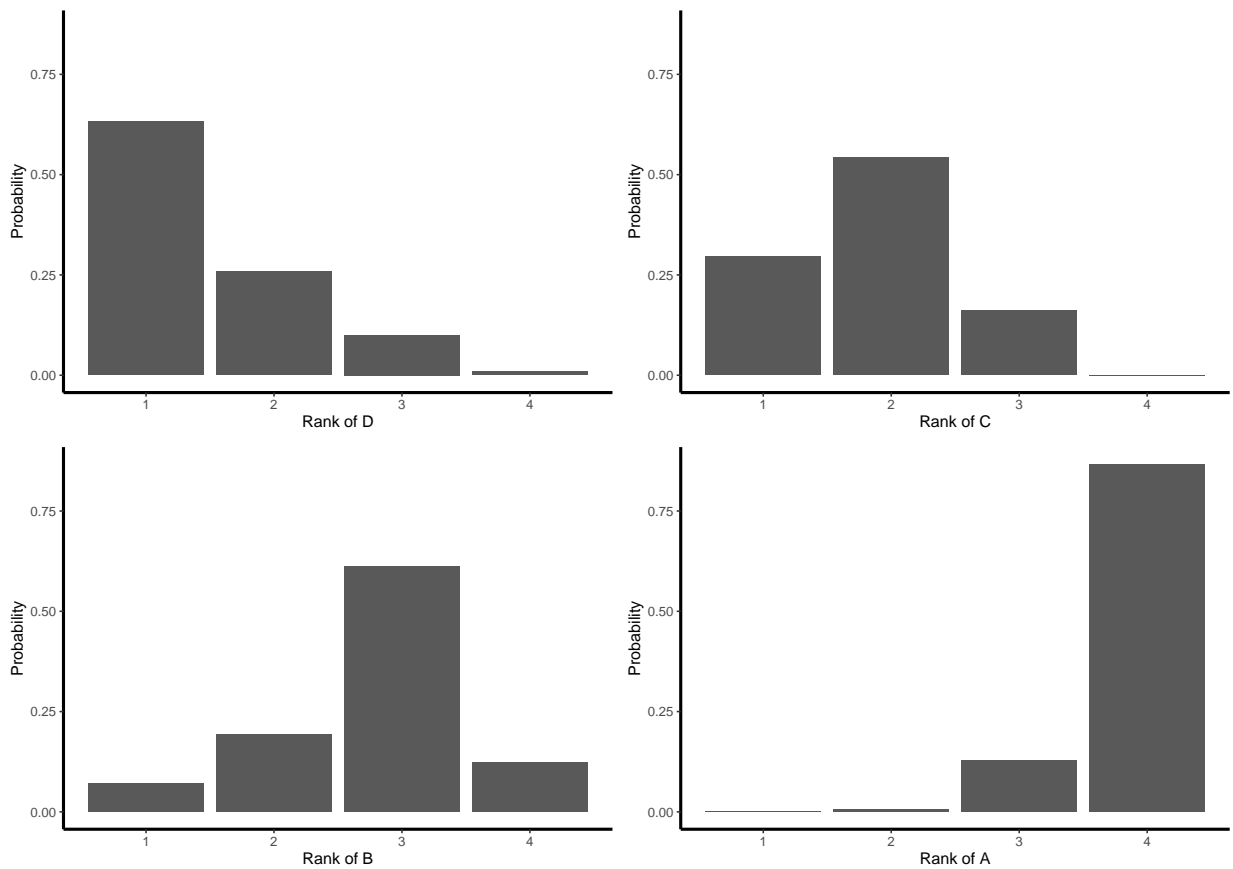
Abbreviations: A = no contact, B = self-help, C = individual counseling, D = Group counseling.

Figure 1.4: Rankogram for the smoking cessation example.

1.2 Transitivity

The validity of any statistical method depends on the plausibility of the assumptions made. The most fundamental assumption in NMA is that of transitivity, also known as similarity assumption (Seitidis et al., 2022a; Salanti, 2012; Donegan et al., 2010; Jansen and Naci, 2013), which refers to the ability to learn through the indirect evidence. We discuss several manifestations of the transitivity assumption. To illustrate the concepts, we will use a network of three interventions forming a closed ABC loop in the sense that all direct comparisons are observed (see Figure 1.1b).

Firstly, the definition of the treatments (nodes) in the network must be similarly defined across trials (Salanti, 2012). For example, in the ABC network of Figure 1.1b, we obtain the indirect estimate of BC using treatment A as common comparator. Hence, treatment A must be similarly defined in AB and AC trials. For example, if treatment A is given as a pill to the participants of AB trials and as an injection to the participants of AC trials, it is quite possible that transitivity is violated.

Secondly, the missing treatments in each trial must be missing at random (MAR) (Salanti, 2012). This implies that the probability of the treatment being missing is independent of the outcome (Lu and Ades, 2006). In practice, it is difficult to defend this assumption, but there are several reasons that indicate a violation. For example, when the purpose of comparing a drug with placebo or other suboptimal interventions is to demonstrate its efficacy. Generally, defending transitivity in sparse networks is more intensive. This is because in sparse networks where interventions are compared with only one common intervention (star networks) or just a few, it requires a judgement whether the missing edges are due to chance or because the trialist expected unwelcome results from them. A further consideration is whether the chronological order of interventions overlaps. Consider a recent intervention. Clearly, it is not possible to evaluate the efficacy of this intervention from past studies, whose results may be affected by the year in which they were published (Mavridis et al., 2016). Transitivity may be also violated in such a case for two potential reasons: 1) substantial changes in the definition of treatments; i.e. changes in the standard of care; 2) studies previously indicated exaggerated effect sizes as a result of low quality and publication bias (Seitidis et al., 2022a).

NMA assumes that all participants could have been randomized to any of the available treatments and the assumption of transitivity is violated if this is not the case (Salanti, 2012). Suppose in our network that treatment A is a first-line treatment, and treatments B and C could be either first or second-line treatments; second-line treatment implies that the participant has received it as a "stronger" treatment because the initial one failed to help

him or had stopped working. Therefore, it is not plausible to assume that participants in BC trials would have a similar treatment effect if they had been assigned to AC trials.

Treatment's efficacy many times depends on several participants' characteristics (effect modifiers) such as age, gender, baseline risk or year of randomization. Transitivity assumes that the distribution of effect modifiers is similar across treatment comparisons (Salanti, 2012; Jansen and Naci, 2013). For example, if age is an effect modifier and the mean age of AC trials differs significantly from the mean age of AB trials, then transitivity assumption is violated. It should be noted that if age differs within AC (or AB) trials, but the mean and the variability of the participant's age between AB and AC trials are similar, transitivity assumption holds.

1.3 Inconsistency

A key assumption of NMA is that of consistency which implies that the direct evidence is in agreement with the indirect evidence. Also, through this assumption we can statistically evaluate the transitivity assumption.

Consider the simple case, where a network is created from studies comparing treatments A, B and C; assume that only two-arm studies are available, and all possible direct comparisons (AB, AC, BC) are observed. Consistency implies that the direct evidence of BC comparison is in agreement with the indirect evidence of BC. This is expressed mathematically as

$$\mu_{BC}^{DIR} = \mu_{BC}^{IND} = \mu_{AB}^{DIR} - \mu_{AC}^{DIR} \tag{1.1}$$

where upper index "DIR" and "IND" refer to the direct and indirect evidence, respectively. Inconsistency arises when direct and indirect evidence differ substantially. It is broadly split in two categories, the loop and the design inconsistency (Higgins et al., 2012).

Loop inconsistency refer to inconsistencies in closed-loops such as the ABC loop of Figure 1.1b. Design inconsistency refer to inconsistencies among studies involving different sets of treatments. Let assume that in the above ABC network there are also multi-arm studies comparing the three treatments. If the effect size of BC comparison, obtained from the multi-arm studies, differs substantially form the equivalent effect size from the BC studies (studies comparing treatments B and C), then design inconsistency in the BC comparison is present.

1.4 Methods to Detect Inconsistency

A common practice for detecting inconsistencies in the NMA model is to use inconsistency factors. For example, in the ABC example of Figure 1.1b we can evaluate the consistency assumption by adding an extra term in the consistency equation (1.1). This expressed mathematically as

$$\mu_{BC}^{DIR} = \mu_{AB}^{DIR} - \mu_{AC}^{DIR} + w \tag{1.2}$$

where w denotes the inconsistency factor for the ABC loop. By evaluating whether $w = 0$, we can test the consistency assumption.

1.4.1 Global Methods

1.4.1.1 Lu & Ades

Lu and Ades (2006) proposed a global method for testing consistency in the network by adding inconsistency factors to each treatment independent closed-loop; a closed-loop is considered independent if it contains at least one edge which is not a part of any other independent closed-loop (Seitidis et al., 2022b). Consistency is examined globally via a chi-square test that evaluates if the sum of the inconsistency factors' absolute values differs significantly from zero. The number of inconsistency factors, which is in addition the number of inconsistency degrees of freedom (ICDF), is calculated as the number of the independent closed-loops. In the case where only two-arm studies are present, the number of the independent closed-loops is specified using the following formula:

$$ICDF = K - T + 1,$$

where K denotes the total number of observed comparisons and T the total number of treatments. In the case where multi-arm studies are also present, there is not any formula to calculate the number of independent closed-loops.

A major drawback of the method is the fact the specification of the inconsistency factors is not unique and different parameterization may lead to different findings (Higgins et al., 2012). The specification of inconsistency factor is performed manually by hand. This could be a limitation of the method, because in complex network with many interventions and comparisons, the specification of the inconsistency factor may be extremely challenging. Furthermore, the method accounts only for loop inconsistencies, ignoring the design inconsistencies in the network.

1.4.1.2 Design-by-Treatment Interaction Model

To free the issue of parameterization in the Lu & Ades model, Higgins et al. (2012) proposed a method that accounts for both types of inconsistency, loop and design inconsistencies, namely design-by-treatment interaction model. Inconsistency factors are added in the designs that includes a pair of treatments whose contrast can be estimated directly from previous designs, or indirectly through the consistency assumption. The number of inconsistency factors can be calculated from the following formula:

$$\#IF_{DBT} = df_{DBT} - df_{NMA} - T + 1,$$

where df_{DBT} denotes the degrees of freedom of the NMA model adjusted for inconsistencies under the design-by-treatment method, df_{NMA} denotes the degree of freedom of the NMA model and T denotes the total number of treatments included in the NMA model. The specification of df_{DBT} is based on the following rationale:

- each design from two-arm studies contributes one point

- multi-arm designs with k arms contribute k-1 points

- df_{DBT} is calculated as the sum of the points from two-arm designs and multi-arm designs.

After specifying the inconsistency factors, consistency is evaluated globally by using the Wald statistic which follows a X^2 distribution with $\#IF_{DBT}$ degrees of freedom (White et al., 2012).

1.4.1.3 Random-effects Implementation of the Design-by-Treatment Interaction Model

A random-effects implementation of the design-by-treatment interaction model was proposed by Jackson et al. (2014). It is an extension of the design-by-treatment model, in which inconsistency factors are modelled by using random-effects, whereas fixed-effects are used in the design-by-treatment model. The main advantage of the method is its ability to incorporate study-level covariate effects that may mitigate the need for inconsistency factors (Jackson et al., 2014). The method evaluates the consistency assumption in the Bayesian framework using Markov Chain Monte Carlo (MCMC), by adding inconsistency factors to each design. Inconsistency is evaluated by comparing the Deviance Information Criterion (DIC) between the "inconsistent" NMA model (model that includes inconsistency factors) and the consistent NMA model (model without inconsistency factors).

1.4.1.4 Q - Statistic

Heterogeneity in pairwise meta-analysis is usually assessed through the Cochran's Q-statistic (Higgins and Thompson, 2002). By decomposing the generalized Q-statistic consistency in NMA can be evaluated globally (Krahn et al., 2013; König et al., 2013; Freeman et al., 2019). More specifically, the overall variability in the network (Q) is split into two sources. The variability due to within-design heterogeneity (denoted as Q^{het}), and between-design heterogeneity (denoted as Q^{inc}) which is also termed design inconsistency. Let assume that $\hat{\theta_{ic}}$ and $\hat{\sigma_{ic}}$ denote the treatment effect estimate and the standard error, respectively, for the treatment comparison of design c where there are i studies available. Also, let $\hat{\theta}_c$ and $\hat{\theta}_{ic}^{NMA}$ denote the summary direct and NMA estimate, respectively, for the treatment comparison of design c. The overall Q-statistic is calculated using the following formula:

$$Q = Q^{het} + Q^{inc}$$

where

$$Q^{het} = \sum_c Q_c^{het} = \sum_c \sum_i \left(\frac{\hat{\theta}_{ic} - \hat{\theta}_c}{\hat{\sigma_{ic}}}\right)^2$$

$$Q^{inc} = \sum_c \left(\frac{\hat{\theta}_c - \hat{\theta}_c^{NMA}}{\hat{\sigma}_c}\right)^2.$$

Consistency is evaluated globally through the Q^{inc} which follow a X^2 distribution with $\#IF_{DBT}$ degrees of freedom. It should be noted that the use of the fixed-effects model within designs may provide a more accurate way of identifying network inconsistencies (Krahn et al., 2013).

1.4.1.5 Unrelated Mean Effects

Dias et al. (2013) proposed a method where the consistency assumption is evaluated by comparing the NMA model with a NMA model that is not based on the consistency assumption. More specifically, in NMA, estimates are derived by selecting one treatment as reference. The comparisons between the reference treatment and the rest treatments are termed basic parameters, while the rest comparisons are termed functional parameters. NMA assumes that the basic parameters are normally distributed and calculates the functional parameters as a linear combination of the basic parameters. For example, in the ABC network of Figure 1.1, if we select treatment A as reference, the NMA model assumes

$$d_{AB} \sim N(0, \sigma^2)$$

$$d_{AC} \sim N(0, \sigma^2)$$

where d_{AB} and d_{AC} denote the treatment effect of comparisons A versus B and A versus C, respectively. The treatment effect for the comparison between treatments B and C is calculated through the consistency assumption:

$$d_{BC} = d_{AC} - d_{AB}.$$

In contrast, the unrelated mean effects (UME) model treats the functional parameters as independent parameters and assumes that are normally distributed. Hence, in the ABC network of Figure 1.1, the treatment effect of BC is derived by assuming $d_{BC} \sim N(0, \sigma^2)$. The implementation of the method is performed in the Bayesian framework, and evaluation of the consistency assumption could be performed either by comparing the DIC values between the two models (NMA model and UME model), or by examining the contribution to the posterior mean residual deviance (deviance contribution plot).

1.4.2 Local Methods

1.4.2.1 Loop Specific

Bucher et al. (1997) proposed a method for identifying inconsistencies in the network by testing for each closed-loop whether the effect of the inconsistency factor is zero. In addition, it is a two-stage method where at the first stage, the closed-loops where direct and indirect evidence are available, are specified. At the second stage, differences between direct and indirect evidence are identified, by adding an inconsistency factor w to the consistency equation and testing via a Z-test whether $w = 0$. For example, in the ABC network of Figure 1.1, this expressed mathematically as

$$d_{BC}^{DIR} = d_{BC}^{IND} + w = d_{AC}^{DIR} - d_{AB}^{DIR} + w \Leftrightarrow$$

$$w = d_{BC}^{DIR} - d_{BC}^{IND} = d_{BC}^{DIR} - d_{AC}^{DIR} + d_{AB}^{DIR} \Rightarrow$$

$$Var(w) = Var\left(d_{BC}^{DIR} - d_{AC}^{DIR} + d_{AB}^{DIR}\right) = Var\left(d_{BC}^{DIR}\right) + Var\left(d_{AC}^{DIR}\right) + Var\left(d_{AB}^{DIR}\right).$$

Consistency is evaluated by using the z-score of comparison BC $z_{BC} = \frac{w}{\sqrt{Var(w)}}$, which follows a standard normal distribution.

1.4.2.2 Separate Indirect from Direct Evidence (SIDE)

A Bayesian local method for identifying inconsistencies in NMA was proposed by Dias et al. (2010), namely node-split method. The method separates the direct and the indirect

evidence and tests whether there are substantial differences among these two sources of evidence. More specifically, for any pair of treatments X and Y, inconsistency is identified by comparing the posterior distribution of d_{XY}^{DIR} and d_{XY}^{IND}; d_{XY}^{DIR} is obtained from the studies directly comparing treatments X and Y, while d_{XY}^{IND} is calculated through the consistency assumption.

The frequentist equivalent of the node-split method has been developed by White (2015), termed side-splitting. A drawback of the model is that the selection of the reference treatment in the NMA model may affect the conclusions about consistency; different parameterization yield to different results. For that reason, the symmetric side-splitting method has been developed, which addresses the issue of parameterization by symmetrically splitting the inconsistency factor in multi-arm studies (White, 2015; Shih and Tu, 2021). Moreover, Efthimiou et al. (2019) suggested a method that involves assessing direct and indirect evidence separately. More specifically, for any pair of treatments X, Y that belong in at least one closed-loop, the direct evidence (d_{XY}^{DIR}) is obtained by meta-analyzing the studies that compare the corresponding treatments, while the indirect evidence (d_{XY}^{IND}) is obtained from the NMA model by excluding the studies that compare directly the treatments X and Y. Inconsistency is located by testing via a Z-test whether $d_{XY}^{DIR} - d_{XY}^{IND} = 0$.

Shih and Tu (2021) mention that the former "node-split" models do not satisfy the principle of independence between direct and indirect evidence; that is (1) changes in data constituting the indirect evidence should not affect the direct evidence estimate (and its standard error) (2) when the model implies no indirect evidence, the NMA estimate (and its standard error) should be identical to the direct evidence estimate. Therefore, they proposed a novel method based on the principle of independence, termed evidence-splitting approach, which is implemented using structure equation modelling.

1.4.2.3 Back - Calculation

A two-phase method has been proposed by Dias et al. (2010) for checking inconsistency in which the indirect evidence is back-calculated by considering the NMA estimates as a weighted average of the direct and indirect evidence. These two sources of evidence are considered independent, because they obtained from separate evidence sources. For any pair of treatments X, Y included in the network, the indirect evidence of comparison X versus Y (denoted as d_{XY}^{IND}) is obtained using the following formula:

$$\hat{d}_{XY}^{IND} = \left(\frac{\hat{d}_{XY}^{NMA}}{V_{XY}^{NMA}} - \frac{\hat{d}_{XY}^{DIR}}{V_{XY}^{DIR}} \right) V_{XY}^{IND}$$

$$\frac{1}{V_{XY}^{IND}} = \frac{1}{V_{XY}^{NMA}} - \frac{1}{V_{XY}^{DIR}},$$

where $\hat{d}_{XY}^{NMA}$ denotes the NMA estimate for the comparison between treatments X and Y, V_{XY}^{NMA} the variance of $\hat{d}_{XY}^{NMA}$, $\hat{d}_{XY}^{DIR}$ the treatment effect estimate of X versus Y obtained from the studies comparing treatments X and Y, V_{XY}^{DIR} the variance of $\hat{d}_{XY}^{DIR}$ and V_{XY}^{IND} denotes the variance of $\hat{d}_{XY}^{IND}$. Consistency is evaluated by testing via a Z-test whether

$$w_{XY} = d_{XY}^{DIR} - d_{XY}^{IND} = 0.$$

1.4.2.4 Net Heat Plot

Net heat plot developed by Krahn et al. (2013), is meant to visualize the sources of inconsistency in the NMA model. It visualizes the Q^{inc} statistic via a quadric-colored matrix. Off-diagonal elements denote the change in inconsistency between direct and indirect evidence in the NMA estimate of the corresponding row, when consistency in the design of the corresponding column is relaxed (Rücker et al., 2022). Diagonal elements represent the contribution of each design to the total inconsistency Q^{inc}. The coloring of the boxes is proportional to the evidence of inconsistency. A dark red color indicates strong evidence of inconsistency, whereas a dark blue indicates that the evidence of the design in the column supports the evidence in the row (consistency). As an additional feature, the contribution of each design to the NMA relative effect estimate is displayed; the larger the size of the grey box, the larger the contribution to the NMA estimate. Consequently, inconsistencies are identified by locating those boxes with intense red colors and large contributions to the NMA estimates. Although the net heat plot is an effective tool for identifying inconsistencies, it may mask inconsistencies in large networks with numerous nodes and designs (Freeman et al., 2019).

1.5 Inverse Variance NMA Model Adjusted for Inconsistency Factors

A commonly used model in NMA is the Inverse Variance (IV) NMA model (Salanti et al., 2008). Let us consider that a systematic review resulted in a total of n studies, where each study $s = 1, \ldots, n$ compares T_s treatments. From each study s, $T_s - 1$ comparisons are included in the NMA model, because the rest can be obtained as a linear combination. Hence, the NMA model ends up with a total of $N = \sum_{s=1}^{n} (T_s - 1)$ comparisons. Let the vector $\mathbf{y} = (y_1, y_2, \ldots, y_N)'$ contain the estimated contrasts across all studies, where each y_i denotes the $u_{is} \in 1, \ldots, T_s - 1$ contrast of study $s \in 1, 2, \ldots, n$. The random-effects IV

NMA model is described from the following equation:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1.3}$$

where $X_{N\times(T-1)}$ denotes the design matrix and $\boldsymbol{\mu}$ contains the basic contrasts (dimension $T-1$). Vectors $\boldsymbol{\beta}$ and $\boldsymbol{\epsilon}$ (dimension N) contain the normally distributed random-effects and sampling errors, respectively, which are described from the following equations:

$$\boldsymbol{\beta} \sim MN_N(\mathbf{0}, \boldsymbol{\Delta})$$

$$\boldsymbol{\epsilon} \sim MN_N(\mathbf{0}, \boldsymbol{\Sigma}).$$

A common assumption in NMA is to assume common heterogeneity across different treatment comparisons, which is reflected from the block diagonal structure of matrix $\boldsymbol{\Delta}$. Covariance matrix $\boldsymbol{\Sigma}$ is assumed known and is obtained from the data (Franchini et al., 2012).

Design matrix $\mathbf{X}$ describes the observed treatment comparisons as a linear combination of the basic contrasts, embedding the consistency assumption. More specifically, the columns and rows of the matrix reflect the basic contrasts and the treatments comparisons, respectively. Hence, each comparison can be written as a combination of the basic contrast by assigning values 1, 0, and -1 to the corresponding columns. For example, in the ABC network of Figure 1.1b let assume that treatment A is used as reference; comparisons AB and AC are considered as basic contrasts, while comparison BC as a functional comparison. Let also assume that $\mathbf{X_1}$ and $\mathbf{X_2}$ corresponds to AB and AC, respectively . Comparisons AB and AC are denoted in the design matrix by assigning 1 to the column that correspond to the corresponding basic contrast and 0 elsewhere, while comparison BC is reflected by setting $X_{31} = -1$ and $X_{32} = 1$. The design matrix ends up with the following form:

$$\mathbf{X} = \begin{array}{c} \\ AB \\ AC \\ BC \end{array} \begin{array}{c} \begin{array}{cc} AB & AC \end{array} \\ \left(\begin{array}{cc} 1 & 0 \\ 0 & 1 \\ -1 & 1 \end{array} \right) \end{array}.$$

The IV NMA model can be adjusted to include inconsistency factors, by adding an extra term to the equation 1.3. In the case where p inconsistency factors are included in the NMA model, it can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\beta} + \mathbf{Zb} + \boldsymbol{\epsilon}, \tag{1.4}$$

where $\mathbf{Z}_{N \times p}$ is an index matrix with elements values of 1, 0, -1 indicating in which comparisons inconsistency factors are added, and $\mathbf{b} = (b_1, b_2, \ldots, b_p)'$ denotes the inconsistency factors' coefficients.

Chapter 2

Treating Inconsistency as a Variable Selection Problem

In this chapter, we propose a novel method for detecting inconsistencies in NMA; differences between direct and indirect effect estimates. The method evaluates network consistency both globally and locally, by treating the consistency assumption as a variable selection problem in a regression model. The chapter is structured as follows. We start by describing the stochastic search variable selection method and the IV-NMA model. Then, we present our novel method for identifying inconsistencies and evaluate its performance in two published NMAs. Lastly, we discuss the benefits of the proposed method.

2.1   Stochastic Search Variable Selection

Stochastic search variable selection (SSVS), introduced by George and McCulloch (1993), performs variable selection for linear regression models in the Bayesian framework, by restricting the predictors' coefficients close to zero according to their inclusion probabilities. Due to its efficiency, it has been adapted for use in a broad range of modelling applications (Dellaportas et al., 2000), including log-linear models (Ntzoufras et al., 2000), generalized linear models (George and McCulloch, 1997; George et al., 1996), multivariate regression (Brown et al., 1998), population pharmacokinetic models (Wakefield and Bennett, 1996), factor analytic models (Mavridis and Ntzoufras, 2014; Dunson et al., 2006), building stock portfolio in financial sector (George and McCulloch, 1996) and combining aggregate and individual participant data (IPD) in component NMA (Efthimiou et al., 2022).

Let assume that we have set of n observations and a linear regression model that includes

predictors $X_1, X_2, \ldots, X_p$, as described in the equation bellow:

$$y = a + b_1 X_1 + b_2 X_2 + \ldots + b_p X_p + \epsilon, \quad \epsilon \sim N(0, \sigma^2) \tag{2.1}$$

where y denotes the dependent variable, $\mathbf{b} = (b_1, b_2, \ldots, b_p)$ the coefficients of the candidate predictors and ϵ the error term. Variable selection is performed by assigning to each predictor an inclusion probability and assuming that predictors' effects are described from a mixture of two normal distributions. This is expressed mathematically as

$$b_\ell \mid \gamma_\ell \sim (1 - \gamma_\ell) N(0, \psi_\ell^2) + \gamma_\ell N(0, c_\ell^2 \psi_\ell^2), \quad \ell = 1, \ldots, p \tag{2.2}$$

where γ_ℓ operates as an indicator variable (taking values 0 or 1) that identifies whether the predictor X_ℓ is included in the model or not, c_ℓ and ψ_ℓ are tuning parameters controlling the mixing ability of the model. The first density ("spike") denotes the effect of the predictor when it is excluded from the model, while the second density ("slab"), when it is included. Tuning parameters c_ℓ and ψ_ℓ are defined in a manner that ensures that, when a predictor is not included in the model ($\gamma_\ell = 0$), its effect is close to the area of zero, and far away from this area when it is included ($\gamma_\ell = 1$). SSVS spike and slab prior setup are illustrated in Figure 2.1, where spike's density is high at values close to zero, while slab's density is flatter, supporting larger values.
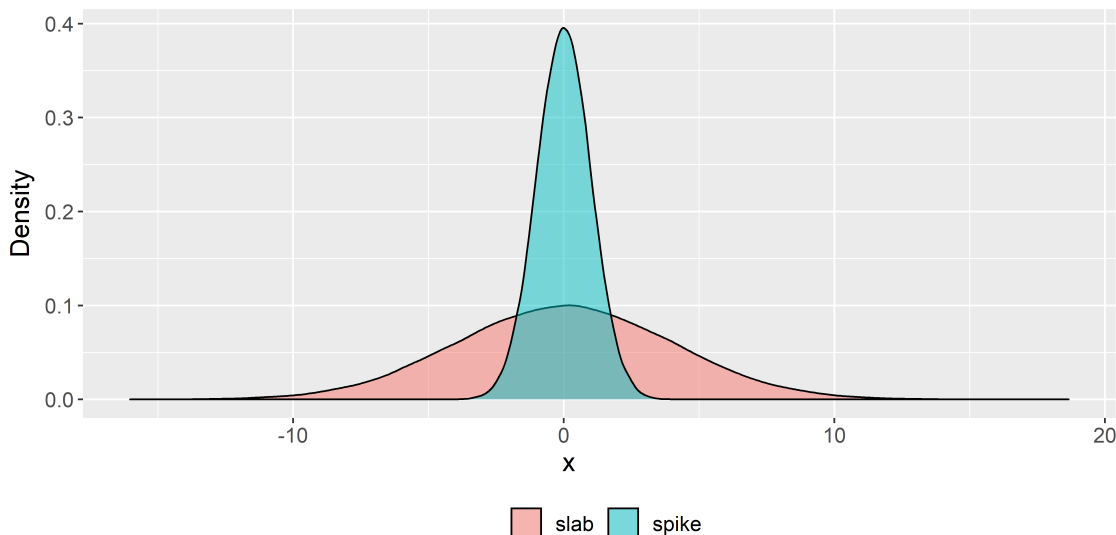
Figure 2.1: Graphical representation of the SSVS spike and slab prior setup.

The mixture of these two densities can be represented in a condensed matrix notation by

$$\mathbf{b} \mid \boldsymbol{\gamma} \sim N(\mathbf{0}, \mathbf{D_\gamma R D_\gamma}) \tag{2.3}$$

where $\mathbf{R}$ is the prior correlation matrix of the elements in $\mathbf{b}$, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_p)$ is a latent vector and

$$\mathbf{D_\gamma} = \begin{pmatrix} a_1\psi_1 & 0 & \ldots & 0 \\ 0 & a_2\psi_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_p\psi_p \end{pmatrix}, \qquad a_\ell = \begin{cases} 1, & \gamma_\ell = 0 \\ c_\ell, & \gamma_\ell = 1 \end{cases}.$$

2.1.1 Prior Setup

The correlation between the predictors is reflected by the matrix $\mathbf{R}$. If $\mathbf{R} = \mathbf{I_p}$, predictors are assumed independent. In practice, this assumption may be implausible since predictors are usually correlated with each other. An established prior setup for variable selection is the Zellner's g-prior (Zellner, 1986), in which prior correlation equals the design correlation multiplied by a scalar. This is expressed mathematically as

$$\mathbf{R} = g(\mathbf{X}'\mathbf{X})^{-1}\sigma^2, \quad \pi(\sigma^2) \propto \frac{1}{\sigma^2} \tag{2.4}$$

where $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p]$ denotes the design matrix and σ^2 the error variance term. This prior setup has the advantage that the general correlation of $\mathbf{b}$ is encapsulated by the matrix $(\mathbf{X}'\mathbf{X})^{-1}$ while we only need to specify one variance parameter, as opposed to the standard formulation which requires the specification of the whole matrix $\mathbf{R}$. Parameter σ is assumed to follow the Jeffreys scale-invariant prior. Alternatively, a proper choice is to assume a non-informative inverse gamma distribution, with equal shape and scale parameters, typically set to 10^{-3} or 10^{-4} (Perrakis and Ntzoufras, 2018). Furthermore, it is important to specify parameter g with cautious to avoid triggering the Jeffreys-Lindley-Bartlett paradox (Lindley, 1957; Bartlett, 1957; Jeffreys, 1961), in which for proper large values of g the posterior odds tend to favour the simplest model. To address this issue, it is recommended to assume g as fixed (Kass and Wasserman, 1995) or specify it using hyper-prior (Liang et al., 2008; Zellner and Siow, 1980).

Prior inclusion probabilities define the probability of including each predictor in the model, a priori. George and McCulloch (1993) suggest that a reasonable choice is to assume

that

$$\gamma_\ell \sim Bernoulli(\pi_\ell), \quad \ell = 1, \ldots, p. \tag{2.5}$$

In the case where $\pi_\ell = 0.5 \; \forall \; \ell = 1, \ldots, p$, each predictor is included/excluded from the model with the same probability, yielding to a model that includes $p/2$ predictors.

2.1.2 Variable Selection

A common variable selection strategy in SSVS is to exclude the predictors that are associated with posterior inclusion probabilities smaller than 0.5 (median probability model). In a Markov Chain Monte Carlo (MCMC) run of M iterations, the posterior inclusion probability of predictor ℓ is estimated as the average number of times the predictor was included in the model ($\gamma_\ell = 1$) in M iterations of the MCMC algorithm, which is expressed mathematically as

$$\hat{f}(\gamma_\ell = 1 \mid \mathbf{y}) = \frac{1}{M-B} \sum_{t=B+1}^{M} I(\gamma_\ell^{(t)} = 1), \tag{2.6}$$

where B is the number of iterations considered as burn-in period, and $\gamma_\ell^{(t)}$ is the inclusion value of predictor l at iteration t.

Alternatively, variable selection can be performed by estimating the posterior model odds and identifying the maximum a posteriori model. Let assume that we are interested on the comparison among two candidate models, m_1 and m_2. The posterior odds of m_1 over m_2 is obtained as

$$PO_{m_1 m_2} = \frac{f(m_1 \mid \mathbf{y})}{f(m_2 \mid \mathbf{y})} = \frac{f(\mathbf{y} \mid m_1)f(m_1)}{f(\mathbf{y} \mid m_2)f(m_2)} = BF_{m_1 m_2} \frac{f(m_1)}{f(m_2)},$$

where $BF_{m_1 m_2}$ is the Bayes Factor of m_1 over m_2 and $f(m_1)$ and $f(m_2)$ denote the prior model probabilities of models m_1 and m_2, respectively. It should be noted that in the case where the prior model probabilities are equal, the posterior model odds and the Bayes Factors coincide. In a MCMC run of M iterations, posterior model odds can be also derived as the ratio of the posterior model probabilities which in SSVS are estimated as

$$\hat{f}(m \mid \mathbf{y}) = \frac{1}{M-B} \sum_{t=B+1}^{M} I(m^{(t)} = m), \tag{2.7}$$

where $m^{(t)}$ is the model indicator in t iteration which transform the $\boldsymbol{\gamma}$ to a unique decimal

number and calculated as

$$m(\boldsymbol{\gamma}) = \sum_{\ell=1}^{p} \gamma_\ell 2^{\ell-1}. \tag{2.8}$$

2.2 Stochastic Search Inconsistency Factor Selection

Let assume an NMA model as described in equation 1.4 that includes $\ell = 1, \ldots, p$ inconsistency factors. Stochastic Search Inconsistency Factor Selection is a two-step method, in which at the first step inconsistency factors are specified, and at the second step, SSVS is applied to inconsistency factors.

2.2.1 Specification of Inconsistency Structure

Several models have been suggested for the specification of the inconsistency factors (matrix **Z**). Lu and Ades (2006) introduced a model that considers only loop inconsistencies and tests consistency in each independent closed-loop of the network (more details in section 1.4.1.1). Implementing the Lu & Ades method in complex networks is challenging because the inconsistency factors are sequentially specified manually. To address this issue, we have developed an algorithm that automatically locates in which comparisons inconsistency factors should be added. Figure 2.2 presents the process flowchart of the algorithm.
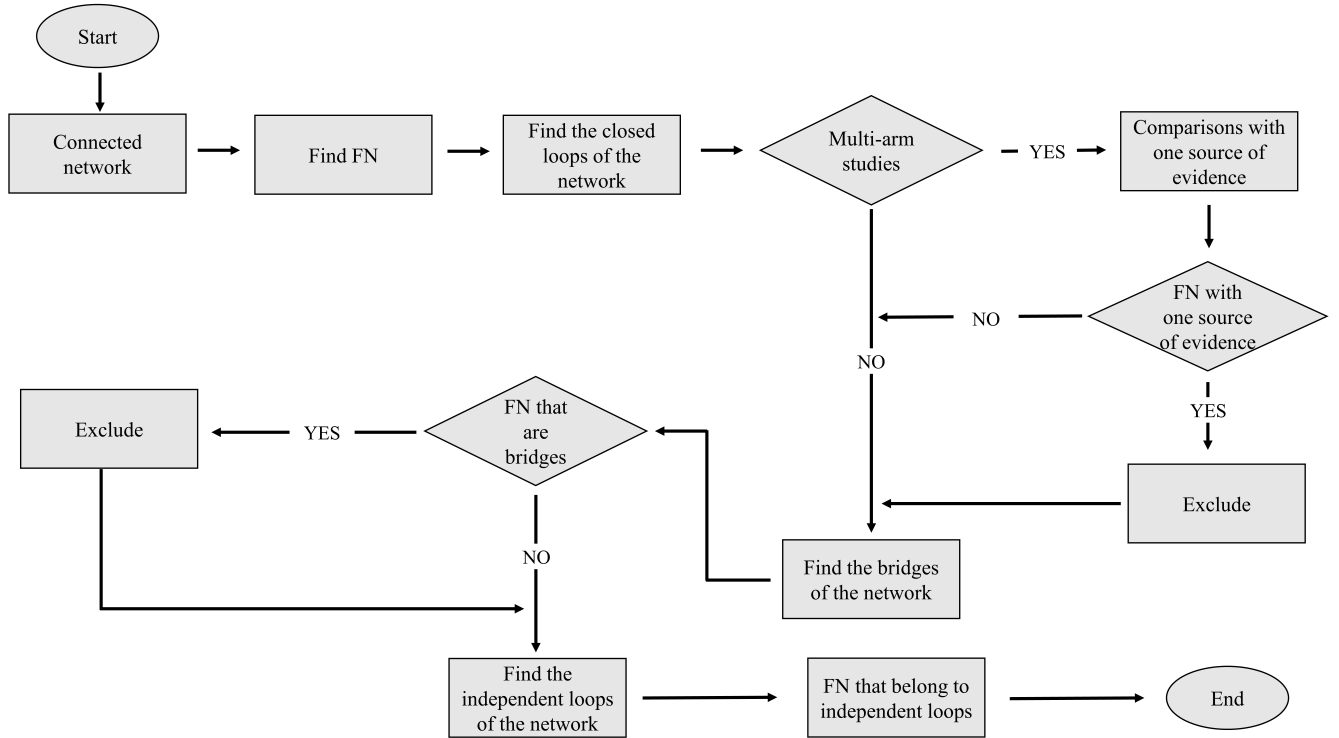
Figure 2.2: Process flowchart of the Lu & Ades algorithm.

The process flowchart can be summarised into the following steps:

- The algorithm firstly checks if the network is connected. In the case that it is not, the algorithm keeps the largest of the two networks by excluding the nodes that do not belong to it.

- Assume that inconsistency factors could be added in all functional parameters.

- Specify the closed-loops of the network by taking all the possible paths that start and end in the same node.

- Check if multi-arm studies are present in the network. In the case where multi-arm studies are present, inconsistency factors from functional parameters with one source of evidence (presented only in one multi-arm trial) are excluded.

- Exclude the bridges of the network by locating which functional parameters are not part of any closed-loop.

- Identify independent loops. Note: A loop is considered independent if it contains at least one edge which is not a part of any other independent loop.

- Exclude the functional parameters that do not belong to any independent loop. Also, the number of inconsistency factors must be equal to the number of independent loops. Thus, for each independent loop we keep only one functional parameter.

- The remaining functional parameters are the comparisons in which inconsistency factor should be added.

It is important to note that the algorithm ignores the treatment ordering (Higgins et al., 2012) and identifies a subset of viable Lu & Ades models. As a result, we suggest using the design-by-treatment model (described in section 1.4.1.2) instead, which includes all of the Lu & Ades models and addresses both sources of inconsistency, design and loop. Additionally, the random-effects implementation of the design-by-treatment interaction (described in section 1.4.1.3) model can be utilized, in which inconsistency factors are added in each design of the network.

2.2.2 Prior Setup

After specifying the $\mathbf{Z}$ matrix, inconsistency factors are treated as predictors in a linear regression model, and variable selection is performed by applying SSVS on them. The prior setting is similar to the setting of equation 2.3 with the only difference that $c_\ell = c$, $\forall$ $\ell = 1, \ldots, p$.

2.2.2.1 Correlation Matrix

Correlation matrix $\mathbf{R}$ describes the prior correlation between the inconsistency factors. Assuming independence between the inconsistency factors (by setting $\mathbf{R} = \mathbf{I_p}$) would be a convenient choice. Independence among inconsistency factors in SSIFS is translated as the inclusion of an inconsistency factor does not affect the inclusion of another. This is a strong assumption to defend and in practice it may be implausible. Instead, to describe the correlation among the inconsistency factors, Zellner g-prior is used as described in equation 2.4. Also, to avoid triggering the Jeffreys-Lindley-Bartlett paradox, which is in our case favouring the consistent NMA model, the unit-information criterion approach (Kass and Wasserman, 1995) is used, which mathematically translates to $g = N$.

2.2.2.2 Prior Inclusion Probabilities

The probability of including inconsistency factors in the NMA model is defined a-priori from the prior inclusion probabilities. These probabilities should be specified aptly to avoid multiplicity issues. Assuming equal inclusion probability for each inconsistency factor is considered as a naïve non-informative approach since it does not account for multiplicity

(Scott and Berger, 2010). An established method for addressing these issues, is to use a Beta-binomial distribution on the prior inclusion probabilities, as described in the equation:

$$\gamma_\ell \sim \ Bernoulli(\pi)$$
$$\pi \sim Beta(\alpha, \beta).$$

This approach has the advantage that it accounts for multiplicity by automatically introducing a penalty that handles multiple testing (Scott and Berger, 2010, 2006). In the case where $\pi \sim Beta(1,1)$, this approach is equivalent to $\pi \sim U(0,1)$. Although this setup may seem enticing, it cannot be used in SSIFS due to its inappropriate favouring of inconsistent NMA models. Consider the simple case where

$$\gamma_\ell \sim Bernoulli(0.5), \quad \ell = 1, 2, \ldots, p.$$

Let $\pi_\ell = P(\gamma_\ell = 1)$ denote the probability that the inconsistency factor is included in the model and $1 - \pi_\ell = P(\gamma_\ell = 0)$ the probability not included. The prior probability of consistency in the NMA model (consistent NMA model) is calculated as

$$\pi_{cons} = \prod_{\ell=1}^{p} P(\gamma_\ell = 0) = 0.5^p.$$

In this manner, SSIFS tends to favour NMA models that include $p/2$ inconsistency factors and the probability of consistency decreases exponentially as the number of inconsistency factors increases. A proper setting should maintain the probability of consistency constant, irrespective of the number of inconsistency factors. For this purpose, the inclusion probability π is obtain by the following formula:

$$\pi_{cons} = P(\gamma_1 = 0, \ldots, \gamma_p = 0) \quad \Leftrightarrow$$
$$\pi_{cons} = (1 - \pi)^p \quad \Leftrightarrow$$
$$\pi = 1 - \pi_{cons}^{1/p}.$$

An important feature of SSIFS is the ability to incorporate past knowledge concerning network consistency. Parameter π_{cons} reflects the researcher's prior belief of having a consistent network. In cases where expertise is available, an informative prior could be used, while in the absent of expert opinion, our prior ignorance could be expressed by setting $\pi_{cons} = 0.5$. A systematic review revealed that among 201 published NMA networks, 44 found to be globally inconsistent (Veroniki et al., 2021). According to this study, we could assume either that $\pi_{cons} = 0.78$ or $\pi_{cons} \sim Beta(157, 44)$; the latter setting assumes that the

mean probability of network consistency is equal to 0.78 with a standard deviation of 0.03. It should be mentioned that every case is unique, and if experts' opinion or prior knowledge is available, the parameter π_{cons} should be informed accordingly. Furthermore, the prior specification of parameter π_{cons} could be performed by modelling the probability of network consistency based on clinical criteria and network's characteristics, such as the number of studies, participants, nodes, etc.

The prior setting strategy for specifying the inclusion probabilities of the inconsistency factors in SSIFS, is quite different compared to the standard approach that is used in Bayesian variable selection. More specifically, prior inclusion probabilities in SSIFS are specified based on the parameter π_{cons}, whereas in the standard approach, they are specified based on the parameter π_ℓ.

2.2.3 Tuning

Tuning in SSIFS is essential to assure the proper use of the method. Parameters c and ψ_ℓ should be defined in a such manner to ensure that the density of b_ℓ is similar to the density described in Figure 2.1. Possible values of ψ_ℓ could be the inconsistency factor's standard deviations obtained from a pilot MCMC run of the full NMA model (Seitidis et al., 2022b; Ntzoufras, 2011). Parameter c reflects the prior odds of $\gamma_\ell = 0$ when $b_\ell \approx 0$. Therefore, the larger the values of c, the less likely is to include inconsistency factors with minor effects. As a general rule, values between 10 and 100 perform well in most cases, but heed is advised according to the characteristics of the data (George and McCulloch, 1993; Perrakis and Ntzoufras, 2015).

Alternatively, tuning could be performed by assuming a value of practical significance for the inconsistency factors. By assigning a minimum value of acceptance (say ω), inconsistency factors with effect larger than ω in absolute values ($\mid b_\ell \mid > \omega$), are a-priori supported by the NMA model. This can be achieved by specifying the intersection points among the consistent and inconsistent NMA models; see Appendix for details. In the case where inconsistency factors assumed independent, the inconsistent NMA model is supported by SSIFS when

$$\mid b_\ell \mid > \psi_\ell \sqrt{\xi(c)}, \quad \xi(c) = \frac{2c^2 \log c}{c^2 - 1},$$

where $\xi(c)$ indicates the number of standard deviations an inconsistency factor should be beyond zero to be deemed significant. For example, let assume that a difference of 0.1 between direct and indirect evidence is deemed important. The tuning parameters of SSIFS

are obtained by solving the equation

$$0.1 = \psi_\ell \sqrt{\xi(c)}.$$

The solution of the equation is not unique. One possible solution is to set $c = 10$ and calculate ψ_ℓ as $\psi_\ell = \frac{0.1}{\sqrt{\xi(10)}} \approx 0.05$. In the case where a correlation between inconsistency factors is assumed, tuning should be performed based on solving the following inequality:

$$\mathbf{b}'[\boldsymbol{D}_{\boldsymbol{\gamma}=\mathbf{0}}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{D}_{\boldsymbol{\gamma}=\mathbf{0}}]^{-1}\mathbf{b} \le gp\sigma^2\xi(c), \quad \xi(c) = \frac{2c^2\log c}{c^2 - 1}.$$

More details about tuning can be found in Ntzoufras et al. (2000) and in Mavridis and Ntzoufras (2014).

2.2.4 Detection of Inconsistency

Inconsistency in SSIFS is detected by estimating the posterior odds or the posterior inclusion probabilities as described in section 2.1.2. Using the posterior odds of the consistent NMA model ($m(\boldsymbol{\gamma}) = 0$) over the rest observed inconsistent NMA models ($m(\boldsymbol{\gamma}) \neq 0$), network consistency is evaluated globally. Also, through the posterior odds we can identify the local sources of inconsistency in the network. Let assume that the posterior odds of a specific inconsistent NMA model over the consistent NMA model, is estimated as larger than 1. The comparisons in which inconsistency factor are added, denote the local sources of inconsistency, which causes global inconsistency to the network. The magnitude of the evidence is based on the estimate of the posterior odds. Estimates close to 1, indicates trivial evidence that the network is inconsistent. To determine whether the network is inconsistent, the Jeffrey's scale (Jeffreys, 1961) could be used in which the magnitude of the evidence is summarized according to the following table.

Table 2.1: Strength of evidence according to the Jeffrey's scale for the comparison between models m_1 and m_2.

Posterior Odds m_1 vs m_2	Strength of evidence (in favour of m_1)
$1 - 3.16$	Barely worth mentioning
$3.16 - 10$	Substantial
$10 - 31.6$	Strong
$31.6 - 100$	Very strong
>100	Decisive

Another approach for identifying the local sources of inconsistency is by using the median probability model, which includes inconsistency factors with $\hat{f}(\gamma_\ell = 1 \mid \mathbf{y}) \geq 0.5$ in the NMA model.

It has been observed that SSVS may not accurately estimate the posterior inclusion probabilities when the model space is large (e.g more than 20 predictors) (Ntzoufras, 2011). Therefore, in complex networks with more than 20 inconsistency factors, to avoid false conclusions it is recommended to mitigate the dimension by excluding the inconsistency factors that have posterior inclusion probability below 0.20 (Seitidis et al., 2022b; Fouskakis et al., 2009).

2.3   Example

To demonstrate the applicability of the method we utilize two published NMAs, one pertaining to the Agency for Health Care Policy and Research (AHCPR) recommendations concerning smoking cessation (Lu and Ades, 2006), and the other regarding the comparative effectiveness of oral phosphodiesterase type-5 inhibitors for erectile dysfunction (Veroniki et al., 2021; Yuan et al., 2013). SSIFS was applied by taking into considerations all approaches concerning the specification of the inconsistency factors (matrix $\mathbf{Z}$). Two scenarios were assumed for the prior correlation matrix $\mathbf{R}$, where at the first inconsistency factors were assumed independent ($\mathbf{R} = \mathbf{I}$), and at the second that were correlated which descripted by a Zellner g-prior by setting $\mathbf{R} = g(\mathbf{Z}'\mathbf{Z})^{-1}\sigma^2$. Parameters c and ψ_ℓ were specified based on the assumption that a value of $\omega = 0.2$ on the log scale is considered as the minimum

value of inconsistency that is of practical significance. Concerning the probability of having a consistent network, we considered two different scenarios. In the first case, to describe our ignorance concerning network consistency we assume that the probability of observing a consistent network is 50%, by setting $\pi_{cons} = 0.5$ (equivalent to uniform prior on model space approach), while in the second case, an informative prior based on historical evidence were assumed by setting $\pi_{cons} \sim Beta(157, 44)$. To assess the robustness of our findings further, we performed a sensitivity analysis in which inconsistency is supported a priori by 75% ($\pi_{cons} = 0.25$). Results from the MCMC run were obtained by using a burn-in period of 50K iterations, and a total of 300K iterations, while two chains were utilized to evaluate convergence.

The implementation of SSIFS was conducted by using the R-package *ssifs* (Seitidis et al., 2023a) which was developed as part of this research. The package is published on CRAN and is also hosted on a GitHub repository (https://georgiosseitidis.github.io/ssifs/). Details about the usage of the package are presented in the package's vignette (https://georgiosseitidis.github.io/ssifs/articles/An_introduction_to_ssifs.html).

2.3.1 Smoking Cessation

The network includes 24 studies (22 two-arm studies and two three-arm studies), 4 nodes and 28 pairwise comparisons, evaluating the effectiveness of no contact (reference), self-help, individual counseling and group counseling on smoking cessation. Figure 1.2 illustrates the geometry of the network.

According to SSIFS the network is globally and locally consistent. More specifically, in each scenario employed the effects of inconsistency factors on the logarithmic scale and the posterior inclusion probabilities estimated close to zero (for more details see Figure 2.3 and Table A1). As a result, the median probability model clearly indicates the absence of significant design or loop inconsistencies in the network. Furthermore, the model without inconsistency factors (consistent NMA model) estimated as the most probable with a posterior probability around 0.56 in the case where prior ignorance concerning network consistency was assumed, and 0.81 when historical evidence was used (for more details see Table 2.2). Hence, the posterior odds favour the consistent NMA model compared to the rest inconsistent NMA models. The hypothesis of network consistency according to Table 2.2 persists favoured, even when the observed inconsistent NMA models are grouped and considered as a single model, since the posterior odds of global consistency versus inconsistency shows marginal evidence (posterior odds from 1.22 to 1.38) when $\pi_{cons} = 0.50$ and substantial evidence when $\pi_{cons} \sim Beta(157, 44)$ (posterior odds from 4.26 to 4.88). Even in the extreme

scenario of $\pi_{cons} = 0.25$, the median probability model continues to imply that there are no significant inconsistencies, and the consistent NMA model was estimated to be the most dominant with a posterior model probability around 0.30 (for additional information, see Figure 2.3 and Table 2.2). Also, the global consistency test shows marginal evidence against the consistency hypothesis, with posterior odds ranging from $1/0.47 \approx 2.13$ to $1/0.43 \approx 2.33$ (see Table 2.2 for more details), which was expected given the high probability that it was assigned to the inconsistent NMA models a-priori. Hence, we can confidently conclude that the network is globally consistent without any important local inconsistency.
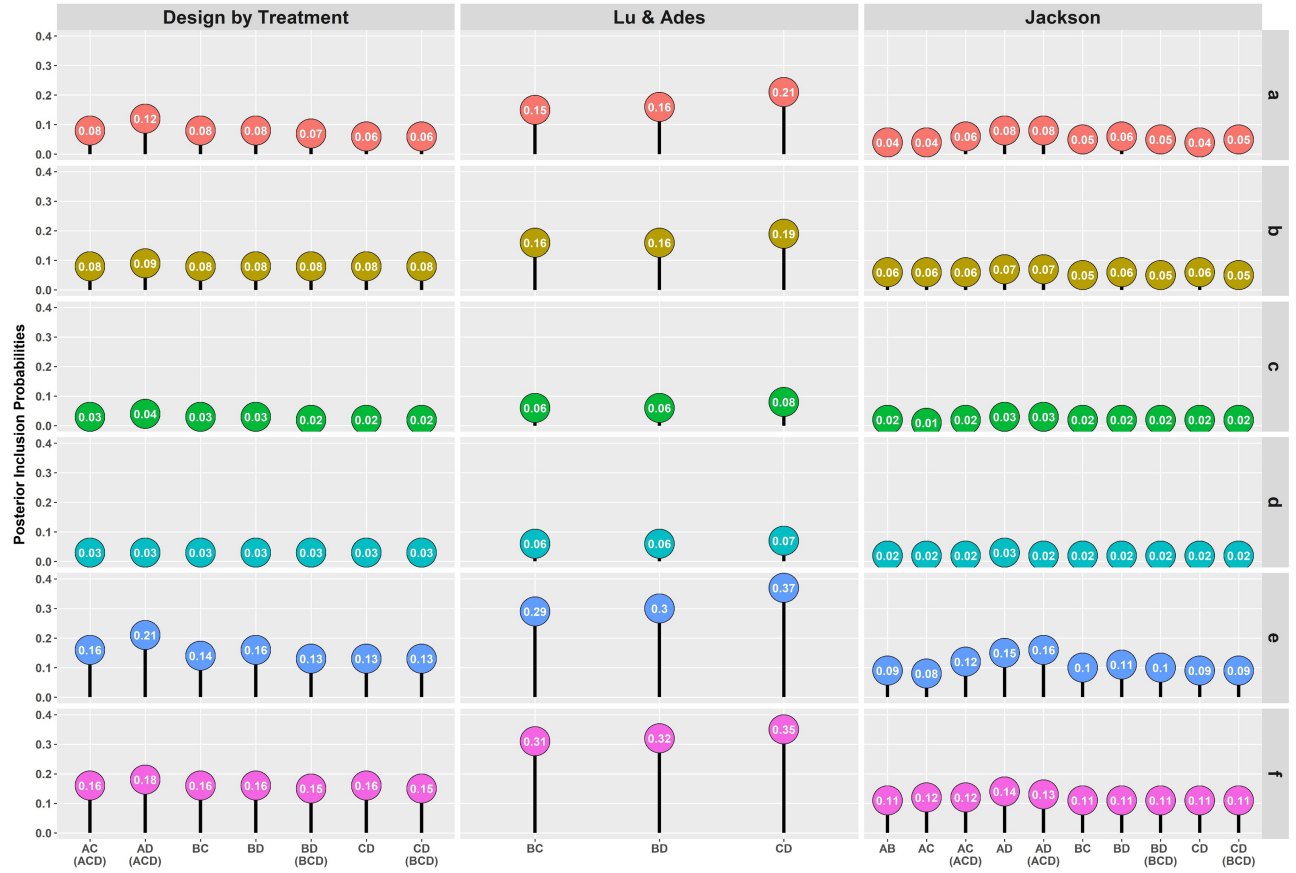
Scenario: a) $\pi_{cons} = 0.5$ & $\mathbf{R} = \mathbf{I}$, b) $\pi_{cons} = 0.5$ & $\mathbf{R} = g(\mathbf{Z}'\mathbf{Z})^{-1}\sigma^2$, c) $\pi_{cons} \sim Beta(157,44)$ & $\mathbf{R} = \mathbf{I}$, d) $\pi_{cons} \sim Beta(157,44)$ & $\mathbf{R} = g(\mathbf{Z}'\mathbf{Z})^{-1}\sigma^2$, e) $\pi_{cons} = 0.25$ & $\mathbf{R} = \mathbf{I}$, f) $\pi_{cons} = 0.25$ & $\mathbf{R} = g(\mathbf{Z}'\mathbf{Z})^{-1}\sigma^2$.

Abbreviations: A = no contact, B = self-help, C = individual counselling, D = group counselling.

Axis x denotes the comparisons in which inconsistency factors were added. For the Design by Treatment and the Jackson model, parenthesis denotes the design of the comparison.

Figure 2.3: Posterior inclusion probabilities for the smoking cessation example for the three different inconsistency models.

Table 2.2: Posterior model odds (PO) of consistent NMA model vs inconsistent NMA models and the corresponding probability of the consistent NMA model ($\hat{m}_{cons}$) for the three different inconsistency modelling approaches.

Inconsistency modelling approach	$\pi_{cons} = 0.50$				$\pi_{cons} \sim Beta(157,44)$				$\pi_{cons} = 0.25$			
	$\mathbf{R} = \mathbf{I}$		$\mathbf{R} = g(\mathbf{Z}'\mathbf{Z})^{-1}\sigma^2$		$\mathbf{R} = \mathbf{I}$		$\mathbf{R} = g(\mathbf{Z}'\mathbf{Z})^{-1}\sigma^2$		$\mathbf{R} = \mathbf{I}$		$\mathbf{R} = g(\mathbf{Z}'\mathbf{Z})^{-1}\sigma^2$	
	$\hat{m}_{cons}$	PO	$\hat{m}_{cons}$	PO	$\hat{m}_{cons}$	PO	$\hat{m}_{cons}$	PO	$\hat{m}_{cons}$	PO	$\hat{m}_{cons}$	PO
Smoking cessation												
DBT	0.56	1.27	0.56	1.27	0.81	4.26	0.82	4.56	0.31	0.45	0.31	0.45
Lu & Ades	0.56	1.27	0.58	1.38	0.82	4.56	0.83	4.88	0.31	0.45	0.32	0.47
Jackson	0.57	1.33	0.55	1.22	0.82	4.56	0.81	4.26	0.32	0.47	0.30	0.43
Erectile dysfunction												
DBT Lu & Ades	0.14	0.16	0.14	0.16	0.36	0.56	0.36	0.56	0.05	0.05	0.05	0.05

Abbreviations: DBT = Design-by-Treatment, PO = Posterior model odds, $\hat{m}_{cons}$ = posterior probability of the consistent NMA model.

Note that besides several scenarios employed assuming different prior inclusion probabilities, the posterior model probability of the consistent NMA model is similar and robust, irrespective of the correlation prior setup, ranging around 0.56 when $\pi_{cons} = 0.50$, 0.81 when $\pi_{cons} \sim Beta(157,44)$ and around 0.31 when $\pi_{cons} = 0.25$. Also, the Lu & Ades model is associated with larger posterior inclusions probabilities compared to rest methods employed for the specification of the $\mathbf{Z}$ matrix, because it includes fewer inconsistency factors and the posterior estimates are mainly driven by the prior set up; e.g when $\pi_{cons} = 0.50$, the Lu & Ades model includes three inconsistency factors with prior inclusion probability equal to

$$\pi_\ell = 1 - (0.50)^{1/3} \approx 1 - 0.79 \approx 0.21,$$

while the design-by-treatment model includes seven inconsistency factors with prior inclusion probability equal to

$$\pi_\ell = 1 - (0.50)^{1/7} \approx 1 - 0.91 \approx 0.09.$$

The conclusions about network consistency are the same when consistency is evaluated by using the standard methods. More specifically, the SIDE method suggests that there is not any significant discrepancy between direct and indirect evidence (see Figure 2.4) and also, the Q-statistic under the assumption of a full design-by-treatment interaction random-effects

model, indicates that the network is globally consistent ($Q = 4.66$, $df = 7$, $p-value = 0.70$). Moreover, the DIC values of the consistent NMA model ($DIC = 333.10$) and the UME model ($DIC = 333.00$) are similar, indicating that the network is globally consistent. Previous work also supports the absence of important local or global inconsistency in the network (Lu and Ades, 2006; Dias et al., 2010). The authors report that we cannot guarantee that there is no inconsistency in the network due to the high level of heterogeneity. SSIFS reflects this statement as well, in the case where we employed a non-informative prior concerning network consistency ($\pi_{cons} = 0.50$), in which marginal evidence exists that the network is consistent.

Comparison	Number of Studies	Direct Evidence	I2	Random Effects Model	OR	95%-CI
B:A						
Direct estimate	3	0.58	32%		1.38	[0.54; 3.57]
Indirect estimate					1.72	[0.57; 5.22]
Network estimate					1.52	[0.74; 3.12]
C:A						
Direct estimate	15	0.90	92%		2.10	[1.34; 3.31]
Indirect estimate					1.90	[0.49; 7.34]
Network estimate					2.08	[1.36; 3.20]
D:A						
Direct estimate	2	0.26	79%		2.44	[0.51; 11.71]
Indirect estimate					2.47	[0.97; 6.34]
Network estimate					2.47	[1.10; 5.52]
B:C						
Direct estimate	2	0.36	0%		1.08	[0.31; 3.70]
Indirect estimate					0.58	[0.23; 1.48]
Network estimate					0.73	[0.35; 1.53]
B:D						
Direct estimate	2	0.53	48%		0.54	[0.16; 1.82]
Indirect estimate					0.71	[0.20; 2.57]
Network estimate					0.62	[0.25; 1.49]
C:D						
Direct estimate	4	0.74	66%		1.08	[0.45; 2.63]
Indirect estimate					0.41	[0.09; 1.86]
Network estimate					0.84	[0.39; 1.82]

Abbreviations: A = no contact, B = self-help, C = individual counselling, D = group counselling.

Figure 2.4: Difference between direct and indirect evidence for the smoking cessation example, using the SIDE method.

2.3.2 Erectile Dysfunction

The network includes 69 two-arm studies, 7 nodes and 69 pairwise comparisons, evaluating the effectiveness of oral phosphodiesterase type-5 inhibitors (Avanafil, Mirodenafil, Sildenafil, Tadalafil, Udenafil, Vardenafil) concerning erectile dysfunction. In the NMA model, placebo was used as a reference node. Figure 2.5 illustrates the geometry of the network.
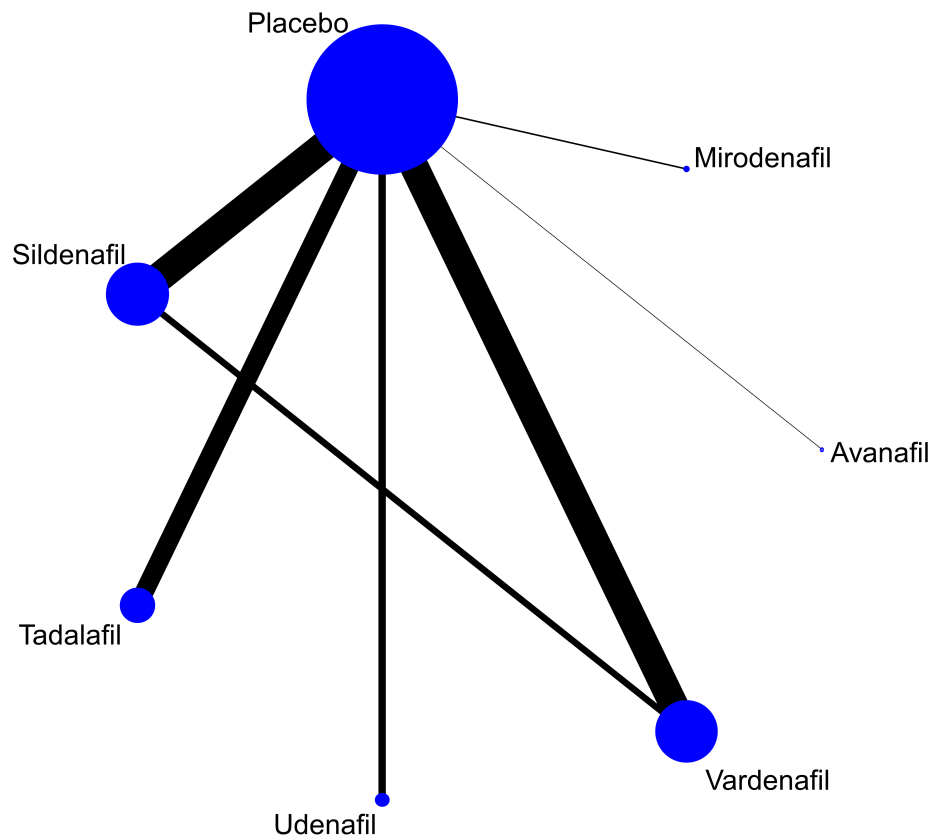
Edge's thickness and node's size is analogous to the number of direct treatment comparisons and participants, respectively.

Figure 2.5: Network geometry of oral phosphodiesterase type-5 inhibitors for erectile dysfunction.

In SSIFS, the design-by-treatment method for the specification of the inconsistency factors produces equivalent results with the Lu & Ades method, because the network includes

only one closed loop and two-arm studies; only in the comparison between Sildenafil and Vardenafil inconsistency factor can be added. Similarly, the Jackson's model will be not utilized for the evaluation of network consistency.

SSIFS indicates that the network is globally and locally inconsistent. More specifically, in the case where $\pi_{cons} = 0.50$ the posterior inclusion probability of the inconsistency factors is 0.86 (see Figure 2.6), larger than 0.50, indicating major local inconsistency in the closed loop between Sildanafil, placebo and Vardenafil. Also, because the network includes only one inconsistency factor, the posterior model probability of the consistent NMA model is calculated as $1 - 0.86 = 0.14$ (see Table 2.2) yielding to a posterior odds of

$$PO = \frac{\text{posterior model probability of the inconsistent NMA model}}{\text{posterior model probability of the consistent NMA model}} = \frac{0.86}{0.14} = 6.14$$

indicating substantial evidence in favour of the inconsistent NMA model. The evidence against consistency is stronger when $\pi_{cons} = 0.25$. In this case, the estimated posterior inclusion probability is 0.95 (see Figure 2.6), which results in a posterior odds of $1/0.05 = 20$ against consistency (see Table 2.2). Even in the case where $\pi_{cons} \sim Beta(157, 44)$, which assumes that the average probability of observing a consistent network is 0.78, the conclusions are similar (see Table 2.2 and Figure 2.6 for more details). Therefore, we can safety conclude that the network is globally inconsistent because of the loop inconsistency.
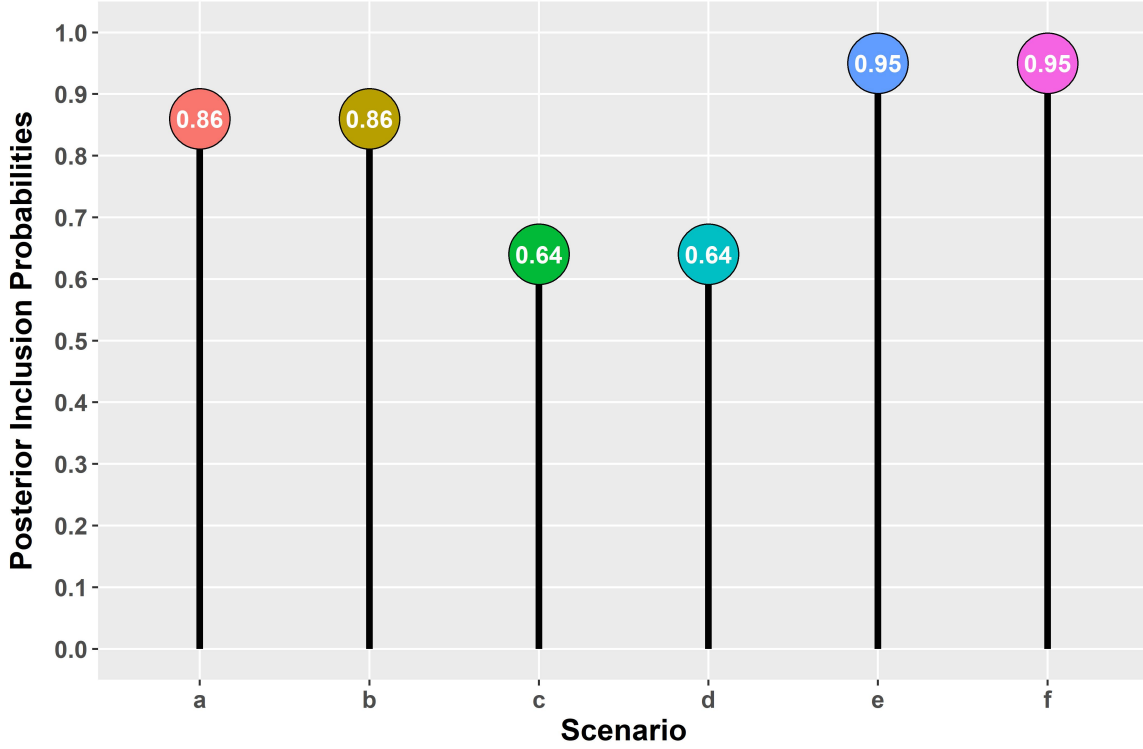
Scenario: a) $\pi_{cons} = 0.5$ & $\mathbf{R} = \mathbf{I}$, b) $\pi_{cons} = 0.5$ & $\mathbf{R} = g(\mathbf{Z}'\mathbf{Z})^{-1}\sigma^2$, c) $\pi_{cons} \sim$ $Beta(157, 44)$ & $\mathbf{R} = \mathbf{I}$, d) $\pi_{cons} \sim Beta(157, 44)$ & $\mathbf{R} = g(\mathbf{Z}'\mathbf{Z})^{-1}\sigma^2$, e) $\pi_{cons} = 0.25$ & $\mathbf{R} = \mathbf{I}$, f) $\pi_{cons} = 0.25$ & $\mathbf{R} = g(\mathbf{Z}'\mathbf{Z})^{-1}\sigma^2$.

Figure 2.6: Posterior inclusion probabilities of the inconsistency factor added in the comparison between Sildenafil and Vardenafil for the erectile dysfunction example.

The conclusions about network consistency are the same when consistency is evaluated by using the standard methods. More specifically, the SIDE method suggests that there is considerable discrepancy between direct and indirect evidence in the closed loop of the network (see Figure 2.7) and also, the Q-statistic under the assumption of a full design-by-treatment interaction random-effects model, indicates that the network is globally inconsistent ($Q = 8.44$, $df = 1$, $p - value = 0.0037$). It should be mentioned that the DIC criterion falsely suggests that the network is globally consistent since the DIC value for the consistent NMA model (DIC = 920.40) is much smaller compared to corresponding DIC value of the UME model (DIC = 940.50).

Comparison	Number of Studies	Direct Evidence	I2	Random Effects Model	OR	95%-CI
Sildenafil:Placebo						
Direct estimate	27	0.93	32%		9.25	[7.82; 10.95]
Indirect estimate					4.41	[2.37; 8.21]
Network estimate					8.80	[7.48; 10.35]
Vardenafil:Placebo						
Direct estimate	17	0.91	24%		8.56	[7.08; 10.34]
Indirect estimate					17.94	[9.70; 33.17]
Network estimate					9.12	[7.61; 10.93]
Sildenafil:Vardenafil						
Direct estimate	1	0.16			0.52	[0.29; 0.93]
Indirect estimate					1.08	[0.84; 1.39]
Network estimate					0.96	[0.76; 1.22]

0.1   0.5 1 2   10

Figure 2.7: Difference between direct and indirect evidence for the erectile dysfunction example, using the SIDE method.

2.4 Benefits of using Stochastic Search Inconsistency Factor Selection

SSIFS is a Bayesian two-step method which evaluates the consistency assumption based on the posterior inclusion probabilities and posterior odds. So far, global evaluation of network consistency in the Bayesian framework, is performed by comparing the consistent NMA model with inconsistent NMA model though the DIC (Daly et al., 2022). There are several limitations associated with DIC (Celeux et al., 2006; Lunn et al., 2012; Millar, 2009; Spiegelhalter et al., 2014), including concerns about discriminatory performance, invariance to reparameterization, inconsistency, lack of a proper predictive criterion and strong theoretical justification, and absence of a threshold for judging important differences among models. Moreover, DIC based on conditional likelihood is invalid for hierarchical modelling (Millar, 2009). Despite its limitations, DIC is commonly used in NMA for model selection. Note also that DIC can be used for evaluating network consistency by comparing each time two competing models; consistent NMA model versus a full/partial inconsistent NMA model. In contrast, SSIFS evaluates network consistency by comparing several inconsistent NMA models at each time without multiplicity issues, and also by incorporating past knowledge

concerning network consistency.

Another important advantage of SSIFS, is the ability to evaluate network consistency based on a value that is of practical significance. This is incredibly helpful, in networks where a pre-specified difference between direct and indirect evidence is considered important. Lastly, SSIFS evaluates the network globally by identifying the local sources of inconsistency, unlike the other methods in which consistency is evaluated either locally or globally.

Chapter 3

Multi-component (Complex) Interventions

In this chapter, strategies for handling multi-component (complex) interventions in NMA are presented. The chapter is structured as follows. Firstly, a brief description of multi-component (complex) intervention is presented. Secondly, standard methods for handling complex interventions are outlined. The chapter ends by proposing novel ways of visualizing NMA results with complex interventions. The applicability of the proposed tools is demonstrated through two real-life networks, and by using the R package *viscomp* (Seitidis et al., 2023b) which is published on CRAN.

3.1   Definition of Complex Intervention

Complex interventions in NMA refer to interventions that consist of multiple (interactive) components (Seitidis et al., 2022c; Craig et al., 2008). These interventions are often multi-faceted, and their effectiveness depends not only on the individual components, but also on how these components interact with each other. Possible components of an intervention could be the characteristics of the intervention, such as the frequency of the intervention, the dosage, the method used for delivering the intervention (e.g. delivered as a pill or as an injection), etc. Other potential components could be the type of encounter (e.g. clinical visits, support sessions or self-guided), delivery mode (e.g. face-to-face or remotely), time of communication (e.g. synchronous, asynchronous), type of recipient (e.g. group, individually), provider (e.g. peers, lay person), location (e.g. hospital, home, community-based care), or even other drugs that is used to treat a medical condition.

3.2 How to Deal with Complex Interventions

When complex interventions are present, the interest usually lies on addressing the three following research questions:

- Do the interventions work in general?

- Which intervention is the best in terms of efficacy?

- Which component(s) is/are working?

To address the first research question, interventions are usually lumped, and compared with a control group (e.g. placebo, usual care) by using the standard meta-analytic model (Nikolakopoulou et al., 2014b). In such cases, heterogeneity is often large since intervention differences among the studies are ignored. The second research question is commonly answered by using the NMA model (Seitidis et al., 2022a; Salanti, 2012; Mavridis et al., 2015), while the third through the Component Network Meta-Analysis (CNMA) model (Rücker et al., 2020; Welton et al., 2009; Tsokani et al., 2022; Veroniki et al., 2022).

Suppose we have a network that compares the effectiveness of 10 interventions (Figure 3.1). From the network plot we see that most interventions result as a combination of the components A, B, C, D and E. The rational of CNMA lies on the decomposition of complex interventions to estimate the effects of their components (Welton et al., 2009; Tsokani et al., 2022). Therefore, by using UC as reference, CNMA will estimate the effectiveness of the five components; assume that d_A, d_B, d_C, d_D and d_E refer to the relative effect of components A, B, C, D and E, respectively. The key assumption of CNMA implies that the effect of each intervention is derived by summing the relevant component effects. For example, the effect of the intervention A+B+C (denoted as d_{A+B+C}) is calculated as

$$d_{A+B+C} = d_A + d_B + d_C. \tag{3.1}$$

This assumption is often encountered in the literature as additivity assumption. Based on this assumption, we can obtain the intervention effect for any combination of the five components. Note also that we could obtain estimates for interventions (component combinations) that are not observed in the network (e.g. d_{B+C+E}).
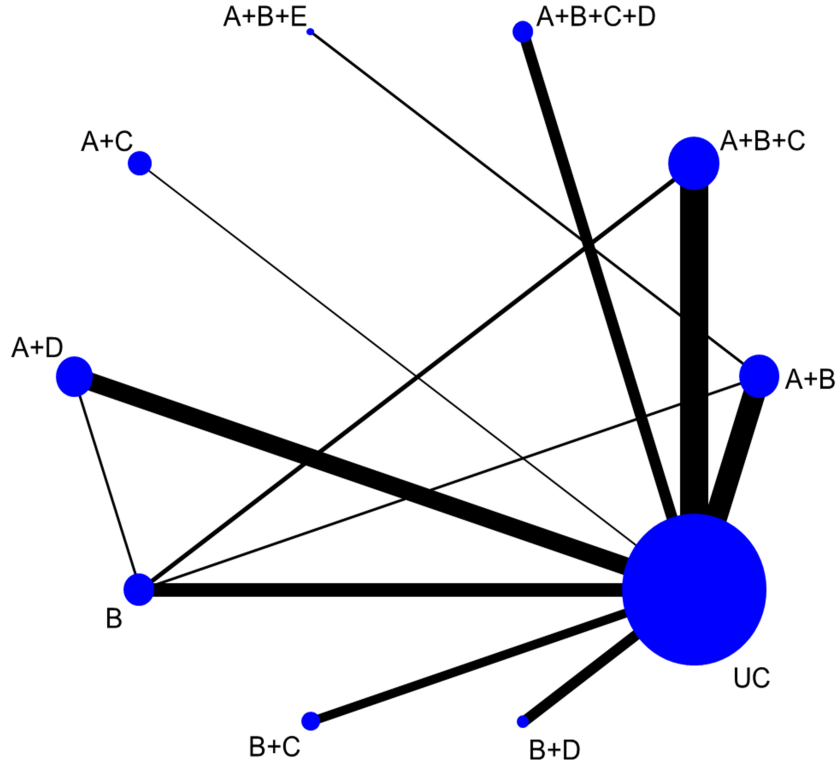
Figure 3.1: Network plot of complex interventions.

Additivity may be difficult to defend as it is a strong assumption, which sometimes appears to be implausible in practice. This is because it does not account for potential interactions between components; this issue could be mitigated by including interactions terms. For example, if components A and B interact within each other, an extra term d_{AB} should be added in the equation 3.1. This procedure can be challenging in practice, as there could be many different components interacting within each other. The hypothesis of additivity is commonly tested by using a chi-square test of the overall heterogeneity/inconsistency in the network between the NMA and CNMA models (Rücker et al., 2020). However, in disconnected networks it remains unclear how to evaluate additivity.

Defending additivity is more intense in cases where the network includes many multi-component interventions (nodes) with few head-to-head comparisons, because transitivity could be challenged due to the sparsity of the network. Furthermore, is quite possible to observe NMA effect estimates that are mainly driven by the study effect estimates or confounded by the study characteristics. In such case, it is also harder to identify the most efficacious interventions and component. Therefore, we have developed evidence synthesis

tools to handle the former issues, that could help on identifying the most promising interventions and components more easily.

3.3 Visualize NMA results to Explore the Behaviour of the Components

In this work, novel tools of visualizing the NMA results of complex interventions in order to explore the behaviour of the components and identify the most promising ones, are presented (Seitidis et al., 2022c).

3.3.1 Data

To demonstrate the usage of the proposed visualization tools two real-life networks are utilized. Both networks compare the effectiveness of self-management interventions in patients with type-II diabetes. The first measures the levels of glycated hemoglobin (HbA_{1c}) which is considered as harmful outcome, while the second measures the improvement of self-management behaviour (SMB) which is consider as beneficial outcome. The latter network was used only to demonstrate the components rank-heat plot (see section 3.3.2.8). Data were collected during the COMPAR-EU project, which aims to compare the effectiveness of self-management interventions in four high priority chronic conditions (type-II diabetes, obesity, heart failure, chronic obstructive pulmonary disease) for adults in Europe (Ballester et al., 2020). Interventions are masked as the findings are about to be submitted to an epidemiological journal and cannot be disclosed in advance (see Table 3.1 for components' abbreviations).

The network for the reduction of HbA_{1c} levels, consists of 461 studies (429 two-arm, 32 multi-arm), 97 nodes, 534 comparisons, 190 designs and 11 components. Interventions are constructed as a combination of the 11 components, yielding to a complex and sparse network since it includes many nodes but few head-to-head comparisons (see Figure 3.2). Moreover, most comparisons involve node A which referred to control group and it was used as reference node. It should be also noted that the NMA model suffers from high levels of heterogeneity ($\tau^2 = 0.09$, $I^2 = 86.50\%$). Figure 3.3 illustrates the geometry of the SMB network. Details about the NMA model for both networks are presented in Table 3.2.

Table 3.1: Components' abbreviations.

Component	Abbreviation
Usual Care	A
1st characteristic	B
Remote (mixed)	C
Usual Care Plus	D
2nd characteristic	E
3rd characteristic	F
Group	G
4th characteristic	H
Peers and lay persons	I
5th characteristic	J
6th characteristic	K

Table 3.2: Description of the networks

Network	Outcome Type	Effect Size	Number of Studies	Number of Patients	Number of Interventions	Number of Components
HbA_{1c}	Continuous	MD	461	79025*	97	11
SMB	Continuous	SMD	41	5585	30	11

*: 23 studies did not report sample size but were included in the NMA

Abbreviations: MD = Mean Difference, SMD = Standardized Mean Difference
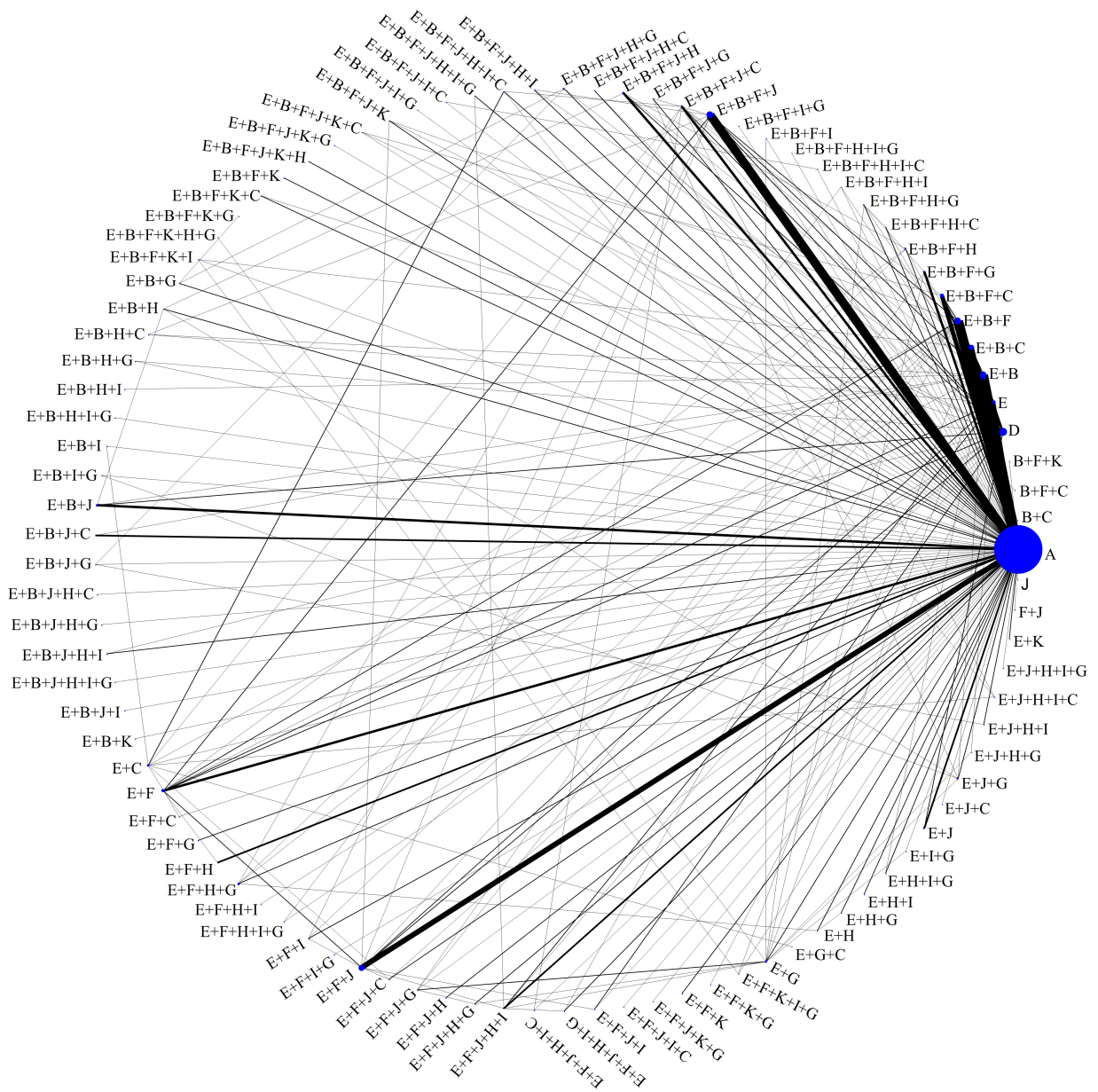
Nodes are proportional to the number of participants randomized to the nodes, while edges to the number of studies involving the corresponding nodes.

Figure 3.2: Network plot of HbA$_{1c}$ example.

Nodes are proportional to the number of participants randomized to the node, while edges to the number of studies involving the corresponding nodes.

Figure 3.3: Network plot of SMB improvement.

3.3.2 Proposed Visualization Tools

The implementation of the following visualization tools could be performed by using the R-package *viscomp* (Seitidis et al., 2023b). The package is available on CRAN and is also hosted on a GitHub repository (https://github.com/georgiosseitidis/viscomp). The highlight of the package is that it offers a friendly user toolkit aiming to identify the most promising components. The major input that is required to use the package is the NMA

model as obtained from the R-package *netmeta* (Rücker et al., 2022). The usage of the functions will be briefly presented in the next sections. More details about the installation and the usage of the package can be found at https://georgiosseitidis.github.io/viscomp/.

To illustrate the package's functions, we will assume that m and $m2$ denote the NMA model for the HbA_{1c} and SMB networks, respectively, both obtained from the R package *netmeta*.

3.3.2.1 Components Descriptive Analysis

Exploring the geometry of the network is essential, especially when dealing with large/sparse networks. Components descriptive analysis can assist in achieving this goal. By visualizing the component frequency in a coloured cross-table the researcher can easily identify the most frequent components (combinations). Elements on the diagonal pertain to components, while outside the diagonal, to the combinations of components. Moreover, the parentheses in the diagonal elements denote the proportion of the study arms in which the component from the corresponding row is present, while in the off-diagonal elements, it denotes the proportion of the study arm that includes both components out of those that include the corresponding component in the row. To identify more easily the most frequent component (combinations), each cell was coloured according to its absolute frequency; dark red colors suggest large percentages, while white colors percentages close to zero.

Figure 3.4 displays the components' cross-table for the HbA_{1c} example. The most frequent components are the E ($\frac{565}{957} = 59.04\%$), B ($\frac{371}{957} = 38.77\%$) and F ($\frac{352}{957} = 36.78\%$) while the most frequent combination is between components E and F (included in 349 study arms). Despite that this combination is the most frequent in the network, it does not have the largest absolute frequency. For example, the element that corresponds to row E and column F indicates that these components were observed in the $\frac{349}{565} = 61.77\%$ of the study arms that include component E, while the off-diagonal elements of column E, indicate that component E was almost always part of the intervention (percentages close to 100%) when combined with the rest components. Note that components A and D were not combined with any components since they refer to control groups.
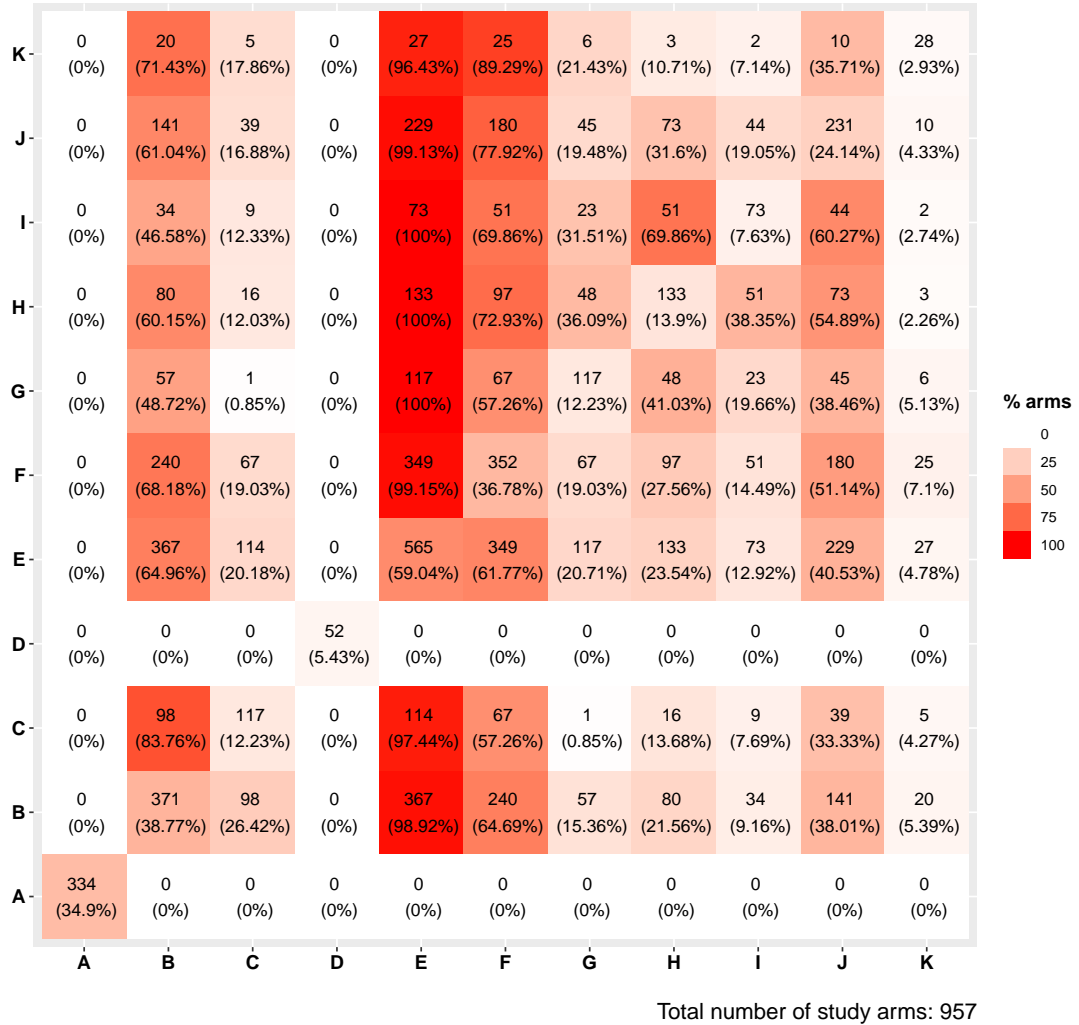
Figure 3.4: Components cross-table for the reduction of HbA$_{1c}$.

Components descriptive analysis can be performed through the function *compdesc*, by typing:

$$compdesc(model = m).$$

The function returns also as an output, a frequency table that provides some additional information regarding the frequency of the components. More specifically, it presents the number of studies in which the underlying component is included in all arms, in at least one arm and not included in any arm. This information could be used on comprehending the CNMA results. Table 3.3 presents the frequency table for the HbA$_{1c}$ example.

Table 3.3: Components' frequency for the HbA$_{1c}$ example.

Component	Number of arms	N_{all}	$N_{all}\%$	$N_{\geq 1}$	$N_{\geq 1}\%$	N_{none}	$N_{none}\%$	Ratio of $N_{all}/N_{\geq 1}$
A	334.00	0.00	0.00	334.00	0.72	127.00	0.28	0.00
B	371.00	28.00	0.06	329.00	0.71	132.00	0.29	0.09
C	117.00	10.00	0.02	101.00	0.22	360.00	0.78	0.10
D	52.00	0.00	0.00	52.00	0.11	409.00	0.89	0.00
E	565.00	72.00	0.16	458.00	0.99	3.00	0.01	0.16
F	352.00	19.00	0.04	309.00	0.67	152.00	0.33	0.06
G	117.00	15.00	0.03	101.00	0.22	360.00	0.78	0.15
H	133.00	5.00	0.01	123.00	0.27	338.00	0.73	0.04
I	73.00	1.00	0.00	67.00	0.15	394.00	0.85	0.01
J	231.00	6.00	0.01	215.00	0.47	246.00	0.53	0.03
K	28.00	0.00	0.00	27.00	0.06	434.00	0.94	0.00

Abbreviations: N_{all} = Number of studies in which the underlying component is included in all study arms, $N_{\geq 1}$ = Number of studies in which the underlying component is included in at least one study arm, N_{none} = Number of studies in which the underlying component is not included in any study arm.

3.3.2.2 Components Network Graph

Another visualization tool for exploring the frequency of the component combinations is the components network graph. Nodes refer to components, while edges refer to different component combinations that include the corresponding components. The thickness of the edges is proportional to the frequency of the combinations. Figure 3.5 visualizes the seven most common component combinations in the network. The thickest edges are the blue and the orange, which correspond to the combination E+B (included in 52 study arms) and E+B+F (included in 51 study arms), respectively.

Combination	# of arms
E+B	52
E+B+F	51
E+B+F+J	46
E+B+C	30
E+F+J	29
E+B+F+C	26
E	24
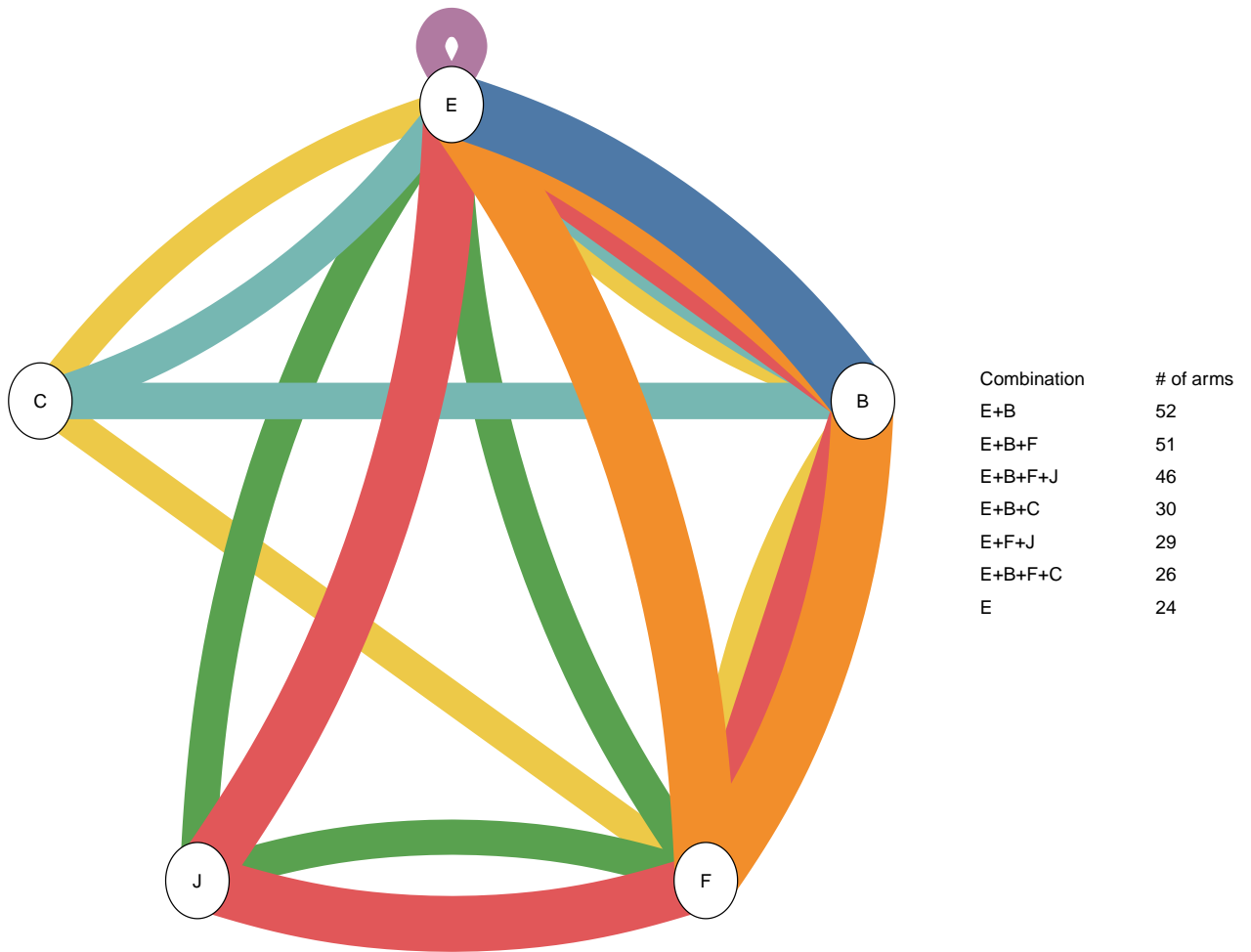
Figure 3.5: Components network graph for the HbA$_{1c}$ example, presenting the seven most frequent combinations of components.

Components network graph can be performed through the function *compGraph* by typing:

$$compGraph(model = m).$$

It should be noted that in this example nodes A and D were excluded since they were not combined with any component. This can be achieved by typing:

$$compGraph(model = m, excl = c("A", "D").$$

3.3.2.3 Leaving One Component Combination Out Scatter Plot

The leaving one component combination out scatter plot provide valuable insight on exploring whether the inclusion of a specific component (combination) affect the efficacy of the intervention. More specifically, it identifies pairs of interventions that vary by a certain component (or at least one component) and visualizes their NMA relative effect estimates in a scatter plot. Axis-x denotes the intervention's effect estimate when the component is part of the intervention, while axis-y when is not. Interventions' effect estimates are obtained by the NMA relative effect estimates. Dots close to the line of equality signify that the inclusion/exclusion of the component do not impact on the outcome. The interpretation of the direction depends on the nature of the outcome. In our example, which is considered as a harmful outcome, if the dots are located above the equality line, it indicates that the inclusion of the underlying component increases the efficacy of the intervention. In the case where dots are located below the equality line, it indicates that the efficacy of the intervention is hindered by the inclusion of the underlying component. The opposite holds for a beneficial outcome. According to Figure 3.6, we may conclude that the inclusion of component E may hamper the reduction of HbA_{1c} since most dots are below the equality line. For the rest components, it is unclear whether the inclusion of a component has a positive or negative impact on the intervention's efficacy.

The scatter plot can be also utilized as a visual method to assess additivity in CNMA. The principle of additivity suggest that the effect of an intervention is derived by summing its relevant components' effects. Hence, it is anticipated that the impact of inclusion or exclusion of a component would be the same in interventions that differentiate by this component. Therefore, a line that runs parallel to the line of equality on the scatter plot indicates visually the fulfilment of the additivity assumption. It is evident from our example that this is not the case, suggesting potential violations of additivity in CNMA.

The leave one component combination out scatter plot can be easily extended to component combinations and adjusted to use z-values instead of relative effects. In the latter case, additivity assumption cannot be evaluated. It can be constructed through the function *loccos*, which by default plots the pair of interventions that differ by one component. If we are interested on a certain component or combination of components, we have to set appropriately the argument *combination*. For example, if we are interested on the interventions that differ by the component B, we should type

$$loccos(model = m, combination = "B").$$
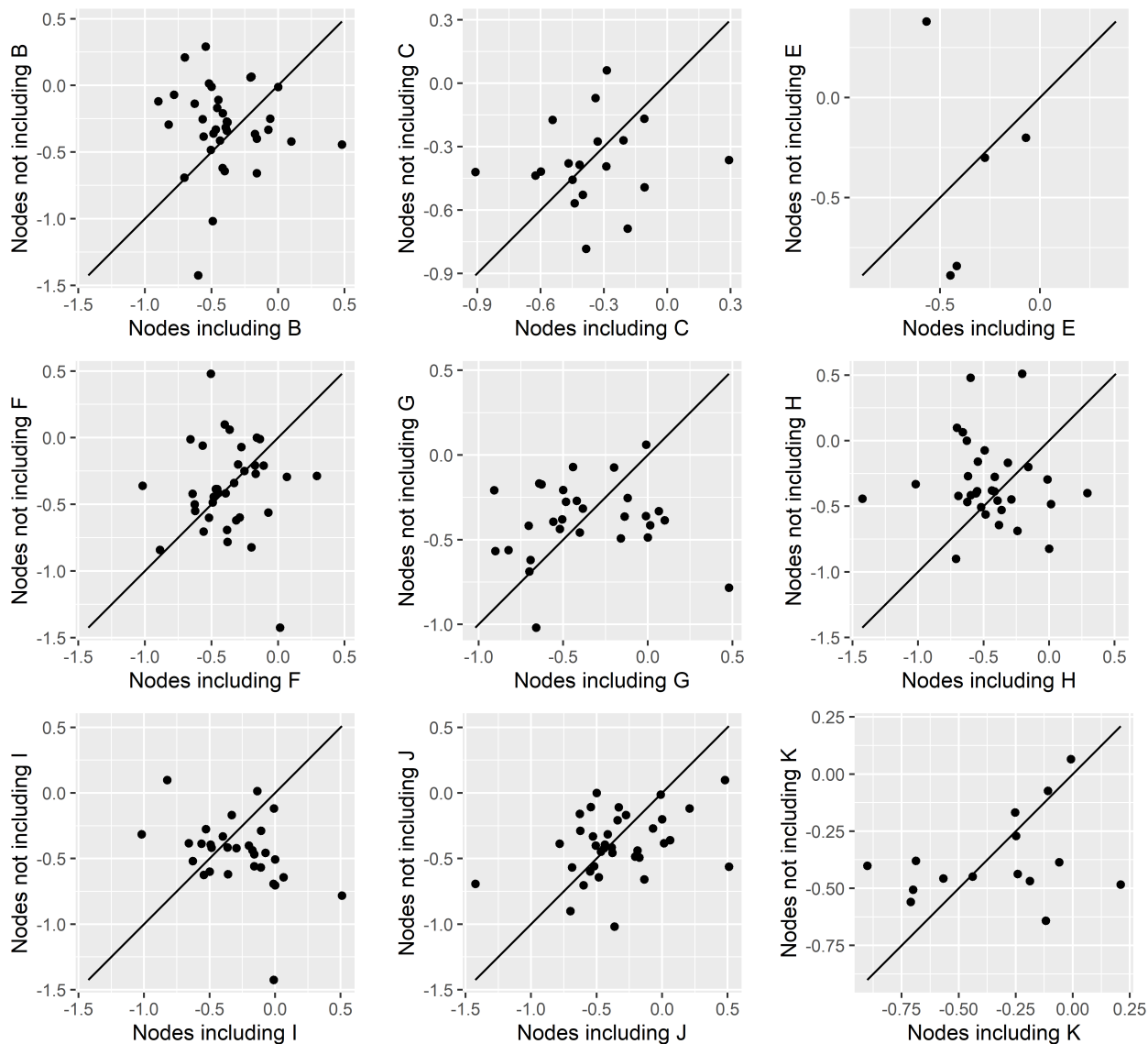
Figure 3.6: Leaving one component combination out scatter plot for each component for HbA$_{1c}$ example, using NMA relative effects.

3.3.2.4 Waterfall Plot

An alternative to the leave one component combination out scatter plot is the water-fall plot, which also visualizes the set of interventions that differ by a specific component (combination) but in a different manner. Axis-y displays the difference of the intervention effects with and without the component (combination). Bars close to zero signify that the inclusion or exclusion of the corresponding component (combination) has a trivial (or no)

impact on the outcome, while bars above or below zero that the inclusion/exclusion impacts on the intervention's effect. The direction of the impact (positive or negative) depends on the nature of the outcome (beneficial or harmful). For a beneficial outcome, a bar above zero denotes that the intervention's effect is larger when the component (combination) is included in the intervention. Hence, the inclusion of the extra component (combination) enhances intervention's efficacy. Conversely, for a harmful outcome, a bar below zero indicates that the intervention's effect is larger (in absolute values) when the extra component (combination) is included in the intervention. Consequently, a bar below zero signifies that the inclusion of the extra component (combination) increases efficacy. Waterfall plot can be also adjusted to use z-values instead of relative effects.

The waterfall plot can be constructed through the function *watercomp*. The rational of the function is the same with the function *loccos*. By typing

$$watercomp(model = m, combination = "K"),$$

we construct a waterfall plot for the interventions that differ by the component K (Figure 3.7).
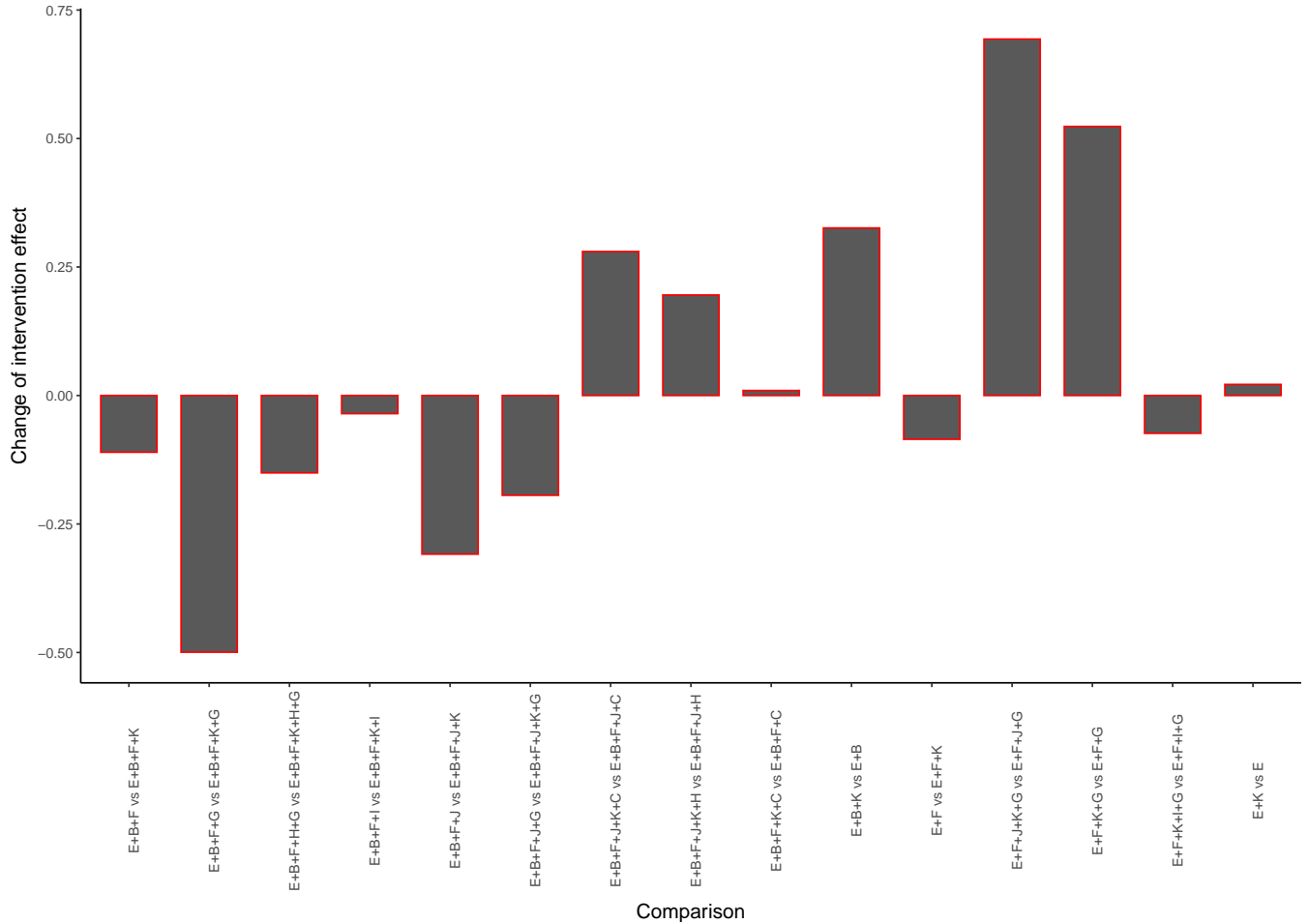
Figure 3.7: Waterfall plot for the comparisons that differ by K in the HbA$_{1c}$ example, using NMA relative effects.

3.3.2.5 Components Heat Plot

Another visualization tool that could help on identifying the most promising components is the components heat plot. By summarizing the intervention effects in a coloured square-table, the analyst can easily identify which component (combination) is the most efficacious. In addition, each cell summarizes the effect of the interventions that include the components that denoted by the corresponding row and column of the table. Combinations that were not observed in the network, are denoted in the heat plot with the letter 'X'. The interventions effects are summarized by using the median (default choice), or the mean instead. To identify more easily whether the underlying component (combination) has a positive or negative impact, the table is coloured according to the magnitude of the component's (combination) effect. Dark green color indicate a strong impact on the desired direction, while dark red

color to the opposite direction; direction of the impact is subtracted by the nature of the outcome (beneficial or harmful). Moreover, to increase the confidence around the summary estimates, the number of interventions in which the corresponding component (combination) was observed, is provided in the parenthesis of each cell, as well as the uncertainty around the summary estimates, which is reflected by the size of the grey box. Large grey boxes suggest considerable uncertainty around the summary estimate.

Component heat plot can be constructed through the function *heatcomp*. By typing

$$heatcomp(model = m)$$

the function produces a heat plot by using the NMA relative effect estimates and the median as a summary measure; z-values can be also utilized instead of relative effects through the argument *z_value*. The nature of the outcome is automatically subtracted from the NMA model m.

Figure 3.8 displays the components heat plot the reduction of HbA_{1c}. From the figure we can see that most of the elements are green, indicating that self-management interventions reduce the levels of HbA_{1c} in general. The most efficacious component combination is between components C and G. However, we should be cautious with the interpretation of this results since the summary estimate is based only in one NMA relative effect estimate. Lastly, we should be moderate confident about the results due to the uncertainty in the estimates, which is reflected by the medium size of the grey boxes.
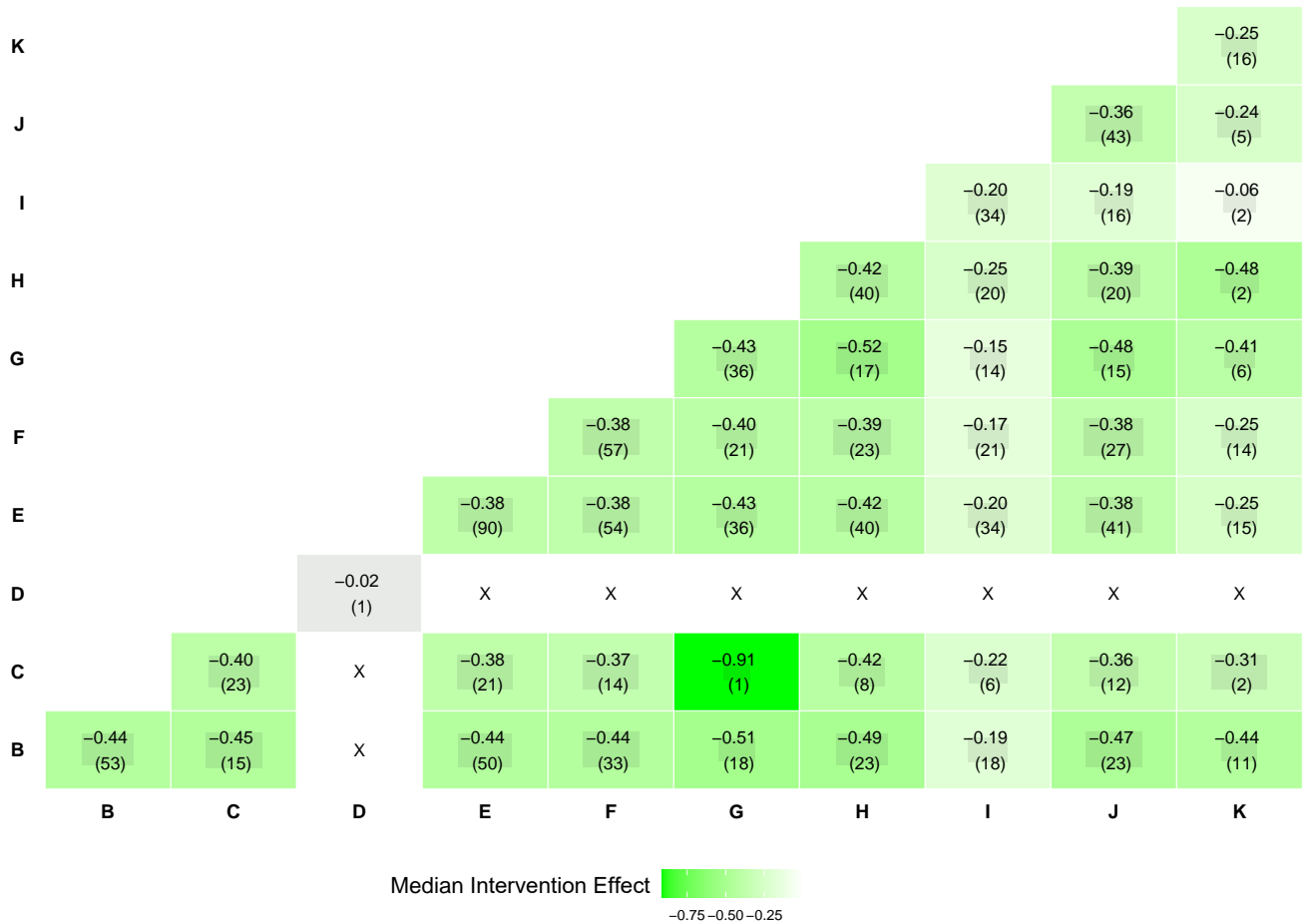
Figure 3.8: Components heat plot for the HbA$_{1c}$ example, using NMA relative effects and the median as summary measure.

	B	C	D	E	F	G	H	I	J	K
K										−0.25 (16)
J									−0.36 (43)	−0.24 (5)
I								−0.20 (34)	−0.19 (16)	−0.06 (2)
H							−0.42 (40)	−0.25 (20)	−0.39 (20)	−0.48 (2)
G						−0.43 (36)	−0.52 (17)	−0.15 (14)	−0.48 (15)	−0.41 (6)
F					−0.38 (57)	−0.40 (21)	−0.39 (23)	−0.17 (21)	−0.38 (27)	−0.25 (14)
E				−0.38 (90)	−0.38 (54)	−0.43 (36)	−0.42 (40)	−0.20 (34)	−0.38 (41)	−0.25 (15)
D			−0.02 (1)	X	X	X	X	X	X	X
C		−0.40 (23)	X	−0.38 (21)	−0.37 (14)	−0.91 (1)	−0.42 (8)	−0.22 (6)	−0.36 (12)	−0.31 (2)
B	−0.44 (53)	−0.45 (15)	X	−0.44 (50)	−0.44 (33)	−0.51 (18)	−0.49 (23)	−0.19 (18)	−0.47 (23)	−0.44 (11)

Median Intervention Effect

−0.75 −0.50 −0.25

3.3.2.6 Component Combination Violin Plot

Identification of the most efficacious components can be also performed by combining the kernel distribution and the boxplot for each component in a violin plot, based on the NMA relative effect estimates. Such a plot can be easily constructed through the function *specc* by typing

$$specc(model = m).$$

This approach can be easily extended to component combinations through the argument combination, or adjusted to use z-values instead or relative effects though the argument *z_value*.

Figure 3.9 visualizes the components violin plot for the HbA_{1c} example, which indicates that in general the components reduce the levels of HbA_{1c}, but it remains unclear which component is associated with largest effect due to overlap between the violins; some overlap is expected, because each component within interventions that comprise multiple components will produce the same relative effect estimate. It should be mentioned that the violin for the component D cannot be constructed since there is only one NMA estimate that contains D; D refers to a control group and was not combined with any other component (for more details see Figure 3.4).
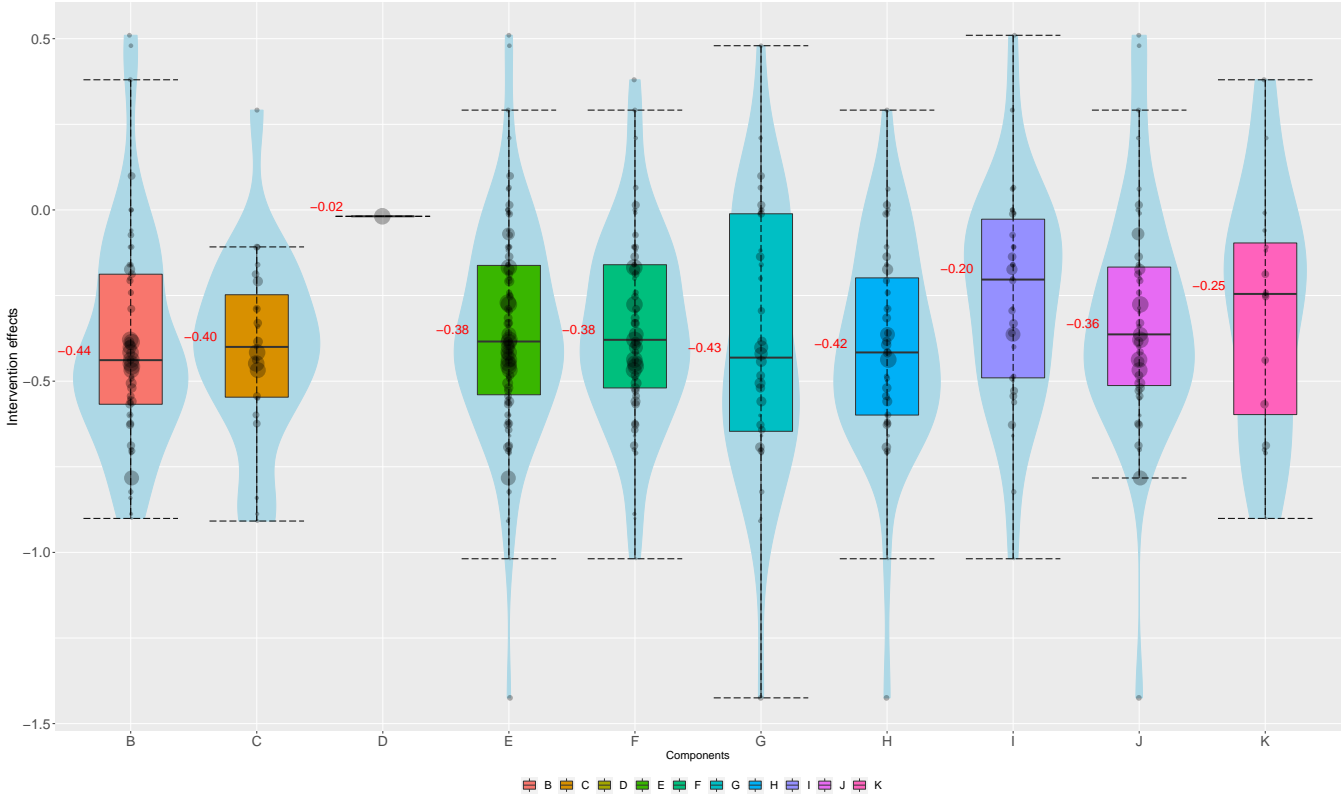
Dots are proportional to the precision of the NMA relative effect estimates.

Figure 3.9: Violin plots evaluating the effectiveness of the components used for the reduction of HbA_{1c}, using NMA relative effects.

By constructing the violins based on the number of components that are included in the interventions, it can be visually explored whether the number of components affects the

intervention's efficacy. This can be easily performed by typing

$$specc(model = m, \quad component\_number = TRUE).$$

In our example, there is a signal that as the number of components are increased, the efficacy is also increased at some point and then it starts decreasing (Figure 3.10). However, the results are not conclusive due to overlap between the violins.
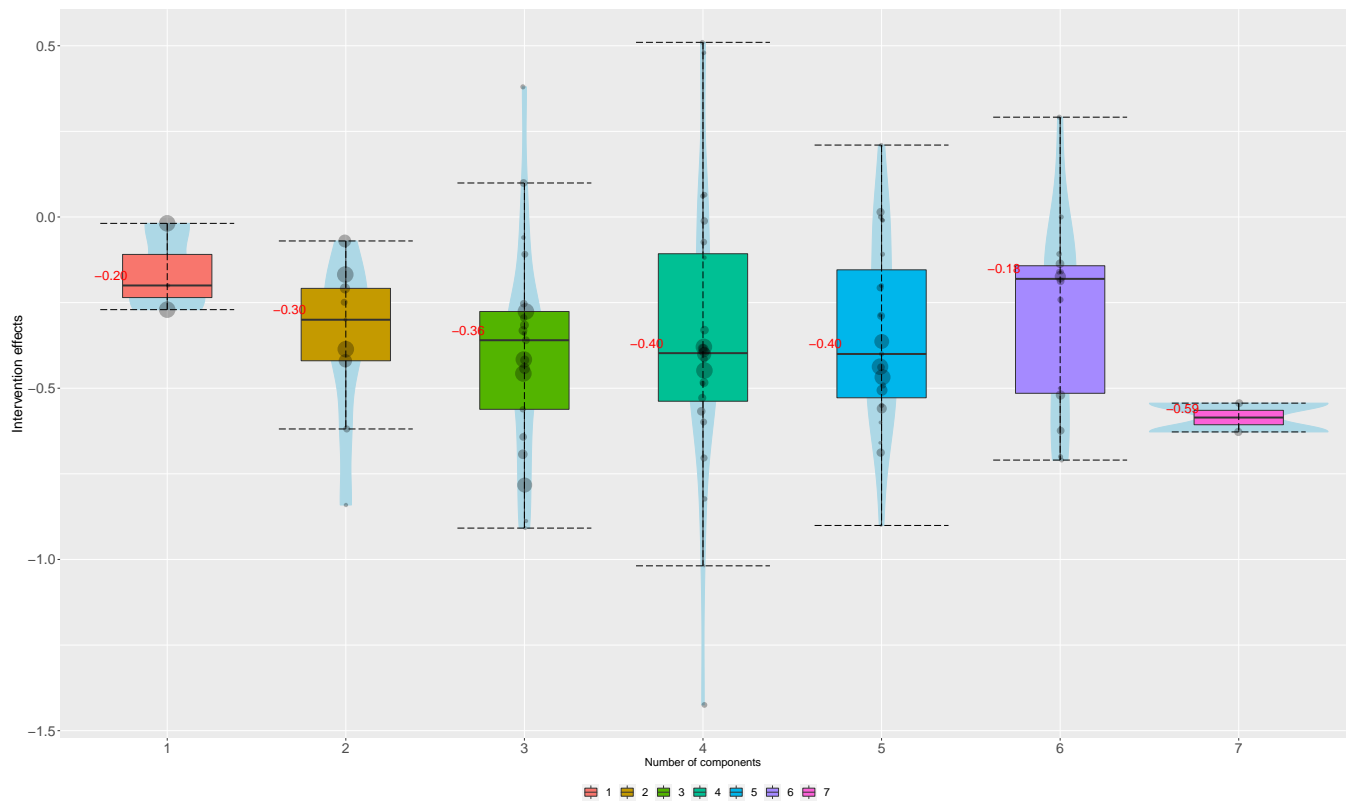
Dots are proportional to the precision of the NMA estimates.

Figure 3.10: Violin plot based on the number of components included in the intervention for the HbA$_{1c}$ example, using the NMA relative effect estimates.

3.3.2.7 Components Density Plot

Another visualization tool that could help on exploring the behaviour of the components is the components density plot, which compares the density derived from the interventions that

include the corresponding component (combination) with the density from the interventions that do not include this component (combination). Therefore, we can explore whether a specific component (combination) is associated with larger effects. Such a density plot can be easily constructed through the function *denscomp*. By typing

$$denscomp(model = m, combination = "B")$$

we construct a density plot based on the interventions that 1) include component B and 2) do not include component B. The function uses the NMA relative effect estimates, but it can be also adjusted to use the z-values instead. The density plot for each component for the HbA_{1c} example is illustrated in Figure 3.11, which indicates that interventions without component I or K exhibit larger efficacy compared to those that include them.
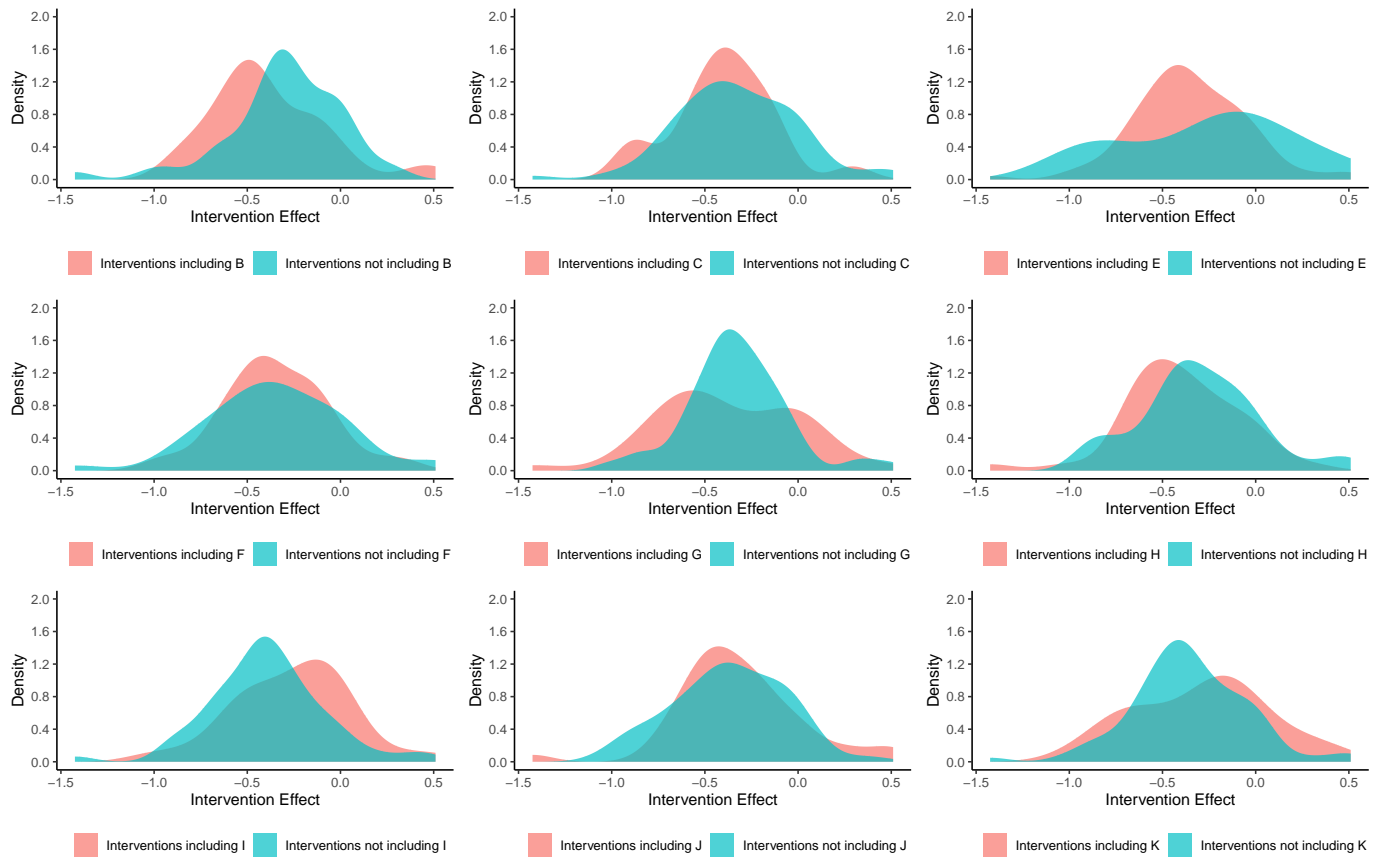

Figure 3.11: Components density plot for the reduction of HbA_{1c}, using NMA relative effect estimates.

3.3.2.8 Components Rank Heat Plot

When multiple outcomes are included in the analysis (for example various medical conditions), the interest usually lies on identifying the intervention that overall performs best. In such cases, the Hasse diagram (Carlsen and Bruggemann, 2014; Rücker and Schwarzer, 2017) or the rank-heat plot could be used on achieving this goal. The former represent a finite partially ordered set by drawing curves between interventions (Seitidis et al., 2022c) while the latter presents the p-scores for each intervention across outcomes (Veroniki et al., 2016). However, there is lack of a visualization tool that could provide insight on identifying which component overall work well across multiple outcomes. Therefore, we have extended the rational of the rank-heat plot to components.

In the components rank-heat plot, components are denoted via rads and outcomes via circles. Each sector denotes the median (or mean) p-score of the interventions that include the corresponding component within the underlying outcome. To identify the "best" components more easily, the colouring of the sectors is proportional to the components' p-scores. Dark green colors indicate p-scores close to 100%, while yellow and dark red colors indicate p-scores close to 50% and 0%, respectively. In the case where a sector is uncoloured, it denotes that the corresponding component in the underlying outcome was not observed.

Components rank-heat plot can be constructed through the function *rankheatplot*. By typing

$$rankheatplot(model = list(m, m2))$$

we construct the component rank-heat plot for the HbA_{1c} and the SMB networks, using the median for calculating the components p-scores (see Figure 3.12). From the figure we can confidently conclude that the worst components are the components A and D as expected since both components refer to control groups. Regarding which component is the best, component B appears to be overall the best. However, the results are not conclusive as there are several components with similar p-score estimates.
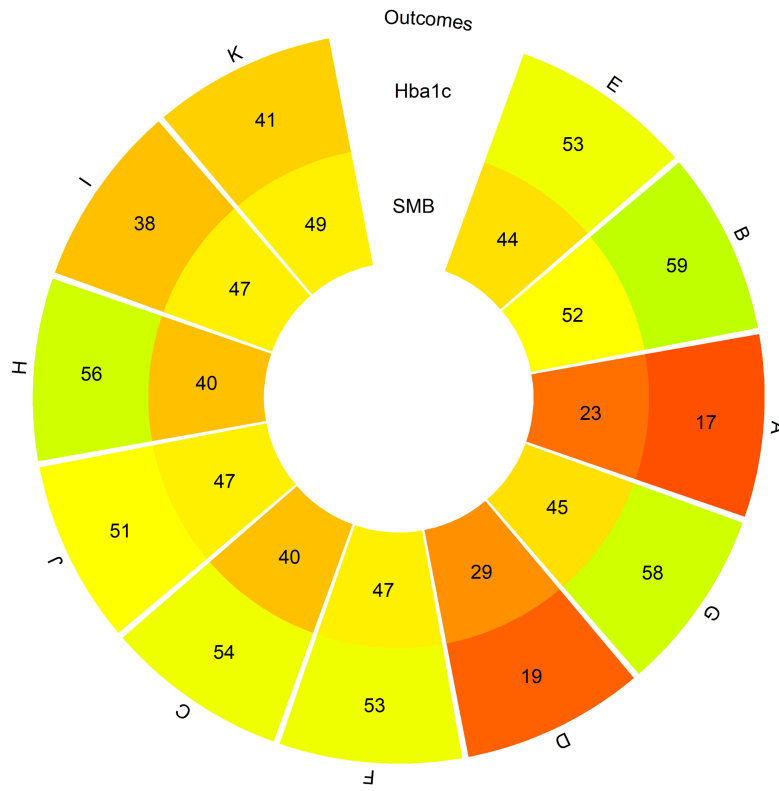
Figure 3.12: Components rank-heat plot for the reduction of HbA$_{1c}$ and the SMB improvement, using the median for calculating the components p-scores.

Chapter 4

Software

In this chapter, we provide a comprehensive guide on the implementation of two essential R packages, namely *ssifs* and *viscomp*, which are published on CRAN. They offer a wide array of functionality to identify inconsistencies in NMA and to enhance the results of NMA when it includes multi-component interventions. By exploring the documentation and leveraging the capabilities of these packages, users can enhance their analysis and gain valuable insights.

4.1   Package *ssifs*

The package "Stochastic Search Inconsistency Factor Selection" was published on CRAN on the 5th of December, 2023. The scope of the package is to evaluate the consistency assumption of NMA both locally and globally, by using variable selection techniques in the Bayesian framework.

A vignette titled "An Introduction to the Stochastic Search Inconsistency Factor Selection method using the *ssifs* package" is provided alongside the R package. This vignette serves as a comprehensive guide on how to utilize the *ssifs* package and its functionalities. It offers a detailed explanation of the SSIFS method and demonstrates its practical application through a published NMA example. You can access the vignette at https://georgiosseitidis.github.io/ssifs/articles/An_introduction_to_ssifs.html.

4.1.1   Function *ssifs()*

The function *ssifs* ("Stochastic Search Inconsistency Factor Selection") identifies inconsistencies in the NMA model by using the SSIFS method, as described in chapter 2. The function uses the IV-NMA model, adjusted to include inconsistency factors (see equation

1.4), and assumes common heterogeneity between different comparisons and no correlation between different studies.

4.1.1.1 Arguments

As input, the function requires the specification of the network's contrasts along with the corresponding effect size estimates and standard errors. More specifically, the following arguments must be specified:

- TE: Treatment effect (e.g. mean difference, log odds ratio)

- $seTE$: Standard error of treatment effect

- $treat1$: Treatment 1

- $treat2$: Treatment 2

- $studlab$: Study label

- ref: Reference treatment.

It is important to note that when a network is disconnected, the function automatically uses the largest sub-network and ignores the comparisons that belong to smaller sub-networks. Also, when multi-arm studies are present, the user must specify all feasible comparisons.

By default, the function uses the Design-by-Treatment method to specify the inconsistency factors of the network (matrix $\mathbf{Z}$). Argument *method* controls the method used for specifying matrix $\mathbf{Z}$. There are three options available to the user:

- "*LuAdes*": Lu & Ades model (see section 1.4.1.1)

- "*DBT*": Design-by-Treatment model, which is the default choice (see section 1.4.1.2)

- "*Jackson*": Jackson's model (see section 1.4.1.3).

Inconsistency factors can be independent or dependent. The default choice of

$$zellner = TRUE$$

assumes dependency between inconsistency factors and describes this dependency through a Zellner g-prior (see 2.2.2.1). In the case where

$$zellner = FALSE,$$

inconsistency factors are assumed independent, and as a result, the inclusion or exclusion of one does not affect the inclusion or exclusion of another.

Regarding the prior inclusion probabilities of the inconsistency factors, they are specified with respect to the probability of observing a consistent network, denoted as π_{cons} (see section 2.2.2.2 for more details). By default, the argument $rpcons$ is set to $TRUE$, which implies that

$$\pi_{cons} \sim Beta(157, 44).$$

This prior choice is quite informative and supports the consistency hypothesis a-priori by 78% on average. By setting

$$rpcons = FALSE,$$

we can express our prior ignorance concerning this probability, by assuming that the probability of network consistency is fixed and equal to 50%. This probability can also be modified through the argument $pcons$. For example, if we set

$$rpcons = FALSE \quad \text{and} \quad pcons = 0.25,$$

the probability of network consistency would be 25%. In this case, we support a-priori network inconsistency by 75%.

SSIFS assumes that the effects of the inconsistency factors are described from a mixture of two normal distributions (see section 2.2.2). The mixing of the method is controlled by the arguments c and psi. These arguments should be defined in such a manner as to ensure a distribution similar to the distribution of Figure 2.1. Argument c reflects the prior probability of including inconsistency factors in the model with trivial effects, while argument psi denotes the standard deviations of the inconsistency factors (ψ_ℓ) when they are not included in the model. By default, the function assumes that $c = 3$ and specifies ψ_ℓ based on a pilot MCMC run of the full NMA model as the standard deviation of the inconsistency factors. The user can also specify the parameters ψ_ℓ manually, by setting the argument psi equal to a vector of positive numbers. For example, if the network includes five inconsistency factors, then by setting

$$psi = c(0.7, \ 0.5, \ 0.3, \ 0.6, \ 0.2)$$

the user assumes that

$$\psi_1 = 0.7, \ \psi_2 = 0.5, \ \psi_3 = 0.3, \ \psi_4 = 0.6 \ \text{and} \ \psi_5 = 0.2.$$

Furthermore, by properly setting arguments c and psi, the user can evaluate network consistency based on a value that is considered of practical significance (see section 2.2.3). For example, if we set $c = 10$ and $psi = 0.09$, the user assumes that a difference of 0.2 between direct and indirect evidence, is of practical significance. Thus, inconsistency factors with effect larger in absolute values than 0.2, are included in the model.

Lastly, through the arguments M, B, n_thin and n_chains the user can control the iterations of the MCMC run, the burn-in period, the thinning interval, and the number of chains, respectively. The default choice is to employ a burn-in period of 10000 iterations and a total of 50000 iterations, using 2 unthinned chains to monitor convergence. These options are also provided for the pilot MCMC run of the NMA model (arguments M_pilot, B_pilot, n_thin_pilot and n_chains_pilot).

4.1.1.2 Output

The function returns as output the following objects:

- *Posterior_inclusion_probabilities*

- *Posterior_Odds*

- *Summary*

- *Bayes_Factor*

- *Z_matrix*

- *MCMC_run*

- *disconnected_studies*

- *n_subnetworks*

- *subnetworks*.

Object *Posterior_inclusion_probabilities* is a data.frame that contains the posterior inclusion probabilities of the inconsistency factors. Columns *Comparison* and *Design* denote in which comparisons inconsistency factors were added, while column *PIP* denotes the corresponding posterior inclusion probabilities; values above 0.5 indicate that the corresponding inconsistency factors should be included in the NMA model and signifies inconsistency. Moreover, the coefficients of the inconsistency factors along with the corresponding 95% credible intervals are reflected by the columns *b*, *b.lb*, and *b.ub*, respectively. Inconsistency factors'

coefficients can also be found in the object *Summary*, which contains the coefficients along with the 95% credible intervals for the functional comparisons, heterogeneity parameter τ^2 and parameter π_{cons}. Raw results of the MCMC run are provided in the object *MCMC_run*; the user can check the convergence of the MCMC run using the classical diagnostics.

Object *Posterior_Odds* is a data.frame that contains the posterior odds and the posterior model probabilities. Column *IFs* denotes the observed NMA models, while columns *Freq* and $f(m|y)$ denote the frequency of the corresponding model and the posterior model probabilities, respectively. Column *PO_IFCONS* denotes the posterior odds of the consistent NMA model (denoted as *NO IFs*) versus the inconsistent NMA model of the corresponding row of column *IFs*. A value above one favours the consistency hypothesis, while below one signifies inconsistency. In the latter case, the network is globally inconsistent, and the sources of inconsistency are the comparisons in which inconsistency factors were added. Note also that in the case where $method = "DBT"$ or $method = "Jackson"$, the inconsistency factors for multi-arm designs are denoted as

$$treatment\ comparison\_Design.$$

For example, if an inconsistency factor is added in the comparisons between treatments B and C of the ABC design, the inconsistency factor's name will be $B; C\_ABC$.

Object *Bayes_Factor* contains a positive number which denotes the Bayes Factor of the comparison between the consistent NMA model and the rest inconsistent NMA models; observed inconsistent NMA models are grouped and their posterior model probabilities are summed. A value above one indicates that the network is globally consistent, while below one suggests that the network is globally inconsistent. In the latter case, the sources of inconsistency should be specified according to posterior odds and posterior inclusion probabilities.

Furthermore, the user can identify the comparisons where inconsistency factors have been added via the object *Z_matrix*. In addition, it is a data.frame where the comparisons between the treatments and the study labels are presented in the first three columns, while the rest columns refer to matrix **Z**. Values of 1 or -1 indicate that an inconsistency factor was added, whereas values of 0 indicate that was not.

Lastly, the function provides information concerning the sub-networks that were excluded before SSIFS was applied. More specifically, the objects *disconnected_studies*, *n_subnetworks*, *subnetworks* denote the studies that belong to the excluded sub-networks, the sub-networks, and the number of sub-networks, respectively.

4.1.2 Function *spike.slab()*

The function *spike.slab* ("Inconsistency Factors' Spike and Slab") visualizes the spike and slab for each inconsistency factor. Thus, the user can check the model's mixing ability. Generally, good mixing indicates that the spike has a high density for values close to zero, while the slab is flat.

4.1.2.1 Arguments

As input, the function requires an object of class *ssifs*.

4.1.2.2 Output

It returns objects of class *ggplot*. The number of objects equals the number of inconsistency factors that were added to the NMA model.

4.1.3 Data

The package includes two datasets of two published NMAs that refer to smoking cessation (Dias et al., 2013) and the effectiveness of brief alcohol interventions for preventing hazardous drinking (Seitidis et al., 2022a).

4.2 Package *viscomp*

The package "Visualize Multi-Component Interventions in Network Meta-Analysis" was published on CRAN on the 16th of January, 2023. The scope of the package is to visualize the results of NMA in order to explore the behaviour of the components and identify valuable patterns between them.

A vignette titled "An introduction to network meta-analysis using the *viscomp* package" is provided alongside the R package. This vignette serves as a comprehensive guide on how to utilize the *viscomp* package and its functionalities. It offers a detailed explanation of the novel visualization tools and demonstrates their practical application through illustrative examples. You can access the vignette at https://georgiosseitidis.github.io/viscomp/articles/viscomp.html.

The package's functions, except for the function *rankheatplot*, require as input an object of class *netmeta* (argument *model*). By default, they identify the components of the network by using the symbol " + " as a separator. The spacing between the components does not

affect the results. For example, if the network includes four interventions, namely "A+B", "A +B", "C+ B", "A + B+C", the function will identify the elements "A", "B", and "C", as components of the network. The separator symbol can be modified through the argument *sep*.

Furthermore, the functions utilize by default the NMA relative effects which are obtained from the random-effects NMA model. By setting the argument

$$random = FALSE,$$

the common-effects NMA model is utilized instead, while by setting

$$z\_value = TRUE,$$

z-values are used instead of relative effects.

4.2.1 Function *compGraph()*

The function *compGraph* ("Components Network Graph") displays the frequency of the components' combinations found in the network.

4.2.1.1 Arguments

By default, the function displays the five most frequent component combinations. The user can change this number through the argument *mostF*. Additionally, the user can exclude specific combinations from the plot through the argument *excl* (character vector). This feature is particularly useful when there are components that pertain to control groups and are not combined with other components.

The plot title can be customized using the argument *title*, which is set by default as "Most frequent combinations of components". The legend size can be adjusted using the argument *size_legend*. The legend can also be excluded by setting the argument

$$print\_legend = TRUE.$$

4.2.1.2 Output

An object of class *qgraph*.

4.2.2   Function *compdesc()*

The function *compdesc* ("Components descriptive analysis") performs a descriptive analysis concerning the frequency of the components.

4.2.2.1   Arguments

By default, the function produces a heat matrix using percentages instead of fraction values; percentages are rounded up to two decimal places (argument $digits = 2$). By setting the argument

$$percentage = FALSE,$$

fraction values are displayed instead of percentages. Also, the user has the option to not plot the heat matrix by setting the argument

$$heatmap = FALSE.$$

4.2.2.2   Output

The function returns as output the following objects:

- *crosstable*

- *frequency*

- *heatmap* (if $heatmap = TRUE$).

Object *crosstable* is a cross-table that contains the frequency of two-by-two component combinations. Each cell refers to the number of study arms that include the corresponding component (combination). Diagonal elements refer to components, while off-diagonal elements refer to the combinations of components.

Object *frequency* is a data.frame that displays the frequency of the components. More specifically, it includes the following columns:

- *Component*: it refers to the components of the network.

- *Frequency*: it refers to the number of arms that include the corresponding component.

- *A*: it refers to the number of studies in which the corresponding component was observed in all arms.

- *A_percent*: it refers to the percentage of studies in which the corresponding component was observed in all arms.

- *B*: it refers to the number of studies in which the corresponding component was observed in at least one arm.

- *B_percent*: it refers to the percentage of studies in which the corresponding component was observed in at least one arm.

- *C*: it refers to the number of studies in which the corresponding component was not included in any arm.

- *C_percent*: it refers to the percentage of studies in which the corresponding component was not included in any arm.

- *A.B*: it refers to the ratio of A/B.

Object *heatmap* is an object of class *ggplot*. It refers to the heat matrix which visualizes the object *crosstable*. The percentage of arms that contain the corresponding component (combination) is also displayed in the parenthesis of each cell. Diagonal elements refer to the percentage of arms that contain the corresponding component, while non-diagonal elements refer to the percentage of arms that include the corresponding component combination out of those that include the component of the corresponding row. Moreover, each cell is coloured according to relative frequency, so the user could identify the most frequent combinations more easily; dark red indicates large percentages.

4.2.3   Function *heatcomp()*

The function *heatcomp* ("Components Heat Plot") creates a heat plot that visualizes the effect of the components (combinations) based on the NMA results.

4.2.3.1   Arguments

By default, the function summarises the NMA relative effect estimates (intervention versus reference) of the interventions that include the corresponding component (combination) by using the median as a summary measure. Also, the number of nodes in which the underlying component (combination) was observed, is printed in the parenthesis of each cell; by setting the argument $freq = FALSE$ frequencies are not printed. Moreover, the uncertainty around the relative effect estimates is reflected by the size of the grey boxes. The larger the size of the box, the more imprecise the summary estimate. In the case where z-values are used, uncertainty around the estimates is not displayed.

The user can use the mean as a summary measure instead of the median by setting

$$median = FALSE,$$

and modify the legend title through the argument *legend_name*.

4.2.3.2 Output

An object of class *ggplot*.

4.2.4 Function *loccos()*

The function *loccos* ("Leaving One Component Combination Out Scatter plot") identifies the set of interventions that differ by a specific set of components. Therefore, the user can explore whether the inclusion or exclusion of a specific component (combination) impacts on the outcome.

4.2.4.1 Arguments

By default, the function identifies sets of interventions that differ by one component. Axis x displays the intervention effect when the component (combination) is included in the intervention, while axis y when is not. Through the argument *combination* we can specify the component (combination) of interest. For example, by setting

$$combination = "A"$$

the function identifies the set of interventions that differ by the component A, and by setting

$$combination = "A + B",$$

identifies the set of interventions that differ by the components "A" and "B". Dots close to the $y = x$ line indicate no impact on the outcome. Dots above the $y = x$ line indicate for a beneficial outcome that the inclusion of the component (combination) hampers efficacy, while for a harmful outcome, that increases the intervention effect. Dots below the $y = x$ line suggest for a beneficial outcome that the inclusion of the component (combination) increases efficacy, while for a harmful outcome, that decreases the intervention effect.

The function can also be used to evaluate visually the additivity assumption. Additivity implies that the intervention effect is obtained as the sum of the relative components' effects. Therefore, the inclusion or exclusion of a component (combination) should have the same

impact on the outcome. This is expressed by a line parallel to the $y = x$ line. In the case where z-values are used, additivity cannot be evaluated.

Along with the scatter plot, a histogram displaying the interventions' effects is printed on each axis. By setting $histogram = FALSE$ the histogram is not printed, while through the argument $histogram.color$ the histogram color can be modified.

4.2.4.2 Output

An object of class *ggplot*.

4.2.5 Function *specc()*

The function *specc* ("Specific Component Combination violin plot") visualizes the components (combinations) effects through violin plots.

4.2.5.1 Arguments

By default, the function visualizes the distribution of interventions' effects for each component. The number of violins is equal to the number of components. Argument *combination* is a character vector and defines the components (combinations) of interest. For example, if

$$combination = c("A", "A + B", "A + B + C"),$$

the function will create three violins based on the:

- interventions that include component "A"

- interventions that include components "A" and "B"

- interventions that include components "A", "B" and "C".

The function can also be modified to construct violins based on the number of components. By setting

$$components\_number = TRUE$$

the function visualizes the interventions' effects based on the number of components that are included in the interventions. For example, if the number of components included in the interventions ranges between one and four, the function will create four violins based on the:

- interventions that include one component

- interventions that include two components

- interventions that include three components

- interventions that include four components.

Through the argument *group* which is a character vector, the user can group the violins. As input it requires integer numbers (e.g. 4 or "4"), range values (e.g. "1-3") and "XX+" values (e.g. "4+"). For example, if

$$components\_number = TRUE \quad \text{and} \quad group = c("1 - 2", 2, "3 + ")$$

the function will construct three violins based on the interventions that include:

- one or two components

- two components

- more than three components.

Moreover, through the arguments *fill_violin*, *color_violin*, *adj_violin*, and *width_violin*, the user can modify the fill color of the violin, the color of the violin, the width of the violin and its adjustment, respectively. Also, by default, an additional boxplot is printed along with the violin. The user can also modify the width of the boxplot and its errorbar through the arguments *width_boxplot* and *errorbar_type*, respectively. The boxplot can be excluded from the plot by setting the argument

$$boxplot = FALSE.$$

Median values are printed by default for each boxplot, and they can be excluded by setting *values* = *FALSE*. Lastly, dots are displayed by default, visualizing interventions' effects/z-values. The shape and position of the dots can be modified through the arguments *jitter_shape* and *jitter_position*, respectively. The dots can be excluded from the plot by setting

$$dots = FALSE.$$

It should be noted that in the case where intervention effects are used, by setting

$$prop\_size = TRUE$$

the size of the dots is proportional to the precision of the estimates.

4.2.5.2 Output

An object of class *ggplot*.

4.2.6 Function *watercomp()*

The function *watercomp* ("Waterfall plot") creates a waterfall plot for interventions that differ by one (specific) component combination. A bar above or below the $y = 0$ line indicates an impact on the outcome. Depending on the nature of the outcome (beneficial or harmful), the direction of the impact (positive or negative) will differ.

4.2.6.1 Arguments

By default, the function identifies sets of interventions that differ by one component. The user can identify sets of interventions that differ by a specific component combination through the arguments *combination*, which behaves the same as in the function *loccos* (for more details see 4.2.4).

4.2.6.2 Output

An object of class *ggplot*.

4.2.7 Function *denscomp()*

The function visualizes the density of the components (combinations).

4.2.7.1 Arguments

By default, the function visualizes the NMA relative effects based on:

- the interventions that include the underlying component (combination)

- the interventions that do not include the underlying component (combination).

The component (combination) of interest is defined by the argument *combination*. In the case where the length of the argument *combination* is larger than one, then the number of densities is equal to its length. For example, if

$$combination = c("A", "A + B", "A + B + C")$$

the function will produce three densities based on:

- the interventions that include component "A"

- the interventions that include components "A" and "B"

- the interventions that include components "A", "B" and "C".

Also, if the argument $violin = TRUE$, the function will use violins instead of density plots. The function can be also adjusted to utilize z-values instead of relative effects.

4.2.7.2 Output

An object of class *ggplot*.

4.2.8 Function *rankheatplot()*

The function *rankheatplot* ("Components Rank Heat Plot") visualizes the components p-scores for multiple outcomes.

4.2.8.1 Arguments

Argument *model* must be a list containing multiple *netmeta* objects. The length of argument *model* must be at least two. By default, the function summarizes the p-scores from the interventions that include the corresponding component, using the median as a summary measure. It can also be modified to use the mean instead, by setting

$$median = FALSE.$$

Moreover, by default, p-score estimates are obtained from the random-effects NMA model. By setting

$$random = FALSE,$$

p-score estimates are obtained for each outcome from the common-effects model. If it is required to use the common-effects model for some outcomes and not for all, the argument *random* must be a logical vector; the length of the vector must be equal to the length of the argument *model*. For example, if three outcomes are included in the argument *model* and the user needs to utilize the common-effects model for the first two outcomes, it must set

$$random = c(FALSE, FALSE, TRUE).$$

In this case, the names of the outcomes by default would be Outcome 1, Outcome 2 and Outcome 3. The outcome's names can be modified through the argument *outcomeNames*; it is a character vector with length equal to the number of outcomes. Lastly, the font size of the components, p-score estimates, and outcomes' names can be modified through the arguments *cex_components*, *cex_values*, *cex_outcomes*, respectively.

4.2.8.2 Output

An invisible rank-heat plot.

4.2.9 Data

The package includes a fictional network of multi-component interventions that compares the effectiveness of interventions on major adverse cardiovascular events.

Chapter 5

Conclusions and Further Research

Evaluating network consistency and handling multi-component (complex) interventions are two prevalent challenges encountered in the practice of NMA. In this work we center on these two research topics, with an emphasis on providing potential solutions to address them.

5.1   Conclusions

Transitivity is a fundamental assumption of NMA which is evaluated statistically through the consistency assumption. The proper evaluation of consistency is vital as the validity of the NMA findings are primarily affected by the satisfaction of the consistency assumption. So far, consistency was mainly evaluated in the Bayesian framework through DIC, which has several limitations and does not address adequately the consistency assumption. In this work, we transform the inconsistency detection into a variable selection problem, by incorporating the SSVS method into the NMA model. Therefore, by treating the inconsistency factors as variables in a generalized linear model, we can evaluate the network globally by identifying the local sources of inconsistency. Advantages of the method include the ability to incorporate historical evidence, evaluation of consistency based on a value that is of practical significance, avoid using improper information criteria such as the DIC, and simultaneously test network consistency both globally and locally. Also, it can be used to evaluate consistency in Individual Participant Data (IPD) NMA by using the two-stage approach. The proposed method, namely SSIFS, was evaluated in two published networks and demonstrated good performance on testing consistency. Implementation of SSIFS could be performed though the R-package *ssifs* which was developed by the authors of this work.

In this work inconsistency factors (matrix $\mathbf{Z}$) were specified using either the Lu & Ades model (section 1.4.1.1), or the design-by-treatment interaction model (section 1.4.1.2), or

the Jackson's model (section 1.4.1.3). It is important to note that the matrix **Z** can also be manually specified when necessary, for example, when inconsistency factors need to be included in certain comparisons or when there are inconsistencies in the network. Thus, in an inconsistent network where the sources of inconsistency are known, by adjusting matrix **Z** appropriately, we can obtain NMA estimates adjusted for inconsistency. However, this feature of SSIFS cannot be applied directly to the former methods used (Lu & Ades model, design-by-treatment interaction model, Jackson's model) for the specification of matrix **Z**, since it is quite likely that the NMA estimates will be biased due to the inconsistency factors terms that are placed to the consistent designs/loops.

Another important issue in NMA is the handling of multi-component (complex) interventions. So far, the identification of the best components is commonly assessed through the CNMA model. In sparse networks with many nodes and few direct evidence, the results of CNMA could be biased since additivity may be implausible. Moreover, NMA and CNMA estimates could be primarily driven by the study estimates or study characteristics. To address these issues and strengthen the NMA, we provide a series of visualization tools that could be used on exploring the geometry of the network and explore the behaviour of the components. Moreover, components (combinations) that are associated with a large impact on the outcome can be visually detected. A drawback of these tools is that it can be applied only in connected networks. Implementation of the visualization tools could be performed though the R-package *viscomp* which was developed by the authors of this work.

Interpretation of the results concerning the effectiveness of the components (combinations) should be treated with caution to avoid false conclusions. It is vital to firstly understand the geometry of the network and then proceed on exploring the components' effectiveness. Suppose that there are two ineffective components in the network, which are observed in a few multi-component interventions that have large effects and always appear together. In such cases, the leave one component out combination scatter plot and the waterfall plot will not illustrate these components due to missing comparators. Also, the component heatplot and the violin plot will suggest that these two components have a considerable impact on the outcome. Despite the large effectiveness of these two components, the results are misleading. This is because the large intervention effects are due to intervention complexity rather than to the effectiveness of the two underlying components. It should be noted, however, that if we had first examined the network geometry through the components crosstable, we would have observed that these components were always observed together and never apart. Lastly, it should be mentioned that both NMA and CNMA should be conducted when complex interventions are present.

5.2 Further Research

In terms of further research, there exist various intriguing research topics that could be relevant to our work. One possibility is to broaden the scope of the SSIFS by exploring alternative variable selection techniques such as the Gibbs Variable Selection (GVS) (Dellaportas et al., 2000, 2002), or the Bayesian Least Absolute Shrinkage and Selection Operator (LASSO) (O'hara and Sillanpää, 2009). Also, as an alternative to the spike and slab approach, the horseshoe prior could be utilized (Carvalho et al., 2010), which mimics the spike and slab approach and is particularly useful when there are many potential predictors but few data (Piironen and Vehtari, 2017). Hence, it may be an efficient alternative for identifying inconsistencies in networks that include many inconsistency factors but few studies (sparse networks). Another potential direction is to tailor SSIFS to the network meta-regression framework or to multivariate NMA, or modify SSIFS to evaluate consistency in networks that incorporate both RCTs and non-RCTs. Finally, it is necessary to conduct a simulation to confirm the method's overall good performance.

Further research also includes the evaluation of consistency using posterior predictive checks (Gelman et al., 1996), and the development of visualization tools that could provide valuable insight concerning NMA and CNMA.

Appendix A

Tuning based on assuming a-priori the minimum value of inconsistency that is of practical significance.

Tuning parameters ψ_ℓ and c in order inconsistency factor to be included in the NMA model when $\mid b_\ell \mid$ is larger than a threshold. In the case where the prior correlation matrix is $\mathbf{R} = \mathbf{I_p}$ we have that:

$$f(b_\ell \mid \gamma_\ell = 0) = f(b_\ell \mid \gamma_\ell = 1) \Leftrightarrow$$

$$\frac{1}{\psi_\ell\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{b_\ell}{\psi_\ell}\right)^2} = \frac{1}{c\psi_\ell\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{b_\ell}{c\psi_\ell}\right)^2} \Leftrightarrow$$

$$-\frac{1}{2}\left(\frac{b_\ell}{\psi_\ell}\right)^2 = -\log c - \frac{1}{2}\left(\frac{b_\ell}{c\psi_\ell}\right)^2 \Leftrightarrow$$

$$c^2 b_\ell^2 = 2c^2\psi_\ell^2 \log c + b_\ell^2 \Leftrightarrow$$

$$b_\ell^2\left(c^2 - 1\right) = 2c^2\psi_\ell^2 \log c \Rightarrow$$

$$b_\ell^2 = \frac{2c^2\psi_\ell^2 \log c}{c^2 - 1} \Rightarrow$$

$$b_\ell^2 = \psi_\ell^2 \xi(c), \quad \xi(c) = \frac{2c^2 \log c}{c^2 - 1}.$$

Inconsistency factors with coefficient $\mid b_\ell \mid > \psi_\ell\sqrt{\xi(c)}$ will have larger density to be included in the NMA model. If we assume that the elements of $\mathbf{b}$ are correlated then,

$$f(\mathbf{b} \mid \boldsymbol{\gamma} = \mathbf{0}) = f(\mathbf{b} \mid \boldsymbol{\gamma} = \mathbf{1}) \Leftrightarrow$$

$$\mid 2\pi\mathbf{D_{\gamma=0}RD_{\gamma=0}}\mid^{-1/2}e^{-\frac{1}{2}\mathbf{b}'(\mathbf{D_{\gamma=0}RD_{\gamma=0}})^{-1}\mathbf{b}} = \mid 2\pi\mathbf{D_{\gamma=1}RD_{\gamma=1}}\mid^{-1/2}e^{-\frac{1}{2}\mathbf{b}'(\mathbf{D_{\gamma=1}RD_{\gamma=1}})^{-1}\mathbf{b}}.$$

We have

$$\mathbf{D_{\gamma=1}} = \begin{pmatrix} c\psi_1 & & \\ & \ddots & \\ & & c\psi_p \end{pmatrix} = c\begin{pmatrix} \psi_1 & & \\ & \ddots & \\ & & \psi_p \end{pmatrix} = c\mathbf{D_{\gamma=0}}.$$

Thus,

$$\mid \mathbf{D_{\gamma=0}RD_{\gamma=0}}\mid^{-1/2}e^{-\frac{1}{2}\mathbf{b}'(\mathbf{D_{\gamma=0}RD_{\gamma=0}})^{-1}\mathbf{b}} = (c^2)^{-\frac{p}{2}}\mid \mathbf{D_{\gamma=0}RD_{\gamma=0}}\mid^{-1/2}e^{-\frac{1}{2c^2}\mathbf{b}'(\mathbf{D_{\gamma=0}RD_{\gamma=0}})^{-1}\mathbf{b}} \Leftrightarrow$$

$$e^{-\frac{1}{2}\mathbf{b}'(\mathbf{D_{\gamma=0}RD_{\gamma=0}})^{-1}\mathbf{b}} = c^{-p}e^{-\frac{1}{2c^2}\mathbf{b}'(\mathbf{D_{\gamma=0}RD_{\gamma=0}})^{-1}\mathbf{b}} \Leftrightarrow$$

$$-\frac{1}{2}\mathbf{b}'(\mathbf{D_{\gamma=0}RD_{\gamma=0}})^{-1}\mathbf{b} = -p\log c - \frac{1}{2c^2}\mathbf{b}'(\mathbf{D_{\gamma=0}RD_{\gamma=0}})^{-1}\mathbf{b} \Leftrightarrow$$

$$\mathbf{b}'(\mathbf{D_{\gamma=0}RD_{\gamma=0}})^{-1}\mathbf{b}\left(\frac{1}{c^2} - 1\right) = -2p\log c \Rightarrow$$

$$\mathbf{b}'(\mathbf{D_{\gamma=0}RD_{\gamma=0}})^{-1}\mathbf{b} = \frac{2pc^2\log c}{c^2 - 1}.$$

Zellner g-prior is assumed for the prior correlation matrix $\mathbf{R}$

$$\mathbf{R} = g\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\sigma^2$$

Hence,

$$\mathbf{b}'\left(\mathbf{D_{\gamma=0}}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{D_{\gamma=0}}\right)^{-1}\mathbf{b} = \frac{2gp\sigma^2c^2\log c}{c^2 - 1}.$$

Thus, it is considered unsignificant when

$$\mathbf{b}'\left(\mathbf{D_{\gamma=0}}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{D_{\gamma=0}}\right)^{-1}\mathbf{b} = gp\sigma^2\xi(c).$$

Appendix B

Tables

Table A1: Smoking cessation example. Posterior inclusion probabilities (γ_ℓ) and inconsistency factors' coefficients (b_ℓ) with the corresponding 95% credible interval (95% Cr.I) are calculated using the design-by-treatment, the Lu & Ades, and the Jackson's model for the specification of the inconsistency factors, and by assuming either that the inconsistency factors are independent ($\mathbf{R} = \mathbf{I_p}$) or not ($\mathbf{R} = g(\mathbf{Z'Z})^{-1}\sigma^2$). Probability to have a consistent network assumed either to be 0.5, or 0.25, or $\pi_{cons} \sim Beta(157, 44)$.

| Comparison (Design) | $\pi_{cons} = 0.5$ | | | | $\pi_{cons} \sim Beta(157,44)$ | | | | $\pi_{cons} = 0.25$ | | | |
| | $\mathbf{R} = \mathbf{I_p}$ | | $\mathbf{R} = g(\mathbf{Z'Z})^{-1}\sigma^2$ | | $\mathbf{R} = \mathbf{I_p}$ | | $\mathbf{R} = g(\mathbf{Z'Z})^{-1}\sigma^2$ | | $\mathbf{R} = \mathbf{I_p}$ | | $\mathbf{R} = g(\mathbf{Z'Z})^{-1}\sigma^2$ | |
	$\hat{\gamma}_\ell$	$\hat{b}_\ell$ [95% Cr.I]	$\hat{\gamma}_\ell$	$\hat{b}_\ell$ [95% Cr.I]	$\hat{\gamma}_\ell$	$\hat{b}_\ell$ [95% Cr.I]	$\hat{\gamma}_\ell$	$\hat{b}_\ell$ [95% Cr.I]	$\hat{\gamma}_\ell$	$\hat{b}_\ell$ [95% Cr.I]	$\hat{\gamma}_\ell$	$\hat{b}_\ell$ [95% Cr.I]
Design by Treatment												
BC (BC)	0.08	0.01 [-0.21, 0.56]	0.08	0.00 [-0.31, 0.56]	0.03	0.01 [-0.19, 0.22]	0.03	0.00 [-0.33, 0.57]	0.14	0.01 [-0.32, 0.90]	0.16	0.00 [-0.31, 0.59]
BD (BD)	0.08	-0.01 [-0.67, 0.21]	0.08	0.00 [-0.58, 0.31]	0.03	-0.01 [-0.23, 0.18]	0.03	0.00 [-0.59, 0.34]	0.16	-0.01 [-1.01, 0.33]	0.16	0.00 [-0.64, 0.31]
CD (CD)	0.06	0.00 [-0.27, 0.23]	0.08	0.00 [-0.31, 0.30]	0.02	0.00 [-0.20, 0.19]	0.03	0.00 [-0.32, 0.33]	0.13	0.00 [-0.57, 0.47]	0.16	0.00 [-0.33, 0.32]
AC (ACD)	0.08	-0.02 [-0.78, 0.18]	0.08	0.00 [-0.65, 0.24]	0.03	-0.02 [-0.26, 0.17]	0.03	0.00 [-0.64, 0.26]	0.16	-0.02 [-1.06, 0.22]	0.16	0.00 [-0.70, 0.24]
AD (ACD)	0.12	0.03 [-0.18, 1.18]	0.09	0.00 [-0.21, 0.91]	0.04	0.02 [-0.17, 0.58]	0.03	0.00 [-0.23, 0.84]	0.21	0.04 [-0.20, 1.43]	0.18	0.00 [-0.21, 1.05]
CD (BCD)	0.06	0.00 [-0.30, 0.22]	0.08	0.00 [-0.42, 0.38]	0.02	0.00 [-0.20, 0.19]	0.03	0.00 [-0.44, 0.41]	0.13	0.00 [-0.61, 0.42]	0.15	0.00 [-0.43, 0.39]
BD (BCD)	0.07	0.01 [-0.21, 0.43]	0.08	0.00 [-0.32, 0.51]	0.02	0.01 [-0.19, 0.21]	0.03	0.00 [-0.33, 0.55]	0.13	0.01 [-0.35, 0.75]	0.15	0.00 [-0.32, 0.55]
Lu & Ades												
BC (-)	0.15	0.02 [-0.26, 0.84]	0.16	0.00 [-0.37, 0.78]	0.06	0.01 [-0.19, 0.43]	0.06	0.00 [-0.38, 0.79]	0.29	0.03 [-0.47, 1.05]	0.31	0.00 [-0.38, 0.80]
BD (-)	0.16	-0.01 [-0.79, 0.48]	0.16	0.00 [-0.62, 0.84]	0.06	-0.01 [-0.32, 0.21]	0.06	0.00 [-0.59, 0.87]	0.30	-0.01 [-0.99, 0.85]	0.32	0.00 [-0.67, 0.90]
CD (-)	0.21	0.03 [-0.22, 1.28]	0.19	0.00 [-0.26, 1.38]	0.08	0.02 [-0.18, 0.85]	0.07	0.00 [-0.27, 1.31]	0.37	0.05 [-0.38, 1.51]	0.35	0.00 [-0.28, 1.53]
Jackson												
AC (AC)	0.04	0.00 [-0.20, 0.21]	0.06	0.00 [-0.12, 0.13]	0.01	0.00 [-0.18, 0.19]	0.02	0.00 [-0.12, 0.13]	0.08	0.00 [-0.23, 0.29]	0.12	0.00 [-0.13, 0.14]
AB (AB)	0.04	0.00 [-0.20, 0.23]	0.06	0.00 [-0.22, 0.28]	0.02	0.00 [-0.22, 0.28]	0.02	0.00 [-0.23, 0.29]	0.09	0.01 [-0.24, 0.46]	0.11	0.00 [-0.22, 0.29]
AD (AD)	0.08	-0.01 [-1.02, 0.20]	0.07	0.00 [-1.03, 0.28]	0.03	-0.01 [-0.24, 0.18]	0.03	0.00 [-0.80, 0.29]	0.15	-0.02 [-1.43, 0.24]	0.14	0.00 [-1.31, 0.26]
BC (BC)	0.05	0.01 [-0.20, 0.33]	0.05	0.00 [-0.31, 0.54]	0.02	0.01 [-0.18, 0.21]	0.02	0.00 [-0.32, 0.54]	0.10	0.01 [-0.25, 0.75]	0.11	0.00 [-0.30, 0.55]
BD (BD)	0.06	-0.01 [-0.46, 0.19]	0.06	0.00 [-0.57, 0.31]	0.02	-0.01 [-0.21, 0.18]	0.02	0.00 [-0.21, 0.18]	0.11	-0.01 [-0.87, 0.23]	0.11	0.00 [-0.58, 0.29]
CD (CD)	0.04	0.00 [-0.22, 0.21]	0.06	0.00 [-0.31, 0.30]	0.02	0.00 [-0.19, 0.19]	0.02	0.00 [-0.32, 0.31]	0.09	0.00 [-0.40, 0.29]	0.11	0.00 [-0.30, 0.28]
AC (ACD)	0.06	-0.02 [-0.60, 0.18]	0.06	0.00 [-0.62, 0.25]	0.02	-0.01 [-0.23, 0.17]	0.02	0.00 [-0.61, 0.26]	0.12	-0.02 [-0.95, 0.20]	0.12	0.00 [-0.64, 0.23]
AD (ACD)	0.08	0.02 [-0.18, 1.00]	0.07	0.00 [-0.22, 0.84]	0.03	0.02 [-0.17, 0.27]	0.02	0.00 [-0.23, 0.75]	0.16	0.03 [-0.19, 1.29]	0.13	0.00 [-0.21, 0.93]
CD (BCD)	0.05	0.00 [-0.23, 0.20]	0.05	0.00 [-0.42, 0.38]	0.02	0.00 [-0.20, 0.19]	0.02	0.00 [-0.43, 0.39]	0.09	0.00 [-0.46, 0.28]	0.11	0.00 [-0.40, 0.35]
BD (BCD)	0.05	0.01 [-0.20, 0.26]	0.05	0.00 [-0.33, 0.49]	0.02	0.00 [-0.18, 0.20]	0.02	0.00 [-0.35 0.51]	0.10	0.01 [-0.25, 0.61]	0.11	0.00 [-0.30, 0.49]

Additional supplementary tables are available at https://github.com/georgiosseitidis/Setidis-G.-Thesis/blob/main/Supplementary%20Tables.pdf.

Bibliography

Ballester, M., Orrego, C., Heijmans, M., Alonso-Coello, P., Versteegh, M. M., Mavridis, D., Groene, O., Immonen, K., Wagner, C., Canelo-Aybar, C., et al. (2020). Comparing the effectiveness and cost-effectiveness of self-management interventions in four high-priority chronic conditions in europe (compar-eu): a research protocol. *BMJ open*, 10(1):e034680.

Bartlett, M. S. (1957). A comment on dv lindley's statistical paradox. *Biometrika*, 44(3-4):533–534.

Booth, A., Clarke, M., Dooley, G., Ghersi, D., Moher, D., Petticrew, M., and Stewart, L. (2012). The nuts and bolts of prospero: an international prospective register of systematic reviews. *Systematic reviews*, 1(1):1–9.

Borenstein, M., Hedges, L. V., Higgins, J. P., and Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons.

Brown, P. J., Vannucci, M., and Fearn, T. (1998). Multivariate bayesian variable selection and prediction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(3):627–641.

Bucher, H. C., Guyatt, G. H., Griffith, L. E., and Walter, S. D. (1997). The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *Journal of Clinical Epidemiology*, 50(6):683–691.

Caldwell, D. M., Ades, A. E., and Higgins, J. P. T. (2005). Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ*, 331(7521):897–900.

Carlsen, L. and Bruggemann, R. (2014). Partial order methodology: a valuable tool in chemometrics. *Journal of Chemometrics*, 28(4):226–234.

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480.

Celeux, G., Forbes, F., Robert, C. P., and Titterington, D. M. (2006). Deviance information criteria for missing data models.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., and Petticrew, M. (2008). Developing and evaluating complex interventions: the new medical research council guidance. *BMJ*, 337.

Daly, C., Downing, B. C., and Welton, N. J. (2022). A practical guide to inconsistency checks in bayesian network meta-analysis.

Deeks, J. J., Higgins, J. P., Altman, D. G., and Group, C. S. M. (2019). Analysing data and undertaking meta-analyses. *Cochrane handbook for systematic reviews of interventions*, pages 241–284.

Dellaportas, P., Forster, J. J., and Ntzoufras, I. (2000). Bayesian variable selection using the gibbs sampler. *Biostatistics-Basel-*, 5:273–286.

Dellaportas, P., Forster, J. J., and Ntzoufras, I. (2002). On bayesian model and variable selection using mcmc. *Statistics and computing*, 12(1):27–36.

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

Dias, S., Welton, N. J., Sutton, A. J., Caldwell, D. M., Lu, G., and Ades, A. (2013). Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. *Medical Decision Making*, 33(5):641–656.

Donegan, S., Williamson, P., Gamble, C., and Tudur-Smith, C. (2010). Indirect comparisons: A review of reporting and methodological quality. *PLOS ONE*, 5(11):1–11.

Druyts, E., Thorlund, K., Humphreys, S., Lion, M., Cooper, C. L., and Mills, E. J. (2013). Interpreting discordant indirect and multiple treatment comparison meta-analyses: an evaluation of direct acting antivirals for chronic hepatitis c infection. *Clinical epidemiology*, 5:173.

Dunson, D. B. et al. (2006). Efficient bayesian model averaging in factor analysis. *Biostatistics Branch, National Institute of Environmental Health Sciences, North Carolina, USA*.

Efthimiou, O., Rücker, G., Schwarzer, G., Higgins, J. P., Egger, M., and Salanti, G. (2019). Network meta-analysis of rare events using the mantel-haenszel method. *Statistics in Medicine*, 38(16):2992–3012.

Efthimiou, O., Seo, M., Karyotaki, E., Cuijpers, P., Furukawa, T. A., Schwarzer, G., Rücker, G., and Mavridis, D. (2022). Bayesian models for aggregate and individual patient data component network meta-analysis. *Statistics in medicine*, 41(14):2586–2601.

Egger, M., Higgins, J. P., and Smith, G. D. (2022). *Systematic reviews in health research: Meta-analysis in context.* John Wiley & Sons.

Fouskakis, D., Ntzoufras, I., and Draper, D. (2009). Bayesian variable selection using cost-adjusted bic, with application to cost-effective measurement of quality of health care.

Franchini, A. J., Dias, S., Ades, A. E., Jansen, J. P., and Welton, N. J. (2012). Accounting for correlation in network meta-analysis with multi-arm trials. *Research Synthesis Methods*, 3(2):142–160.

Freeman, S. C., Fisher, D., White, I. R., Auperin, A., and Carpenter, J. R. (2019). Identifying inconsistency in network meta-analysis: Is the net heat plot a reliable method? *Statistics in Medicine*, 38(29):5547–5564.

Gelman, A., Meng, X.-L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica sinica*, pages 733–760.

George, E. I. and McCulloch, R. E. (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889.

George, E. I. and McCulloch, R. E. (1996). Stochastic search variable selection", markov chain monte carlo in practice, eds. wr gilks, s. richardson and dj spiegelhalter.

George, E. I. and McCulloch, R. E. (1997). Approaches for bayesian variable selection. *Statistica sinica*, pages 339–373.

George, E. I., McCulloch, R. E., and Tsay, R. (1996). Two approaches to bayesian model selection with applications. *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner*, 309:339.

Hedges, L. V. and Olkin, I. (2014). *Statistical methods for meta-analysis.* Academic press.

Hedges, L. V., Pustejovsky, J. E., and Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods*, 3(3):224–239.

Higgins, J. P., Jackson, D., Barrett, J. K., Lu, G., Ades, A. E., and White, I. R. (2012). Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Research Synthesis Methods*, 3(2):98–110.

Higgins, J. P., Li, T., and Deeks, J. J. (2019a). Choosing effect measures and computing estimates of effect. *Cochrane handbook for systematic reviews of interventions*, pages 143–176.

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., and Welch, V. A. (2019b). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.

Higgins, J. P. and Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11):1539–1558.

Jackson, D., Barrett, J. K., Rice, S., White, I. R., and Higgins, J. P. (2014). A design-by-treatment interaction model for network meta-analysis with random inconsistency effects. *Statistics in Medicine*, 33(21):3639–3654.

Jansen, J. P. and Naci, H. (2013). Is network meta-analysis as valid as standard pairwise meta-analysis? it all depends on the distribution of effect modifiers. *BMC medicine*, 11(1):1–8.

Jeffreys, H. (1961). Theory of probability, ed. 3 oxford university press. *Figure Click here to access/download*.

Kass, R. E. and Wasserman, L. (1995). A reference bayesian test for nested hypotheses and its relationship to the schwarz criterion. *Journal of the american statistical association*, 90(431):928–934.

Krahn, U., Binder, H., and König, J. (2013). A graphical tool for locating inconsistency in network meta-analyses. *BMC medical research methodology*, 13(1):1–18.

König, J., Krahn, U., and Binder, H. (2013). Visualizing the flow of evidence in network meta-analysis and characterizing mixed treatment comparisons. *Statistics in Medicine*, 32(30):5414–5429.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g priors for bayesian variable selection. *Journal of the American Statistical Association*, 103(481):410–423.

Lindley, D. V. (1957). A statistical paradox. *Biometrika*, 44(1/2):187–192.

Lu, G. and Ades, A. (2004). Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine*, 23(20):3105–3124.

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

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

Lunn, D., Jackson, C., Best, N., Thomas, A., and Spiegelhalter, D. (2012). *The BUGS book: A practical introduction to Bayesian analysis.* CRC press.

Mavridis, D., Giannatsi, M., Cipriani, A., and Salanti, G. (2015). A primer on network meta-analysis with emphasis on mental health. *BMJ Ment Health*, 18(2):40–46.

Mavridis, D. and Ntzoufras, I. (2014). Stochastic search item selection for factor analytic models. *British Journal of Mathematical and Statistical Psychology*, 67(2):284–303.

Mavridis, D., Palmer, S. C., and Strippoli, G. F. (2016). Comparative superiority of ace inhibitors over angiotensin receptor blockers for people with ckd: Does it matter? *American Journal of Kidney Diseases*, 67(5):713–715.

Mavridis, D., Porcher, R., Nikolakopoulou, A., Salanti, G., and Ravaud, P. (2020). Extensions of the probabilistic ranking metrics of competing treatments in network meta-analysis to reflect clinically important relative differences on many outcomes. *Biometrical Journal*, 62(2):375–385.

Millar, R. B. (2009). Comparison of hierarchical bayesian models for overdispersed count data using dic and bayes' factors. *Biometrics*, 65(3):962–969.

Mills, E. J., Bansback, N., Ghement, I., Thorlund, K., Kelly, S., Puhan, M. A., and Wright, J. (2011). Multiple treatment comparison meta-analyses: a step forward into complexity. *Clinical epidemiology*, 3:193.

Nikolakopoulou, A., Chaimani, A., Veroniki, A. A., Vasiliadis, H. S., Schmid, C. H., and Salanti, G. (2014a). Characteristics of networks of interventions: a description of a database of 186 published networks. *PloS one*, 9(1):e86754.

Nikolakopoulou, A., Mavridis, D., Furukawa, T. A., Cipriani, A., Tricco, A. C., Straus, S. E., Siontis, G. C. M., Egger, M., and Salanti, G. (2018). Living network meta-analysis compared with pairwise meta-analysis in comparative effectiveness research: empirical study. *BMJ*, 360.

Nikolakopoulou, A., Mavridis, D., and Salanti, G. (2014b). Demystifying fixed and random effects meta-analysis. *BMJ Ment Health*, 17(2):53–57.

Ntzoufras, I. (2011). *Bayesian modeling using WinBUGS.* John Wiley & Sons.

Ntzoufras, I., Forster, J. J., and Dellaportas, P. (2000). Stochastic search variable selection for log-linear models. *Journal of Statistical Computation and Simulation*, 68(1):23–37.

O'hara, R. B. and Sillanpää, M. J. (2009). A review of bayesian variable selection methods: what, how and which.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., et al. (2021). The prisma 2020 statement: an updated guideline for reporting systematic reviews. *International journal of surgery*, 88:105906.

Page, M. J., Sterne, J. A., Higgins, J. P., and Egger, M. (2022). *Investigating and Dealing with Publication Bias and Other Reporting Biases*, chapter 5, pages 74–90. John Wiley & Sons, Ltd.

Perrakis, K. and Ntzoufras, I. (2015). Stochastic search variable selection (ssvs). *Wiley StatsRef: Statistics Reference Online*, pages 1–6.

Perrakis, K. and Ntzoufras, I. (2018). Bayesian variable selection using the hyper-g prior in winbugs. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(6):e1442.

Petropoulou, M., Nikolakopoulou, A., Veroniki, A.-A., Rios, P., Vafaei, A., Zarin, W., Giannatsi, M., Sullivan, S., Tricco, A. C., Chaimani, A., Egger, M., and Salanti, G. (2017). Bibliographic study showed improving statistical methodology of network meta-analyses published between 1999 and 2015. *Journal of Clinical Epidemiology*, 82:20–28.

Piironen, J. and Vehtari, A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, 11(2):5018 – 5051.

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

Rücker, G., Petropoulou, M., and Schwarzer, G. (2020). Network meta-analysis of multi-component interventions. *Biometrical Journal*, 62(3):808–821.

Rücker, G. and Schwarzer, G. (2015). Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC medical research methodology*, 15(1):1–9.

Rücker, G. (2012). Network meta-analysis, electrical networks and graph theory. *Research Synthesis Methods*, 3(4):312–324.

Rücker, G., Krahn, U., König, J., Efthimiou, O., Davies, A., Papakonstantinou, T., and Schwarzer, G. (2022). *netmeta: Network Meta-Analysis using Frequentist Methods*. R package version 2.6-0.

Rücker, G. and Schwarzer, G. (2017). Resolve conflicting rankings of outcomes in network meta-analysis: Partial ordering of treatments. *Research Synthesis Methods*, 8(4):526–536.

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

Salanti, G., Ades, A., and Ioannidis, J. P. (2011). Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *Journal of Clinical Epidemiology*, 64(2):163–171.

Salanti, G., Higgins, J. P., Ades, A., and Ioannidis, J. P. (2008). Evaluation of networks of randomized trials. *Statistical methods in medical research*, 17(3):279–301.

Schmidt, F. L., Oh, I.-S., and Hayes, T. L. (2009). Fixed- versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology*, 62(1):97–128.

Scott, J. G. and Berger, J. O. (2006). An exploration of aspects of bayesian multiple testing. *Journal of statistical planning and inference*, 136(7):2144–2162.

Scott, J. G. and Berger, J. O. (2010). Bayes and empirical-bayes multiplicity adjustment in the variable-selection problem. *The Annals of Statistics*, pages 2587–2619.

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

Seitidis, G., Nikolakopoulos, S., Ntzoufras, I., and Mavridis, D. (2022b). Inconsistency identification in network meta-analysis via stochastic search variable selection. *arXiv preprint arXiv:2211.07258*.

Seitidis, G., Nikolakopoulos, S., Ntzoufras, I., and Mavridis, D. (2023a). *ssifs: Stochastic Search Inconsistency Factor Selection*. R package version 1.0.0.

Seitidis, G., Tsokani, S., Christogiannis, C., Kontouli, K.-M., Fyraridis, A., Nikolakopoulos, S., Veroniki, A. A., and Mavridis, D. (2022c). Graphical tools for visualizing the results of network meta-analysis of multicomponent interventions. *Research Synthesis Methods*.

Seitidis, G., Tsokani, S., Christogiannis, C., Kontouli, K. M., Fyraridis, A., Nikolakopoulos, S., Veroniki, A. A., and Mavridis, D. (2023b). *viscomp: Visualize Multi-Component Interventions in Network Meta-Analysis*. R package version 1.0.0.

Shih, M.-C. and Tu, Y.-K. (2021). An evidence-splitting approach to evaluation of direct-indirect evidence inconsistency in network meta-analysis. *Research Synthesis Methods*, 12(2):226–238.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(3):485–493.

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

Tsokani, S., Seitidis, G., and Mavridis, D. (2022). Component network meta-analysis in a nutshell. *BMJ Evidence-Based Medicine*.

Veroniki, A. A., Seitidis, G., Nikolakopoulos, S., Ballester, M., Beltran, J., Heijmans, M., and Mavridis, D. (2022). *Modeling Multicomponent InterventionsMulticomponent interventions in Network Meta-Analysis*, pages 245–261. Springer US, New York, NY.

Veroniki, A. A., Straus, S. E., Fyraridis, A., and Tricco, A. C. (2016). The rank-heat plot is a novel way to present the results from a network meta-analysis including multiple outcomes. *Journal of clinical epidemiology*, 76:193–199.

Veroniki, A. A., Tsokani, S., White, I. R., Schwarzer, G., Rücker, G., Mavridis, D., Higgins, J. P., and Salanti, G. (2021). Prevalence of evidence of inconsistency and its association with network structural characteristics in 201 published networks of interventions. *BMC Medical research methodology*, 21:1–10.

Wakefield, J. and Bennett, J. (1996). The bayesian modeling of covariates for population pharmacokinetic models. *Journal of the American Statistical Association*, 91(435):917–927.

Welton, N. J., Caldwell, D., Adamopoulos, E., and Vedhara, K. (2009). Mixed treatment comparison meta-analysis of complex interventions: psychological interventions in coronary heart disease. *American Journal of Epidemiology*, 169(9):1158–1165.

White, I. R. (2015). Network meta-analysis. *The Stata Journal*, 15(4):951–985.

White, I. R., Barrett, J. K., Jackson, D., and Higgins, J. P. (2012). Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Research Synthesis Methods*, 3(2):111–125.

Yuan, J., Zhang, R., Yang, Z., Lee, J., Liu, Y., Tian, J., Qin, X., Ren, Z., Ding, H., Chen, Q., et al. (2013). Comparative effectiveness and safety of oral phosphodiesterase type 5 inhibitors for erectile dysfunction: a systematic review and network meta-analysis. *European urology*, 63(5):902–912.

Zellner, A. (1986). On assessing prior distributions and bayesian regression analysis with g-prior distributions. *Bayesian inference and decision techniques*.

Zellner, A. and Siow, A. (1980). Posterior odds ratios for selected regression hypotheses. *Trabajos de estadística y de investigación operativa*, 31:585–603.