Indirect Comparisons with Population Adjustment Methods using Single-arm Studies in Health Technology Assessment

Nahid Sultana

Supervisors: Professor Dr. Tracey A Young, Dr. Kate Ren and Dr. Jean Hamilton

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in the Department of School of Medicine and Population Health.

in the

Faculty of Health

January, 2024

Declaration

As required by the University's Regulations and Code of Practice for Research Degree Programmes, I declare that this dissertation was completed in accordance with those requirements. The work is the candidate's own work and the work has not been submitted for any other academic award. All sentences or ideas quoted in this thesis from other people's work have been specifically acknowledged by clear cross-referencing to author, work and page(s).

Name: _____

Signature: _____

Date: _____

Abstract

In the development process of pharmaceuticals, a frequent step is that a therapy is administered to all patients within a study; which is known as a single-arm study. A particular feature of single-arm studies is that they provide no direct estimate of treatment effects owing to the lack of a comparator arm. Therefore, estimation of treatment effects from single-arm studies involves reference to an external comparator (unanchored indirect treatment comparison). Even though single-arm studies can be completed faster than randomised control trials (RCTs), they add complexity to indirect comparisons as both prognostic and effect-modifier variables need to be balanced to obtain a valid relative treatment effect estimate.

In health technology assessment (HTA), when companies analyse their intervention treatment from a single-arm study with comparator/comparators, access to individual patient data (IPD) in all studies of interest is a rare situation as sharing of clinical data is often limited. A middle-ground situation is more realistic where the company has access to IPD for its own study and aggregate data (AgD) for the comparator studies. Moreover, the companies often have to estimate relative treatment effects of a single-arm study treatment against multiple comparator treatments in a larger disconnected network of evidence. Therefore, the fundamental objective of this thesis was to assess whether the population-adjustment method matching adjusted indirect comparison (MAIC) is suitable to implement for a larger disconnected network of evidence or not.

This thesis starts with a review on National Institute for Health and Care Excellence (NICE) single technology appraisal (STA)s to evaluate the methods with the single-arm study. Unanchored MAIC and simulated treatment comparison (STC) were found to be frequently used methods to estimate relative treatment effects with single-arm studies. It was found that unanchored MAIC was applied multiple times to estimate the relative treatment effect of a single-arm study intervention in a larger disconnected network of evidence. The relative effect estimates from this multiple MAICs were described as if the MAIC estimates made a set of coherent relative effect estimates ignoring the fact that these estimates were from different target populations. Additionally, using the IPD several times for conducting multiple MAICs breaks the independence of the unit of analysis assumption. In order to assess the impact of this, a simulation study was designed with multiple MAIC estimates in a fixed and a random effects network meta-analysis (NMA).

The major impact of performing an MAIC-adjusted NMA was seen in the coverage of the NMA estimates where the coverage dropped below nominal level (95%). The violation of the independence assumption together with the sandwich estimator had a repercussion

on the NMA estimate coverage. The deviation from the nominal level of coverage was more pronounced for the larger compared to a smaller disconnected network of evidence. Double-bootstrapping with MAIC was found to solve the problem of undercoverage both for fixed and random effects NMA. However, the biases were found to be comparatively high with low-overlap scenarios and a smaller sample size. The proposed double-bootstrapping method was also applied in a case study with asthma. The case study illustrates how to make multiple comparisons simultaneously using double-bootstrapped MAIC-adjusted NMA where a correct level of coverage for the NMA estimates can be preserved with the use of double-bootstrapping. Therefore, this thesis recommends MAIC-adjusted NMA with double-bootstrapping approach when there exists a sufficient level of overlap between studies together with a satisfactory sample size.

Acknowledgements

It had always been my passion to enroll in a PhD program, but I was finally able to start in 2020. Starting a PhD journey in a COVID lock-down situation was very tough, however, by the grace of Allah I am very close to my goal. Whatever the situation was, I tried to do my best. In this grueling journey, my supervisors Dr. Tracey A. Young, Dr. Kate Ren and Dr. Jean Hamilton acted as my backbone. I am also hugely thankful to my former supervisor Dr. John W. Stevens. My supervisor team's patience and inspiration throughout my journey provided me with excellent motivation to achieve my goal. They taught me how I should be in my future academic journey.

This journey also would not have been possible without my family. First of all my husband Anam whose constant support and steadfastness helped me survive through this journey. He also acted as an excellent listener which helped me greatly. My mother, in-laws, sister Nargis, Nipa, brother Reza, my nieces, friends Farhana and Tanni, the University of Sheffield, my home University (SUST), and UK Bangladeshi community also supported me tremendously. I dedicate this thesis to my father who left us in 2018, if he were here, he would be the happiest person to witness me achieve this. I also want to acknowledge my PhD colleagues Saleema, Ghadha, Norah, Azzah, Elvis, Bright, Ray, Rami, Chloe and James who have supported me by giving me mental support to overcome all my stress throughout the process. Last but not least, special thanks to GlaxoSmithKline (GSK) who supported this thesis as this project is jointly funded by University of Sheffield and GSK.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The efficacy of pharmaceuticals is usually evaluated through randomised control trials (RCTs) for marketing authorisation. In general, RCTs guarantee a high-level unbiased assessment of efficacy and safety of treatments with such reliability that no other methodology can compete with that. It gives RCTs the highest standard for the evaluation of new treatments. In general, RCTs possess some important characteristics to generate an unbiased comparison between treatments which can be summarised as: random allocation of patients to treatment groups to minimise potential confounders, double blinding of patients and clinicians about treatment allocation to avoid biased assessment of outcomes, analysing patients within the group with which they were allocated to preserves the benefits of randomisation, and the aim of estimating the size of the difference in predefined outcomes. At the start of a study, when participants are assigned to different treatments, if systematic differences exist, it can distort the treatment effect. Randomisation aims to alleviate both measured and unmeasured confounders (NICE, 2013).

In the development process of pharmaceuticals, a single-arm study is conducted by assigning a treatment to all patients within a study. By design, single-arm studies do not have a comparator arm which prevents them from providing a direct relative estimate like an RCT, therefore, indirect comparison is the only option with single-arm studies. There can be various reasons for implementing a single-arm study. It may be conducted when unprompted improvement in patients is not expected, placebo effects are minimal, and assigning patients to a placebo arm is ethically not justified. Single-arm studies are commonly executed in oncology, such as rare cancers where the target patient group is often very small (Evans, 2010). They are commonly used for outcomes like tumor response rate in order to assess earlier efficacy. A product can be submitted to regulatory agencies for licensing with a single-arm study despite the fact that it may not be a preferred approach due to the absence of a concurrent comparator arm. Even though in some cases they can be completed faster than an RCT, they add complexity to relative treatment effect estimation due to the inability to provide head-to-head treatment comparisons (Agrawal et al., 2023).

European union member states, Canada, UK, Japan, and Australia, which cover most of the drug markets worldwide, have health-care systems where the cost of a prescribed medicine is almost totally covered by social security agencies or third-party public payers such as the

National Health Service (NHS) (Eichler et al., 2010). Gaining regulatory consent is important for a drug to get market access, however, gaining positive reimbursement or coverage decisions from payers is also necessary for drugs to be available to patients. Health technology assessment (HTA), refers to a collaborative system where the economic, social, organizational, and ethical issues associated with a health technology are evaluated. The main objective of HTA is to perform an assessment that can help in policy decision-making (WHO, 2020).

HTA bodies, such as National Institute for Health and Care Excellence (NICE) which produce recommendations on public funding of health care technologies in England and Wales, need to find out the treatment with the highest efficacy of all available options to reimburse given that the availability of resources is limited. In doing this, often a new treatment from a single-arm study needs to be compared with one or multiple relevant comparators. As no direct comparison can be made with single-arm studies, indirect treatment comparison (ITC) is a standard approach companies rely on for their HTA reimbursement submissions. ITCs are a valid approach as long as they are not biased due to differences in patient characteristics (Signorovitch et al., 2010). Hence, adjustments need to be considered when making indirect comparisons to ensure the validity of the results. An indirect treatment comparison is usually performed in a connected network of evidence. A network of evidence can be in a connected form when pair-wise comparisons link each treatment to every other treatment, that is, for each treatment, there is a chain of pair-wise comparisons (Sutton et al., 2008). The situation becomes more complex when no connected network of evidence is present which makes the comparison between treatments more challenging.

This chapter begins with a discussion of indirect comparison in Section 1.1 which continues with a description of biases in indirect comparison in Section 1.2. Comparative effectiveness with single-arm studies is discussed in Section 1.3 more elaborately. The existence of single-arm studies in the evolution of new medication is thoroughly discussed in Section 1.4. In addition, in many HTA processes, it happens recurrently that individual patient data (IPD) is available for a company's own study and aggregate data (AgD) is available for comparator studies. This is discussed in more detail in Section 1.5. Section 1.6 sheds light on statistical methodology with single-arm studies. At the conclusion of this chapter, the research questions that will be addressed in this thesis have been discussed in Section 1.7.

1.1 Indirect comparisons

Figure 1.1 illustrates a number of network diagrams corresponding to different evidence structures of indirect comparisons. The nodes with capital letters indicate treatments and the solid lines mean treatments in head-to-head clinical studies have been compared directly. A broken line between nodes means no direct comparison exists between treatments.

Suppose the outcomes of the studies are binary and odds ratios (ORs) have been used to estimate relative effectiveness between studies. In Figure 1(a), two treatments A and C can be directly compared to B, which is the common comparator. No direct comparison exists between A and C. An indirect comparison between A and C, i.e. d_{AC} can be achieved by $OR_{AB}/OR_{BC} = OR_{AC}$ where OR_{AB} is the odds ratio of A versus B, OR_{BC} is the odds

ratio B versus C and d_{AC} is the relative effect between the treatments. This is termed as "standard adjusted indirect comparison" as it takes into account the randomisation of the corresponding study (Phillippo et al., 2016). The comparative effect of d_{AC} will be called "unadjusted indirect comparison" if it is estimated by simply comparing the A arm of the AB study with the C arm of the BC study as OR_{AC}. An unadjusted or naive indirect comparison has been discouraged as it assumes results of particular arms from different studies assuming that within-study randomisation is preserved like an RCT study (Sutton et al., 2008). Usually, the common comparator treatment is a standard treatment or a placebo. Diagram (a) depicts the simplest form of indirect comparison, but with the addition of treatments and studies, the network can become complex and evidence about relative effects can come from multiple sources such as diagrams (b) and (c).

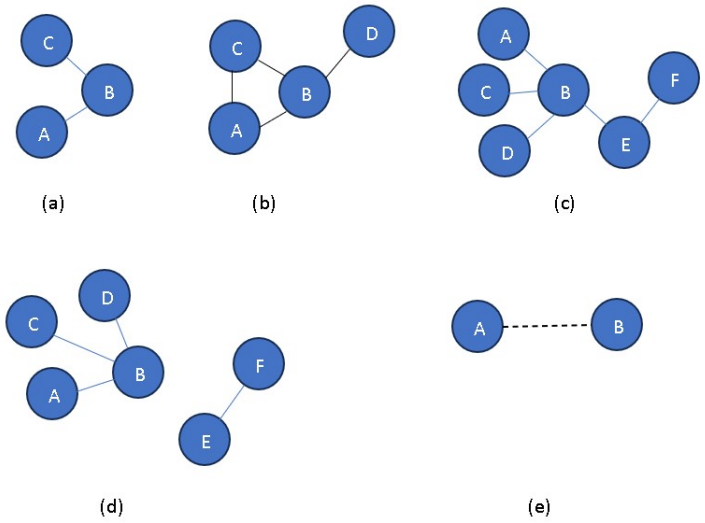


Figure 1.1: *Different forms of a network in indirect comparison*

Diagrams (a),(b) and (c) in Figure 1.1 can be termed as a larger connected network of evidence that can be comprised of multiple treatments or multiple studies per treatment comparison or both. In a larger connected network of evidence, a path or edge exists that connects every treatment directly or indirectly to all other treatments. This is not the case for a disconnected network of evidence. In diagram (d) a larger disconnected network of evidence is depicted where there is no direct or indirect path to connect treatments E and F with treatments A, B, C, and D. In diagram (d), the disconnected network of evidence is comprised of two connected networks (network A, B, C, D and network E, F), however, the network can also include single/multiple single-arm studies. A larger disconnected network of evidence is not suitable for making comparisons between all possible treatments.

Connected networks are exhibited in diagrams (a), (b), and (c), and data from such networks could be analysed using indirect comparison and mixed treatment comparison (MTC). In practice, MTC is also known as network meta-analysis (NMA). A MTC or NMA is an expansion of a standard pairwise meta-analysis, which is capable of comparing multiple treat-

ments altogether in studies sharing at least one treatment in common. In MTC, both direct and indirect comparisons are used. In diagram (b), treatments A and B can be compared using both direct and indirect evidence. Direct evidence comes from the AB study whereas indirect evidence comes from the BC and AC studies. Hence, relative effectiveness for A versus B is a mixture of direct and indirect evidence. This is a simple example of the MTC network. The MTC model allows the inclusion of all the evidence which in turn decreases the uncertainty in the pooled estimate. In addition, including both direct and indirect comparisons provides a chance to judge the consistency of the network of evidence (Sutton et al., 2008). An indirect treatment comparison is used to describe a comparison that does not contain any loops as in diagrams (a) and (c), whereas a mixed treatment comparison is used to describe a comparison that does contain loops as in diagrams (b)(Shim et al., 2017; Fleetwood, 2020).

The validity of MTC or NMA depends on three vital assumptions known as similarity, transitivity, and consistency (Jansen et al., 2008; Donegan et al., 2013; Shim et al., 2017). Under the assumption of similarity, the studies included for analysis need to be similar with respect to the methodology used in the studies. That means the population, intervention, comparison, and outcome among the studies need to be similar. The assumption requires that the studies should be homogeneous with respect to any criteria that may impact the treatment effect. While similarity explains the methodological equivalence between studies, transitivity is needed to ensure the validity of inference. For instance, suppose in a particular health condition there exist three treatments named Trt 1, Trt 2, and Trt 3 that have been investigated in a head-to-head setting, and it was found that Trt 1 is more efficacious than Trt 2, and Trt 2 is more efficacious than Trt 3, then logically Trt 1 should be more efficacious than Trt 3. Transitivity needs to be held for all cases in an NMA. Consistency measures the transitivity objectively. It refers to the equivalence of direct and indirect evidence. Consistency is also known as coherence.

Relative effects that are estimated indirectly with a common comparator are also known as "anchored comparison". In Figure 1.1, diagram (a), (b), (c) can be taken as an anchored form of comparison where treatment B is the common comparator. Since the requirement of MTC is a connected network, discontinuity in the network makes the application of MTC infeasible. In Figure 1.1, diagrams (d) and (e) both illustrate disconnected networks. In diagram (d), there is no study that connects treatments E and F to the rest of the network and in diagram (e), treatments A and B come from two single-arm studies, so no head-to-head comparison is possible. An indirect comparison that can be made between treatment A and E in diagram (d) or treatment A and B in diagram (e) can also be termed as an "unanchored" form as no common comparator exists between them.

These disconnected networks depicted in diagrams (d) and (e) are the most difficult ones to be inferred on. Disconnected networks may occur due to various reasons such as the debatable use of a placebo, a major shift in the treatment paradigm, designating a product as an orphan, or when there are many accepted standards of care. Estimating the relative effectiveness as well as safety of new drugs against alternatives in disconnected networks becomes a challenge to analysts (Goring et al., 2016).

Adjusted indirect comparisons implement direct comparisons with a treatment arm that is common to the network for the sake of overcoming criticism against naive comparison. Whereas standard indirect comparison relies on connected evidence networks, different situations can arise if treatments are not linked via connected networks. When there is no connected evidence, an "unanchored" indirect comparison needs to take place which is illustrated in diagrams (d) and (e).

1.2 Different forms of biases in indirect comparison

Estimation of treatment effects relies on four core assumptions. These assumptions have been summarised by Phillippo et al. (2016) and have been discussed by several authors (Stuart et al., 2011; Hartman et al., 2015). These assumptions are true for any method of indirect comparison. These are:

i. Whether or not an individual is assigned to a study, outcomes are the same on treatment and control. This is called the homogeneity of outcomes.

ii. Stable unit treatment value. The outcomes of a participant are not dependent on any other participants.

iii. Strongly ignorable treatment assignment. The distribution of prognostic variable or effect-modifier variable will be balanced if treatment allocation is random and does not depend on sample selection from the target population given the observed covariates.

iv. Strongly ignorable sample assignment. Both sample selection and outcome do not relate to any unmeasured variables and given observed covariates, each participant in the target population has a substantial probability of being selected into the study sample.

Although these assumptions are needed for making indirect comparisons, bias can hamper the validity of comparisons whether the comparison is performed directly or indirectly. When the relative effect of treatments is estimated in different populations or settings, if there are any differences in the collection, interpretation, analysis, and publication phase that can dilute the true response of treatment, then it is called bias. Biases can be categorised as being internal or external (Turner et al., 2012).

Figure 1.2 summarises different forms of biases that can hamper the validity of a study. A component of biases is termed as "selection bias" which means at baseline, treatment, and control groups are not similar with respect to patient characteristics. "Performance bias" arises as a consequence of the absence of blinding of participants or caregivers whereas "attrition bias" emerges from an imbalance in exclusion and drop-outs. Biases related to the assessment of the outcome are called "outcome bias". Again externally different kinds of biases can challenge study validity. "Population bias" arises when the idealised study population and target population differ substantially. Difference between the study and target interventions causes "intervention bias" and "control bias" stems from differences between the study and target control strategies. "Publication bias" arises when authors, editors of journals, or reviewers grow a tendency only to publish studies with significant/favourable results. The focus of this thesis is to explore the methods that are capable of handling selection bias as other forms of biases are not adjustable with the existing methods.

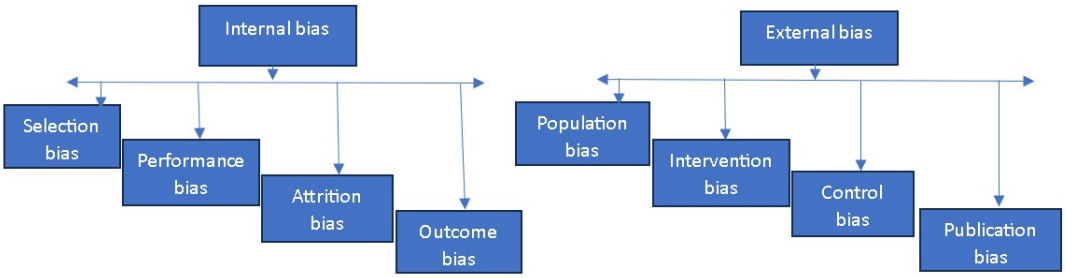


Figure 1.2: *Different forms of biases in indirect comparison*

1.3 Indirect comparisons in disconnected networks with single-arm study

Although a MTC or NMA is capable of comparing multiple treatments simultaneously, when a disconnected network of studies arises, which means studies have different comparator arms or lack a control arm completely (single-arm study), it becomes more difficult to provide relative treatment efficacy. Single-arm studies, which are designed without a comparator arm, are unable to provide a direct relative treatment effect estimate. Although the use of single-arm study has been discouraged in all cases, it is still being used due to various reasons. RCTs are not always possible due to realistic or moral issues. In life-threatening diseases, randomising patients to a placebo arm that is assumed to be inefficacious is unethical, or in the case of rare diseases, it can be problematic to find a sufficient amount of patients for the two arms of an RCT to get a significant difference. Moreover, for some advanced diseases, such as some advanced cancer, it can be that no established comparator treatment is available. In these situations, single-arm studies can be the only way to get available evidence.

In Figure 1.3, different forms of disconnected networks with single-arm study have been illustrated where the nodes indicate treatments, solid lines connecting nodes indicate direct comparison and a dashed line indicates an indirect comparison. In diagram (a), both treatments A and B come from a single-arm study and an unanchored comparison needs to be made between them whereas in diagram (b), treatment A comes from a single-arm study and B comes from an RCT. In diagram (c), a single-arm study treatment A needs to be compared with a network of evidence where the network is composed of both RCTs and single-arm study (treatment G). In diagram (c), although only two single-arm studies are included (treatment A, G), however, the network can consist of several single-arm studies also. Additionally, diagrams (b) and (c) can be termed as a larger disconnected network of evidence as they consist of a large number of treatments (more than two) that need to be compared simultaneously, and the network is disconnected due to the presence of single-arm studies which obstructs the connection of each treatment to every other treatment in the

network.

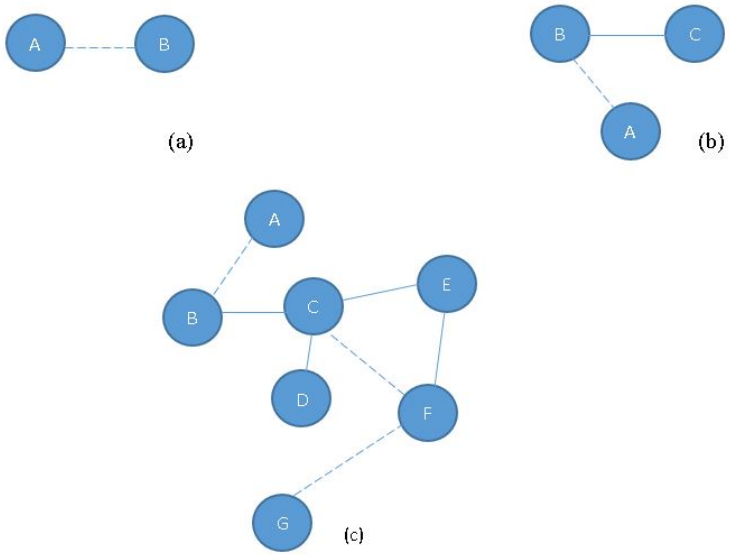


Figure 1.3: *Different forms of disconnected network with single-arm study*

Standard indirect comparisons and NMAs use a connected network of evidence with the constancy of relative effects. This assumption requires that the effect-modifying variables are distributed similarly across studies so that relative effects are constant. The only source of error is sampling error which occurs as a consequence of variation in study sizes. For the case of single-arm studies, a crude "unadjusted" comparison with comparator treatments can be partitioned into two components, such as sampling error and systematic error (Phillippo et al., 2016). The sources of systematic error in single-arm studies come from disparity in both prognostic and effect-modifier variables. Therefore, when estimating relative treatment effects with single-arm studies, it should be ensured that estimates are not biased due to differences in prognostic and effect-modifier variables using an adjusted comparison. Over the last decade, "population-adjusted indirect comparisons" have become a commonly used approach for submissions to reimbursement agencies which can address the problem of unanchored indirect comparison (Phillippo et al., 2018). To calculate the relative effect in unanchored form, population-adjusted indirect comparisons made the assumption of "conditional constancy of absolute effects" where it is assumed that all effect-modifiers and prognostic variables are known and at any given level of the variables, the absolute treatment effects are constant. In practice, this assumption is difficult to meet due to the presence of unobserved prognostic and effect-modifier variables that causes unanchored comparisons affected by an unknown amount of residual bias (Phillippo et al., 2019a).

1.4 Presence of single-arm study in drug development

Generally, single-arm studies are desirable when the availability of patients is limited which restricts randomisation to a control arm. Such designs may be considered when spontaneous improvement in participants is not expected, placebo effects are not large, and randomisation to placebo may not be ethical (Evans, 2010).

The European Medical Agency (EMA) and US Food and Drug Administration (FDA), both acknowledge that there can be situations when RCT is infeasible in terms of operationally and ethically (Goring et al., 2019). An in-depth study was made by Hatswell et al. (2016) on the approval of new pharmaceuticals from 1 January 1999 to 8 May 2014. This study investigated that over the past 15 years, to what extent uncontrolled clinical studies were used for drug approvals by the EMA and the FDA. A factor that was common in these uncontrolled studies was that they lacked a control arm. It showed that without undergoing an RCT, a huge number of treatments got licensed. Seventy-six unique indications were accepted without RCT results of which 44 were granted by the EMA and 60 by the FDA. Primarily, oncology was the disease area where most of the uncontrolled studies took place, with 66% being either solid tumour or haematological oncology. The most common was for haematological malignancies (34), following oncology (15) and metabolic conditions (15).

Another review was made by Goring et al. (2019) to update the review by Hatswell et al. (2016). A systematic search was conducted on EMA and FDA regulatory submissions which summarises the characteristics of non-randomised studies. The review was conducted spanning the years 2005–2017. Non-randomised evidence was used for 43 indication-specific products that were submitted to the FDA (n=41) or the EMA (n=34). The indications were haematological cancer, conditions associated with stem cell transplantation, other haematological conditions or rare metabolic disease. Of the 96 unique studies for the 43 indication-specific products, the most common was single-arm studies which covers 67% of the studies. Of the indication-specific products, 37% provided evidence from external control: 28% used aggregate-level controls, and the rest of the products (9%) involved IPD for external controls that have similar patient characteristics in the intervention studies. External control groups were not mentioned for the remaining 63% indications. The most common endpoint was the objective response rate (ORR). Out of the 43 indication-specific products, only 5% reported a hazard ratio (HR) to compare overall survival (OS). This study concluded that there is an increase in acceptance processes based on non-RCT evidence, specifically for the broader context of oncology indications.

A systematic review on anticancer drugs approved by FDA from 1973 through 2006 showed that out of 68 oncology drugs, 31 drugs which do not include hormone therapy and supportive care, were granted without an RCT. Except for ORR, which was the most common endpoint, other endpoints, such as disease-free survival, were also used; the median response rate was 33% (range, 11% to 90%) (Tsimberidou et al., 2009).

Griffiths et al. (2017) performed a review including NICE, the Canadian Agency for Drugs and Technologies in Health (CADTH), and the German Institute for Quality and Efficiency in Health Care (IQWiG) in order to evaluate the importance of non-randomised evidence

in HTA decision-making. The time frame for the review was January 2010 to December 2015. Single-arm studies were not the only ones that were considered non-randomised but also dose-ranging studies, registry studies [1], compassionate use programs [2], and uncontrolled extension studies[3] were considered. The main bulk of the non-randomised evidence was single-arm studies. Among the total 549 submissions during this time, 38% (45 of 118) of NICE submissions, 13% (34 of 262) of CADTH submissions and 12% (20 of 169) of IQWiG submissions were from non-comparative studies that were used as supporting evidence and only 4% (5 of 118) of NICE appraisals, 6% (16 of 262) of CADTH appraisals, and 4% (6 of 169) of IQWiG appraisals were solely on non-comparative evidence. The disease area that was most prevalent was cancer or infection (specifically hepatitis C) followed by an orphan disease.

Djulbegovic et al. (2018) conducted a study from 1995 to 2015 to evaluate how frequently non-randomized studies have been approved to authorise drugs by EMA. They concluded that of the 723 newly invented drugs that obtained market access, 92.94% (672 drugs) were granted based on RCTs, and 7% (51/723) were granted based on non-randomized data. Out of the 71 drug-indication pairs that were granted based on non-randomized data, 58% (41/71) were for treating leukemias and lymphomas, rare diseases covered 27% (19/71), followed by chronic diseases (8%; 6/71) and other health problems (7%; 5/71). Sixty-one percent (43/71) of the studies used predefined response criteria such as overall response, cytogenetic response, and objective response as the primary outcome. Six percent (4/71) of the studies used survival as the primary endpoint. Seventy-six percent (54/71) were single-arm studies without mentioning control; 7% (5/71) were single-arm studies that involved external controls; and 17% (12/71) used non-comparative studies.

A review by Phillippo et al. (2019a) focuses on population-adjustment methods in submissions to NICE from 1st January 2010 and 20th April 2018. Eighteen TAs were found to have used population-adjustment methods, 16 of them (89%) used unanchored comparisons, and 83% were in oncology. Thirteen (72%) of this TAs have used survival outcomes (for instance progression free survival (PFS), overall survival (OS)). Other than oncology, 16.7% were in hepatology, and 3.6% percent appraisals were in rheumatology. A recent review by Agrawal et al. (2023) has assessed the impact of single-arm studies for FDA in oncology that covers between January 1, 2002, and December 31, 2021. During this duration, 176 hematology and oncology indications were found in single-arm studies. Out of these 176 single-arm studies, 98% (173 of 176) used response rate as an outcome to support approval in these single-arm studies.

From the previous discussion, it is evident that over the past few years, a huge escalation has been seen in the use of single-arm studies, specifically in oncology and haemato-oncology

[1] A registry study is a structured way to collect clinical and other data using observational methods to assess specified outcomes for a population characterised by a condition.

[2] Compassionate use programs permits the use of an unapproved medicine to groups of patients who have a health condition with no satisfactory approved medicine and are unable to enter clinical studies.

[3] Extension phase of a comparative study consist with only active treatment arm. All patients from the control group in the parent RCT are moved to the active treatment in a long-term extension. Moreover, uncontrolled extensions study allow patients the opportunity to move from placebo to the active treatment group, and early access to a promising new medication for an extended period. However, uncontrolled extensions face difficulty in estimating causal inferences about efficacy and safety.

(Phillippo et al., 2016; Cucherat et al., 2020; Agrawal et al., 2023). In oncology, in addition to PFS and OS, one of the key outcomes for establishing clinical effectiveness is often a binary outcome (such as overall response rate (ORR) and complete response rate (CRR)). Therefore, concentration is specially made on application to binary outcomes in single-arm studies in this thesis. Furthermore, despite the fact that most of the applications of single-arm studies have been seen in oncology, the presence of single-arm studies in other disease areas is not completely rare, such as in asthma (Pilette et al., 2019; Buhl et al., 2020; Lugogo et al., 2022). Therefore, this thesis also includes a case study of single-arm study with asthma.

1.5 Single-arm study with limited access to IPD

Consider the diagram (a) in Figure 1.3, where treatments A and B come from two single-arm studies. In diagram (a), the nodes indicate treatments, and the dashed line indicates that there exists no direct comparison between them. Suppose, the relative efficacy of treatments needs to be assessed with binary outcomes. An ORs can be estimated to calculate the relative treatment effect between treatment A and B. However, this does not address the whole problem of making comparisons in the case of a single-arm study. The problem of relative comparison with single-arm study has several aspects. Although comparative evidence in diagram (a) can be calculated as ORs, there can be an imbalance in prognostic and effect-modifying variables between the studies. In this case, adjustment needs to be made for both studies and the relative treatment effects can be estimated in the form of an ORs. This particular setting of single-arm study where an unanchored comparison in disconnected networks needs to be addressed is very difficult to handle.

The issue of indirect comparison with single-arm studies that has been described in Section 1.3 becomes denser with the additional complexity of partial availability of IPD. The presence of partial IPD is prevalent which can be seen from the review by Goring et al. (2019). The review was on non-randomised studies where they classified their analysis into three categories. The first category was IPD-based external controls where studies used IPD to adjust intervention and external control groups. Second, aggregate-level external controls, which include studies that did not make any adjustment to correct for the imbalance between intervention and control groups, and in the third category, studies where control groups were not mentioned explicitly. The findings of this review reveal that the existence of aggregate-level data on external controls is frequent.

In the presence of full IPD from all studies in a network, the "IPD network meta-regression" is capable of adjusting all effect-modifiers and considered to be the gold-standard (Lambert et al., 2002; Riley et al., 2010; Dias et al., 2011b). However, the availability of full IPD is rare. On the contrary, partial availability of IPD is a common scenario in a technical appraisal (TA) context when a pharmaceutical manufacturer prepares for a submission to a reimbursement agency using a single-arm study with single or multiple comparator treatments. A manufacturer usually has access to IPD on their own study/studies, but only published aggregate data on their comparator study/studies. As no head-to-head comparison is possible with a single-arm study, the assessment of comparative efficacy is done indirectly. However, techniques like propensity score matching (PSM) or propensity score weighting (PSW) are not

possible as they require access to IPD for all studies. There can be various reasons for this partial availability of IPD, such as sharing IPD can be hampered by confidentiality issues, it can be owned by a comparator company or the data can be unavailable due to the passage of time (Hatswell et al., 2020).

Due to the restricted access to IPD, when relative treatment effect has to be made with single-arm study, usually IPD from AgD studies need to be reconstructed, particularly for time-to-event data to estimate hazard ratio (HR). For binary outcomes, AgD will usually be a published estimate of the proportion of patients with an event. An ORs can be estimated to calculate the relative treatment effect between intervention and comparator study. To calculate the ORs, 0/1 values for the outcome variable need to be reconstructed using the proportion reported in AgD studies. For time-to-event data, IPD needs to be reconstructed for AgD studies that require a simulation of individual outcomes. However, the simulation cannot provide covariate information for AgD studies. Therefore, reconstructing 0/1 values (binary outcomes) or IPD (time-to-event outcomes) from a comparator study is not the same as having access to full information of the AgD studies.

1.6 Statistical methods with single-arm study

The amount of guidance provided by HTA bodies on the use of single-arm studies is very limited. The IQWiG says it can consider ITC to evaluate cost-benefit relations after taking into account the lower reliability of results. It accepts solely the adjusted indirect comparisons and rejects the use of unadjusted indirect comparisons (i.e. the naive use of a single-arm study). It accepts non-randomised studies only for justified exceptional cases, such as if an RCT is unachievable or when non-randomised studies produce results with adequate reliability (IQWiG, 2016). According to NICE, when inference needs to be made from non-RCT evidence about relative treatment effect, then one must be more prudent, and biases with non-RCT evidence should be identified and adjusted (NICE, 2022). CADTH claims if an indirect comparison is requisite then proofs to support this indirect comparison should be provided as much as possible. Resubmission must be based on efficacy data from an RCT. In the absence of RCT, case-control or cohort studies can be counted on if the new information is related to better safety (CADTH, 2014).

Single-arm studies suffer from biases because of various reasons such as confounding, lack of blinding, and incomplete follow-up. Identifying biases with single-arm studies and adjusting them is recommended by NICE (2009), but there is no guidance on what could be the eligible methods to estimate relative treatment effect with single-arm studies. The use of unsuitable methods to estimate relative treatment effectiveness with single-arm study may have detrimental consequences both on the clinical and cost-effectiveness decision on health technologies. Faria et al. (2015) have summarised commonly implemented methods to estimate treatment effect using comparative IPD from non-RCTs to inform NICE TAs and have proposed guidance to enhance the quality for future assessments. The document gives guidance on single-arm studies when IPD is available both on intervention and comparator studies but mentions that IPD on the intervention and AgD on comparator study is out of the scope of the document.

Phillippo et al. (2016) explore methods on population-adjusted ITC, in which differences in patient characteristics are adjusted using IPD in one or more studies. This technical support document (TSD) summarises recently proposed methods on population-adjustment, looks at the theory behind them, makes a review of published applications, and describes alternative methods briefly. Unanchored methods such as matching adjusted indirect comparison (MAIC) and simulated treatment comparison (STC) with the single-arm study were considered problematic and their use was discouraged if anchored methods could be applied. After the publication of the TSD, several systematic reviews have been done where methods with limited IPD have been discussed but no review has come to a conclusion on the validity and preference of these methods (Phillippo et al., 2016; Stevens et al., 2018; Phillippo et al., 2018). Subsequently, a number of simulation studies have been done for the anchored case to explore the appropriateness of the population-adjustment methods (Ishak et al., 2015b; Belger et al., 2015a; Kühnast et al., 2017; Leahy and Walsh, 2019; Hatswell et al., 2020; Phillippo et al., 2020b; Remiro-Azócar et al., 2020). MAIC was found to be most favoured methodology both in its application and in the simulation studies when it satisfies all the assumptions (Remiro-Azócar et al., 2020; Hatswell et al., 2020; Jiang and Ni, 2020; Wang, 2021). Modification on both MAIC and STC have also been discussed (Remiro-Azócar, 2022b; Remiro-Azócar et al., 2022). Still, there are uncertainty in these methods when they deviate from their assumptions. Therefore, it is crucial to explore in addition to MAIC and STC, what other methods are available in the literature with single-arm studies and also explore the suitability of these methods for different situations.

1.7 Research questions

The use of single-arm studies in drug development seems to be prevalent and non-trivial despite methodological limitations. Methods for example unanchored MAIC and STC are frequently used for estimating treatment effects with single-arm studies which are believed to satisfy the conditional constancy of absolute effects. In the case of non-compliance with the assumptions, these methods can suffer from residual bias. When a single-arm study is compared to a comparator treatment, the network of evidence needs to be taken into account also. Therefore, in addition to these methods, it is necessary to evaluate the validity and reliability of other methods in the literature that can give relative treatment effects with single-arm studies.

This PhD aims to address the following research questions:

1. What approaches have been used for conducting unanchored indirect treatment comparisons to any outcome with single-arm studies in STAs submitted to NICE?
2. Reviewing the literature:
    - What methods currently exist in the literature for conducting unanchored indirect treatment comparisons with single-arm studies?
    - Which methods are suitable in particular situations?

3. Is the population-adjustment method MAIC suitable to implement for a larger disconnected network of evidence?

4. What recommendations should be provided for unanchored MAIC with single-arm study when a larger disconnected network of evidence exists?

Chapter two of this thesis describes a review that was done to explore the approaches/methods that have been used for conducting unanchored indirect treatment comparisons to any outcome with single-arm studies in NICE. The timeline of the review was 2018-2021 which includes the most recent STAs. Chapter three also describes a review that focuses on what methods are available in the literature for unanchored comparisons with single-arm studies and discusses the appropriateness of the methods. After conducting both reviews, a simulation study was designed based on the findings of the reviews. The simulation study and its results are described in Chapters four and five. The aim of the simulation study was to investigate the suitability of unanchored MAIC for a larger disconnected network of evidence both for fixed and random effects NMA. MAIC was chosen for the simulation study due to its widespread applicability that was found in NICE STA review (Chapter 2). Chapter six proposes a novel method called "double-boootstrapping" for MAIC with single-arm studies in a disconnected network of evidence to overcome problems found in the simulation study. Chapter seven discusses the application and practical problems of MAIC-adjusted NMA for a real-world case study with asthma and the concluding chapter discusses the findings from previous chapters, limitations, and finishes with future recommendations.

NICE is one of the three HTA bodies in the UK that provide recommendations and suggestions on how public health and welfare workers can support people by meticulous, independent assessment of complex evidence. In addition to being highly respected, NICE recommendations have made a significant impact on public policy (Scullard et al., 2011). It is important to understand how and what methods have been implemented in NICE to assess clinical effectiveness with single-arm studies. The next chapter will describe the findings of a systematic review on NICE to understand the methods with single-arm studies and also inquire about the legitimacy of these methods under different situations.

Chapter 2

Review of Methods used to Estimate Treatment Effects against Relevant Comparators using Evidence from Single-Arm Studies in NICE Single Technology Appraisals

2.1 Introduction

In HTA, the process of making a decision on the use of a new intervention in the absence of concurrent comparators is not always straightforward. In reality, information on relevant treatments is available from multiple studies, and direct comparison is often nonexistent. Besides, studies that need to be compared can also differ on various aspects like populations, designs, protocols, different doses or timings of delivery of drugs, and so on. HTA is highly reliant on ITCs, which are considered to be in the second highest rank in the hierarchy of evidence for decision making on reimbursement when RCT is not possible (Dias et al., 2013).

An indirect comparison is in anchored form when the comparison is conducted via common comparators which are depicted in diagrams (a) and (b) in Figure 1.1. An unanchored indirect comparison takes place when treatments are compared across studies without considering randomisation within studies. This happens when a common comparator does not exist across studies as depicted in diagrams (d) and (e) in Figure 1.1. In diagram (d), the disconnected network of evidence is comprised of two connected networks, therefore, any pair of treatments taken one from each of these networks will be called an unanchored indirect comparison.

The assessment of the efficacy and safety of a drug is determined by a regulatory body that permits marketing approval to pharmaceutical companies. In HTA, when companies

analyse their intervention treatment that comes from a single-arm study with comparator/comparators, access to IPD in all studies of interest is a rare situation as sharing of clinical data is often limited. A middle-ground situation is more realistic where the pharmaceutical company has access to IPD for its own study and AgD for the comparator studies.

Single-arm studies involve a group of patients who are given the same treatment and it uses unanchored ITC to estimate the relative treatment effect. A particular feature of single-arm studies is that they provide no direct estimate of treatment effects owing to the lack of a comparator arm. Therefore, the use of single-arm studies involves reference to an external comparator. Using an external comparator to estimate a treatment effect can induce several limitations. First, the selection of the comparator is often post-hoc, so the probability that it was intentionally selected to favor the new treatment is non-trivial; second, the use of an external comparator can suffer from biases as no assurance can be given that compared groups are comparable in terms of patients characteristics (Cucherat et al., 2020).

NICE, which assembles guidance based on evidence and suggestions for health, public health and social care practitioners, is a non-departmental public body of the Department of Health and Social Care in England. A single technology appraisal (STA) that relates to a single technology for a single indication facilitates recommendations on making use of existing and newly invented treatments within the NHS (NICE, 2009). As a part of this process, a company is expected to make a comparison, direct or indirect, between its treatment and standard of care or other relevant comparators. It does so using evidence about the benefits, harms, and resources associated with its treatment and comparator by applying statistical methods that it considers appropriate.

An appraisal comprises several documents, including committee papers, appraisal consultation document (ACD), final appraisal document (FAD). Committee papers include company submission (CS), external assessment group (EAG) reports. The ACD is a document that includes the appraisal committee's provisional recommendations to NICE and the FAD produces the appraisal committee's final recommendations to NICE whether the technology is recommended or not. The committee papers discuss the clinical effectiveness and cost-effectiveness of a treatment. The clinical effectiveness section discusses the clinical evidence based on the studies that have been conducted. The cost-effectiveness section discusses the value of the new intervention based on an estimate of the incremental cost per quality-adjusted life year (QALY) gained. Hence, an estimate of the incremental clinical benefit of the new treatment compared to relevant comparators is required.

The aim of this review of NICE STAs was to determine how comparisons against relevant comparators have been performed using evidence about new treatments obtained from single-arm studies when IPD is available partially. Additionally, the appropriateness of the methods that were identified in the review was also appraised. NICE has been selected to conduct this review as it is currently considered to be one of the most pivotal HTA agencies in the world (Sealey et al., 2014). Their decisions on HTA are influential and the analyses are considered valuable.

A systematic review conducted by Phillippo et al. (2019a) characterised the use of population-adjustment methods in TAs submitted to NICE from 2010 to 2018. They focused on how these methods were implemented for any outcome in both anchored and unanchored situations without restricting their search for single-arm studies. Moreover, in the review, 7% (18/268) of the TAs were found to have used population-adjustment methods. Though the search was not restricted to unanchored cases, most of the population-adjustment methods were found for unanchored comparison (89%, 16/18). Eighty-three percent (15/18) of the identified applications were in oncology where 89% (16/18) used MAIC and 17% (3/18) used STC. The criteria that were used for the inclusion of covariates in the adjustment were effective sample size (ESS), expert opinion, availability, cross-validation, or statistical significance. Fifty-six percent (10/18) of the comparisons were conducted for a larger network of evidence where comparisons were made for multiple comparators and/or multiple aggregate study populations.

The current review has focused only on unanchored comparisons with single-arm studies. The time frame of this review was 2018 to 2021. This time frame has been chosen to include more recent STAs information. In addition to the information extracted by Phillippo et al. (2019a), this review has extracted information on additional issues including how prognostic and effect-modifier variables have been identified and how survival extrapolations have been conducted. From the review of Phillippo et al. (2019a), it was evident that most of the application was in oncology with survival outcomes, therefore, information on survival extrapolation was also extracted to assess whether or not proper adjustments had been made. Section 2.2 of this chapter narrates how the review was conducted and Section 2.3 describes the results. The chapter concludes with a discussion of the results in Section 2.4.

2.2 Methods

2.2.1 Inclusion criteria

STAs based on single-arm study and published on the NICE website from 1st January 2018 to 31st December 2021 have been included in this review.

2.2.2 Data extraction

All STAs listed on the NICE website (https://www.nice.org.uk/guidance/published) from 2018 to 2021 were screened one by one to identify relevant STAs involving single-arm studies. The appraisals that had access to IPD from all included studies were discarded from the review and appraisals with partial access to IPD were included. Relevant documents such as committee papers, ACD and FAD have been used to extract information. Committee papers were the main document that has been used for data extraction. If committee papers were not available then ACD and FAD have been used for screening purposes.

A data extraction spreadsheet was created using Microsoft Excel (available in the Appendix A with Table A.1, A.2, A.3, A.4, A.5, A.6, A.7) which was used to collect relevant information about the methods that have been used to make an indirect comparison between different

comparators using single-arm studies. Information has been extracted on the following questions:

1. Name/TA number.
2. Publication date of the appraisal.
3. Protocol number and/or name of the pivotal study.
4. Therapeutic area.
5. What types of outcome measures were indirectly compared?
6. What type of network is being considered?
7. What methodology was used to conduct unanchored indirect comparison?
8. How were covariates included in the model?
9. How many variables were included in the model?
10. Were all identified prognostic and effect-modifier variables included in the model?
11. If not, what was the reason for excluding variables?
12. Other than prognostic and effect-modifier variables, were other variables also included?
13. Were second-order terms included?
14. What were the original sample sizes?
15. What was the subsequent effective sample size?
16. If NMA was conducted, was any attempt made to check if any inconsistencies were found in the connected part of the network?
17. Was heterogeneity among studies assessed?
18. If yes, what was the amount of heterogeneity identified?
19. Were at least two studies available on each contrast for the heterogeneity parameter?
20. Along with the chosen method, were other methods also discussed?
21. Was any justification given for the chosen method?
22. How many events were available for the time-to-event outcome?
23. What approach was used for the extrapolation of time-to-event data?
24. What adjustment was made for time-to-event data?
25. Is overlapping between weighted and reconstructed K-M been checked/ commented on?
26. If not, what procedure was taken to ensure overlapping between weighted IPD and reconstructed IPD?
27. Was the population for the extrapolation clearly defined?
28. Treatment effects were estimated for which population?
29. Had any justification given for transportable treatment effects if they are estimated for IPD population?
30. If PH assumption was made, was it tested for both unadjusted and adjusted comparison?
31. What procedure has been taken to measure uncertainty?
32. What attempt was made to quantify residual bias?

2.3 Results

A total of 260 TAs have been found which have been published between January 2018 to December 2021. Of these identified TAs, 27 appraisals were identified where the pivotal study/studies were single-arm studies. Of these 27 identified TAs, 7 of these had access to IPD from all included studies, so they were excluded and 20 TAs were identified with only partial IPD. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram in Figure 2.1 demonstrates the selection process.

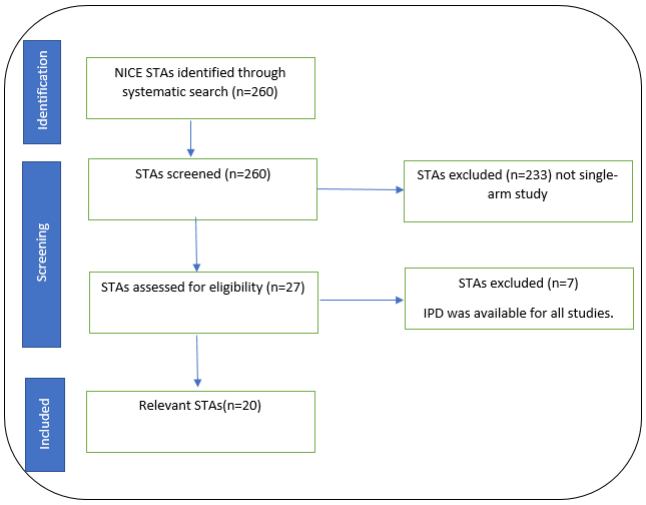


Figure 2.1: *PRISMA diagram for inclusion of relevant NICE STAs with single-arm studies*

2.3.1 Clinical area and outcomes of the published TAs

All the TAs included in this review were found to be in oncology (20 out of 20, 100%). Of these TAs, 16 (80%) have used population-adjustment methods and four TAs did not make any kind of adjustment. As all the TAs were on oncology, most of the outcomes that were indirectly compared were time-to-event outcomes. Only TA 756 (TA756, 2021) used binary outcomes spleen volume reduction (SVR) and total symptom score (TSS). PFS and OS were found to be the most common outcome types used in population-adjusted analyses in the included TAs.

2.3.2 Application of population-adjustment methods

MAIC was the most widely used population-adjustment method (13 out of 20, 65%). STC was also used as a population-adjustment method, but the frequency of using STC was lower than MAIC. STC was used in 5 out of 20 appraisals (25%). Two appraisals used both MAIC and STC (TA592, 2019; TA756, 2021). The application of different methods has been summarised in Table 2.1.

Table 2.1: *Application of different methods for comparison*

Applied method/methods	Number of appraisals
Only MAIC	11
Only STC	3
Both MAIC and STC	2
No adjustment	4
Total	20

Of these 13 TAs which have used MAIC, only 5 (TA510, 2018; TA571, 2019; TA592, 2019; TA604, 2019; TA756, 2021)(38.45%) of them have reported their ESS. Of these, the median ESS was 67.1 (range: 3.8 to 84), with a median reduction in ESS from the original sample size of 50.3% (range: 43.24% to 94.73%). A huge decrease in ESS suggested a scarcity of overlap between the IPD and AgD studies. It means the resulting comparisons were conducted on a restricted number of participants in the IPD study and may be erratic. Estimates become unstable in the sense that, over repeated sampling, the resulting estimates could exhibit high variance as the estimates are placing high weights on a few observations that vary over different samples. The estimates are less reliable as they are effectively based on a few observations. If there exists inadequate overlap between study populations, it can be difficult to obtain reliable estimates of the weights. Moreover, a detailed description of the propensity score model was missing for all the TAs. None of the TAs mentioned whether the propensity score model for MAIC only includes first-order terms or second-order terms were also included. In the propensity score model, the first and second-order terms refer to whether the model was built with only linear or with squaring terms respectively.

Of the identified TAs, 4 did not attempt an indirect comparison (TA529, 2018; TA567, 2019; TA630, 2020; TA644, 2020). In TA529 (TA529, 2018), instead of using evidence from a single-arm study, two RCTs were used to make relative treatment effect estimates as the company claimed that the treatments from the RCTs were similar to the intervention from the single-arm study. In TA567 (TA567, 2019), the company considered it reasonable to assume the study populations homogeneous for making comparisons of outcomes without adjustment which was strongly discarded by EAG as important differences were found in prognostic factors. In TA630 (TA630, 2020), no published data were available for making an indirect comparison as there was no comparator treatment available for the disease. Therefore, the company considers a population for comparison as the comparator arm that is in line with the existing standard of care. In TA644 (TA644, 2020), the intervention study was a basket study where investigators and manufacturers categorised cancer patients with respect to their common genomic alterations by conducting tumour-agnostic studies across multiple solid tumours. Owing to the notable diversity between patient and disease characteristics, tumour types, and potential comparator therapies, a conventional indirect treatment comparison was regarded as infeasible.

Of the thirteen appraisals that have applied population-adjustment with MAIC, more than half of them (10 out of 13, 77%) mentioned that the weighted Cox proportional hazard model has been used to estimate relative effectiveness (TA510, 2018; TA540, 2018; TA554, 2018; TA571, 2019; TA592, 2019; TA628, 2020; TA643, 2020; TA704, 2021; TA722, 2021; TA742,

2021). Despite the fact that Cox proportional models do not specify the baseline hazard, still it is widely used. This unspecified baseline hazard makes the Cox model a semiparametric model. Although the Cox model is popular because of its non-reliance on the distributional assumptions for the outcome variable, in a parametric model, the outcome (survival time) follows a known distribution. When survival estimates are derived from parametric models, the survival plots conform more with a theoretical survival curve in contrast with a Cox-adjusted survival curve where the survival plot is estimated using nondistributional methods. A parametric model allows a complete specification of the survival and hazard functions which in turn give more precise estimates compared to the Cox model.

Other than the weighted Cox model, the application of other measures has also been seen. In TA604 (TA604, 2019), a weighted K-M survival function was used for estimating relative effectiveness. In TA716 (TA716, 2021), mean survival was estimated both for the intervention study and each comparator study independently. This was done by extrapolation of the K-M function using parametric survival curves and calculation of the area under the curve. Mean survival has been estimated as the proportional hazard (PH) assumption was very unlikely to hold for comparisons of the intervention and the comparator's treatment. Other than the weighted Cox proportional hazard model, weighted risk difference has also been used for treatment comparison (TA756, 2021). In TA 756, a weighted binomial model with a logit link was used to combine data from the intervention and comparator study. For the comparator study, IPD was simulated using the published number of events and non-events. The binomial model assigned MAIC weights to the intervention study patients and unit weights to the comparator study patients. Subsequently, the proportion of comparator and intervention study events was predicted from the fitted model, and the difference of these proportions was used as a weighted risk difference.

Three TAs (TA522, 2018; TA525, 2018; TA530, 2018) have used STC with fractional polynomial network meta-analysis (FP NMA). In order to do the analysis, first, the company conducted a STC by incorporating bootstrapping to produce estimates of variability. A bootstrap sample is a random sample with replacement generated from the IPD in the intervention study. Several competing models were estimated with each bootstrap sample and the parameters of each model were estimated. The company states that on average about 1/3 of the patients were not included in each bootstrap sample and called these patients out-of-bag (OOB). The OOB patients were used for cross-validation. Cross-validation was done by comparing the outcomes of the OOB patients with the outcomes predicted from the estimated models. The Cox proportional hazards model was used to develop the regression model informed by baseline covariates. The company simulated a large number of hypothetical individuals based on the reported marginal distribution of the covariates of interest and the correlation from the intervention study. The company also generated the predicted log hazards. The mean of the predicted log hazard and the variance of the log hazard from bootstrap samples were used in the FP NMA model to produce time-varying hazard ratios for each comparator. The different kinds of effect measures using MAIC and STC have been summarised in Table 2.2.

All the appraisals of this review include unanchored comparisons without any common treat-

ment, so the comparisons made were dependent on unexplored amounts of residual bias. Out of 16 TAs that have attempted unanchored comparison, only TA530 (TA530, 2018) has attempted to quantify residual bias by using "out sample " method which was one of the recommended methods by Phillippo et al. (2016).

Table 2.2: *Application of different effect measures*

Effect measure for relative comparison	**Number of appraisals**
Weighted Cox proportional model (using MAIC weights)	10
Weighted risk difference (using MAIC weights)	1
Weighted K-M survival function (using MAIC weights)	1
Mean survival from parametric model (using MAIC weights)	1
Adjusted time varying HR from FP NMA (using STC)	3

2.3.3 Presence of larger networks

Of these 20 appraisals, 11 of them (55%) have a larger disconnected network of evidence, i.e. either an intervention treatment has been compared to more than one comparator treatment where the comparator treatments have come from different studies or an intervention treatment has been compared to one comparator treatment but from multiple studies.

Of these 11 studies that have larger networks (TA522, 2018; TA525, 2018; TA530, 2018; TA510, 2018; TA554, 2018; TA529, 2018; TA571, 2019; TA643, 2020; TA644, 2020; TA704, 2021; TA716, 2021), six of them have used multiple MAICs. Despite having a larger network, TA644 (TA644, 2020) did an unadjusted indirect comparison as it claimed that no adjustment is needed due to the similarity between studies. The TA's that have used multiple MAICs to estimate relative treatment effect with multiple comparators, most of them used the MAIC estimates as stand-alone estimates. TA571 (2019) did a standard pairwise meta-analysis using the MAIC estimates to estimate an overall, pooled estimate.

Three appraisals (TA522, 2018; TA525, 2018; TA530, 2018) used STC to construct a predicted treatment arm for each single-arm study. This predicted arm forms a newly-connected network and then analyses were made with FP NMA. This approach gives the opportunity to make a coherent set of relative effect estimates, but, it adds another additional assumption that no difference exists in prognostic and effect-modifiers among the single-arm studies included in the NMA.

The number of comparator treatments for these larger networks ranges from 1 to 6. Apart from the single-arm study of the intervention treatment, no connected network of evidence

was found for the comparator treatments. The reasons for not getting a connected network were either that most of the comparator studies were also single-arm studies or that no common comparator was found from the RCTs.

2.3.4 Uncertainty estimates in the population-adjustment methods

The TAs included in this review either have conducted unanchored MAIC or STC for population-adjustment. Most of the TAs did not mention what methods were used to take into account uncertainty. Five TAs that have used MAIC, either mentioned sandwich estimator (TA510 (2018)) or bootstrap (TA592 (2019); TA628 (2020); TA704 (2021); TA722 (2021)) to capture uncertainty. Three TAs that have used STC (TA525 (2018); TA522 (2018); TA592 (2019)), mentioned bootstrap as their method to capture uncertainty. Both methods estimate uncertainty from the data and discard the robust assumption about the weights by not treating the weights as fixed and known. However, in bootstrapping, the main challenge is it is computationally very intensive.

2.3.5 Preference of other methods over the chosen method

Of the included TAs, three of them have discussed other methods apart from the method that has been applied. In TA522 (TA522, 2018), STC was applied with FP NMA. MAIC was discussed as an alternative to STC but it was not implemented. The rationale for not conducting MAIC was that the MAIC method requires access to relatively large and complete data sets; often this level of data granularity is not reported within published articles. The company states that one approach is not necessarily favoured over the other between STC and MAIC; STC with bootstrap has the benefit of allowing cross-validation and assessment of the model performance. In TA592 (TA592, 2019), both MAIC and STC were performed as scenario analysis for a single comparison, and for the base case, a naive indirect comparison was applied. The company states that the naive or unadjusted comparison was deemed reasonable for the base case analysis considering the uncertainties associated with both of the indirect comparisons. In TA716 (TA716, 2021), MAIC was chosen over STC because though STC can be applied to a large number of comparators, for multiple outcomes, an outcome equation needs to be determined for each comparison which is often problematic, especially for time-to-event data. As there was a single comparator but many outcomes to be compared, MAIC was a better option as after estimating the weights once, they could be used for multiple outcomes.

2.3.6 Adjustment made in survival extrapolation

For survival extrapolation, 12 TAs (TA525 (2018); TA530 (2018); TA522 (2018); TA510 (2018); TA628 (2020); TA643 (2020); TA554 (2018); TA571 (2019); TA704 (2021); TA716 (2021); TA722 (2021); TA756 (2021)) estimated the absolute survival effects directly based on the unadjusted K-M functions of the intervention study. The absolute effects for comparator treatments were estimated in the cost-effectiveness model by applying the adjusted HR to the intervention K-M function. For STC, time-varying HRs which were estimated from the FP NMA model were applied to the K-M function of the intervention study. In all these TAs, the target population of extrapolation was not clearly defined. Only one TA (TA604 (2019))

applied a two-stage method, where, in the first stage, adjusted K-M survival function was digitised and pseudo IPD were reconstructed for intervention study and in the second stage, parametric models were fitted separately for intervention and comparator study.

2.3.7 Identification and inclusion of covariates

All the adjustment methods in this review were in unanchored form, so all prognostic and effect-modifier variables need to be included in the propensity score model or regression model for MAIC and STC respectively. Out of the 20, for 10 TAs, the strategy for identification of variables was either a literature search or a clinical expert opinion or a combination of both (TA530, 2018; TA510, 2018; TA567, 2019; TA571, 2019; TA522, 2018; TA592, 2019; TA704, 2021; TA716, 2021; TA742, 2021; TA756, 2021). The availability of baseline characteristics reported in both studies was also one of the common approaches for variable selection (TA604, 2019; TA540, 2018; TA716, 2021). Most of the appraisals did not discuss whether the identified variables were prognostic or effect-modifiers. Four TAs (TA643, 2020; TA554, 2018; TA529, 2018; TA722, 2021) did not mention how they identified the important covariates.

In TA571 (2019), the company gathered feedback from five clinicians through interviews and questionnaires where clinicians were asked to point out and rank the variables that they believed to be important in survival outcomes. In TA628 (2020) and TA756 (2021), variables were identified using both clinical feedback and Cox regression for univariate and multivariate settings. Three TAs that used STC for unanchored comparison, used different statistical techniques to choose covariates. Out of several competing models with different combinations of covariates and their interaction terms, the final model was chosen based on the best predictive performance or by the stepwise model selection algorithm or based on the OOB predictive performance.

Except for two appraisals (TA522, 2018; TA525, 2018), the included TAs did not include all the identified prognostic and effect-modifier variables in the final model. The reason for not including all the variables was almost always a lack of availability issue. However, without including all prognostic and effect-modifying variables in the adjustment, the estimates will remain biased. Other than the availability issue, non-convergence of the estimated model was also mentioned as a reason for not including all variables (TA592, 2019).

2.4 Discussion

The objective of the review in this chapter was to identify the methods that have been used in NICE STAs to estimate relative treatment effects with single-arm studies. It is important to assess the appropriateness of the adjustment techniques used by the identified methods. Recent reviews identified that the implementation and reporting of population-adjustment methods in HTA awfully diverse and substandard (Serret-Larmande et al., 2023; Truong et al., 2023).

MAIC was found to be the most preferred and used population-adjustment method for estimating relative treatment effect with single-arm studies. Other than MAIC, STC was

also found to be frequently used with single-arm studies. A larger disconnected network of evidence was found to be a common scenario in HTA which occurs when the intervention treatment from a single-arm study needs to be compared with multiple comparator treatments or the intervention treatment needs to be compared with a single comparator from multiple studies. In both of these cases, the intervention treatment needs to be compared multiple times for which MAIC or STC were implemented several times. Ignoring the fact that both of those methods were developed to infer on the relative treatment effect for a pair of studies, they were used several times. Moreover, few TAs mentioned a robust sandwich estimator and bootstrap to estimate the uncertainty of estimates, however, it was not clear for most of the TAs, how the uncertainty was estimated.

Although MAIC was found to be the most applied population-adjustment method, a considerable reduction was seen in ESS. Adjusting for all imbalanced baseline variables was often discarded in order to avoid a substantial loss in sample size. The issue of small ESS depends on the original sample size. If the original sample size is big, then a reduction of effective sample size may still be enough to conduct a MAIC. However, there is no clear guideline on how much reduction can be termed as "unacceptable". Very small ESS indicates that there was a serious non-overlapping between the IPD and AgD populations which makes the MAIC results unreliable. When there is a lack of overlap, weighting methods like MAIC are not able to extrapolate beyond those observed in the IPD, and may produce an estimate that remains biased. Moreover, the propensity score model in MAIC needs to be correctly specified in order to balance the covariate distributions between intervention and comparator study. Correct specification of propensity score should be done by including all relevant main effects and higher-order terms (Remiro-Azócar, 2022b). None of the MAICs found in this review mention whether the propensity score model includes second-order terms or not.

The uncertainty estimates of the MAIC method also remain unclear as most of the TAs did not provide information on this. Robust sandwich estimator and bootstrapping were mentioned as uncertainty estimators with MAIC for a few TAs. Though both methods are used to estimate uncertainty, bootstrapping is computationally more time-consuming than a robust sandwich estimator. In MAIC without the sandwich estimator or bootstrapping, weights are typically considered case weights. Case weights reflect the number of underlying subjects represented by a data point and are considered fixed and known. This may induce several problems. First, some software packages may round these to whole numbers. The variance would be sensitive to the scale of the weights, i.e., multiplying all the case weights by a constant would (falsely) decrease the variance of the estimator because the calculations would interpret this as representing more subjects in the dataset. As a result, variance will be incorrectly calculated. The use of bootstrapping or sandwich estimator takes the uncertainty into account by assuming that weights are estimated rather than known and fixed.

When STC is applied for any link function other than the identity link, the treatment effect is non-collapsible which means the marginal and conditional effects do not coincide even if there is no confounding and the distribution of covariates is balanced. As a result, STC will produce a systematic bias (Remiro-Azócar et al., 2020). TA525 (2018) that have used STC and FP NMA to deal with a larger network of evidence, a further issue was identified

which relates to the choice of comparator studies. The level of heterogeneity was found to be moderate to high in this appraisal based on I-squared value.

It was found that population-adjustment methods (MAIC and STC) were not always used as the company's base-case analysis when the relative treatment effect was estimated using a single-arm study. They were included sometimes as supportive evidence. Some TAs conducted naive indirect comparisons despite important differences among the prognostic factors. In most of the TAs, the EAG expressed their concern about the fact that important baseline characteristics were not adjusted because of poor reporting by the studies. All identified prognostic and effect-modifiers were not included in the final model either due to their lack of availability in the included studies or convergence issues. This may cause residual confounding, which indicates that the populations being compared may still be considerably imbalanced and the impact of these imbalances on the survival estimates induces substantial uncertainty.

Several TAs found in this review deal with multiple comparator studies with AgD, which means multiple comparisons were required. Current MAIC and STC originally target a single comparison and if a larger network needs to be dealt with, an additional assumption called "shared effect modifier" needs to be satisfied which means an evidence network consisting of a set of treatments will be affected by the effect-modifiers in the same way (Phillippo et al., 2016). TAs that have used MAIC for larger disconnected networks of evidence, conducted multiple MAICs for comparing several comparator treatments. One TA did a standard pairwise meta-analysis using the MAIC estimates to get an overall, pooled estimate. When multiple MAICs are done, each of these comparisons is valid for different target populations and none of the TAs had tried to justify a coherent analysis with a shared effect modifier assumption. Most importantly, during the multiple MAICs, IPD from the intervention study was used multiple times which created additional complexity as independence between the studies was broken. Repeated use of IPD created a correlation between the estimates of treatment effects which remain unaddressed. TAs that used STC and then NMA, have tried to make a coherent synthesis which requires further assumptions in the process, that is, prognostic, as well as effect-modifiers, are balanced across the studies. This assumption was questionable when moderate to high heterogeneity was found (TA525, 2018). The issues found with a considerable reduction of MAIC ESS and dealing with a larger network of evidence are similar to the findings by Phillippo et al. (2019b). Although Phillippo et al. (2016) have identified that current MAIC and STC are unable to be implemented for larger networks without making the additional shared effect modifier assumption, some TAs are still using them for a larger network of evidence.

All the MAIC and STC included in this review were in an unanchored form which assumes that there are no unmeasured prognostic and effect-modifier factors. This assumption is very hard to justify and therefore some suggestions by Phillippo et al. (2016) were given for quantifying residual bias due to unmeasured confounding. Except for TA 530, no other TAs had tried to follow these suggestions. TA 530 has used the "out-sample" method to quantify the residual bias. Moreover, a question was included in the data extraction sheet to highlight whether included studies had "immature data" or not, to assess the appropriate-

ness of extrapolation and the ability to adjust for covariates. No data were available from the included TAs but in most of the TAs, EAG has expressed concern about data immaturity.

The most common strategy of extrapolation found in this review was to fit a parametric model with the unadjusted IPD of the intervention study. The fitted model will give survival probabilities of the intervention treatment in the population of the intervention study. The survival extrapolation for comparator studies was estimated in the cost-effectiveness model by applying MAIC/STC adjusted HR to these survival probabilities. It provides the survival probabilities for the comparator treatment in the population of the intervention study. When extrapolation is done directly to the unadjusted K-M function of the intervention study and comparator treatment extrapolation is obtained by applying HR, the population of extrapolation is the IPD population. This can only be done if treatment effects are transportable. None of the TAs included in this review has given any justification for the transportable treatment effect. Using this extrapolation approach, a PH/AFT assumption is made which assumes a constant treatment effect. It is very important to consider whether this is likely to hold over the entire time horizon. None of the TAs have made any attempt to justify this assumption. As this is unlikely to be true, therefore this approach is not appreciated generally. Only one TA (TA604 (2019)) had applied a two-stage extrapolation approach where MAIC adjusted K-M survival function for intervention study was digitised and IPD data were reconstructed and in the second stage, parametric models were fitted separately for intervention and comparator study.

Many of the findings of this review have similarities as well as dissimilarities with the review by Phillippo et al. (2019a). Similar to this review, most of the applications of population-adjustment in Phillipo's review were also found to be in oncology where survival outcomes (e.g. PFS, OS) were found to be the most common outcome type. MAIC and STC were found to be mostly used methods in both reviews. A substantial decrease in ESS has been found in both of the reviews which in turn made the comparisons dependent on very few numbers of individuals in the IPD study. Similar to Phillipo's review, the application of population-adjustment methods was found to be very prevalent for a larger network of evidence for this review. Furthermore, the range for covariate adjustment is also found to be similar. The number of covariates adjusted for are 2 to 14 and 1 to 13 for this and Phillipo's review respectively.

Whilst the review by Phillippo et al. (2019a) found that the most common approach to identifying prognostic and effect-modifier variables was their availability in the included studies, this review found that the majority of the TAs have used literature review and clinical expert opinion to identify variables. This shows that TAs are trying to follow guidance on variable selection by Phillippo et al. (2016). In Phillipo's review, no TAs were found that had attempted to adjust the residual bias but in this review, one TA was found which has attempted "out sample" method. No information was extracted for survival extrapolation in Phillipo's review, therefore, all the issues identified with survival extrapolation were found in addition to Phillipo's review.

In NICE TAs, adjustment methods like unanchored MAIC and STC are commonly used as

a result of increasing acceptance of single-arm studies for granting marketing authorisations. These methods assume no unmeasured effect-modifiers and prognostic factors which is difficult to satisfy in reality and in turn can produce residual bias. In addition, these methods target the comparator population which is often not the population of interest. Moreover, these methods were frequently used for larger disconnected networks of evidence despite the fact that they were not developed to handle multiple comparator treatments simultaneously. These methods have tried to make a consistent synthesis using the meta-analysis technique without making shared effect modifier assumptions. Therefore, it is important to identify if other methods also exist in the literature that can estimate relative treatment effects using single-arm studies which are also able to address all the issues identified in this review. For this, a second review has been done to find all methods in the literature to estimate relative treatment effects using single-arm studies. The description and findings of the review will be discussed in the next chapter.

Chapter 3

A Systematic Literature Review of Methods for Unanchored Indirect Comparisons with Single-Arm Study

3.1 Introduction

In this chapter, a review has been undertaken to identify methods in the scientific literature for conducting unanchored indirect treatment comparisons with single-arm studies to estimate a relative treatment effect. The interest lies in understanding how and under what situations the methods can be used. As the target of this review is to identify all the relevant methods, a systematic literature review (SLR) has been done which aims to locate and assess all relevant literature on a topic for answering specific questions (Dewey and Drahota, 2016).

Single-arm studies only contain intervention treatment arm, so to estimate relative treatment effect with a comparator, it needs to rely on external studies for the comparator treatment. As no head-to-head comparison is feasible with a single-arm study, an indirect comparison is the only option. Estimating treatment effects poses the danger of bias since the population of interest can be different between single-arm and comparator studies. When relative treatment effects are estimated with a single-arm study, it needs a statistical approach with strong modelling assumptions of conditional constancy of absolute effect. This assumption is very hard to meet as it assumes that being in the intervention study or in the comparator study does not carry over any information about the absolute effect, once conditioning is done on the prognostic variable and effect-modifiers.

In HTA, there exists a data restriction when the relative treatment effect needs to be determined. A pharmaceutical company has access to its own study with IPD but only AgD measures are available for comparator study/studies. Therefore, methods that need access to IPD for all studies are not feasible options. For this reason, the interest lies in methods that can deal with a mixture of IPD and AgD or only AgD when estimating the treatment effect.

When a drug manufacturer company uses a single-arm study to assess the effectiveness of its treatment, the target population is the population of the single-arm study. Additionally, the target is to estimate the marginal treatment effect, which is the average effect at the population level. Conditional treatment effects which are the average effect at the patient level can be estimated also. It may be important in some contexts as they can identify the benefits of treatment for a particular patient with a specific covariate value. The benefits or harms of treatment can be overestimated if conditional estimates are used. As a consequence, justified use of limited resources becomes questionable.

In the previous chapter a review on NICE STAs was done where unanchored MAIC and STC were found to be used in estimating relative treatment effect with single-arm studies. In addition to these two methods, there might exist other methods that can be used for single-arm studies. The objective of this chapter is to identify those methods with a systematic literature search. Section 3.2 of this chapter describes how the review was conducted and Section 3.3 describes the results. Section 3.4 gives details of the methods that could be used for unanchored comparison with a single-arm study. The chapter concludes with a discussion of the methods in Section 3.5.

3.2 Systematic literature review (SLR) methods

3.2.1 Inclusion criteria

Statistical methods that could be used with single-arm studies to estimate relative treatment effects have been included in this review.

3.2.2 Search strategy

In order to identify methods to estimate relative treatment effects using single-arm studies, a 2-stranded approach was implemented. In the first approach, a keyword search was conducted on MEDLINE via OvidSP (1946-present including MEDLINE In-Process) on April 5, 2021. In the second approach citation searching (pearl growing based) was conducted on key journal articles. The aim of the literature search was to identify methods that can deal with single-arm studies in estimating relative treatment effects. In order to make the search broad, instead of restricting to any particular kind of outcome, the search was conducted for any outcome within a single-arm study.

The main consideration when implementing the keyword search was to find out methods that will estimate treatment effect with single-arm study from disconnected networks. Specific phrases including *single-arm study*, *treatment effect*, *observational study*, *indirect comparison* were used in combination with *disconnected networks*, *no head-to-head*, *historical control*, and *network meta-analysis*. The description is provided in Figure B.1 of Appendix B about the search strategy.

3.3 Systematic literature review (SLR) results

3.3.1 Keyword searching

During the keyword searching, 682 references were retrieved. After limiting the search to the English language, 677 references remained. When duplicates were removed, 674 references remained. To identify relevant articles, the titles of the articles were examined and the abstract of each article was examined if the titles were relevant. After removing all the irrelevant articles, 11 articles remained. Of these 11 articles, one was identified as a key paper which was a review article that describes how disconnected networks of evidence were handled in estimating relative treatment effectiveness (Stevens et al., 2018). A pearl-growing literature search was conducted with these 11 articles and eventually, 13 articles were identified as relevant (Béliveau et al., 2017; Caro and Ishak, 2010; Cucherat et al., 2020; Collignon et al., 2020; Goring et al., 2016; Hatswell et al., 2020; Jiang and Ni, 2020; Leahy et al., 2019; Schmitz et al., 2018; Remiro-Azócar et al., 2020; Schmitz et al., 2013; Signorovitch et al., 2010; Thom et al., 2015).

Of these 13 articles, 4 methods were identified that could be used with single-arm study to estimate the relative treatment effect. Random baseline NMA method was discussed in Thom et al. (2015); Goring et al. (2016) and Béliveau et al. (2017) and NMA with matching was identified in Leahy et al. (2019). Matching adjusted indirect comparison (MAIC) was identified from several articles including Signorovitch et al. (2010); Cucherat et al. (2020); Jiang and Ni (2020); Remiro-Azócar et al. (2020). Simulated treatment comparison (STC) was identified from Caro and Ishak (2010); Remiro-Azócar et al. (2020). One method was identified by Collignon et al. (2020) but it was discarded as IPD was needed for all included studies for this method.

3.3.2 Citation searching

In order to avoid missing any relevant articles, a citation search was conducted on the 11 articles redeemed during a keyword search with Google Scholar. One hundred and three cited references were retrieved and one of them was found to be relevant which is a journal article by Phillippo et al. (2020a). Eventually, the method found by citation searching was discarded as the method in Phillippo et al. (2020a) only considered anchored comparison. The PRISMA flowchart is depicted in Figure 3.1 which demonstrates the selection process both with “keyword” and “citation” searching.

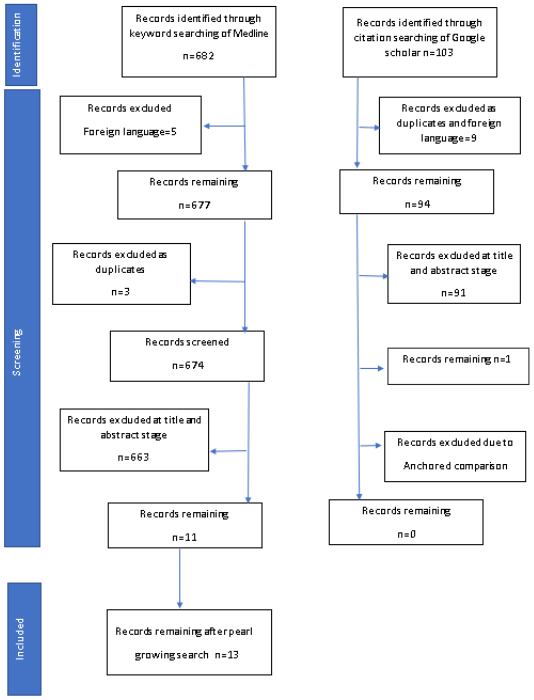


Figure 3.1: *PRISMA flowchart of "keyword" and "citation" searching for single-arm study*

3.4 Methods for conducting ITCs using single-arm studies

In this section, methods are described which were identified with a systematic search to estimate relative treatment effects with single-arm study. Throughout this chapter, methods were described for single-arm studies with binary outcomes but they can be used for other outcomes also. The availability of methods in the literature for different kinds of outcomes has been summarised in Table 3.1. These methods also differ depending on the availability of IPD and AgD. Throughout this thesis, single-arm study/intervention study and comparator study/AgD study will be used interchangeably.

Table 3.1: *Availability of methods for different outcomes*

	MAIC	**STC**	**Random baseline NMA**	**NMA with matching**
Binary Data	x	x	x	x
Continuous data	x	x	x	
Time-to-event data	x	x		

3.4.1 Unanchored matching adjusted indirect comparison (MAIC)

Matching adjusted indirect comparison (MAIC) is a method to balance differences in baseline covariates between an intervention (with IPD) and comparator study (with AgD) (Signorovitch et al., 2010; Cucherat et al., 2020; Jiang and Ni, 2020; Remiro-Azócar et al., 2020). In this section, unanchored MAIC has been described based on Jiang and Ni (2020). The difference is Jiang and Ni (2020) described MAIC for time-to-event data, whereas, here it has been described for binary data. An unanchored MAIC can be applied when there is no treatment arm that is common to connect an intervention to other treatments. The comparator can come from another single-arm study or from one arm of an RCT. MAIC is basically developed to work in a pairwise study setting where the intervention study has IPD and the comparator study has AgD. To apply MAIC, IPD from the single-arm study is re-weighted so that it matches the mean baseline patients characteristics of the comparator study.

The first step in applying MAIC is to calculate MAIC weights, Signorovitch et al. (2010)

$$w_i = \frac{Pr(X_i{=}0|\mathbf{Z_i})}{Pr(X_i{=}1|\mathbf{Z_i})},$$

where w_i represents the odds that patient i gets the comparator treatment versus the intervention treatment. $\mathbf{Z}$ is a vector of baseline patient characteristics, and X specifies the treatment received (X = 1 for intervention study and X = 0 for comparator study). A logistic propensity score model is used to estimate the weights taking into account all known effect-modifiers and prognostic variables. That is,

$$log(w_i) = \beta_0 + \beta_1 Z_i,$$

where β_0 and β_1 are the regression coefficients. In calculating weights, patients who are more likely to have received the comparator treatment versus intervention treatment will be up-weighted to overcome their under-representation in the intervention study; similarly, patients who are less likely to have received the comparator treatment versus intervention treatment will be down-weighted to overcome for their over-representation in the intervention study. The consequence of this weighting is that the covariate distribution of the intervention study will more closely resemble those of the comparator study.

In order to estimate the weight, the maximum likelihood approach needs IPD from both studies. However, when IPD is not available for comparator study (i.e. $X_i = 0$), the likelihood approach is not applicable to estimate the parameters of this model. Method of Moment (MoM) (Signorovitch et al., 2010) or Entropy Balancing (EB) can be used to estimate MAIC weights. MoM optimises an objective function where the covariates in the intervention study are centered on the mean value of the comparator study. The objective function that is minimised is as follows:

$$Q(\beta_1) = \sum_{i=1}^{n} exp(Z_i\beta_1),$$

where $Q(\beta_1)$ is a convex function that can be minimised using any conventional algorithm which will give a unique finite solution. Here, n is the number of participants in the intervention study. After optimising the objective function, the weights are estimated as:

$$\hat{w_i} = exp(z_i\hat{\beta_1}).$$

Other than MoM, EB is also plausible to estimate weights. In estimating weights, EB adds additional constraints so that the estimated weights should be close to unit weights as well as penalise for calculating extreme weight (Hainmueller, 2012). Though EB is another option to estimate weights, it has been proven that estimating weights via either of the methods is mathematically equivalent (Phillippo et al., 2020c).

Eventually, the mean outcome of the intervention treatment in the comparator study is estimated as a weighted average,

$$\hat{Y_i} = \frac{\sum_{i=1}^{n} Y_i\hat{w_i}}{\sum_{i=1}^{n} \hat{w_i}},$$

where Y_i is the outcome for i^{th} patient in the single-arm study. In unanchored cases, the focus is to calculate the absolute outcome for the single-arm treatment. After estimating the weights, pseudo-patient data from the comparator study is needed. This pseudo-patient data need to be simulated and this step is not required for the estimation of weight. Comparator studies usually report the percentage of patients with events that can be used to simulate binary outcome 0/1 to fit a logistic regression model. After simulating the outcomes, a relative treatment effect can be estimated by fitting a weighted logistic model where the weights for the intervention treatment are estimated with MoM/EB and for the comparator treatment the weight is assigned as 1.

A robust sandwich estimator can be used to estimate standard error (SE) for MAIC estimates. The robust sandwich estimator takes into account the fact that weights are not known during the estimation of the uncertainty for weights. Frequentist bootstrapping is a possible alternative that also takes into account all kinds of uncertainty in MAIC weights but it is computationally highly intensive.

In MAIC, ESS can be estimated which represents how many independent individuals are required to provide an estimate with the same accuracy as the weighted sample estimate. In MAIC, the ESS is approximated by (Signorovitch et al., 2010),

$$ESS = \frac{(\sum_{i=1}^{n} \hat{w_i})^2}{\sum_{i=1}^{n} \hat{w_i^2}}.$$

It is likely that this approximation underestimates the true ESS as the weights are correlated with outcome and also they are not known (Phillippo et al., 2016). However, if ESS is small, with respect to the actual sample size, it is an indication that the weights may fluctuate greatly which may cause the estimate to be unstable. Though it is necessary to include all prognostic and effect-modifying variables in the propensity model, the inclusion of too many covariates can result in extreme weights which can cause a reduction in ESS.

To implement the propensity method, there should be an overlap between the adjusted variables in AgD and IPD study. Lack of overlap restricts MAIC to extrapolating beyond the

IPD study which may result in failure to produce an estimate. MAIC assumes that for the AgD study, sufficient information is available for joint covariate distribution which is often unlikely in practice as only marginal covariate distributions are published. The joint covariate distribution can provide information on covariate correlations which then can be balanced by including them in the weighting model. However, in the absence of joint distribution information, they are considered to be identical to the correlations observed amongst covariates in the IPD study which can be wrong. MAIC implicitly specifies an outcome model, so if there is any interaction between treatment and effect-modifiers (one-way or multi-way), misspecifying the correlation between covariates will give a biased indirect comparison (Phillippo et al., 2016).

3.4.2 Unanchored simulated treatment comparison (STC)

Simulated treatment comparison (STC) is a technique that is based on regression adjustment (Caro and Ishak, 2010; Remiro-Azócar et al., 2020). As in MAIC, STC is also applicable to a pair of studies with IPD and AgD. The regression model in STC could be of any kind such as it could be linear, logistic, or any time-to-event regression model if the outcome is continuous, binary, and time-to-event respectively. In the single-arm study, where IPD is available, a regression model of outcomes is fitted as a function of baseline covariates. All prognostic and effect-modifiers need to be incorporated into the model as covariates. Though there are different kinds of formulation of STCs (Caro and Ishak, 2010; Phillippo et al., 2016, 2018), in this section, STC has been described following Phillippo et al. (2016) for binary outcomes.

Suppose, an intervention needs to be compared in a single-arm study with IPD to a comparator treatment with AgD for a binary outcome. A logistic model can be fitted for the single-arm study as follows:

$$g\left(\eta_i^*\right) = \beta_0 + \left(\boldsymbol{Z}_i - \overline{\boldsymbol{Z}}_{agd}\right)\boldsymbol{\beta_1} + \left(\boldsymbol{Z}_i^{(\boldsymbol{EM})} - \overline{\boldsymbol{Z}}_{agd}^{(\boldsymbol{EM})}\right)\boldsymbol{\beta_2},$$

where $g(\cdot)$ is a logit link function for binary outcomes with η_i^* as the expected outcome on the natural outcome scale, β_0 is the intercept, $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ are a vector of K regression coefficients for the prognostic and effect-modifying variables respectively. The covariates are centered at the published mean values from the comparator study. The fitted model can give predicted log odds for intervention treatment in the comparator study. A log odds ratio then can be estimated to get the relative treatment effect. The conventional formula of the variance of a log odds ratio can be used to estimate the variance of the estimate. Detailed information on variance calculation for STC is not available either in the NICE guidance or by the authors (Caro and Ishak, 2010) of STC. However, recent articles of STC show that bootstrapping is being used for variance calculation (Ishak et al., 2015a,b; Remiro-Azócar et al., 2022).

STC needs a good fit of the outcome model similar to any regression adjustment method. In an unanchored STC, all the prognostic and effect-modifier variables need to be included and including or omitting both of these variables incorrectly can produce biased estimates. Furthermore, specifying the true relationship between outcome and covariates is also crucial. However, by assessing the justified relationship between the outcome and included variables, unlike MAIC, STC can extrapolate beyond the observed values of the intervention study.

This conventional STC estimates a conditional treatment effect on linear predictor scale because the regression models the conditional outcome mean on the baseline covariates included as predictors whereas HTA are mainly interested in the marginal effects of treatments. In an unanchored STC, when the predicted log odds from the intervention treatment are estimated using an outcome model, it represents a conditional estimate. However, the log odds from the comparator study is often a marginal estimate and therefore when they are compared with a log odds ratio, the estimate is systematically biased whenever $g(\cdot)$ is not the identity function.

Recently a modification on STC has been proposed known as "parametric G-computation" which simulates the individuals from the comparator study to estimate the treatment effect (Remiro-Azócar et al., 2022). This method predicts the conditional probabilities (i.e., the potential outcomes on the natural scale) for each simulated subject, averages to get a marginal probability prediction, and back-transform to the linear predictor scale (log-odds ratio scale) to perform a comparison with the comparator treatment in that scale.

In this approach, for the purpose of characterising the AgD population, the joint distribution of AgD covariates is approximated under certain parametric assumptions. If the covariate in the comparator study is continuous, a multivariate normal distribution can be used to simulate the values using the mean value from the comparator study and the correlation structure that was found in the intervention study. Correlation needs to be taken from the intervention study because this information is usually not available from published data. A regression model is then fitted to the IPD data without centering the covariates and the coefficients from this fitted model are used to make the absolute prediction of the probabilities of the comparator study patients. These probabilities would have been observed if the intervention treatment was implemented in the comparator study population. These probabilities are estimated in the natural outcome scale and then they are back-transformed to the linear predictor scale by calculating the log odds for the intervention treatment. After calculating the log odds for the intervention treatment, an odds ratio can be estimated which will give the relative treatment effect between the single-arm and comparator study.

3.4.3 Network meta-analysis with random baseline

Network meta-analysis (NMA) is a generalisation of pairwise meta-analysis where the benefits and safety of multiple treatments are compared by combining results of multiple studies (Lu and Ades, 2004; Dias et al., 2011b; O'Connor et al., 2013). An NMA is intended to give more accurate estimates of the treatment effects as opposed to a single direct or indirect estimate (Cooper et al., 2011; Caldwell et al., 2015). NMA is usually performed using a contrast-based approach where one or more treatments are evaluated with respect to a study-specific baseline (Dias et al., 2013). In an NMA, the effect of a treatment is expressed in a model where the relative effect is added to the baseline treatment effect in that study. The baseline treatment effect is regarded as a fixed and nuisance parameter to estimate.

An NMA can either be a fixed or random effects analysis. A fixed effect NMA assumes every study is calculating an estimate of the same parameter whereas a random effects NMA assumes that each study calculates an estimate of a study-specific treatment effect. In the

random effects model, an exchangeability assumption is made which assumes that the study-specific treatment effect is not equal across studies but they are similar and come from a common distribution. In this section, first, a standard NMA will be described following the random baseline NMA and then a recent modification of the random baseline NMA known as the "reference prediction" will be discussed as well.

3.4.3.1 Standard NMA

First, let us discuss a conventional Bayesian random effects NMA for binomial data. In a standard NMA, the effect of a treatment can be modelled as follows:

$$
\begin{aligned}
& r_{jk} \sim \text{Bin}\left(p_{jk}, n_{jk}\right) \\
& \text{logit}\left(p_{jk}\right) = \begin{cases} \mu_k & \text{if } j = 1 \\ \mu_k + \delta_{jk} & \text{otherwise} \end{cases} \\
& \delta_{jk} \sim N\left(d_{t_{1k}t_{jk}}, \sigma^2\right),
\end{aligned}
$$

where r_{jk} and n_{jk} are the number of events and patients in the j^{th} treatment arm of the k^{th} study respectively and p_{jk} is the probability of an event. The quantity μ_k represents the log odds for the baseline treatment in study k, and δ_{jk} is the log odds ratio for treatment j relative to the baseline treatment. δ_{jk} is modelled under a random study effect where σ^2 expresses heterogeneity of treatment effects across studies.

$d_{t_{1k}t_{jk}}$ is modelled by relating them to basic parameters as follows:

$$d_{t_{1k}t_{jk}} = d_{1t_{jk}} - d_{1t_{1k}}.$$

This equation is also known as the consistency equation. In order for the consistency equation to work, a connected network of RCT evidence is needed and the transitivity assumption needs to be satisfied. Treatment 1 is considered as the reference treatment here, relative to which the basic log odds ratios for all treatments are defined. Normally, the reference treatment is assumed to be the most common treatment across the network of evidence. The baseline effects μ_k are considered to be nuisance parameters that are eventually canceled out so that information on the relative treatment effects is only coming within studies. Thus using an independent baseline maintains randomisation within studies. In the Bayesian approach, vague prior distributions are assumed for basic parameters for each treatment and for the between-study variance. Non-informative or weakly informative priors can be assumed for baseline (Dias et al., 2011b). Priors are updated using the data to give the posterior distribution of the parameters.

3.4.3.2 Random baseline NMA

Random baseline NMA is a technique that can be applied to compare an intervention in a disconnected RCT or from a single-arm study with multiple comparators (Thom et al., 2015; Goring et al., 2016; Béliveau et al., 2017). When the response to treatment in a single-arm

study is compared to multiple comparators, its inclusion in an NMA creates a disconnected network as a single-arm study does not have a comparator arm that can connect it with the rest of the network. Random baseline NMAs can include a single-arm study by using a reference treatment as a baseline treatment in all the studies included in the network. This allows the contrast between disconnected treatments to estimate the posterior distribution which will be able to update itself from the prior distribution. Random baseline NMA will be described here from a Bayesian approach combining IPD and AgD although it can be conducted with a frequentist approach also.

Statistical model for the data The model has been described in two connected parts: the modelling of the IPD has been described in part I and modelling of the AgD in part II.

Part I: Modelling the IPD
IPD is available from the single-arm study which could be modelled as:

$$y_{ijk} \sim \text{Bernoulli}\left(p_{ijk}\right),$$

where y_{ijk} is the i^{th} patient in the j^{th} treatment arm of the k^{th} study and p_{ijk} is the individual probability of an event.

Part II: Modelling the AgD
The AgD which is available for the rest of the network could be modelled as:

$$r_{jk} \sim \text{Bin}\left(p_{jk}, n_{jk}\right),$$

where r_{jk} and n_{jk} are the number of events and patients in the j^{th} treatment arm of the k^{th} study respectively and p_{jk} is probability of an event. In random baseline NMA, a random effect is placed on the reference/baseline treatment, such that any arm of a connected RCT, disconnected RCT, or single-arm study can be connected to the reference treatment.

For binary outcomes, the log odds of event on the reference treatment is modelled as

$$\begin{aligned}\text{logit}(p_{jk}) &= \mu_{1k} + \delta_{jk} \\ \mu_{1k} &\sim N(\mu, \sigma_{\mu}^2),\end{aligned}$$

where μ_{1k} is the overall reference treatment 1 rather than study-specific baseline treatment and δ_{jk} is the treatment effect in j^{th} treatment arm in k^{th} study. It is not necessary to make treatment 1 the baseline treatment, it could be any treatment that is common to the network of evidence. As a prior, uniform distribution can be defined for σ^2 and normal distribution can be assumed for both $d_{t_{1k}t_{jk}}$ and μ. This method is built on considering the exchangeability of reference log odds across studies that contradict the independent baseline model.

Across study covariate adjustment on baseline effect
The model described above can be extended to account for between-study heterogeneity by including adjustment of covariates for baseline which can result in an improved assessment of the baseline effect for the single-arm study.

$$\begin{aligned}
&\text{logit}(p_{jk}) = \mu_{1k} + \zeta \bar{z}_{jk} + \delta_{jk} \\
&\mu_{1k} \sim N(\mu, \sigma_\mu^2) \\
&\delta_{jk} \sim N\left(d_{t_{1k}t_{jk}}, \sigma^2\right),
\end{aligned}$$

where ζ is the effect of the mean covariate value $\bar{z}_{jk}$ that takes into account between study diferences. Prior distributions for the parameters μ, σ_μ^2 and $d_{t_{1k}t_{jk}}$ can be assigned as described before while a vague normal distribution can be assigned as a prior for the coefficient ζ.

Across study and within study covariate adjustment on treatment effects
Covariate adjustment on treatment effect or including treatment covariate interaction enables adjustment for effect-modifier variables. Such models allow for estimating the efficacy of treatment in patient subgroups. For AgD studies, the model can be defined as follows:

$$\begin{aligned}
&\text{logit}(p_{ijk}) = \mu_{1k} + \zeta \bar{z}_{jk} + \delta_{jk} \\
&\mu_{1k} \sim N(\mu, \sigma_\mu^2) \\
&\delta_{jk} \sim N\left(d_{t_{1k}t_{jk}} + \phi \bar{z}_{jk}, \sigma^2\right),
\end{aligned}$$

where ϕ is the effect of a continuous covariate z_{jk} on treatment. z_{jk} is a study-level covariate that enters the model through random effects and is able to estimate the extent to which it can account for the variability of the treatment effect.

For the single-arm study for which IPD is available, the covariate adjustment for treatment effects can be done within the study level as follows:

$$\begin{aligned}
&\text{logit}(p_{jk}) = \mu_{1k} + \zeta \bar{z}_{jk} + \psi(z_{ijk} - \bar{z}_{jk}) + \delta_{jk} \\
&\mu_{1k} \sim N(\mu, \sigma_\mu^2) \\
&\delta_{jk} \sim N\left(d_{t_{1k}t_{jk}} + \xi(z_{ijk} - \bar{z}_{jk}), \sigma^2\right),
\end{aligned}$$

where ξ is the effect of the covariate z_{jk} on treatment within the study level for which a vague normal distribution can be assigned as prior.

Although the use of random baseline NMA can be a way to connect a single-arm intervention with a set of comparator treatments to estimate the relative treatment effect, its main drawback is the assumption of exchangeability that assumes similarity of the reference arm throughout the network which can interfere with randomisation. The mean of the random baseline is drawn towards a common mean which can generate a biased estimate of the relative treatment effects (Dias and Ades, 2016).

In the random baseline model, the absolute effect of reference treatment is modelled which can be utilized to evaluate absolute effects for any treatment in the network. Information on the baseline is utilized onto the relative effect parameters. When the absolute effects are misspecified, it can make the relative effect estimates misleading. As entry of less severe

patients is permitted in recent studies, it will make the outcome of the control arm look better but it can be that the relative effect has stayed constant. When both baseline and relative effects are estimated simultaneously by placing two random effects on them, the baseline effects in previous studies will be shrunk upwards, which will underestimate relative effects. The baseline effects in recent studies will be shrunk downwards which will overestimate relative effects in recent studies (López-López et al., 2017).

3.4.3.3 Reference prediction model

Reference prediction, introduced by Thom et al. (2022) is a modifications of random effects on baseline model. This method will be described in the Bayesian framework. In this reference prediction model, studies are categorised into different groups. RCTs that are connected to the reference treatment make one group ($k = 1, 2, ..., n_s$), RCTs that are not connected to reference treatment make another group ($k^{'} = 1, 2, ..., n_s^{'}$) and single-arm studies are formed into another group ($k^{''} = 1, 2, ..., n_s^{''}$). RCTs that are connected to the reference treatment are modelled using standard contrast-based NMA with independent baselines as follows:

$$\text{logit}(p_{jk}) = \mu_k + \delta_{jk}.$$

RCTs that are connected to the reference treatment have a second use where arms on the reference treatment are used in a meta-analysis for prediction of the reference arm.

$$\begin{aligned} r_{1k}^{ref} &\sim \text{Bin}(p_{1k}^{ref}, n_{1k}^{ref}) \\ \text{logit}(p_{1k}^{ref}) &= \mu_k^{ref} \\ \mu_k^{ref} &\sim N(\mu, \sigma_\mu^2). \end{aligned}$$

With this modelling approach, the connected network is prevented from being biased. The prediction for reference treatment from this model is then used as baseline treatment for single-arm and disconnected RCTs.

$$\begin{aligned} &\text{logit}(p_{jk''}) = \mu_{1k''} + \delta_{jk''}, \quad \text{for single-arm study} \\ &\text{logit}(p_{jk'}) = \mu_{1k'} + \delta_{jk'}, \quad \text{for disconnected RCTs.} \end{aligned}$$

If a random effects model is used for treatment effects, the impact of the single-arm studies or disconnected RCTs is avoided on the estimation of heterogeneity parameter σ^2 in connected RCTs. This modification was done in order to prevent biased estimates in connected RCTs by estimating separate estimates of σ^2 for connected RCTs, single-arm studies and disconnected RCTs. It also uses the σ from connected RCTs as an informative prior in estimating σ for disconnected RCTs and single-arm studies. This prior is a normal centered on σ, truncated below at 0 and standard deviation equal to that of the Markov Chain Monte Carlo (MCMC) samples of σ from the independent baselines NMA.

Inclusion of covariates in reference prediction model
An additional modification was also included by Thom et al. (2022) in reference prediction which is the inclusion of covariates in predicting the reference log odds as follows:

$$\begin{aligned} r_{1k}^{ref} &\sim \text{Bin}(p_{1k}^{ref}, n_{1k}^{ref}) \\ \text{logit}(p_{1k}^{ref}) &= \mu_k^{ref} \\ \mu_k^{ref} &\sim N(a_1 + \boldsymbol{\beta} z_i^{ref}, \sigma_\mu). \end{aligned}$$

For single-arm studies, this could be defined as:

$$\begin{aligned} \mu_{1,k''} &\sim N\left(a_1 + \boldsymbol{\beta} \boldsymbol{z}''_{k''}, \sigma_\mu\right) \\ \text{logit}\left(p_{k''}\right) &= \mu_{1,k''} + \delta_{k''}. \end{aligned}$$

For disconnected RCT studies, this could be defined as:

$$\begin{aligned} \mu_{1,k} &\sim N\left(a_1 + \boldsymbol{\beta} \boldsymbol{z}'_{jk'}, \sigma_\mu\right) \\ \text{logit}\left(p_{j,k'}\right) &= \mu_{1,k'} + \delta_{j,k'} \text{ for all } t_{i'k}, \end{aligned}$$

where $\boldsymbol{z}_{ik}$, $\boldsymbol{z}''_{i''}$, $\boldsymbol{z}'_{i'k}$ are covariates from connected RCTs, single-arm studies and disconnected RCTs, respectively and β is a vector of regression coefficients each with vague priors $\beta_l \sim N(0, 10^2)$.

3.4.4 Network meta-analysis by matching

In order to estimate the relative treatment effect of an intervention in a single-arm study with multiple comparator treatments, an NMA can be performed with connected evidence where the single-arm study can be included in the NMA by matching. In order to include the single-arm study, it can be matched to any arm of the connected network with similar patient characteristics including both effect-modifier and prognostic variables. When a matched arm is identified, then the single-arm and matched arm are treated as if they came from the same study (Schmitz et al., 2018; Leahy et al., 2019). In this section, first, NMA with matching will be described and then a recent modification of the method known as the "aggregate level matching" (ALM) will be discussed as well.

Let an intervention in a single-arm study which needs to be compared with a set of comparators. The network of comparators can include both RCTs and single-arm studies and this method currently uses AgD data from all the studies. In order to implement the method, the model that has been described in Leahy et al. (2019) is as follows:

$$\begin{aligned} r_{jk} &\sim \text{Bin}\left(p_{jk}, n_{jk}\right) \\ \text{logit}\left(p_{jk}\right) &= \begin{cases} \mu_k & \text{if } j = 1 \\ \mu_k + \delta_{jk} & \text{otherwise} \end{cases} \\ \delta_{jk} &\sim N\left(d_{t_{1k}t_{jk}}, \sigma^2\right), \end{aligned}$$

where r_{jk} and n_{jk} are the number of events and patients in the j^{th} treatment arm of the k^{th} study respectively and p_{jk} is the probability of an event and treatment 1 is considered as the reference treatment. The quantity μ_k represents the log-odds for the baseline treatment in study k and δ_{jk} is the log odds ratio for treatment j relative to the baseline treatment. δ_{jk} is modelled under a random study effect where σ^2 expresses heterogeneity of treatment effects across trials. Instead of selecting vague priors for the parameters such as $N\left(0, 100^2\right)$, the priors can be chosen as $\mu_i \sim N\left(0, 1.83^2\right), d_k \sim N\left(0, 1.83^2\right)$, and $\sigma \sim \text{Unif}\,(0, 2)$ for two arm studies (Leahy et al., 2019). The reason for discarding the vague prior is that the vague priors are actually not vague on the inverse logit scale, as most of the distribution is close to either 0 or 1 and the chosen one will have an approximate uniform distribution on the log odds ratio (Leahy et al., 2019).

When selecting the matched arm for the single-arm study, it is important that the patients for the chosen comparator arm should be as similar as the single-arm study otherwise it could introduce bias in estimating relative effect. The matching arm for the single-arm study can be any arm from the network but the treatment should not match with the treatment in the single-arm study. Let M be the number of prognostic and effect-modifier variables that were considered for matching, let z_{m_l} be the patient proportion (for binary covariate) or mean value (for continuous covariate) related with the covariate in the single-arm study with treatment l, and let $z_{m_{jk}}$ be the patient proportion or mean value related with the covariate in arm j of study k. The difference can be estimated as, $\Delta_{jk,l} = \sum_{m=1}^{M} \left| z_{m_{jk}} - z_{m_l} \right|$. The arm with which this difference is minimum can be selected as the matched arm.

The approaches that can be used for the inclusion of a single-arm study in this NMA are:

Pooled model: This is the simplest approach where the single-arm study and its matched arm are included in the NMA as if they form a RCT evidence. No distinction is made between different forms of evidence i.e. the model cannot tell which evidence is single-arm and which evidence is RCT. Using the notation from Leahy et al. (2019), the matched part can be written as,

$$\text{logit}\left(p_l\right) = \mu_{\,\text{ChosenRCT[k]}} + \delta_l.$$

ChosenRCT[k] indicates the chosen study to match to single-arm study l.

Hierarchical model: In this case, in addition to considering the single-arm study and its matched arm as an RCT, a hierarchy is placed on the treatment effects. Let d indicate the overall effect of treatment relative to the reference treatment.
The extra level can be modelled as,

$$\begin{aligned} d_{\text{RCT}[t]} &\sim N\left(d_{[t]}, \sigma^2_{\text{design}}\right) \\ d_{\text{MATCHED}[t]} &\sim N\left(d_{[t]}, \sigma^2_{\text{design}}\right), \end{aligned}$$

where $d_{\text{RCT}[t]}$ and $d_{\text{MATCHED}[t]}$ are the actual effect of treatment t with respect to the reference treatment for RCT and matched studies, σ_{design} indicates the variability between the RCT and matched studies. The idea is to estimate the treatment effect at various levels and

then estimate a pooled effect of the overall treatment effect. In a hierarchical model, it is also possible to down-weight specific evidence type and when σ_{design} is set to 0 then it becomes a naive pooling model. The previous prior distributions are also applicable here. In a hierarchical model, inflating the variance with a multiplicative factor can control overprecision for the matched arm.

Plug-in estimator model: For the plug-in estimator model, it is assumed that for the single-arm study and its matched arm, the log odds of the reference treatment is equal. Unlike the hierarchical model, this model also assumes no difference between the different designs of study. For the RCT part, the model can be written as

$$\text{logit}\left(p_{jk}\right) = \mu_k + \delta_{jk},$$

and the matched part is written as

$$\text{logit}\left(p_l\right) = \mu_{\text{ChosenRCT[k]}} + \delta_l.$$

The pooling method fails to take into account uncertainty by assuming no difference between study designs. The hierarchical modelling approach acknowledges this uncertainty and model difference between study designs by random effects. For this reason, the hierarchical modelling approach is better than naive pooling. The arm that has the minimum difference with the single-arm study with respect to baseline characteristics needs to be chosen as the matched arm but it can be that all arms do not have information on selected characteristics. In that case, an appropriate match may not be found. Recently, a modification has been made to this method which will be discussed in the next section.

3.4.4.1 Aggregate level matching (ALM)

A modification has been done on NMA by matching method by Thom et al. (2022). In this modified method, the reference arm for the single-arm study is chosen from an RCT which is the closest to the single-arm study with respect to patient characteristics as before but Euclidian distance has been used to assess the similarity between the intervention and comparator arm. In addition, the heterogeneity parameter has been disconnected from the connected part and the estimate of the heterogeneity parameter from the connected part is used as an informative prior for the disconnected part. This ensures that the connected part of the network is not affected by the disconnected part and randomisation is kept intact in the connected part. Furthermore, as the matched reference arm for a single-arm study is conducted at the modelling level, unlike reference prediction, data from the connected part is not used multiple times (Thom et al., 2022).

3.4.5 Comparison among methods

3.4.5.1 Assumptions

Each of the methods described in this chapter makes different kinds of assumptions which are depicted in Table 3.2. The idea of constructing this table was taken from (Phillippo et al., 2016). All of the methods make the assumption of conditional constancy of absolute

effects. This assumption is not testable and it is difficult to satisfy in practice. The exclusion of unmeasured prognostic and effect-modifiers can introduce residual bias. Random baseline NMA and NMA with matching both make additional assumptions of transitivity and consistency among studies which assume that prognostic and effect-modifiers are evenly distributed across studies. In addition to conditional constancy of absolute effects, random baseline NMA also makes exchangeability of placebo effect which relies on expert opinion and is not testable statistically (Thom et al., 2015). NMA with matching has to satisfy the assumption of exchangeability between matched arms which means discrepancies between single-arm and other study/studies should be small. The accuracy and precision of the estimates are affected when the difference between studies increases.

Table 3.2: *Assumptions required by different methods*

	Methods			
Assumptions	**MAIC**	**STC**	**Random baseline NMA**	**NMA with matching**
Conditional constancy of absolute effects	x	x	x	x
Transitivity			x	x
Consistency			x	x
Exchangeability of placebo effect			x	

3.4.5.2 Differences in estimands

When a method is applied with a single-arm study to estimate the relative treatment effect, the objective is always to mimic what would have happened if the comparator treatment was compared with the single-arm treatment in a head-to-head study. After trying to imitate an RCT, the next question is what kind of estimand should be estimated. In an RCT, both marginal and conditional estimand can be estimated. A marginal treatment effect is the average effect of treatment on the population whereas a conditional treatment effect is the average effect of treatment on the individual level (Austin, 2011). The difference between the mean outcome of the randomised groups in an RCT is assumed to be a marginal effect as it is believed that the difference in the estimate is the result of moving all the patients from one treatment to another. On the contrary, conditional effect describes how the mean outcome of a treatment changes across patients who have the same patient characteristics.

It is often the case that marginal and conditional estimates are thought to be synonymous with unadjusted and adjusted estimates respectively, which is not always true. A marginal effect can also be a covariate-adjusted outcome model (Remiro-Azócar, 2021). Conditional treatment effects may be important in some contexts as they can identify the benefits of treatment for a particular patient with a specific covariate value, but HTA is mainly interested in the marginal effects of treatments. In addition to the variation in estimands, the collapsibility of the effect measure is another issue that needs to be emphasized. An effect

measure is collapsible when the population effect measure can be estimated as a weighted average of the subgroup-specific measures in the absence of confounding bias (Miguel et al., 2023). Collapsibility is mainly related to the linearity issue, which is why the risk ratio and the risk difference are collapsible effect measures but this is not the case for the binary outcome or time-to-event outcome as both of them need to use link function to assume linearity (Greenland, 1987; Austin, 2014, 2011).

MAIC and STC both are covariate-adjusted indirect comparisons but MAIC uses a weighted regression of outcome whereas STC uses a multivariable regression. In an unanchored MAIC, a covariate-adjusted marginal estimate is typically compared with an unadjusted marginal estimate of the comparator study. The version of STC that is described in Phillippo et al. (2016), targets a conditional treatment effect and will be a biased estimate for the marginal treatment effect due to the non-collapsibility issue. The version of STC that uses parametric G-computation or model-based standardisation can produce a marginal effect estimate which is covariate-adjusted Remiro-Azócar (2021).

When multiple studies are synthesised in an NMA, it is quite possible that there exists variability between the study-specific estimands. There could be variability in outcome definition and treatment implementations, for example, in the formulation of dosing, mechanism of delivery, co-treatment regimens, and so on. There may be variability in collecting patient characteristics between studies as well as published reports can be very vague in explaining their target estimand. Due to these inconsistencies, it is difficult to come up with a conclusion for estimand in an NMA (Remiro-Azócar, 2022a; Russek-Cohen, 2022).

3.4.5.3 Target population

Population-adjustments methods like MAIC and STC are formulated in such a way that they target a particular population which is the comparator population. This may create problems as often the target population is the intervention study population which may differ from the comparator population. These methods want to predict what would have happened if the intervention treatment had been applied to the comparator study population. This nature of MAIC and STC can often lead to contradictory results. Suppose there are two pharmaceutical companies who want to get a relative treatment effect of their drug with the other company. Both companies have access to IPD for their own study but AgD from the other comparator study. It is possible that they conduct a MAIC or STC where the target study is the other company's study and they can get a completely opposite result (Phillippo et al., 2016). In standard NMA, studies are included for a joint synthesis where relevant studies are identified using the PICO criteria in a systematic literature search. Therefore, in standard NMA, the target population is the sample of participants who were enrolled in the included studies and the estimates from an NMA are not interpretable outside the NMA studies (Dahabreh et al., 2019).

3.4.5.4 Adjustment of prognostic and effect-modifier variables

Identification and adjustment of all prognostic and effect-modifier variables is an important assumption for the methods used to undertake ITCs with single-arm studies. Adjustment of

all prognostic and effect-modifier variables is not always easy to meet as difficulty can arise in specifying effect-modifier status with new treatments for which clinical knowledge and empirical evidence may not be sufficient. In MAIC/STC all prognostic and effect-modifier variables are adjusted using the propensity/regression model, otherwise, residual bias will affect relative treatment comparison.

In the random baseline NMA, prognostic and effect-modifier variables can be adjusted by conducting a within-study and between-study covariate adjustment on treatment but it is often the case that an adequate number of studies are not available for each treatment effect for conducting the between-study covariate adjustment. Between-study covariate adjustment/interaction effect is harder to detect than within-study interaction as the former needs to differentiate the interaction effect from the random noise whereas the latter only needs to differentiate the interaction effect from sampling error (Dias et al., 2011a). Also, between-study covariate adjustment or meta-regression can suffer from ecological bias where the coefficient of linear regression from patient-level data and study-level data can be very different and sometimes can be completely opposite (Hoaglin et al., 2011). Additionally, though meta-regression allows investigating the effect of both continuous and categorical factors on effect measure, the evidence quality from meta-regression is equivalent to non-randomised studies (Dias et al., 2011a).

In NMA by matching, this adjustment is done by finding an arm that matches with the single-arm study for both effect-modifiers and prognostic variables. However, this matching can be difficult if information on prognostic and effect-modifier variables is not present across the network. In addition, the reference prediction model and ALM model cannot adjust for treatment effect-modifiers, they can only take into account prognostic variables.

3.4.5.5 Appropriateness of methods in different situations

3.4.5.5.1 Two competing treatments

MAIC and STC target in comparing one intervention with a comparator treatment. Both the intervention and comparator can come from a single-arm study or the intervention treatment could be from a single-arm and the comparator treatment from one arm of an RCT. When two competing treatments need to be evaluated, MAIC is a commonly used method that needs sufficient overlap of the input space. STC is a regression adjustment method, therefore it can extrapolate beyond the observed covariate values observed in the IPD population using the linearity assumption or other appropriate assumptions about the input space.

3.4.5.5.2 Multiple competing treatments

Random baseline NMA and NMA with matching can compare an intervention with a single-arm study to a larger network of multiple comparators. The larger network can include both RCT and single-arm studies. NMA-based methods (random baseline NMA and NMA with matching) can deal with larger networks of evidence, given those covariates among studies are balanced. They include non-randomised evidence in the network, but intact the randomisation of the available RCTs. MAIC/STC was originally designed to compare two competing treatments. They are not appropriate for larger networks of evidence as they focus on a different target population in every single comparison, which is the population of the

comparator study. However, they can be extended for a larger network of evidence by making an additional assumption called "shared effect modifier". This assumes that the competing treatments belong to the same class with similar clinical properties and they share the same set of treatment effect-modifiers which allows relative treatment effects to be interpreted into any population. This assumption is hard to meet in practice and is untestable.

3.4.5.6 Situations for the suitability of the methods

In the review of the NICE STA (Chapter 2), it was found that a common scenario in HTA submissions is a larger disconnected network of evidence. The advantage of NMA-based methods is that they are able to handle a larger network of evidence with multiple comparator treatments. Random baseline NMA and NMA with matching can both estimate treatment effects using single-arm studies. A recent study by Thom (2020) found that random baseline NMA may be 'safer' with single-arm studies as it is more conservative. However, the validity of NMA-based methods depends on various issues. Though random baseline NMA is found to be better than NMA with matching, its main criticism lies in the exchangeability assumption which could generate a biased estimate. Additionally, if there exist only a few studies with small sample sizes for each treatment in the network, then random baseline NMA can produce inconclusive results with a wide confidence interval (Thom et al., 2015). Furthermore, in NMA with matching, an appropriate match may not be found for the intervention of a single-arm study if all other arms in the network do not have information on prognostic and effect-modifier variables. Moreover, in terms of matching, the arm with a minimum difference is considered to be a good match, but there are not sufficient guidelines on how much similarity can be considered a good match.

NMA-based methods require that the larger disconnected network of evidence should consist of a connected and a disconnected part with multiple studies per comparison for both parts. In HTA with single-arm studies, this requirement is not easy to meet. The performance of ALM greatly depends on the matched study whereas that of reference prediction greatly on the similarity of the connected and disconnected evidence. However, despite the similarity between studies, reference prediction can still give non-informative and highly variable treatment effect estimates (Thom et al., 2022). Furthermore, random baseline NMA is able to use IPD and AgD but reference prediction is still not available for IPD and AgD. Both Reference prediction and NMA with matching can only use aggregate data from all the studies. In addition to assuming transitivity and consistency assumptions, the NMA-based methods also assume that each effect-modifier modifies the relative treatment effect in the exact same way across all treatments which is often not practical and untestable (Harari et al., 2022).

When an NMA is conducted, the heterogeneity variance needs to be estimated which represents the extent of variation between study-level relative effects on each comparison. With NMA-based methods, an estimate of the between-study heterogeneity parameter is only reasonable when several external studies of each comparator treatment are available. When the number of external studies is very few, estimates of between-study variability need to be justified by using an informative prior.

When a relative treatment effect needs to be estimated between a pair of studies, MAIC

and STC are the preferred approaches. MAIC and STC cannot handle a larger network of evidence unless the added assumption shared effect modifier is made. If this assumption can be justified, both methods can be used for a larger network of evidence. The main problem in implementing MAIC is that as it is a reweighting method, therefore, it is unable to extrapolate when there is an absence of overlap. MAIC produces biased estimates unless the IPD study contains the AgD study completely within it. This is often not the case in practice, so MAIC performs poorly (Phillippo et al., 2020a). Therefore, ensuring good overlap between studies is an essential pre-requisite for MAIC. Although STC can work well with extrapolation as it is a regression adjustment method, any outcome other than continuous, STC produces a systematic bias due to the non-collapsibility issue. However, the modification of STC is able to overcome this issue by simulating covariates from the comparator study and making predictions on outcome scale (Remiro-Azócar et al., 2022). Another issue with MAIC/STC is that these methods estimate comparative evidence for the AgD study population which is often not the target population in HTA.

3.5 Discussion

The aim of this chapter was to identify methods from the literature that could be implemented with single-arm studies to estimate relative effectiveness with one or multiple comparator treatments. From the literature, four methods were identified which could be used along with their advantages and disadvantages. No single method turned out to be appropriate for making comparisons in all situations as every method is suitable for different scenarios.

Unanchored MAIC and STC are useful for estimating treatment effects in a pair of studies but they are not suitable for a larger network of evidence without making additional assumptions. The validity of MAIC strongly depends on the extent of overlap between studies. MAIC is unable to extrapolate when low overlap between study variables exists which can result in complete failure to produce any weights to compare treatments. The main requirement of STC is to fit a regression model correctly with appropriate estimands. Moreover, failure to adjust for all effect-modifier and prognostic variables can result in residual bias. Therefore, if a comparison needs to be made between an intervention and a single comparator, MAIC and STC are preferred methods when low overlap, the non-collapsibility issue, and inclusion of all prognostic as well as effect-modifying variables can be addressed properly. They can be extended to a larger network of evidence if the shared effect modifier assumption can be considered practical.

Two methods were found that could be used with a larger disconnected network of evidence with single-arm studies but those methods are not ideal in every situation. NMA-based methods need to satisfy a lot of assumptions called transitivity, consistency along with the stronger assumption of conditional constancy of absolute effects. Although random baseline NMA was found to be safer than NMA with matching, its exchangeability assumption of the reference treatment is often not practical. Information on prognostic and effect-modifier variables for all arms in a network is essential for NMA with matching/ALM. Failure to do so can produce biased results. Additionally, for NMA with matching/ALM, there is no clear

consensus on the threshold of similarity between variables.

From the literature it was found that the performance of NMA-based methods in a larger disconnected network of evidence greatly depends on the presence of a connected and a disconnected part along with the size of the network. In both the NMA-based methods, the prerequisite for the inclusion of single-arm studies in a larger disconnected network of evidence was that the network should contain a larger connected and disconnected evidence part with multiple studies per comparison in both parts. In NMA-based methods, the inclusion of single-arm studies is acceptable as information on relative treatment effects from the connected network of evidence flows through the whole network which helps to estimate relative treatment effects with single-arm study treatment. This is a crucial condition that the NMA-based methods need to satisfy. However, in the NICE review, it was found that the relative treatment effects of a single-arm study intervention need to be compared in a larger disconnected network of evidence where several single-arm studies mostly formed the larger disconnected network of evidence. Therefore, the presence of a connected part with a sufficient number of studies per comparison was found to be very rare.

In Chapter 2 of NICE STA review, no application of NMA-based methods was found perhaps due to the strict conditions with NMA-based methods. In NICE review chapter, population-adjustment methods MAIC and STC were found to be applied multiple times to obtain relative treatment effects and then used in an NMA to generate combined effects of treatments simultaneously in a larger disconnected network of evidence. Between MAIC and STC, the use of MAIC was found more frequent than STC where MAIC has been used multiple times with multiple comparators or with a single comparator from multiple studies in a larger disconnected network of evidence. The relative effect estimates from these multiple MAICs have been used in a meta-analysis and were described as if the MAIC estimates made a set of coherent relative effect estimates without assuming transitivity of these estimates from different target populations. Additionally, using the IPD several times for conducting multiple MAICs should be taken into account as it breaks the independence of the unit of analysis assumption.

Various reviews can be found in the literature which has assessed the rationality of population-adjustment methods, however, no clear consensus can be seen on the superiority of a single method (Phillippo et al., 2018; Stevens et al., 2018). As a result, a number of simulation studies have been published where MAIC has been appraised against standard ITC in a connected network of evidence. However, it has been found that MAIC is greatly affected by varying covariate overlap between studies, small sample size, and effect-modifier levels (Belger et al., 2015b; Kühnast et al., 2017; Phillippo et al., 2019a; Petto et al., 2019; Hatswell et al., 2020). Due to the frequent use of MAIC in HTA to estimate the relative treatment effect of a single-arm study with multiple comparators or with a single comparator from multiple studies, a simulation study has been designed to assess the appropriateness of MAIC in a larger disconnected network of evidence. Additionally, the simulation study has assessed the impact of varying covariate overlap between studies, sample sizes, and effect-modifier levels mentioned earlier in addition to varying prognostic variables and correlation in covariates in an NMA setting with MAIC. The results of the simulation study will be discussed in the next two subsequent chapters.

Chapter 4

Simulation Study with a MAIC-Adjusted Fixed Effect NMA

4.1 Introduction

In the review of NICE STA (Chapter 2), it was found that methods like MAIC and STC are used to estimate treatment effects for a larger network of evidence from disconnected studies. The review shows that in more than half of the appraisals (55%), the intervention treatment from the single-arm IPD study has been used to make multiple comparisons. In some appraisals, multiple MAICs have been done to estimate the treatment effect of the new intervention with multiple comparators where the comparator treatments have come from different studies. In other appraisals, the new intervention has been compared to a single comparator but the comparator treatment was from different studies which required multiple MAICs to be done. Unanchored MAIC or STC was the only option to estimate relative treatment effects for a disconnected network of evidence due to the fact that the intervention treatment came from a single-arm study and only aggregate form of data were available from the comparator studies.

In some of these TAs after performing multiple MAIC or STC, the next step was to perform an NMA to get overall treatment effects. The company with the IPD study tried to make a coherent synthesis with the MAIC or STC estimates in an NMA. Although MAIC or STC are capable of adjusting prognostic as well as effect-modifier variables in a single comparison, there is no certainty that the balance of these variables can be maintained across the studies especially when moderate to high heterogeneity can be present. Furthermore, the repeated use of the IPD from the single-arm study also violates the independence among the studies. In order to understand the consequence of performing a population-adjusted ITC in an NMA setting, a simulation study was designed for binary outcomes. MAIC was chosen as the population-adjusted ITC for the simulation study. MAIC was found to be the frequently used method in the review of NICE STAs (13 out of 20 TAs, 65%). It was observed that companies were using unanchored MAIC with single-arm studies very often. They were using the estimates from MAIC to explain the treatment effect of an intervention with single or multiple comparators. With multiple comparators, MAIC estimates are frequently used as stand-alone estimates, or sometimes a meta-analysis was performed to get a coherent synthesis. Without

making the shared effect modifier assumption which assumes that all the effect-modifiers are similar across studies, it is not advisable to use MAIC for multiple comparisons as the target population differs for each MAIC. This simulation study aims to explore the situation using MAIC when this shared effect modifier assumption is met. In this chapter, the simulation study is presented following the ADEMP (Aims, Data-generating mechanisms, Estimands, Methods, Performance measures) structure as described by Morris et al. (2019).

The setting of the simulation in this chapter was done by assuming that there is a common treatment effect for each comparison. Conventionally, an NMA can be performed either assuming a fixed effect or a random effects model. In a fixed effect model, the true treatment effect is assumed to be fixed but shared by all the included studies. All studies are presumed to estimate the same effect size. The only reason that the effect size can vary between studies is because of sampling error. The true treatment effect is calculated as a weighted average of the individual studies where more weight is given to large studies. The combined effect that is estimated after performing an NMA is considered to be the estimate of the true treatment effect. This assumption of a fixed effect NMA is simpler in nature compared to a random effects NMA. The main reason to choose a fixed effect NMA is for analytic simplicity. Nevertheless, this simulation study will be conducted for a random effects NMA in the next chapter.

Section 4.2 of this chapter describes the aim of the simulation study. Section 4.3 to Section 4.6 describe how the simulation was designed and conducted and Section 4.7 describes the findings from the simulation study. The chapter concludes with a discussion of the simulation results in Section 4.8.

4.2 Aims

The goal of the simulation study was to assess the appropriateness of MAIC estimates, one of the widely used population-adjustment methods in NICE TAs, in a fixed effect NMA.

4.3 Data generating mechanism (DGM)

In this simulation study, data were generated for four settings that differ according to the connection of the network and according to the amount of evidence informing the network (network size). With respect to connection, two settings were termed as connected and disconnected. A connected network of evidence means a collection of RCTs where data were generated with respect to a common treatment, on the other hand, a disconnected network of evidence refers to a collection of studies without any common treatment. Two network sizes were simulated termed as smaller and larger networks of evidence. A smaller network of evidence refers to a network of studies where there is only one study per comparison whereas a larger network of evidence refers to multiple studies per comparison. Therefore, according to the network connection and size, the simulated data were termed as “connected smaller network”, “connected larger network”, “disconnected smaller network”, and “disconnected larger network”.

The objective of performing the simulation both for a smaller and larger disconnected network of evidence was to assess the impact of MAIC-adjusted NMA when the number of studies varies per treatment comparison. The objective for the inclusion of the connected smaller and larger network of evidence was to assess how much MAIC-adjusted NMA estimates differ in comparison to the connected NMA estimates. Figure 4.1 describes the process of data generation where first data were generated for a smaller and for a larger connected network of evidence. NMA estimates were calculated from both of this connected network of evidence. The target was to convert both the smaller and larger connected network of evidence into a smaller and a larger disconnected network of evidence by dropping treatment arms from each RCT and then apply MAIC to this disconnected network of evidence to estimate treatment effects. The estimates were then used to perform an NMA. This NMA from the MAIC estimates is termed as "MAIC-adjusted NMA". The discrepancy between estimates from connected NMA and MAIC-adjusted NMA will provide insight into the validity of using MAIC both in a smaller and larger disconnected network of evidence.

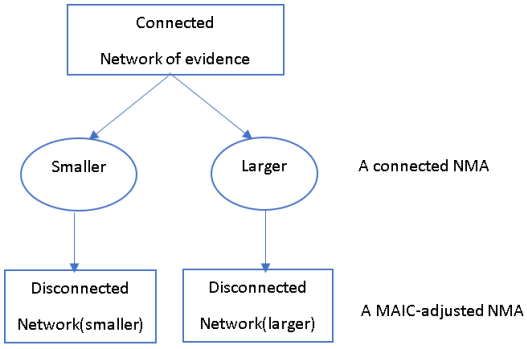


Figure 4.1: *Subsequent steps of data generation*

4.3.1 DGM for a smaller connected and a smaller disconnected network

The first step of the simulation study was to generate data for a smaller connected network of 3 RCTs with 3 treatments for binary data. From the introduction chapter, it is evident that though time-to-event outcomes act as one of the most common outcomes for efficacy assessments of medical interventions, binary outcomes also were used to assess efficacy. Therefore, to start the simulation study with a simpler setting, binary outcomes were chosen. The simulation was conducted by setting the following data properties: the number of nodes in the network, the sample size per study, network density, and the nature of prognostic and effect-modification in the network.

In the simulation settings, the number of nodes means the number of treatments in the network which was kept at 3, network density refers to the number of studies per comparison. All studies consisted of two arms RCTs of equal sizes. A sparse network i.e. a triangular network was built with no closed loops which is depicted in Figure 4.2. The network of evidence consists of 3 studies where 2 studies with treatments 2 and 1 and one study with

treatments 3 and 1. Treatment 1 was considered the new intervention/common treatment in each study. Data was generated with two continuous covariates for each study where one of them was considered prognostic and the other one was an effect-modifying variable. In previous simulation studies it was observed that with MAIC, overlap between studies with continuous covariates is crucial compared to binary covariates and low-overlap can occur with continuous covariates in much less extreme situations (Phillippo et al., 2020b). The R code for data generation process is given in C.1 of Appendix C.

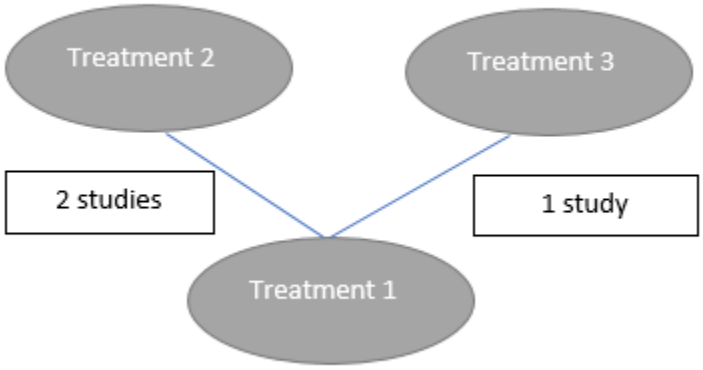


Figure 4.2: *Network diagram for a smaller network of evidence*

The underlying model to generate data for each study in the network of evidence is,

$$y_{ijk} \sim \text{Bernoulli}(P_{ijk})$$

$$logit(P_{ijk}) = \begin{cases} \mu + \beta_1 x_{1ijk} + \beta_1 x_{2ijk} & \text{if } t_k = 1 \\ \mu + \beta_1 x_{1ijk} + \beta_1 x_{2ijk} + \delta_{jk} + \beta_2 x_{1ijk}\ \mathbb{1}(t \neq 1) & \text{if } t_k \neq 1. \end{cases}$$

In the above equation, each outcome y_{ijk} comes from individual i, in study j with treatment k from a Bernoulli distribution. P_{ijk} is the probability of the event which is modelled on the logit scale. μ is the study intercept which was kept at 0.85 for each study. The intercept value was chosen arbitrarily and it is the log odds of having the outcome when all the covariates are zero. x_{1ijk} is the effect-modifying variable and x_{2ijk} is the prognostic variable with coefficient β_1. The coefficient (β_1) was kept identical for both the covariates to keep simplicity in the model. δ_{jk} is the treatment effect in study j and treatment k. The relative effect or log odds ratio of treatments 2 and 3 with treatment 1 was 0.40 and -1.77 respectively. These values were opposite in direction and they were chosen to make the effects as different as possible. β_2 is the coefficient for the effect-modifying variable. The two continuous covariates in each study were generated from a multivariate normal distribution where the standard deviation (SD) was kept at 0.4. This value of SD was chosen arbitrarily. In the network of studies, the mean value of covariates in the first study was kept to 0.60 and 0.50 and the values of the covariate mean for the rest of the studies were varied in such a way that a certain

amount of covariate overlap can be maintained between the first and the rest of the network. Moreover, to emulate a real NMA scenario where mean covariate values from different studies differ, the mean of each covariate in each study was drawn from a normal distribution with a pre-specified value and SD 0.05 except for study 1. Extracting the covariate in this way will maintain a slight difference between covariates in the network.

After generating data for a smaller connected network of evidence, a connected NMA was performed. The next step was to transform this connected network of evidence into a disconnected network of evidence. To do this, the first study in the network was considered as the IPD study and the rest of the network as AgD studies. Therefore, after generating IPD for all the studies in the network, all the studies were converted to AgD except study 1. The objective was to replicate a common scenario in HTA, where the company has IPD on outcome and covariates for their own study and only AgD values for comparator studies. The connected network of evidence was transformed into single-arm studies by dropping one arm from each study. New intervention treatment or treatment arm 1 was kept from the first study and for the rest of the studies, all arms were dropped except arms 2 and 3. Figure 4.3 illustrates the process where each oval shape node represents the study arm with treatment number and an RCT is depicted by joining two nodes with a solid line. Figure 4.3 shows that IPD was generated for 3 connected RCTs and then one arm was dropped from each study to convert them into single-arm studies. The treatment arm that was dropped from each study is depicted by striking through the treatment number. IPD was kept only for the first study and all other studies were converted into AgD studies by estimating mean and variances for covariates and proportion of events for the outcome variable. Then two MAICs were conducted using the IPD from the first study which has been depicted by an arrow line in Figure 4.3. These MAIC estimates gave relative effects of intervention treatments with treatments 2 and 3. An NMA was performed with the MAIC estimates to assess the difference between connected NMA estimates and MAIC-adjusted NMA estimates.

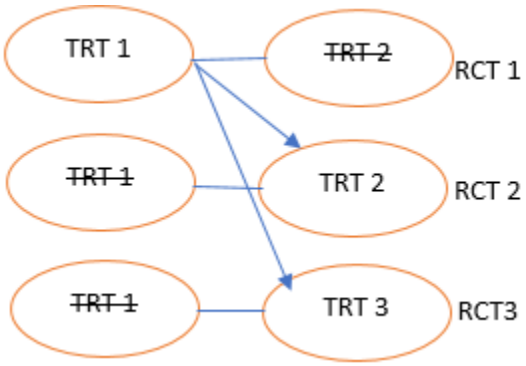


Figure 4.3: *Making of a disconnected network of evidence from a smaller connected network of evidence*

4.3.2 DGM for a larger connected and a larger disconnected network

The next step of the simulation study was to generate a connected network of evidence for 10 RCTs with 3 treatments for binary data. The DGM was identical to Section 4.3.1 except for the fact that now the data has been generated for a larger network. Figure 4.4 shows the network of evidence for 10 studies where 6 studies compared treatments 2 and 1 and 4 studies compared treatments 3 and 1. Figure 4.5 illustrates the process which shows that IPD was generated for the 10 connected RCTs and then one arm was dropped from each study to convert each study into a single-arm study. Individual patient data (IPD) was kept only for the first study and all other studies were converted into AgD studies. Then 9 unanchored MAICs were computed using the IPD from the first study which has been depicted by the nine arrow lines. These MAIC estimates gave relative effects of treatment 1 compared to treatments 2 and 3. A connected NMA as well as a NMA with the MAIC estimates was performed to evaluate the difference between connected NMA estimates and MAIC-adjusted NMA estimates.

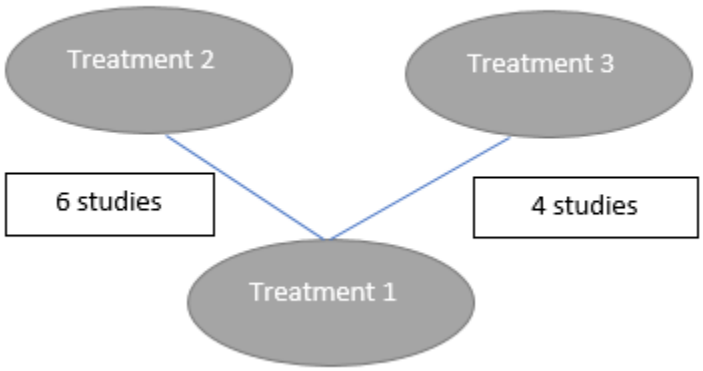


Figure 4.4: *Network diagram for a larger network of evidence*

The simulation study evaluates the change in five factors in a full factorial design. Taking two values from each factor results in a 2x2x2x2x2=32 scenarios. The values of the different levels of factors are depicted in Table 4.1. Table 4.1 shows different combinations of all the factors except sample size. When these 16 scenarios are again combined with sample sizes 150 and 500 per arm, it results in 32 scenarios. The choice of the factors as well as their values were inspired by previous simulations (Phillippo et al., 2020b; Remiro-Azócar et al., 2020). The values of the factors were chosen to observe their effect with two extreme ends. The values are as follows:

- The overall sample size was varied with values of 150 and 500 with a 1:1 randomization for intervention and comparator treatment within each study.
- The correlation coefficient between covariates of each study was varied by either 0.2 (weak correlation) or 0.8 (strong correlation).

- The strength of effect-modification was tested by varying their values to either {-log(0.78)=0.25 or -log(0.40)=0.92} which were considered as weak and strong effect-modifications respectively.

- Varying the strength of prognostic variable relationship with outcome by either {-log(0.67)=0.40 or -log(0.33)=1.11} which were considered as weak and strong prognostic effects respectively.

- Varying the between study overlap. In order to maintain the covariate overlap between the first and the rest of the studies, the mean values of the two covariates in the first study were kept fixed at 0.60 and 0.50 i.e $x_{i11} \sim N(0.60, 0.4^2)$ and $x_{i12} \sim N(0.50, 0.4^2)$ for individual i from study 1 with covariate 1 and 2 respectively. The mean values for the rest of the study covariates were varied by {(0.45,0.48),(0.15,0.20)} with SD 0.05. The scenario where the mean covariates was (0.45,0.48), gave a much bigger overlap between the first study with the rest of the network of evidence compared to the scenario (0.15,0.20). Cohen's d was used to measure the amount of overlap between studies. It is a widely used standardized effect size to measure the difference between the control and treatment groups (Cohen, 2013). It can be characterized as a signal-to-noise ratio where the difference between the two groups is divided by their pooled SD. A large value of the d indicates the difference between the two groups is large i.e. the signal is greater than the noise which in turn means that the amount of overlap is less between the groups. The first pair of values with (0.45,0.48) corresponds to Cohen's d= {0.375, 0.30} which means 85.3% and 88.1% overlap exist between covariate values. The second pair of values with (0.15,0.20) corresponds to Cohen's d= {1.12, 1.0} which means 57.5% and 61.7% overlap exist between covariate values.

Covariate overlap is a crucial issue which is also found in the previous chapter of the NICE review. In the review, it was found that of 13 TAs which have used MAIC, only 5 (35.71%)of them have reported their ESS. The median ESS was 67.1 (range: 3.8 to 84), with a median reduction in ESS from the original sample size of 50.3% (range: 43.24% to 94.73%). A large amount of reduction in ESS indicates a lack of overlap between the IPD and AgD studies which may affect the MAIC estimates by making them unstable and unreliable.

In addition to simulating both a smaller and larger connected network of evidence, a connected network of evidence was also simulated with a big sample size (1 million). This was done to perform an NMA that will produce the true relative treatment effect estimates of treatments 2 and 3 with the common treatment 1 for the connected network of evidence. The treatment effect estimates generated from this big sample size were then used to estimate performance measures both for the connected NMA and MAIC-adjusted NMA.

4.4 Estimands

In this simulation study, first, an NMA was performed after generating data from a smaller and larger connected network of evidence. These connected networks were then transformed into disconnected networks by dropping one arm from each of the RCT. These disconnected networks of evidence were again made into connected networks by applying multiple MAICs.

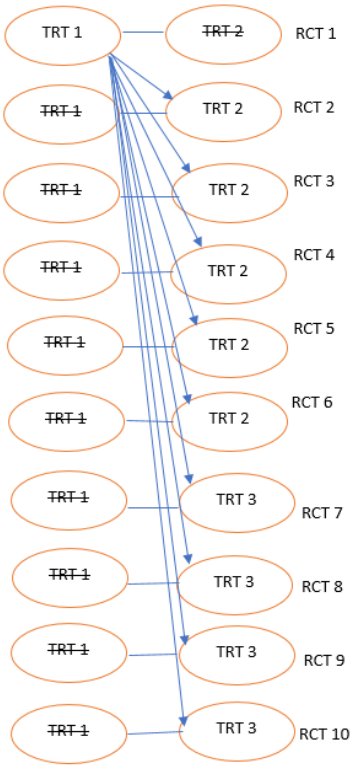


Figure 4.5: *Making of a disconnected network of evidence from a larger connected network of evidence*

A second NMA was performed using these MAIC-adjusted treatment effects. The estimands of interest in the simulation study were the overall treatment effect estimates from the connected NMA and the overall treatment effect estimates from the MAIC-adjusted NMA for treatment 2 and 3.

The design of the simulation study was done in such a way that it satisfied the shared effect modifier assumption as during the data generation process, all the treatments in a network of evidence had the same effect-modifier coefficient β_2. This was done as this assumption is needed to apply the MAIC estimates in an NMA to make a coherent synthesis of evidence.

4.5 Methods

The following methods were applied to the data generated during the simulation exercise:

- A fixed effect NMA was applied to the simulated data of a smaller and a larger connected networks of evidence. The estimation of the NMA was done using the Bayesian approach and using R package `multinma` (Phillippo, 2021). Though R package `multinma` is specially developed to apply the multi level network meta-regression (ML-NMR) method in an NMA setting, this package can also be used to run conventional NMA us-

ing Stan software [1] which is a relatively new program for conducting Bayesian network meta-analyses. For the prior distributions of the treatment effects and study-specific intercepts, a normal distribution was used with $N(0, 100^2)$.

- MAIC was applied to match discrepancy between individuals from different studies with respect to patient demographics or covariate values (Signorovitch et al., 2010). MAIC uses weights which are calculated by the MoM/EB that gives more importance/weight to individuals in the IPD study who are more alike to the AgD study and less importance if they differ between studies. A logistic regression was used to assign the weights which in turn make the study individuals as similar as possible. As the MAICs were applied in unanchored form, all effect-modifiers and prognostic variables need to be included in the model. MAICs were applied upto the second moment i.e. balancing of covariate was done both for mean and standard deviation. After applying MAIC in a smaller and larger disconnected network of evidence, a MAIC-adjusted fixed effect NMA was performed and robust SE was estimated for the NMA estimates. The R package `MAIC` was used to perform the MAIC (`https://rdrr.io/github/Roche/MAIC/`).

Table 4.1: *Parameter values for different simulation scenarios*

Simulation scenarios	**Correlation coefficient**	**Strength of effect modification**	**Strength of prognostic variable**	**Mean of first covariate**	**Mean of second covariate**
1	0.2	0.25	0.40	0.45	0.48
2	0.8	0.25	0.40	0.45	0.48
3	0.2	0.92	0.40	0.45	0.48
4	0.8	0.92	0.40	0.45	0.48
5	0.2	0.25	1.11	0.45	0.48
6	0.8	0.25	1.11	0.45	0.48
7	0.2	0.92	1.11	0.45	0.48
8	0.8	0.92	1.11	0.45	0.48
9	0.2	0.25	0.40	0.15	0.20
10	0.8	0.25	0.40	0.15	0.20
11	0.2	0.92	0.40	0.15	0.20
12	0.8	0.92	0.40	0.15	0.20
13	0.2	0.25	1.11	0.15	0.20
14	0.8	0.25	1.11	0.15	0.20
15	0.2	0.92	1.11	0.15	0.20
16	0.8	0.92	1.11	0.15	0.20

[1] Using the NUTS sampler (Hoffman and Gelman 2012), Stan provides posterior simulations for user-specified models and data using Bayesian inference.

4.6 Performance Measures

In order to compare the performance of MAIC both in a smaller and larger disconnected network of evidence, performance measures bias, model SE, empirical SE, and coverage probability were used.

- Bias: Statistical bias in simulation study gives an estimate of the systematic discrepancy between the true parameter and expected values of the results obtained from each simulated dataset. It can be defined as: $Bias = E[\hat{\theta}] - \theta$. In this simulation study, a connected NMA was estimated with a big sample size (1 million), and the relative treatment effect estimates from this NMA were used as the true parameter value θ. In the bias formula, $E[\hat{\theta}]$ was calculated by taking the mean of an estimate from all the MCMC samples. The bias can be considered questionable when its absolute size is bigger than one-half of the estimate's SE (Schafer and Graham, 2002).

- Empirical SE: Empirical SE is the dispersion measure of the estimator in a simulation study. It represents the precision of an estimator as well as its true variability. An estimator is expected to have low variance when it is applied to multiple datasets. It can be defined as: $EmpSE = \sqrt{Var(\hat{\theta})}$

- Model SE: In a simulation study, when a method is applied to multiple datasets, the measure of the average of the SE reported by the method is known as model SE. It can be defined as: $ModelSE = \sqrt{E[\hat{se}(\hat{\theta})^2]}$. It is desired that the empirical SE is small which shows that the estimator is precise and the model SE is equal to empirical SE.

- Coverage probability: In a simulation study, coverage probability refers to the statistical technique where a percentage/proportion is calculated which shows how many confidence intervals include the true parameter value which is expected to be at $(100\times(1\text{-}\alpha))\%$ nominal level. It is common to fix the value of α at 0.05 i.e. at 95%.

The simulation study simulated 3000 repetitions/ datasets of each simulation scenario.

4.7 Results

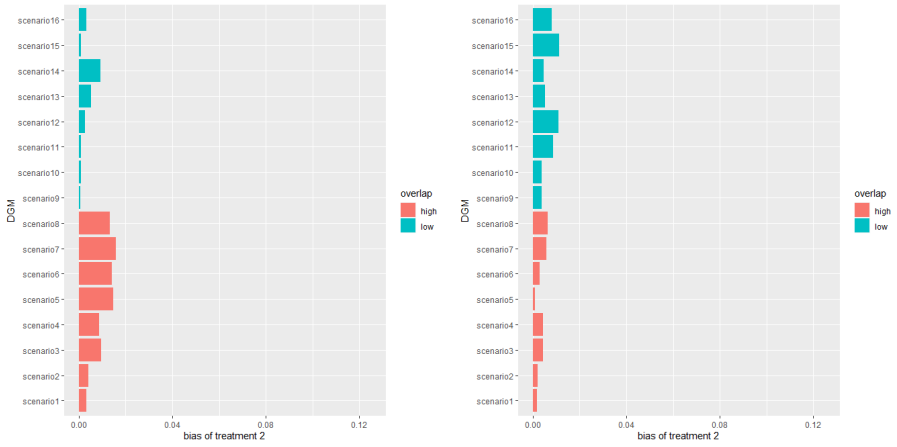
The result section describes the findings from both the connected and disconnected network of evidence. The results for both smaller and larger connected networks of evidence are described in subsections 4.7.1.1 to 4.7.2.8. Subsections 4.7.3.1 to 4.7.4.8 describe results for the disconnected network of evidence.

4.7.1 Results of simulation scenarios for the smaller connected network of evidence

Data were generated for the smaller connected network of evidence as described in Section 4.3.1 and in Figure 4.2 where at first 3 RCTs were generated and then a fixed effect NMA was performed.

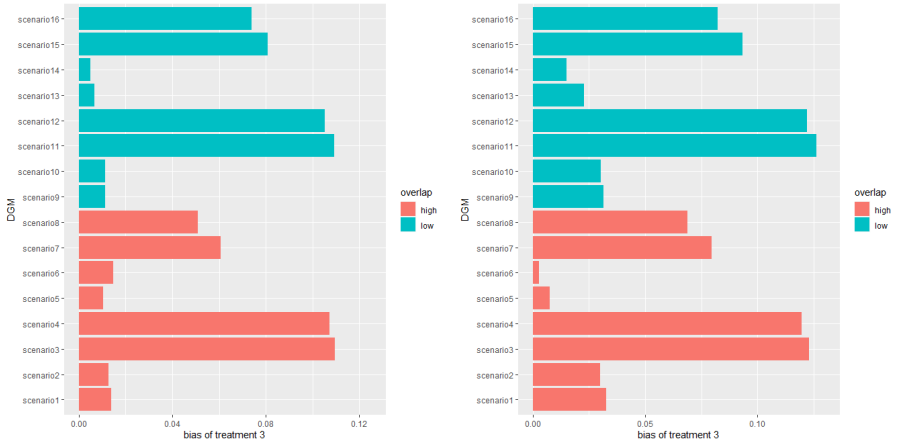
4.7.1.1 Bias by overlap

Figure 4.6 and 4.7 show the amount of bias for each DGM for treatment 2 and 3 respectively. The amount of bias was not found to be related to the amount of overlap. In Figure 4.6 (a), large biases can be seen for high overlap scenarios and the opposite can be found for 4.6 (b). In Figure 4.7 (a),(b), the biases seem to be similar both for high and low overlap scenarios. Overall, higher biases were found for scenarios 3, 4, 7, 8, 11, 12, 15, and in scenario 16 for treatment 3. One reason could be that the effect-modifier variable was high in these scenarios.



(a) *bias by overlap for sample size 150* **(b)** *bias by overlap for sample size 500*

Figure 4.6: *Bias of treatment 2 for different sample sizes (connected smaller network of evidence)*

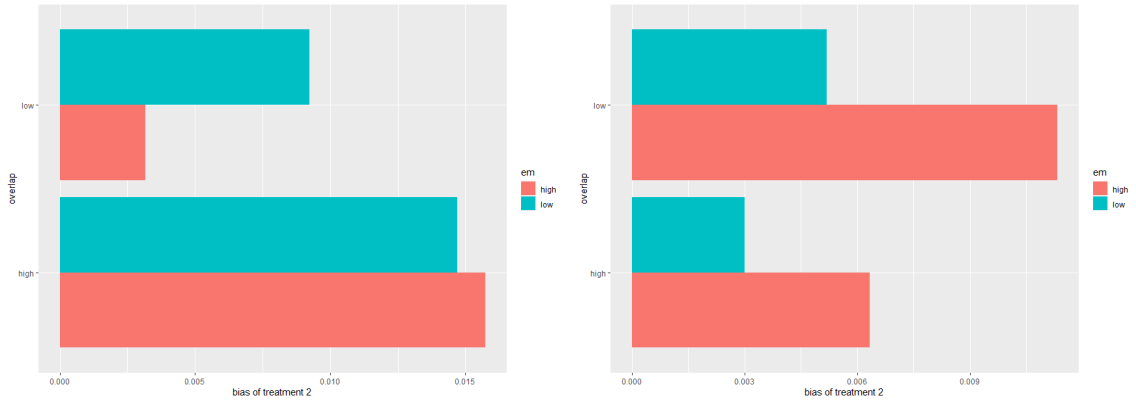


(a) *bias by overlap for sample size 150* **(b)** *bias by overlap for sample size 500*

Figure 4.7: *Bias of treatment 3 for different sample sizes (connected smaller network of evidence)*

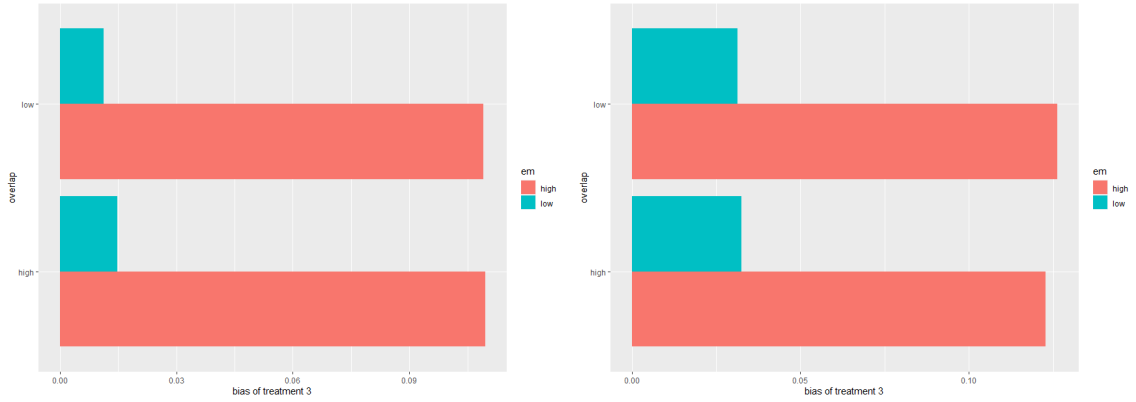
4.7.1.2 Bias by overlap with effect-modifiers

Figure 4.8 and 4.9 show the amount of bias by overlap for different levels of effect-modifier for treatments 2 and 3 respectively. Biases do not reduce with high overlap scenarios, but within each level of overlap, low biases were found with low effect-modifier levels.



(a) *bias by EM for sample size 150* **(b)** *bias by EM for sample size 500*

Figure 4.8: *Bias by overlap with EM levels for treatment 2 (connected smaller network of evidence)*

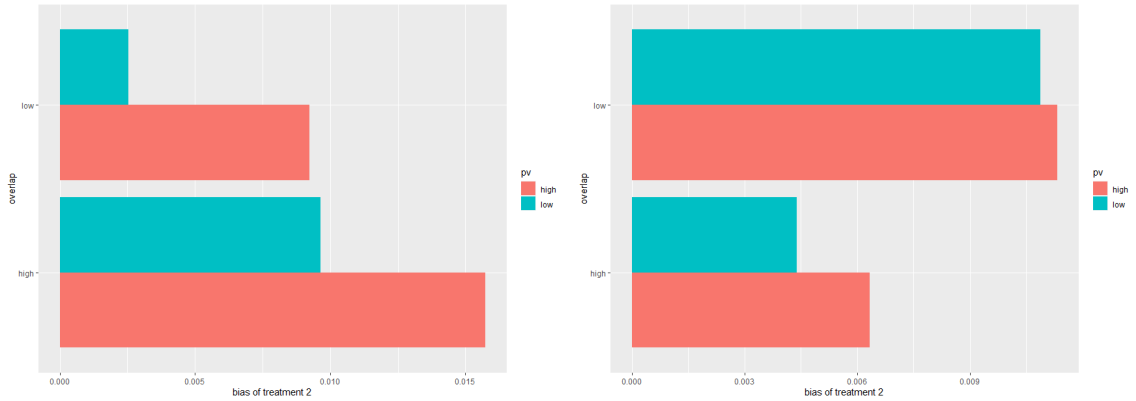


(a) *bias by EM for sample size 150* **(b)** *bias by EM for sample size 500*

Figure 4.9: *Bias by overlap with EM levels for treatment 3 (connected smaller network of evidence)*

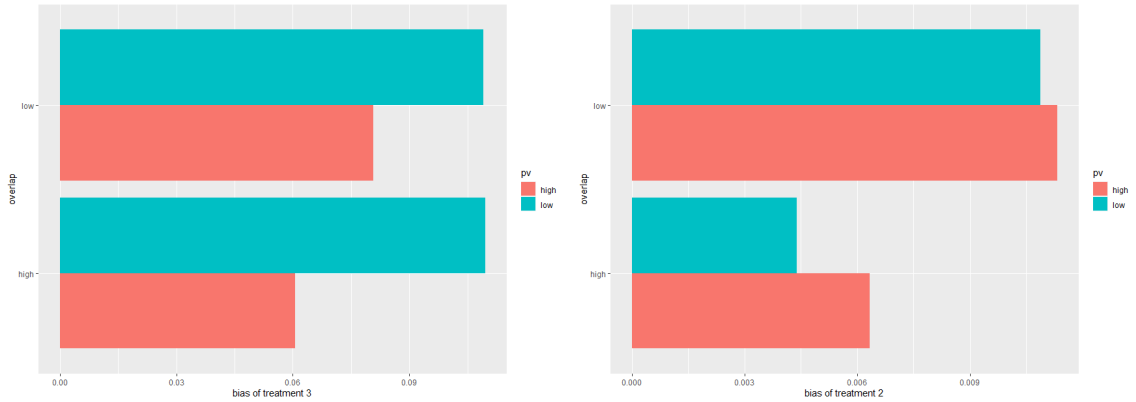
4.7.1.3 Bias by overlap with prognostic variable

Figure 4.10 and 4.11 show the amount of bias by overlap for different levels of prognostic variables for treatments 2 and 3 respectively. Biases do not reduce with high overlap but within the level of overlap, prognostic variables seem to be related to bias. In Figure 4.10 (a),(b), within each level of overlap, bias reduces with low prognostic variable level but the opposite can be seen for treatment 3 in Figure 4.11 (a),(b).



(a) *bias by PV for sample size 150* **(b)** *bias by PV for sample size 500*

Figure 4.10: *Bias by overlap with PV levels for treatment 2 (connected smaller network of evidence)*



(a) *bias by PV for sample size 150* **(b)** *bias byPV for sample size 500*

Figure 4.11: *Bias by overlap with PV levels for treatment 3 (connected smaller network of evidence)*

4.7.1.4 Empirical SE and model SE for the simulation scenarios

The simulation results of performance measures for the smaller connected network of evidence are summarised in Table 4.3 and Table 4.4 for sample sizes 150 and 500 respectively. Table 4.3 and Table 4.4 show the relative estimates of treatments 2 and 3 with treatment 1 from connected NMA for both sample sizes. The overall relative estimates of treatment 2 with the new intervention treatment 1 is termed as d1 and that of treatment 3 with treatment 1 is termed as d2. The empirical SE and model SE are similar or close to each other for each DGM. A quantity was estimated that calculates the difference between the empirical SE and model SE. The highest value of this difference was found to be 0.009 for the smaller connected network of evidence.

Tables 4.3 and 4.4 are colour-coded to understand when the coverage was low and the magnitude of biases. Three colour-coding was used for showing bias and two colour-coding for

coverage. If the coverage was at the nominal level, it was coloured blue otherwise red. For bias three colour-coding were used which are described in the following table:

Table 4.2: *colour-coding for bias*

Amount of bias	0-0.03	0.03-0.05	0.05-onwards
colour	Blue	Yellow	Red

The blue, yellow and red indicate low, moderate, and high biases respectively. From the colour-coding it is evident that none of the coverage deviates from the nominal level. However, the biases were high for those scenarios where the effect-modifying variable was high.

Table 4.3: *Performance measure estimates of a smaller connected network of evidence for sample size 150 (fixed effect)*

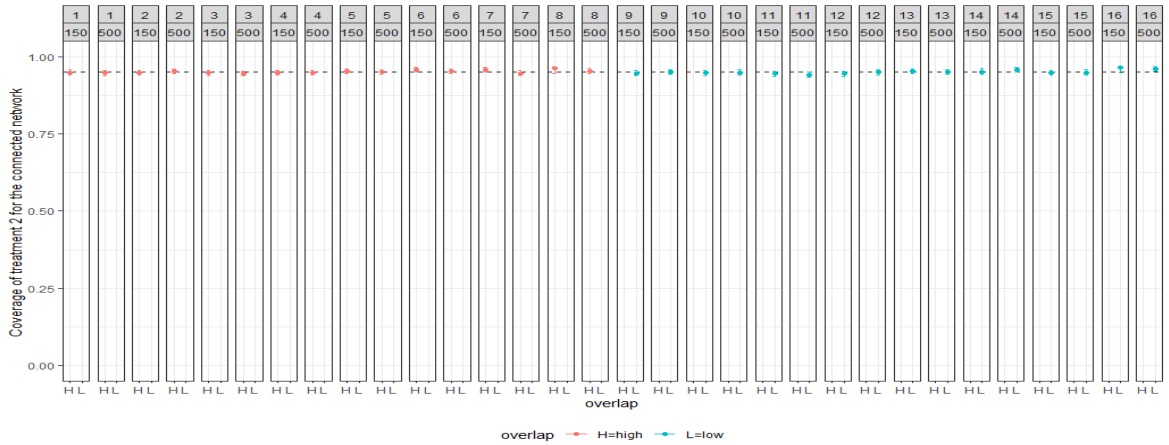
DGM	COVERAGE d1	BIAS d1	EMP SE d1	MODEL SE d1	DIFFERENCE d1	COVERAGE d2	BIAS d2	EMP SE d2	MODEL SE d2	DIFFERENCE d2
Scenario 1	0.949	0.003328	0.221544	0.216095	0.005449	0.948333	0.013868	0.262569	0.258873	0.003696
Scenario 2	0.948333	0.004142	0.219577	0.215133	0.004445	0.944667	0.012598	0.263484	0.258324	0.005159
Scenario 3	0.947333	0.009644	0.239047	0.229875	0.009173	0.935333	0.109601	0.259698	0.256528	0.00317
Scenario 4	0.947667	0.008603	0.235241	0.227759	0.007482	0.941333	0.107289	0.256207	0.256285	-7.8E-05
Scenario 5	0.953667	0.014711	0.264518	0.26475	-0.00023	0.954	-0.01051	0.281557	0.286351	-0.00479
Scenerio 6	0.957667	0.014015	0.250455	0.255835	-0.00538	0.959	-0.01484	0.271633	0.281825	-0.01019
Scenario 7	0.958667	0.015736	0.27757	0.2781	-0.00053	0.954	0.060656	0.281472	0.288237	-0.00677
Scenario 8	0.961333	0.013381	0.257798	0.265619	-0.00782	0.956333	0.050945	0.269702	0.283388	-0.01369
Scenario 9	0.946667	-0.00068	0.21126	0.206657	0.004603	0.946	0.011352	0.25869	0.255473	0.003217
Scenario 10	0.947	0.001032	0.209912	0.205849	0.004064	0.953	0.011129	0.256561	0.255052	0.001509
Scenario 11	0.944333	-0.00087	0.220695	0.21345	0.007245	0.934333	0.109098	0.255917	0.252833	0.003084
Scenario 12	0.945	-0.00254	0.217101	0.211892	0.005209	0.940333	0.105157	0.256173	0.252428	0.003745
Scenario 13	0.952	0.005108	0.230315	0.231213	-0.0009	0.953667	0.006734	0.250375	0.255812	-0.00544
Scenario 14	0.961	0.009238	0.215623	0.225718	-0.01009	0.961667	0.004954	0.23942	0.253819	-0.0144
Scenario 15	0.947667	0.001062	0.233711	0.23528	-0.00157	0.954	0.080783	0.245567	0.254678	-0.00911
Scenario 16	0.964	0.003153	0.216597	0.228273	-0.01168	0.961333	0.074004	0.235319	0.252629	-0.01731

Table 4.4: *Performance measure estimates of a smaller connected network of evidence for sample size 500 (fixed effect)*

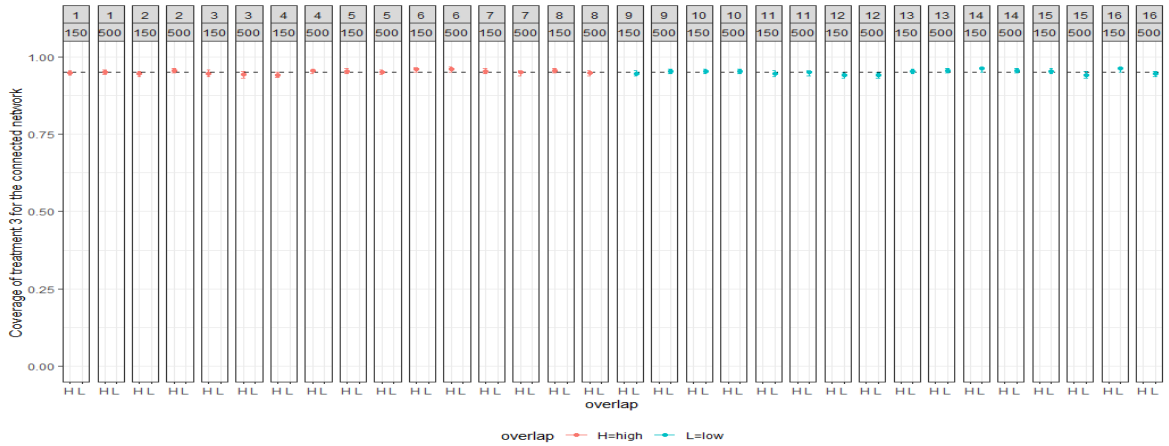
DGM	COVERAGE d1	BIAS d1	EMP SE d1	MODEL SE d1	DIFFERENCE d1	COVERAGE d2	BIAS d2	EMP SE d2	MODEL SE d2	DIFFERENCE d2
Scenario 1	0.947333	-0.00171	0.11833	0.117337	0.000993	0.951	0.03257	0.140139	0.140603	-0.00046
Scenario 2	0.952667	-0.00213	0.117141	0.116814	0.000327	0.954333	0.029926	0.140319	0.140435	-0.00012
Scenario 3	0.9458	-0.0044	0.127869	0.124494	0.003375	0.93	0.122729	0.146642	0.139503	0.007139
Scenario 4	0.9476	-0.00428	0.125744	0.123417	0.002326	0.9348	0.119671	0.144333	0.139278	0.005055
Scenario 5	0.950667	-0.00106	0.142649	0.142962	-0.00031	0.951	0.007417	0.149427	0.154932	-0.0055
Scenario 6	0.952333	-0.003	0.135113	0.138217	-0.0031	0.960667	0.002605	0.143329	0.152564	-0.00924
Scenario 7	0.946333	-0.00591	0.153082	0.149781	0.003301	0.937	0.079572	0.151566	0.156011	-0.00445
Scenario 8	0.954	-0.00634	0.141944	0.143284	-0.00134	0.937	0.068848	0.145897	0.153414	-0.00752
Scenario 9	0.949667	-0.00374	0.113065	0.112343	0.000722	0.952333	0.03141	0.13667	0.13884	-0.00217
Scenario 10	0.949	-0.00378	0.112291	0.111905	0.000386	0.953667	0.030377	0.134775	0.138629	-0.00385
Scenario 11	0.9416	-0.00863	0.120117	0.115847	0.00427	0.936	0.12623	0.142159	0.137563	0.004596
Scenario 12	0.949333	-0.01088	0.118138	0.115081	0.003058	0.9404	0.121909	0.139881	0.137346	0.002534
Scenario 13	0.950667	-0.00519	0.1243	0.125483	-0.00118	0.955667	0.022671	0.132738	0.139141	-0.0064
Scenario 14	0.957667	-0.00456	0.117167	0.122437	-0.00527	0.955667	0.015093	0.132189	0.138223	-0.00603
Scenario 15	0.948667	-0.01133	0.128634	0.127551	0.001083	0.930667	0.093378	0.13429	0.138562	-0.00427
Scenario 16	0.959667	-0.008	0.121242	0.123913	-0.00267	0.934333	0.082247	0.131859	0.137554	-0.0057

4.7.1.5 Coverage for the simulation scenarios

The coverage of d1 and d2 from the connected NMA is at the nominal level for both sample sizes. Figure 4.12 (a),(b) shows the confidence intervals of the coverage estimates for varying levels of sample sizes for treatments 2 and 3 respectively. Red-coloured figures represent confidence intervals for the high overlap scenarios and blue-coloured figures represent confidence intervals for the low overlap scenarios. In the figures, the black horizontal line represents the nominal level of coverage. From the figures, it can be seen that the confidence intervals touch the nominal level for all 16 scenarios irrespective of sample sizes and overlap.



(a) coverage by overlap for treatment 2

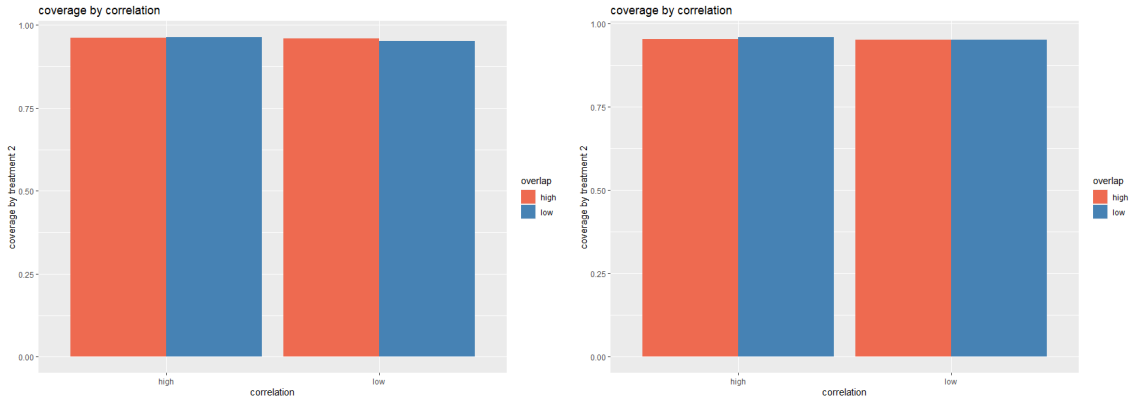


(b) coverage by overlap for treatment 3

Figure 4.12: *Coverage of treatments 2 and 3 for different sample sizes and overlap*

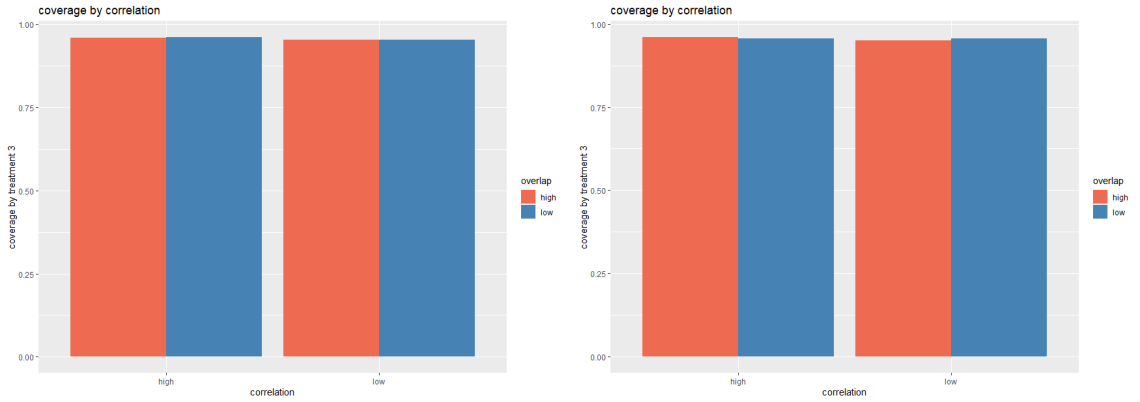
4.7.1.6 Coverage by correlation with overlap

Figure 4.13 and Figure 4.14 show the coverage with different levels of overlap and correlation for treatments 2 and 3 respectively. From the figures, it can be seen that the coverage is not found to be affected by different levels of correlation and overlap.



(a) *coverage by correlation for sample size 150* **(b)** *coverage by correlation for sample size 500*

Figure 4.13: *Coverage by correlation of treatment 2 for different sample sizes (connected smaller network of evidence)*

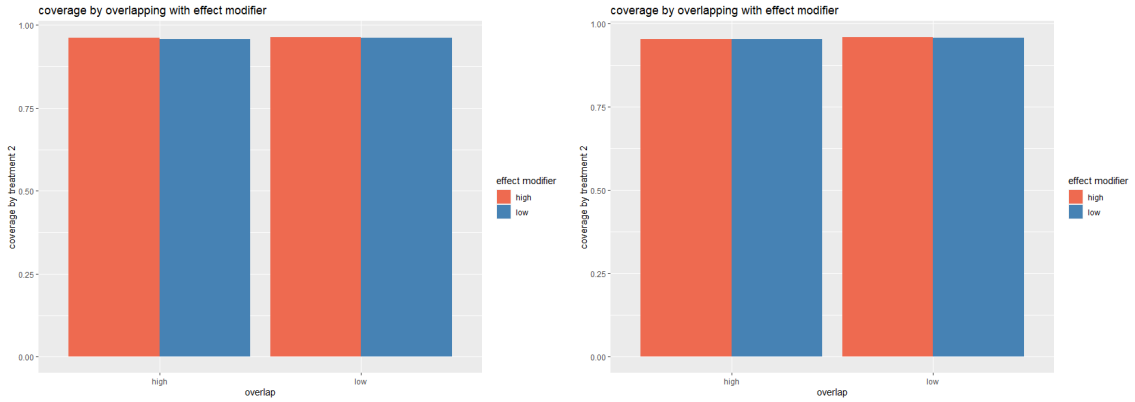


(a) *coverage by correlation for sample size 150* **(b)** *coverage by correlation for sample size 500*

Figure 4.14: *Coverage by correlation of treatment 3 for different sample sizes (connected smaller network of evidence)*

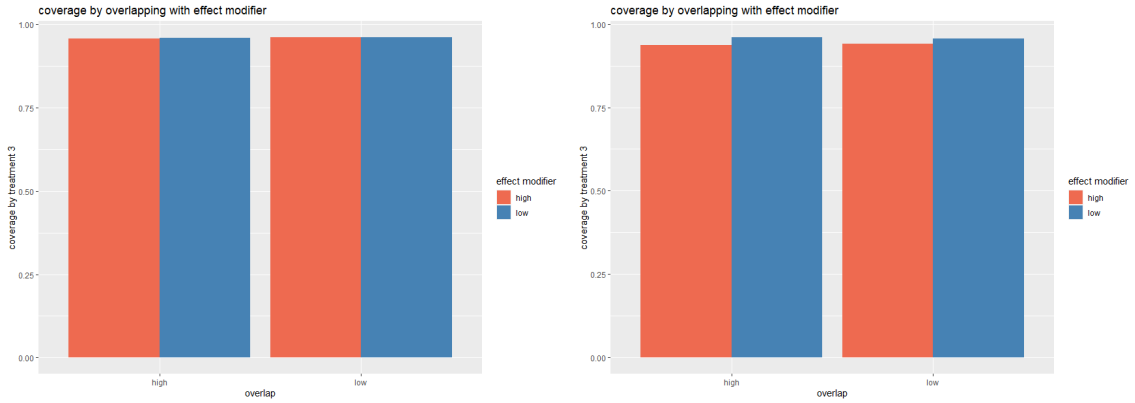
4.7.1.7 Coverage by overlap with effect-modifiers

Figure 4.15 and 4.16 show the coverage by different levels of overlap and effect-modifier for treatments 2 and 3 respectively. For both sample sizes and treatments, the coverage was similar with varying levels of overlap and effect-modifier variable.



(a) *coverage by EM for sample size 150* **(b)** *coverage by EM for sample size 500*

Figure 4.15: *Coverage by overlap with EM levels for treatment 2 (connected smaller network of evidence)*

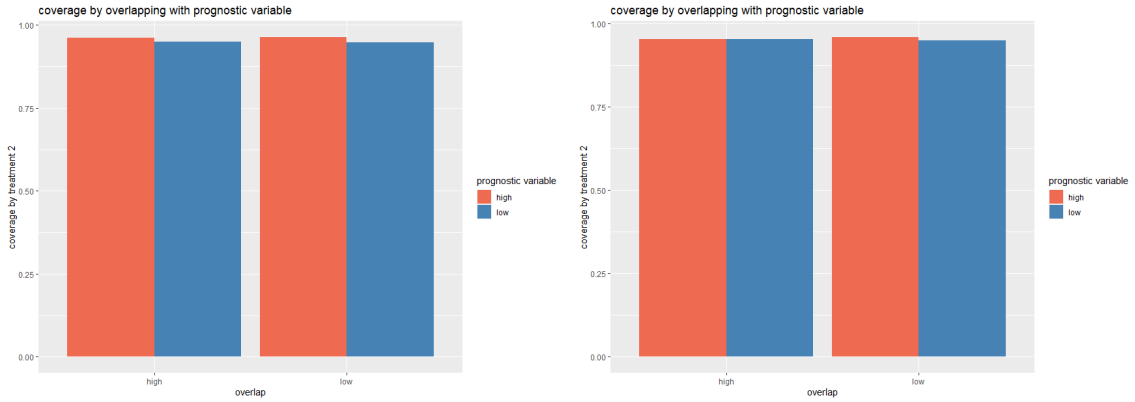


(a) *coverage by EM for sample size 150* **(b)** *coverage by EM for sample size 500*

Figure 4.16: *Coverage by overlap with EM levels for treatment 3 (connected smaller network of evidence)*

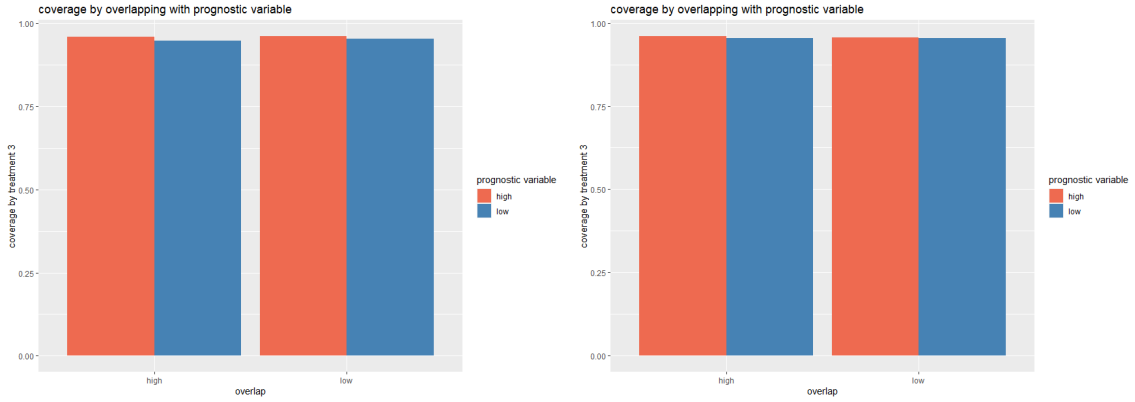
4.7.1.8 Coverage by overlap with prognostic variable

Figure 4.17 and 4.18 show the amount of coverage by different levels of overlap and prognostic variables for treatment 2 and 3 respectively. For both sample sizes and treatments, the coverage was found to be similar with varying levels of overlap and prognostic variables.



(a) *coverage by PV for sample size 150* **(b)** *coverage by PV for sample size 500*

Figure 4.17: *Coverage by overlap with PV levels for treatment 2 (connected smaller network of evidence)*



(a) *coverage by PV for sample size 150* **(b)** *coverage by PV for sample size 500*

Figure 4.18: *Coverage by overlap with PV levels for treatment 3 (connected smaller network of evidence)*

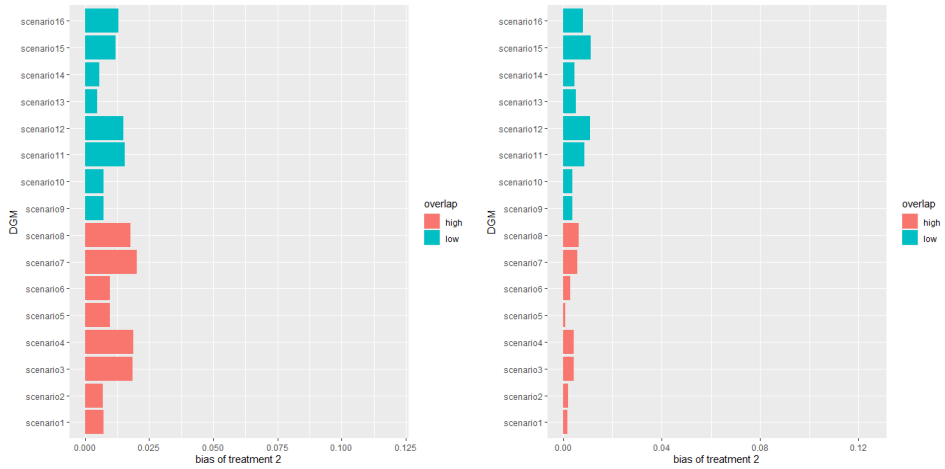
4.7.2 Results of simulation scenarios for the larger connected network of evidence

Data were generated for the larger connected network of evidence as described in Section 4.3.2 and in Figure 4.4 where at first 10 RCTs were generated and then a fixed effect NMA was performed.

4.7.2.1 Bias by overlap

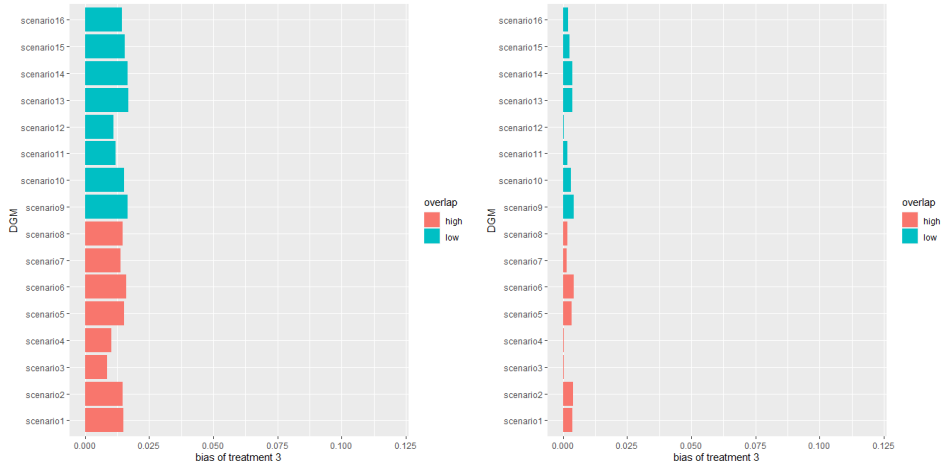
Figure 4.19 and 4.20 show the amount of bias for each DGM for treatment 2 and 3 respectively. Bias and overlap were not found to be related. In Figure 4.19 (a), large biases can be seen for high overlap but the opposite can be seen in Figure 4.19 (b). In Figure 4.20 (a),(b), a similar amount of bias can be seen for both levels of overlap. Although the finding is similar

to the finding from the small connected network of evidence, the magnitude of biases for the larger connected network of evidence is small compared to the smaller network of evidence.



(a) *bias by overlap for sample size 150* **(b)** *bias by overlap for sample size 500*

Figure 4.19: *Bias of treatment 2 for different sample sizes (connected larger network of evidence)*



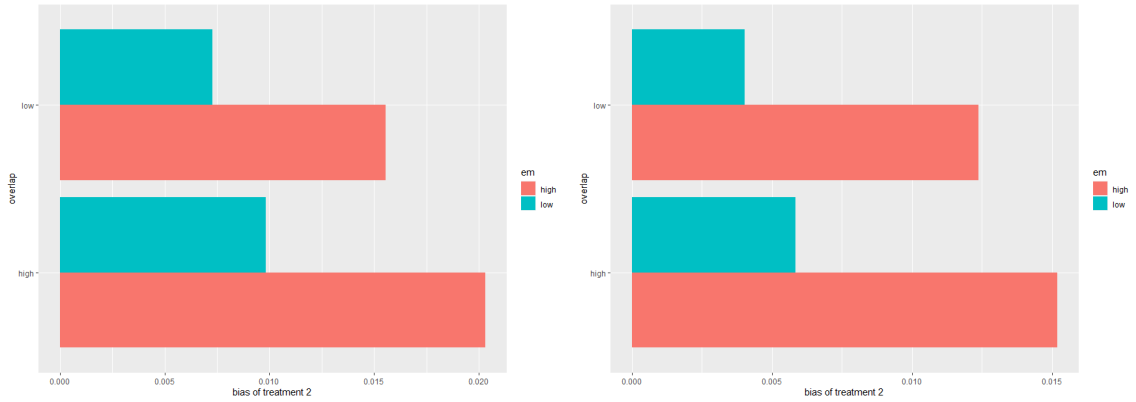
(a) *bias by overlap for sample size 150* **(b)** *bias by overlap for sample size 500*

Figure 4.20: *Bias of treatment 3 for different sample sizes (connected larger network of evidence)*

4.7.2.2 Bias by overlap with effect-modifiers

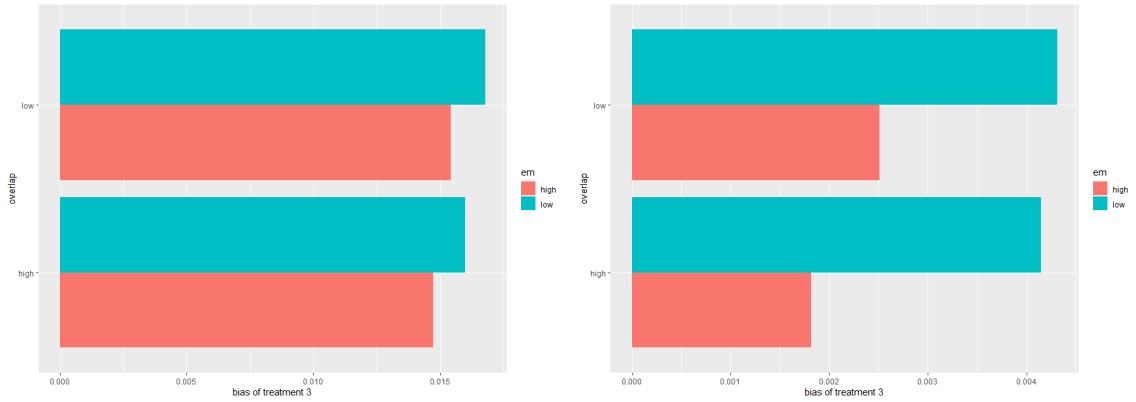
Figure 4.21 and 4.22 show the amount of bias by overlap for different levels of effect-modifier for treatments 2 and 3 respectively. Biases were found to be low for low overlap level for treatment 2 in Figure 4.21 but the opposite can be seen for treatment 3 in Figure 4.22. In

Figure 4.21, for treatment 2, bias was low with low effect-modifier levels whereas the opposite can be seen for treatment 3 in Figure 4.22.



(a) *bias by EM for sample size 150* **(b)** *bias by EM for sample size 500*

Figure 4.21: *Bias by overlap with EM levels for treatment 2 (connected larger network of evidence)*

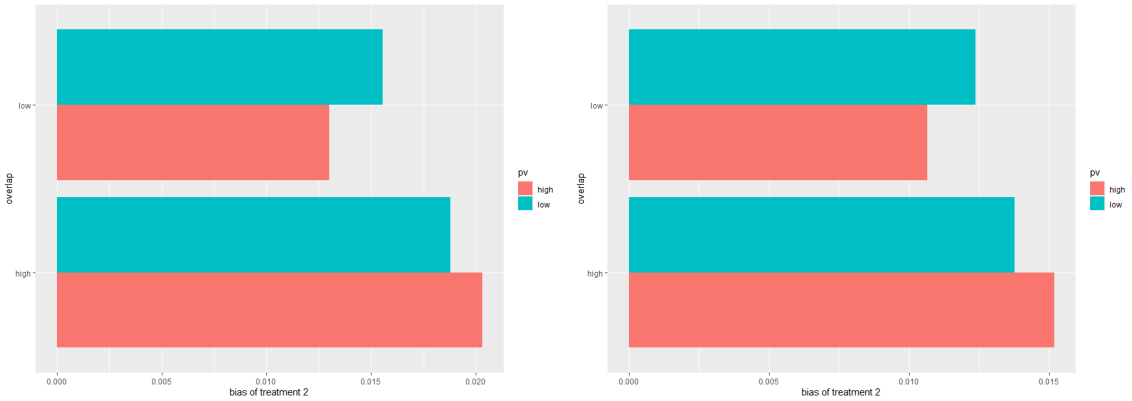


(a) *bias by EM for sample size 150* **(b)** *bias by EM for sample size 500*

Figure 4.22: *Bias by overlap with EM levels for treatment 3 (connected larger network of evidence)*

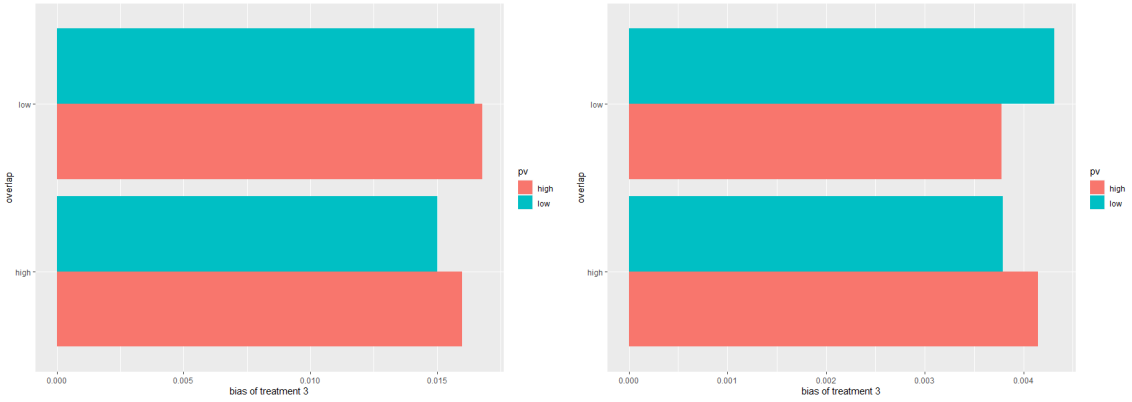
4.7.2.3 Bias by overlap with prognostic variable

Figure 4.23 and 4.24 show the amount of bias by overlap for different levels of prognostic variables for treatments 2 and 3 respectively. In Figure 4.23 (a),(b), biases were found to be low with low overlap and high prognostic variable level. Biases were high in high overlap level but within high overlap level, lower biases were seen with low prognostic level. In Figure 4.24 (a),(b), bias seems to be slightly low with a high overlap level, and within levels of overlap, biases were comparatively low with a low prognostic level.



(a) *bias by PV for sample size 150* **(b)** *bias by PV for sample size 500*

Figure 4.23: *Bias by overlap with PV levels for treatment 2 (connected larger network of evidence)*



(a) *bias by PV for sample size 150* **(b)** *bias by PV for sample size 500*

Figure 4.24: *Bias by overlap with PV levels for treatment 3 (connected larger network of evidence)*

4.7.2.4 Empirical SE and model SE for the simulation scenarios

The simulation results of performance measures for the larger connected network of evidence are summarised in Table 4.5 and Table 4.6 for sample sizes 150 and 500 respectively. Table 4.5 and Table 4.6 show the relative estimates of treatments 2 and 3 with treatment 1 from connected NMA for both sample sizes. The empirical SE and model SE are similar or close to each other for each DGM. The highest value of the difference between the empirical SE and model SE was found to be 0.008 for the larger connected network of evidence. From the color-coding, it can be seen that all were blue which means none of the scenarios shows undercoverage and the amount of biases were below 0.03.

Table 4.5: *Performance measure estimates of a larger connected network of evidence for sample size 150 (fixed effect)*

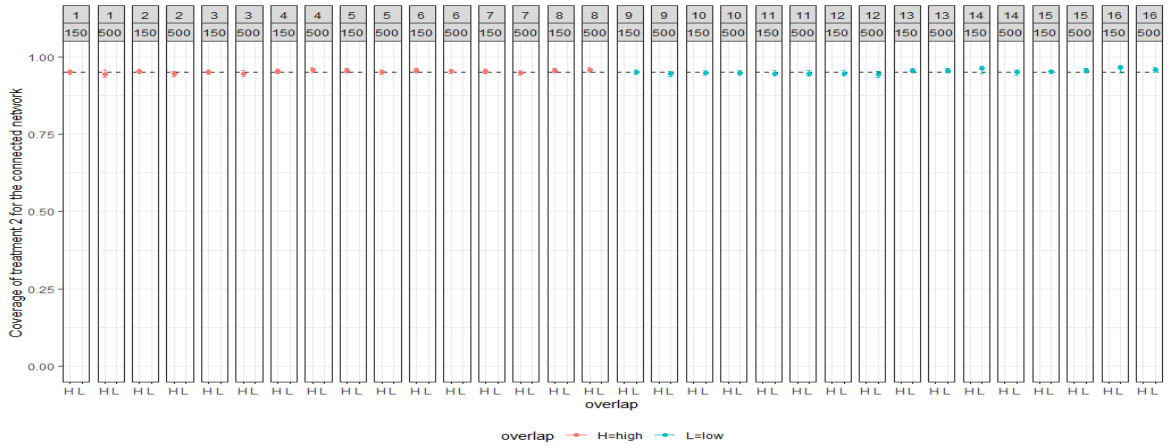
DGM	COVERAGE d1	BIAS d1	EMP SE d1	MODEL SE d1	DIFFERENCE d1	COVERAGE d2	BIAS d2	EMP SE d2	MODEL SE d2	DIFFERENCE d2
Scenario 1	0.950333	0.007241	0.122518	0.122821	-0.0003	0.948	-0.015	0.130341	0.128747	0.001594
Scenario 2	0.953667	0.006925	0.121425	0.122357	-0.00093	0.948333	-0.01478	0.130273	0.128644	0.001629
Scenario 3	0.949667	0.018603	0.12984	0.129224	0.000616	0.932667	-0.00868	0.130791	0.127704	0.003087
Scenario 4	0.952	0.018789	0.12822	0.128129	9.12E-05	0.944	-0.01015	0.129847	0.127544	0.002304
Scenario 5	0.956	0.009819	0.145215	0.147073	-0.00186	0.955333	-0.01518	0.139297	0.141548	-0.00225
Scenario 6	0.954667	0.009589	0.145215	0.147073	-0.00186	0.961333	-0.01599	0.139297	0.141548	-0.00225
Scenario 7	0.953333	0.020311	0.151555	0.152894	-0.00134	0.951333	-0.01389	0.140026	0.142443	-0.00242
Scenario 8	0.954333	0.017607	0.14159	0.146526	-0.00494	0.956333	-0.01472	0.134783	0.140238	-0.00545
Scenario 9	0.95	0.007155	0.114617	0.114642	-2.6E-05	0.951333	-0.01648	0.128425	0.127196	0.001229
Scenario 10	0.948333	0.007289	0.113311	0.114239	-0.00093	0.946333	-0.01514	0.127272	0.127031	0.000241
Scenario 11	0.946	0.015552	0.117286	0.116068	0.001218	0.938	-0.01192	0.128031	0.126031	0.001999
Scenario 12	0.946333	0.014864	0.116401	0.115479	0.000922	0.939	-0.011	0.127539	0.125787	0.001753
Scenario 13	0.956333	0.004678	0.117829	0.121381	-0.00355	0.947667	-0.01677	0.124314	0.127443	-0.00313
Scenario 14	0.962667	0.005594	0.112547	0.119119	-0.00657	0.954667	-0.01669	0.121848	0.126532	-0.00468
Scenario 15	0.953667	0.011851	0.119149	0.121689	-0.00254	0.948667	-0.01542	0.123329	0.12688	-0.00355
Scenario 16	0.964667	0.013006	0.110627	0.119064	-0.00844	0.954667	-0.01441	0.120069	0.125989	-0.00592

Table 4.6: *Performance measure estimates of a larger connected network of evidence for sample size 500 (fixed effect)*

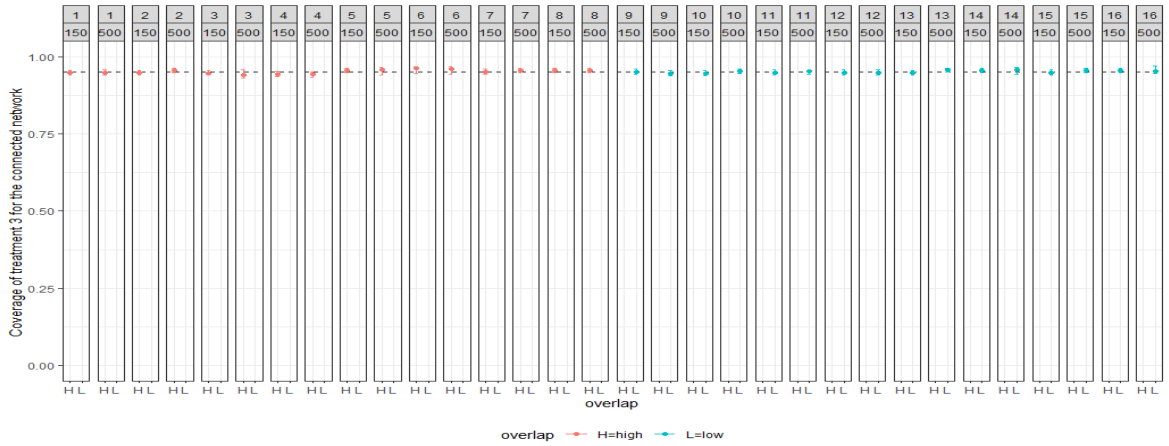
DGM	COVERAGE d1	BIAS d1	EMP SE d1	MODEL SE d1	DIFFERENCE d1	COVERAGE d2	BIAS d2	EMP SE d2	MODEL SE d2	DIFFERENCE d2
Scenario 1	0.942333	0.004471	0.068184	0.067002	0.001182	0.948667	-0.00357	0.068884	0.070229	-0.00134
Scenario 2	0.944333	0.00415	0.067495	0.066719	0.000775	0.955333	-0.00379	0.068377	0.070111	-0.00173
Scenario 3	0.934333	0.013772	0.073431	0.070451	0.00298	0.938333	0.000296	0.071729	0.069609	0.00212
Scenario 4	0.936667	0.013538	0.072184	0.06987	0.002314	0.942667	-0.0004	0.06996	0.069513	0.000447
Scenario 5	0.950667	0.005628	0.079405	0.08013	-0.00073	0.958667	-0.00325	0.074153	0.077133	-0.00298
Scenario 6	0.953	0.005837	0.075562	0.077634	-0.00207	0.96	-0.00415	0.072212	0.075935	-0.00372
Scenario 7	0.947667	0.015188	0.082653	0.083278	-0.00063	0.955667	-0.00149	0.075899	0.077639	-0.00174
Scenario 8	0.957	0.013234	0.077132	0.079876	-0.00274	0.956	-0.00182	0.073277	0.076401	-0.00312
Scenario 9	0.944667	0.002999	0.063554	0.062563	0.000992	0.946667	-0.00431	0.069381	0.069308	7.37E-05
Scenario 10	0.947667	0.003271	0.063127	0.062346	0.000781	0.953333	-0.00319	0.069072	0.069228	-0.00016
Scenario 11	0.936	0.012376	0.065731	0.063378	0.002353	0.932	-0.00177	0.070386	0.068709	0.001677
Scenario 12	0.942333	0.011541	0.064623	0.063016	0.001607	0.938	-0.0004	0.070768	0.068586	0.002182
Scenario 13	0.955667	0.00402	0.064407	0.066211	-0.0018	0.966667	-0.00378	0.065587	0.069469	-0.00388
Scenario 14	0.957333	0.003503	0.06197	0.064975	-0.003	0.968333	-0.00352	0.0636	0.068972	-0.00537
Scenario 15	0.954333	0.010648	0.064726	0.066467	-0.00174	0.954667	-0.00251	0.067169	0.069164	-0.002
Scenario 16	0.958667	0.010183	0.061768	0.06498	-0.00321	0.962333	-0.00186	0.06412	0.068696	-0.00458

4.7.2.5 Coverage for the simulation scenarios

The coverage of d1 and d2 from the connected NMA is at the nominal level for both sample sizes. Figure 4.25 (a),(b) shows the coverage for treatments 2 and 3 for varying levels of sample size and overlap. From the figures, it can be seen that the coverage was at the nominal level for the larger connected network of evidence for all 16 scenarios irrespective of sample sizes and overlap.



(a) coverage by overlap for treatment 2

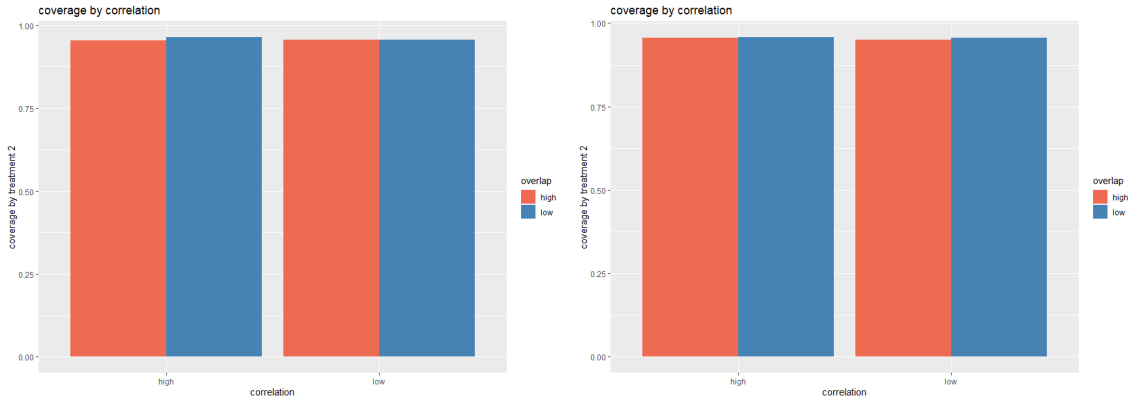


(b) coverage by overlap for treatment 3

Figure 4.25: *Coverage of treatments 2 and 3 for different sample sizes and overlap*

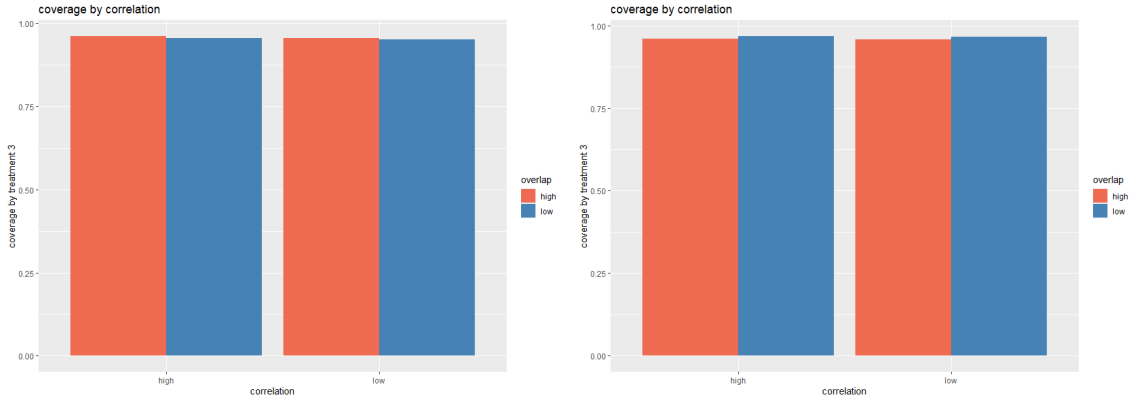
4.7.2.6 Coverage by correlation with overlap

Figure 4.26 and Figure 4.27 show the coverage with different levels of overlap and correlation for treatments 2 and 3 respectively. From the figures, it can be seen that the coverage is not found to be affected by different levels of correlation and overlap.



(a) *coverage by correlation for sample size 150* **(b)** *coverage by correlation for sample size 500*

Figure 4.26: *Coverage by correlation of treatment 2 for different sample sizes (connected larger network of evidence)*

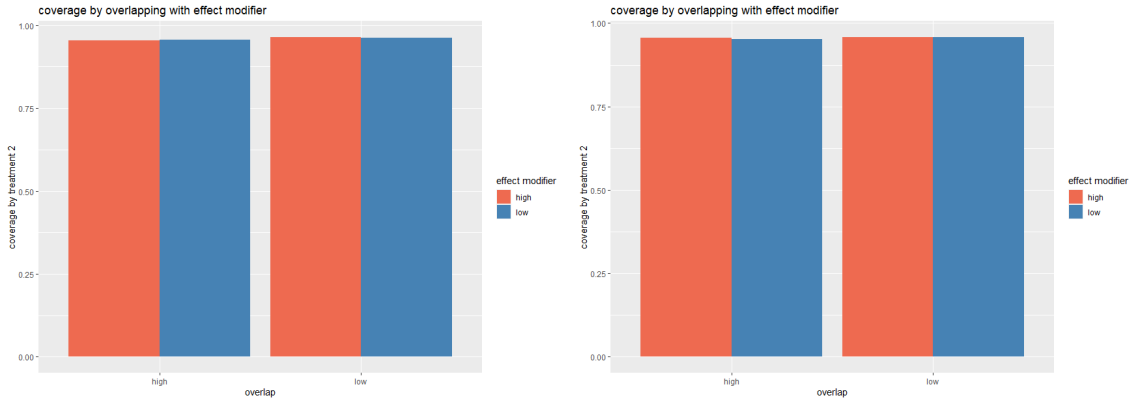


(a) *coverage by correlation for sample size 150* **(b)** *coverage by correlation for sample size 500*

Figure 4.27: *Coverage by correlation of treatment 3 for different sample sizes (connected larger network of evidence)*

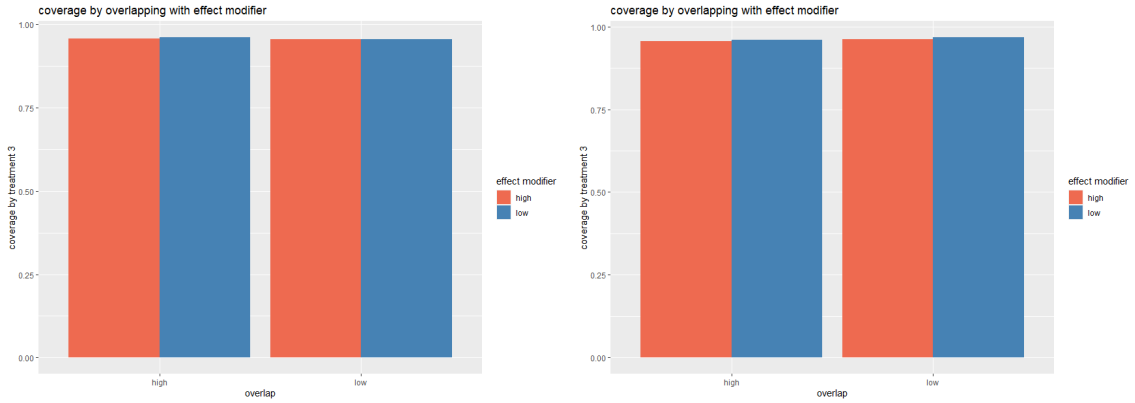
4.7.2.7 Coverage by overlap with effect-modifiers

Figure 4.28 and 4.29 show the coverage by different levels of overlap and effect-modifier for treatments 2 and 3 respectively. For both sample sizes, the coverage was found to be similar with varying levels of overlap and effect-modifier variables.



(a) *coverage by EM for sample size 150* **(b)** *coverage by EM for sample size 500*

Figure 4.28: *Coverage by overlap with EM levels for treatment 2 (connected larger network of evidence)*

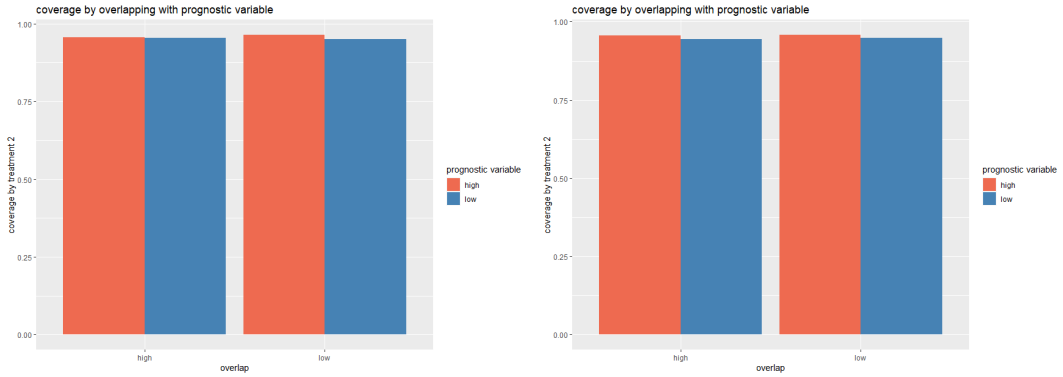


(a) *coverage by EM for sample size 150* **(b)** *coverage by EM for sample size 500*

Figure 4.29: *Coverage by overlap with EM levels for treatment 3 (connected larger network of evidence)*

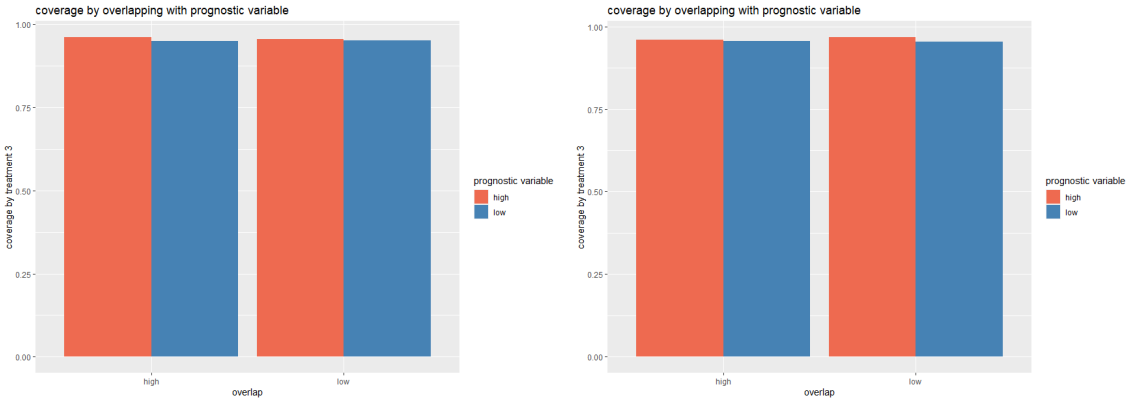
4.7.2.8 Coverage by overlap with prognostic variable

Figure 4.30 and 4.31 show the amount of coverage by different levels of overlap and prognostic variables for treatment 2 and 3 respectively. For both sample sizes, the coverage was similar with varying levels of overlap and prognostic variables.



(a) *coverage by PV for sample size 150* **(b)** *coverage by PV for sample size 500*

Figure 4.30: *Coverage by overlap with PV levels for treatment 2 (connected larger network of evidence)*



(a) *coverage by PV for sample size 150* **(b)** *coverage by PV for sample size 500*

Figure 4.31: *Coverage by overlap with PV levels for treatment 3 (connected larger network of evidence)*

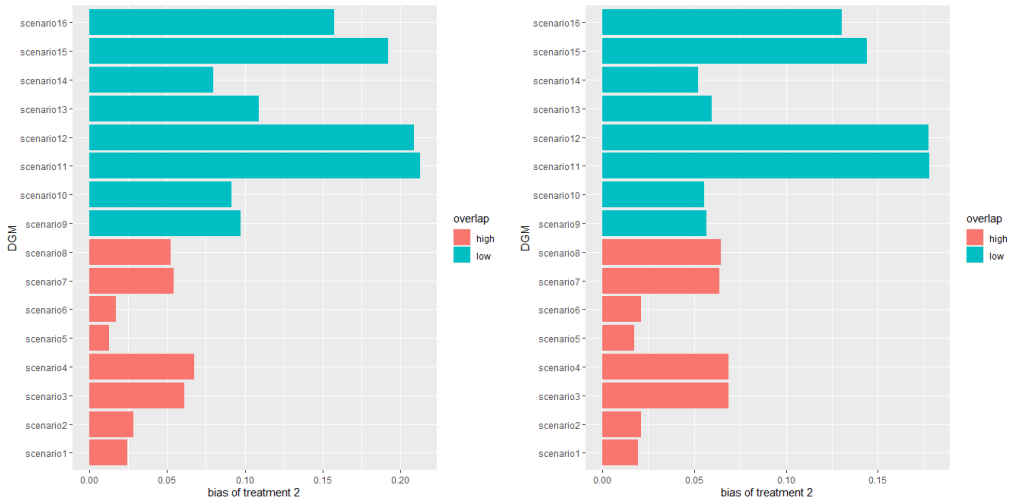
4.7.3 Results of simulation scenarios for the smaller disconnected network of evidence

Data were generated for the smaller disconnected network of evidence as described in Section 4.3.1 and in Figure 4.3 where at first 3 RCTs were generated and then they were made disconnected artificially to perform MAIC. The MAIC estimates were then used to perform a NMA. The following sections describe the simulation results from the MAIC-adjusted NMA for the smaller network.

4.7.3.1 Bias by overlap

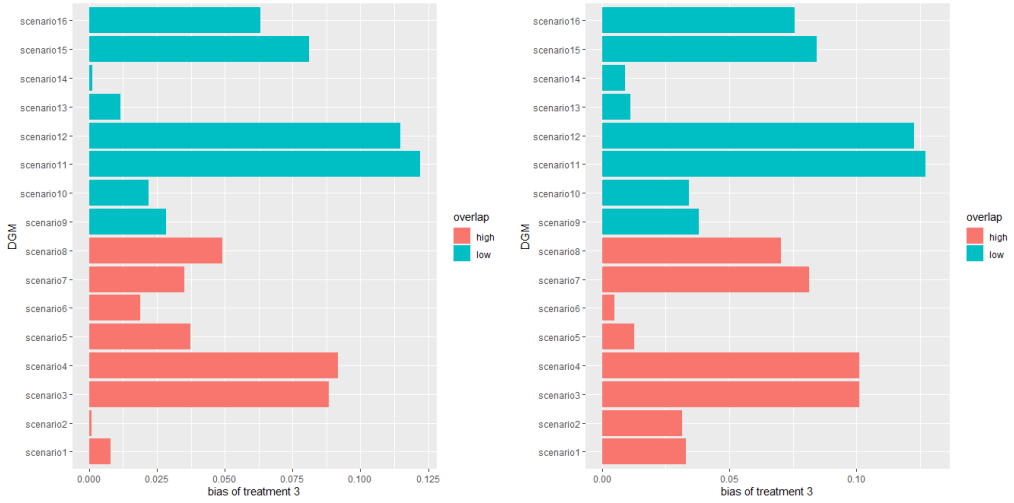
Figure 4.32 and 4.33 show the amount of bias for each DGM for treatment 2 and 3 respectively. Bias was found to be related to overlap. The biases were low for high-overlap scenarios and high for low-overlap scenarios and it was found true for both sample sizes. Figure 4.32 (a),(b)

show the biases for treatment 2 for sample sizes 150 and 500. For both sample sizes, with high overlap, the amount of biases was high for scenarios 3, 4, 7, 8, and that was high for scenarios 11, 12, 15, 16 for low overlap. The effect-modifier variable level was high in these scenarios. Figure 4.33 (a),(b) show the biases for treatment 3 for sample sizes 150 and 500. For treatment 3 the bias was high for scenarios 3, 4, 7, 8, 11, 12 for both sample sizes. Similar to treatment 2, the effect-modifier variable level was also high in these scenarios.



(a) *bias by overlap for sample size 150* **(b)** *bias by overlap for sample size 500*

Figure 4.32: *Bias of treatment 2 for different sample sizes (disconnected smaller network of evidence)*

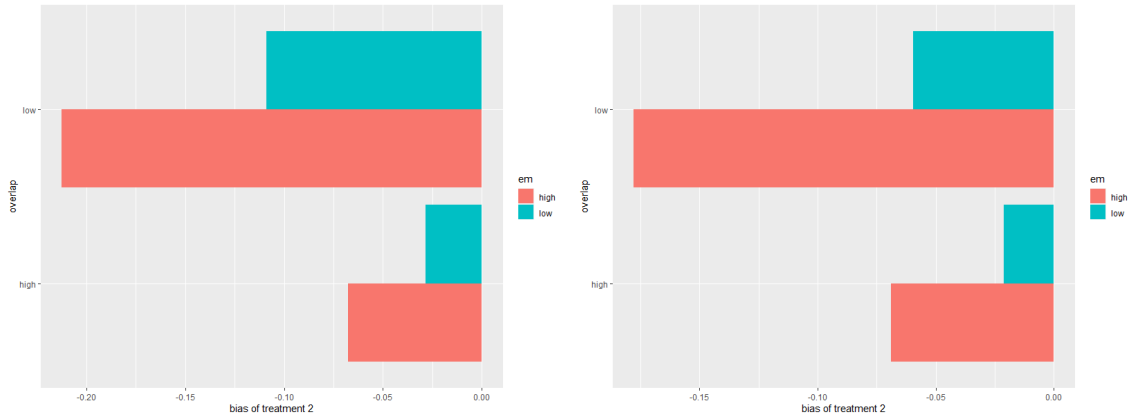


(a) *bias by overlap for sample size 150* **(b)** *bias by overlap for sample size 500*

Figure 4.33: *Bias of treatment 3 for different sample sizes (disconnected smaller network of evidence)*

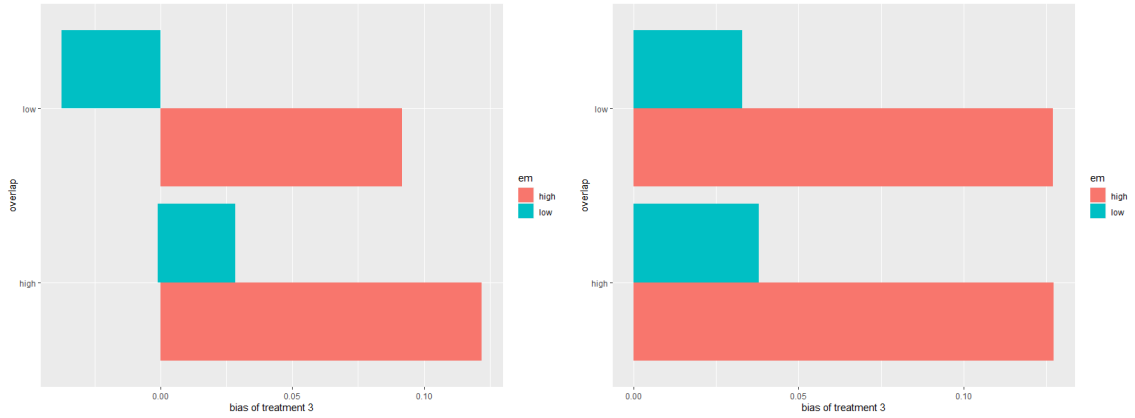
4.7.3.2 Bias by overlap with effect-modifiers

Figure 4.34 and 4.35 show the amount of bias by overlap for different levels of effect-modifier for treatments 2 and 3 respectively. In Figure 4.34 (a),(b), the bias reduces with high overlap level, and the reduction seems to depend on effect-modifier levels also. For treatment 2, the bias is low with the low level of effect-modifier and this reduction is more prominent for the sample size of 500. In Figure 4.35 (a),(b), for treatment 3, a lower bias was found with a lower effect-modifier level.



(a) *bias by EM for sample size 150* **(b)** *bias by EM for sample size 500*

Figure 4.34: *Bias by overlap with EM levels for treatment 2 (disconnected smaller network of evidence)*



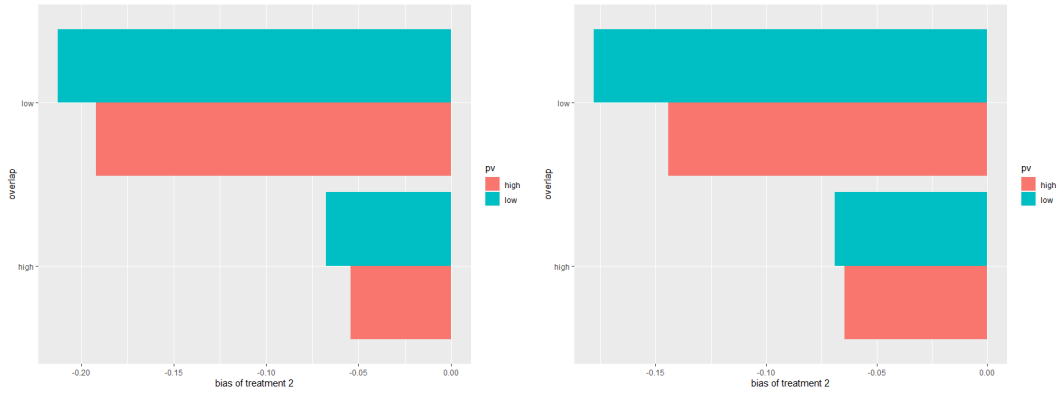
(a) *bias by EM for sample size 150* **(b)** *bias by EM for sample size 500*

Figure 4.35: *Bias by overlap with EM levels for treatment 3 (disconnected smaller network of evidence)*

4.7.3.3 Bias by overlap with prognostic variable

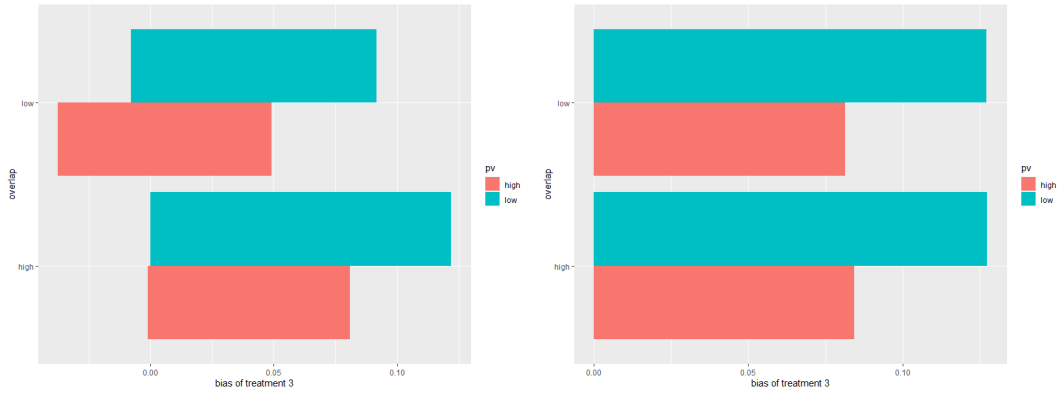
Figure 4.36 and 4.37 show the amount of bias by overlap for different levels of prognostic variables for treatment 2 and 3 respectively. In Figure 4.36 (a),(b), the bias reduces with high

overlap but within the levels of overlap, biases increase with the low level of the prognostic variable. In Figure 4.37 (a),(b) for treatment 3, for both levels of overlap, a bigger bias was found for the lower level of the prognostic variable.



(a) *bias by PV for sample size 150* **(b)** *bias by PV for sample size 500*

Figure 4.36: *Bias by overlap with PV levels for treatment 2 (disconnected smaller network of evidence)*



(a) *bias by PV for sample size 150* **(b)** *bias by PV for sample size 500*

Figure 4.37: *Bias by overlap with PV levels for treatment 3 (disconnected smaller network of evidence)*

4.7.3.4 Empirical SE and model SE by simulation scenarios

The simulation results of performance measures for the smaller disconnected network of evidence were summarised in Table 4.7 and Table 4.8. Table 4.7 and Table 4.8 show the relative estimates of treatments 2 and 3 with treatment 1 from the disconnected NMA for both sample sizes. The empirical SE and model SE are similar or close to each other for the high overlap scenarios where the coverage is at a nominal level. The difference between the empirical SE and model SE starts to increase for the low-overlap scenarios. The empirical SE was bigger than the model SE for low-overlap scenarios. The highest value of the difference between the empirical SE and model SE was found to be 0.09 for the smaller disconnected network of

evidence.

From Table 4.7 and Table 4.8 it can be seen that the coverage starts to decrease from the low-overlap scenario which starts from scenario 9. Most of the higher biases were seen for those low-overlap scenarios also. Additionally, the presence of moderate biases was seen for some scenarios. Most of the time, the presence of moderate or higher biases was seen for higher level of effect-modifying variable.

Table 4.7: *Performance measure estimates of a smaller disconnected network of evidence for sample size 150 (fixed effect)*

DGM	COVERAGE d1	BIAS d1	EMP SE d1	MODEL SE d1	DIFFERENCE d1	COVERAGE d2	BIAS d2	EMP SE d2	MODEL SE d2	DIFFERENCE d2
Scenario 1	0.952333	-0.0244	0.328452	0.323663	0.004789	0.951667	-0.00785	0.281444	0.283762	-0.00232
Scenario 2	0.949667	-0.02859	0.349496	0.3446	0.004896	0.953333	-0.00101	0.306072	0.308316	-0.00224
Scenario 3	0.942333	-0.06121	0.351777	0.338818	0.012959	0.947	0.088515	0.278731	0.281524	-0.00279
Scenario 4	0.94	-0.06754	0.369571	0.35743	0.012141	0.951667	0.091708	0.303173	0.306511	-0.00334
Scenario 5	0.950667	-0.01265	0.396919	0.395676	0.001243	0.961667	-0.03738	0.313231	0.32604	-0.01281
Scenario 6	0.955	-0.01703	0.404454	0.407685	-0.00323	0.954667	-0.01877	0.336248	0.348738	-0.01249
Scenario 7	0.954	-0.05442	0.409406	0.408864	0.000542	0.959667	0.035176	0.31413	0.327798	-0.01367
Scenario 8	0.956	-0.05257	0.41313	0.416418	-0.00329	0.956333	0.049196	0.336585	0.349956	-0.01337
Scenario 9	0.919333	-0.09748	0.547898	0.501499	0.046399	0.920333	0.028385	0.529931	0.49293	0.037001
Scenario 10	0.927	-0.09158	0.510497	0.478942	0.031556	0.933667	0.02195	0.494912	0.470394	0.024517
Scenario 11	0.897333	-0.2125	0.555401	0.502674	0.052727	0.917333	0.121993	0.524681	0.491603	0.033078
Scenario 12	0.902	-0.20872	0.516642	0.479734	0.036908	0.927	0.11461	0.49305	0.468853	0.024197
Scenario 13	0.92	-0.10913	0.646827	0.551519	0.095308	0.927	0.011583	0.593211	0.538513	0.054699
Scenario 14	0.941	-0.07966	0.532749	0.521642	0.011107	0.946667	-0.00107	0.515587	0.509615	0.005972
Scenario 15	0.913	-0.19203	0.642614	0.551548	0.091066	0.924667	0.081001	0.59296	0.537627	0.055334
Scenario 16	0.935	-0.15757	0.530698	0.521174	0.009524	0.944	0.063101	0.516608	0.509379	0.007229

Table 4.8: *Performance measure estimates of a smaller disconnected network of evidence for sample size 500 (fixed effect)*

DGM	COVERAGE d1	BIAS d1	EMP SE d1	MODEL SE d1	DIFFERENCE d1	COVERAGE d2	BIAS d2	EMP SE d2	MODEL SE d2	DIFFERENCE d2
Scenario 1	0.943	-0.01938	0.177358	0.173135	0.004223	0.945333	0.032875	0.152248	0.150998	0.00125
Scenario 2	0.945667	-0.02112	0.187332	0.184201	0.003131	0.948667	0.031449	0.163125	0.163431	-0.00031
Scenario 3	0.9356	-0.06872	0.18932	0.180768	0.008552	0.935	0.0711	0.155628	0.14997	0.005658
Scenario 4	0.9384	-0.06893	0.196717	0.190844	0.005873	0.9464	0.0712	0.167482	0.162474	0.005008
Scenario 5	0.956	-0.01766	0.207112	0.210516	-0.0034	0.958	0.012686	0.164911	0.173069	-0.00816
Scenario 6	0.952667	-0.02105	0.212287	0.217435	-0.00515	0.958333	0.004816	0.177369	0.185362	-0.00799
Scenario 7	0.945667	-0.06378	0.216627	0.217138	-0.00051	0.929333	0.081444	0.168115	0.173966	-0.00585
Scenario 8	0.949667	-0.06452	0.216736	0.2218	-0.00506	0.949333	0.070406	0.177912	0.18613	-0.00822
Scenario 9	0.935667	-0.05685	0.271275	0.262045	0.00923	0.942	0.037927	0.260442	0.253935	0.006507
Scenario 10	0.935333	-0.05571	0.258057	0.249451	0.008605	0.937667	0.034216	0.247495	0.240897	0.006599
Scenario 11	0.89	-0.12781	0.273546	0.262851	0.010695	0.922333	0.123118	0.263165	0.253168	0.009997
Scenario 12	0.896667	-0.12754	0.260683	0.250007	0.010676	0.923667	0.122623	0.249648	0.239914	0.009734
Scenario 13	0.940333	-0.05956	0.290407	0.293059	-0.00265	0.951	0.01103	0.274034	0.282947	-0.00891
Scenario 14	0.947	-0.05238	0.270738	0.275116	-0.00438	0.956667	0.008946	0.261074	0.265079	-0.00401
Scenario 15	0.928667	-0.14413	0.290586	0.293143	-0.00256	0.944333	0.084374	0.275561	0.282642	-0.00708
Scenario 16	0.933	-0.13033	0.270151	0.274752	-0.0046	0.948	0.075615	0.262201	0.264718	-0.00252

4.7.3.5 Coverage by simulation scenarios

The coverage of d1 and d2 from the disconnected NMA is at the nominal level for the high overlap scenarios. Figure 4.38 and Figure 4.39 illustrate the impact of MAIC-adjusted NMA on coverage for treatments 2 and 3 respectively. The confidence intervals for the high-overlap scenarios are represented in red colour and that for the low-overlap scenarios in blue colour. Figure 4.38 and 4.39 show that when the overlap between studies was high, the coverage was at the nominal level for both sample sizes 150 and 500. When the overlap was low, the coverage for some scenarios started to decrease slightly. For treatment 2, the lowest coverage was 90% and 89% for sample sizes 150 and 500 respectively. For treatment 3, the lowest coverage was 92% for both sample sizes. Within low overlap, for the scenarios where the correlation was high, a slight increase was seen for the coverage and this was true for both sample sizes.

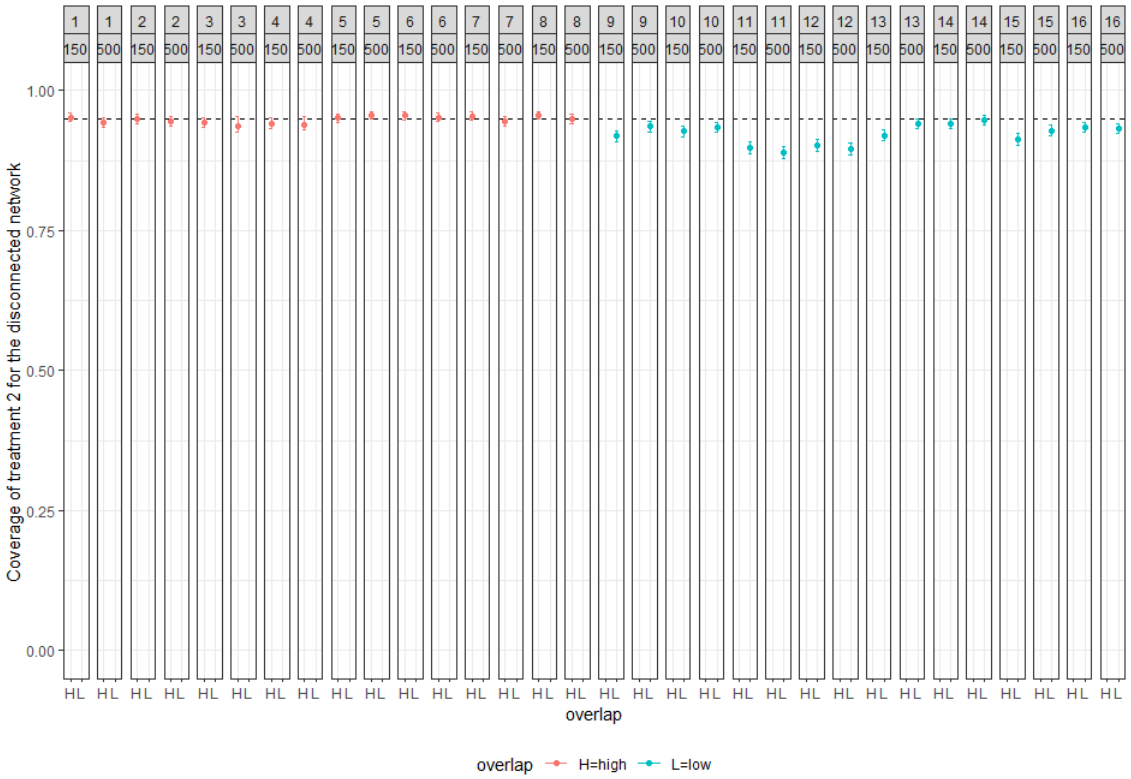


Figure 4.38: *Coverage by overlap for treatment 2*

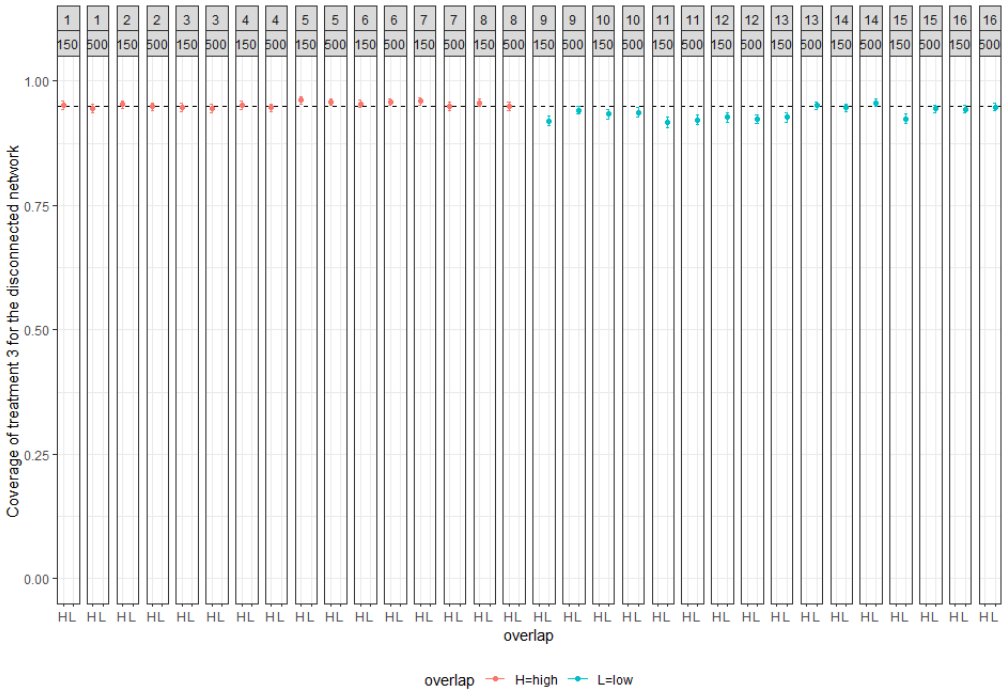
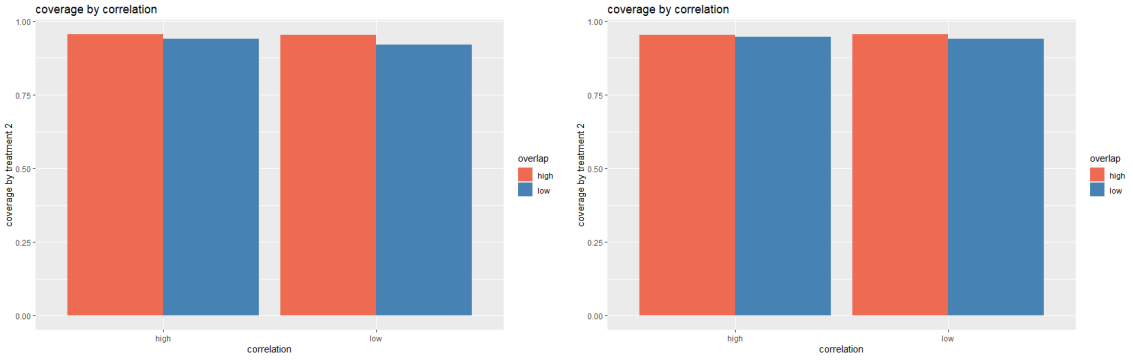


Figure 4.39: *Coverage by overlap for treatment 3*

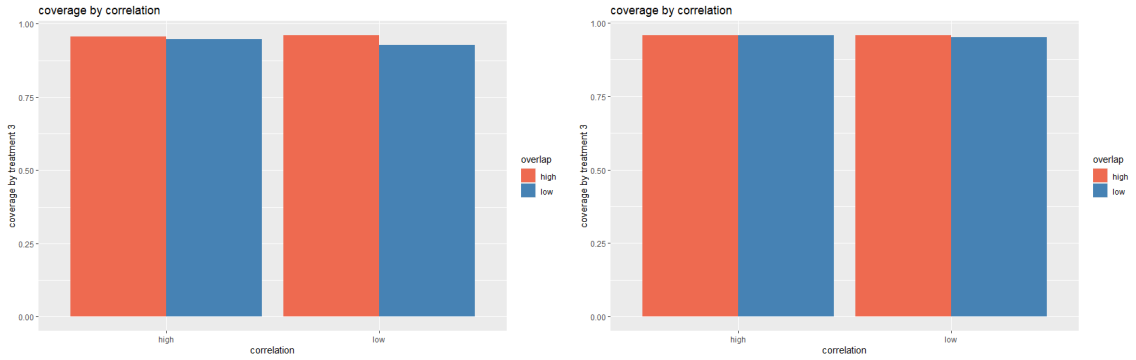
4.7.3.6 Coverage by correlation with overlap

Figure 4.40 and Figure 4.41 show the coverage with different levels of overlap and correlation for treatments 2 and 3 respectively. In Figure 4.40 (b) and Figure 4.41 (b), for sample size 500, the coverage is similar with different levels of overlap and correlation for both treatments. However, with sample size 150 and with a low correlation value, the coverage is slightly higher with high overlap.



(a) *coverage by correlation for sample size 150* **(b)** *coverage by correlation for sample size 500*

Figure 4.40: *Coverage by correlation of treatment 2 for different sample sizes (disconnected smaller network of evidence)*

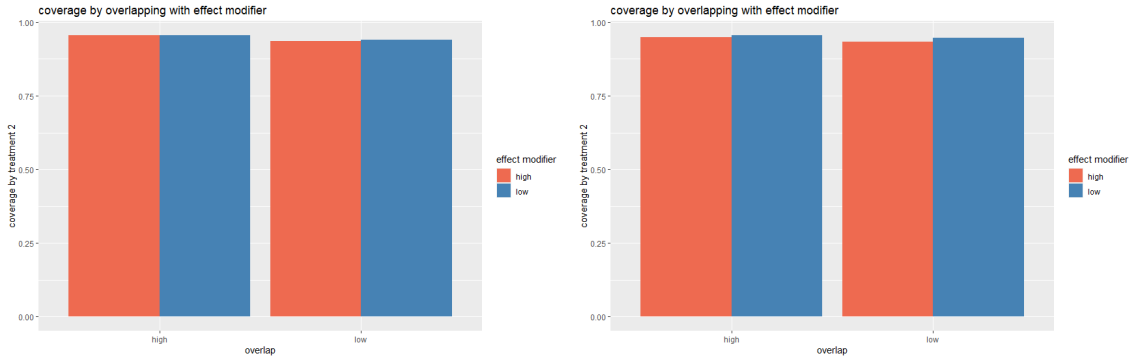


(a) *coverage by correlation for sample size 150* **(b)** *coverage by correlation for sample size 500*

Figure 4.41: *Coverage by correlation of treatment 3 for different sample sizes (disconnected smaller network of evidence)*

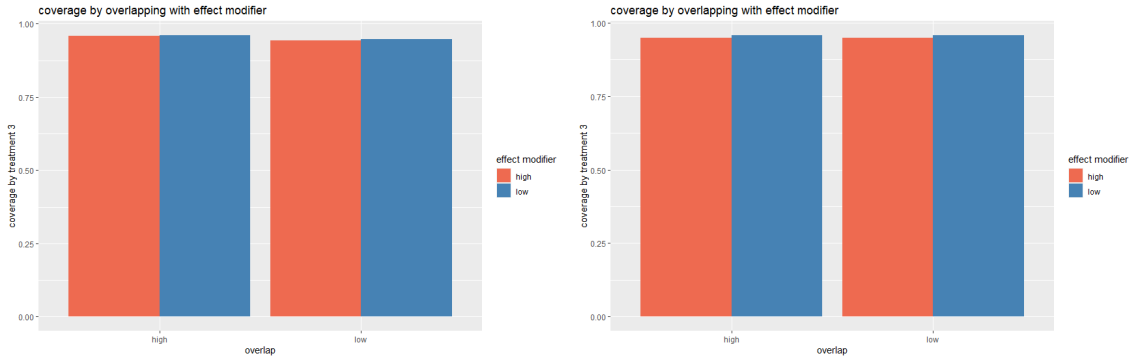
4.7.3.7 Coverage by overlap with effect-modifiers

Figure 4.42 and 4.43 show the coverage by different levels of overlap and effect-modifiers for treatments 2 and 3 respectively. In Figure 4.42 (a) and 4.43 (a), for both treatments 2 and 3, with sample size 150, the coverage decreases slightly from the nominal level with low overlap, but effect-modifiers show no impact on coverage irrespective of sample size and treatments.



(a) *coverage with EM for sample size 150* **(b)** *coverage with EM for sample size 500*

Figure 4.42: *Coverage by overlap with EM levels for treatment 2 (disconnected smaller network of evidence)*

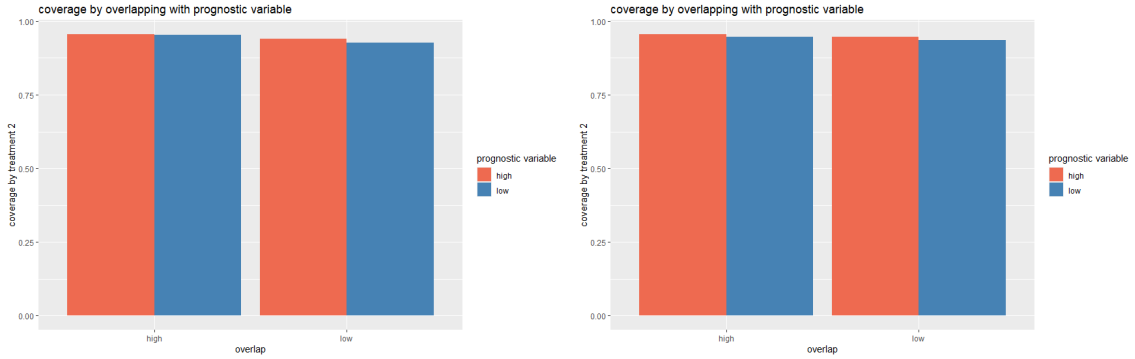


(a) *coverage with EM for sample size 150* **(b)** *coverage with EM for sample size 500*

Figure 4.43: *Coverage by overlap with EM levels for treatment 3 (disconnected smaller network of evidence)*

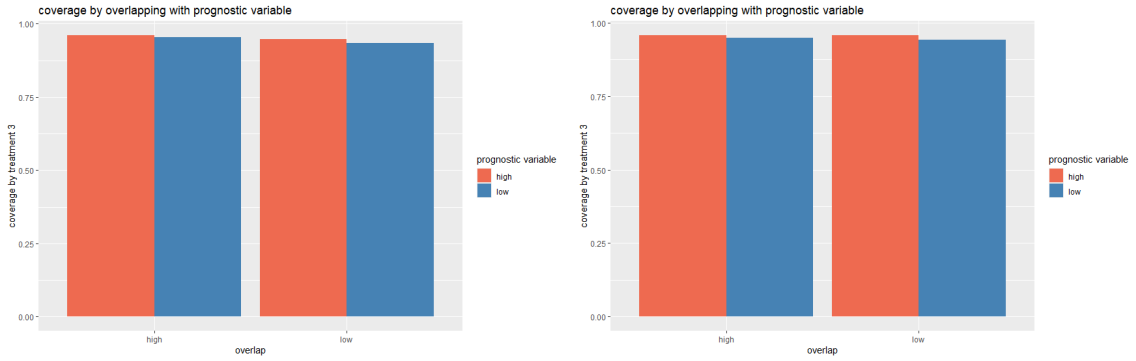
4.7.3.8 Coverage by overlap with prognostic variable

Figure 4.44 and 4.45 show the amount of coverage by different levels of overlap and prognostic variables for treatment 2 and 3 respectively. For both sample sizes, the coverage was similar with varying levels of overlap and prognostic variables.



(a) *coverage by PV for sample size 150* **(b)** *coverage by PV for sample size 500*

Figure 4.44: *Coverage by overlap with PV levels for treatment 2 for different sample sizes (disconnected smaller network of evidence)*



(a) *coverage by PV for sample size 150* **(b)** *coverage by PV for sample size 500*

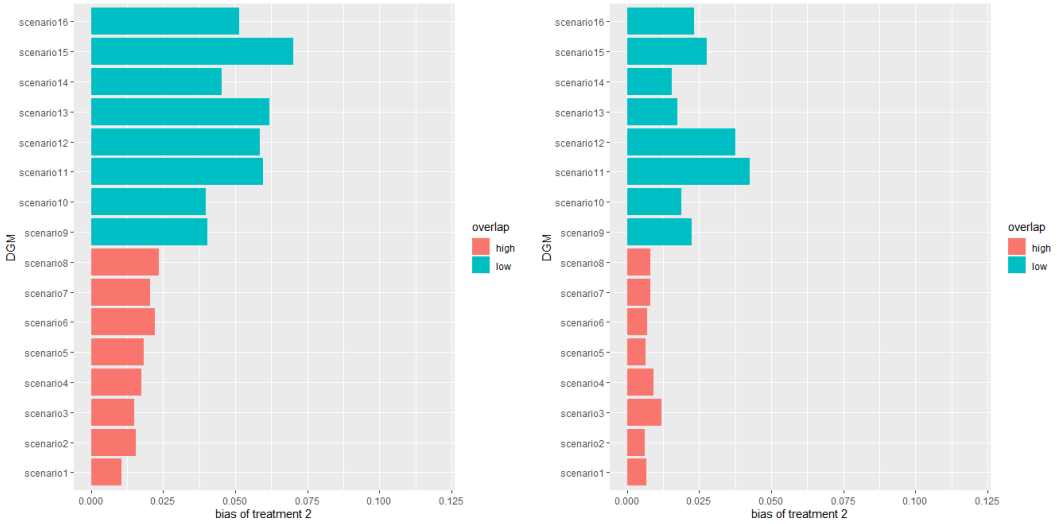
Figure 4.45: *Coverage by overlap with PV levels of treatment 3 for different sample sizes (disconnected smaller network of evidence)*

4.7.4 Results of simulation scenarios for the larger disconnected network of evidence

Data were generated for the larger disconnected network of evidence as described in Section 4.3.2 and in Figure 4.5 where at first 10 RCTs were generated and then these RCTs were made disconnected artificially to implement MAIC. The MAIC estimates were then used to perform NMA again. The following sections describe the simulation results from the MAIC-adjusted NMA.

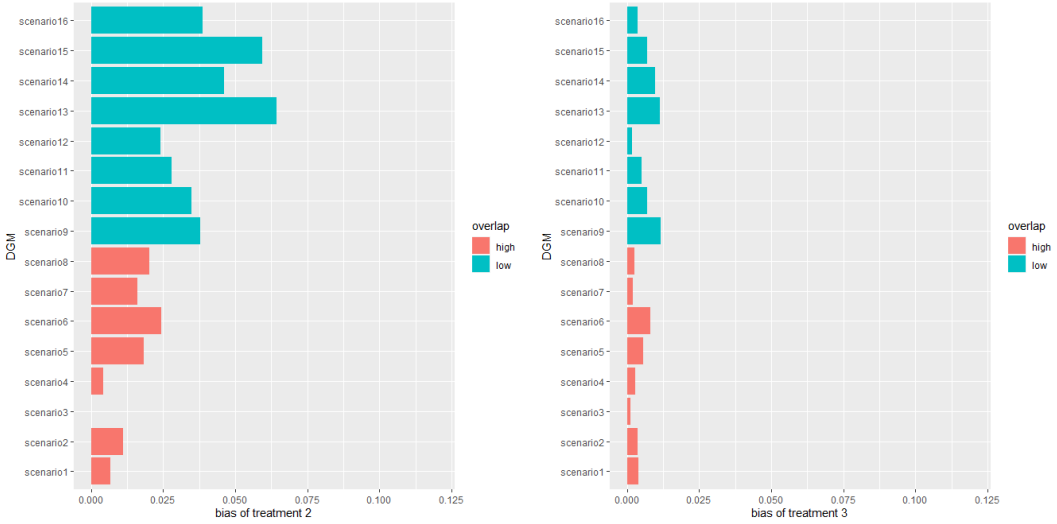
4.7.4.1 Bias by overlap

Figure 4.46 and 4.47 show the amount of bias for each DGM. The amount of bias was found to decrease with the amount of overlap. The biases were low for high-overlap scenarios and high for low-overlap scenarios and it was found true for both sample sizes. Overall, the scenarios where the bias was found to be high were 11, 12, 13, 15, 16. Effect-modifier variables were high in these scenarios except scenario 13. In scenario 13, the coefficient of the effect-modifier was low but the level of the prognostic variable was high and the level of the correlation was low.



(a) *bias by overlap for sample size 150* **(b)** *bias by overlap for sample size 500*

Figure 4.46: *Bias of treatment 2 for different sample sizes (disconnected larger network of evidence)*



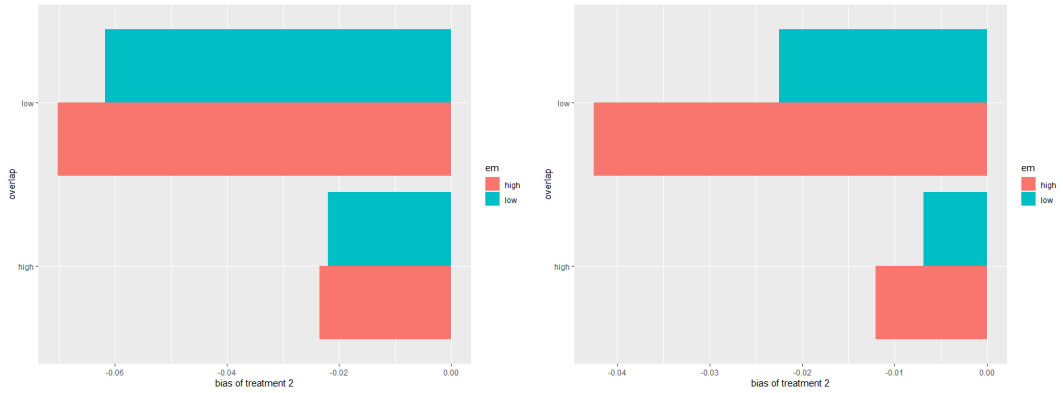
(a) *bias by overlap for sample size 150* **(b)** *bias by overlap for sample size 500*

Figure 4.47: *Bias of treatment 3 for different sample sizes (disconnected larger network of evidence)*

4.7.4.2 Bias by overlap with effect-modifiers

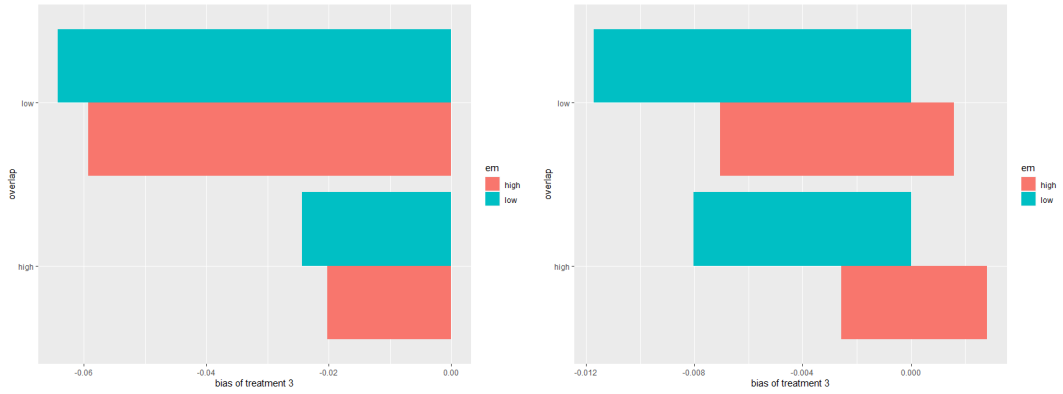
Figure 4.48 and 4.49 show the amount of bias by overlap for different levels of effect-modifier for treatments 2 and 3 respectively. In Figure 4.48 (a),(b), the bias reduces with high overlap, and the reduction seems to depend on effect-modifier levels also. For treatment 2 in Figure 4.48 (a),(b) the bias is low with the low level of effect-modifier and this reduction is more prominent for sample size 500. In Figure 4.49 (a), for treatment 3, the bias reduces with high

overlap but within levels of overlap, more bias is found with a low level of effect-modifier for sample size 150. However, 4.49 (b) with a sample size of 500, an increase in bias was found for low overlap and low effect-modifier level.



(a) *bias by EM for sample size 150* **(b)** *bias by EM for sample size 500*

Figure 4.48: *Bias by overlap with EM levels for treatment 2 (disconnected larger network of evidence)*

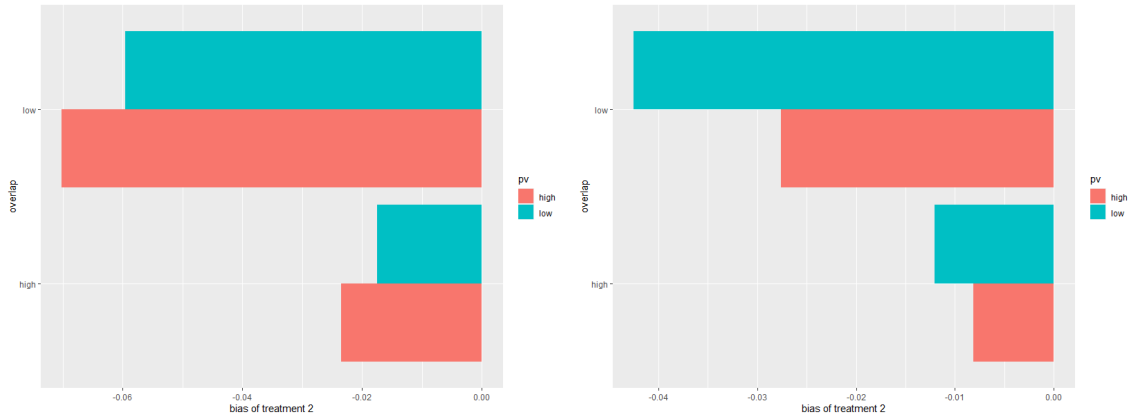


(a) *bias by EM for sample size 150* **(b)** *bias by EM for sample size 500*

Figure 4.49: *Bias by overlap with EM levels for treatment 3 (disconnected larger network of evidence)*

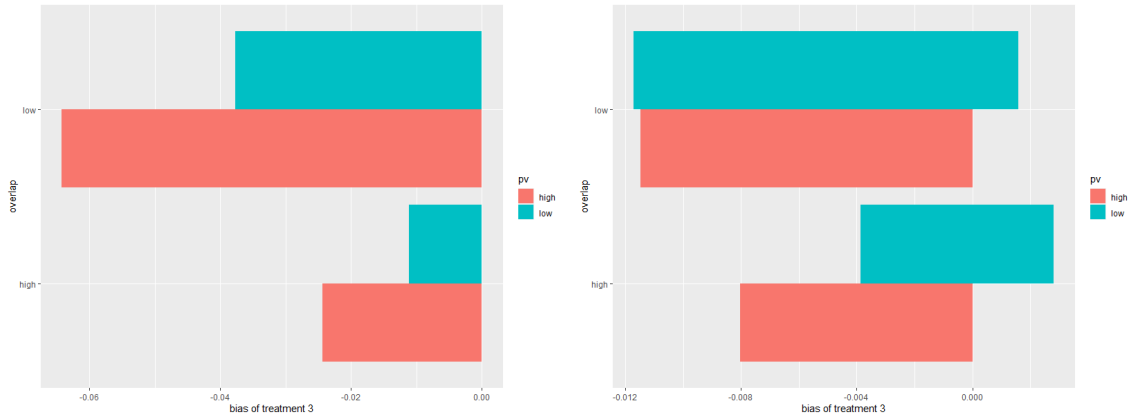
4.7.4.3 Bias by overlap with prognostic variable

Figure 4.50 and 4.51 show the amount of bias by overlap for different levels of prognostic variables for treatments 2 and 3 respectively. In Figure 4.50 (a), the bias reduces with high overlap but increases with the high level of prognostic variable coefficient. The opposite trend was seen for treatment 2 in sample size 500 in Figure 4.50 (b). In Figure 4.51 (a) for treatment 3, with sample size of 150, a higher bias is seen for low overlap and high level of the prognostic variable whereas in Figure 4.51 (b) with the sample size of 500, a higher bias was seen for low overlap and low level of the prognostic variable.



(a) *bias by PV for sample size 150* **(b)** *bias by PV for sample size 500*

Figure 4.50: *Bias by overlap with PV levels for treatment 2 (disconnected larger network of evidence)*



(a) *bias by PV for sample size 150* **(b)** *bias by PV for sample size 500*

Figure 4.51: *Bias by overlap with PV levels for treatment 3 (disconnected larger network of evidence)*

4.7.4.4 Empirical SE and model SE by simulation scenarios

The simulation results of performance measures for the larger disconnected network of evidence were summarised in Table 4.9 and Table 4.10. The empirical SE is always bigger than the model SE which shows undercoverage for all the scenarios. The highest value of the difference between the empirical SE and model SE was found to be 0.28 for the larger disconnected network of evidence. Additionally, all the simulation scenarios show deviation from the nominal level of coverage which is depicted by the red color. For sample size 150, moderate level of biases were seen mainly for low-overlap scenarios, however, overall the magnitude of biases was small for the bigger sample size.

Table 4.9: *Performance measure estimates of a larger disconnected network of evidence for sample size 150 (fixed effect)*

DGM	COVERAGE d1	BIAS d1	EMP SE d1	MODEL SE d1	DIFFERENCE d1	COVERAGE d2	BIAS d2	EMP SE d2	MODEL SE d2	DIFFERENCE d2
Scenario 1	0.766	-0.01049	0.242265	0.144621	0.097644	0.767667	-0.00675	0.234665	0.141232	0.093434
Scenario 2	0.751333	-0.01542	0.262395	0.152425	0.10997	0.760667	-0.01119	0.254919	0.150967	0.103951
Scenario 3	0.776667	-0.01495	0.24708	0.151026	0.096054	0.75	-4.14E-05	0.237374	0.140221	0.097153
Scenario 4	0.756	-0.01747	0.26729	0.158135	0.109155	0.751667	-0.00431	0.257252	0.150036	0.107216
Scenario 5	0.763667	-0.01828	0.293299	0.176485	0.116813	0.756333	-0.01836	0.27812	0.162527	0.115593
Scenario 6	0.764	-0.022	0.30474	0.180363	0.124377	0.758333	-0.02439	0.28971	0.171114	0.118596
Scenario 7	0.777	-0.02061	0.299154	0.1822	0.116954	0.754667	-0.01592	0.279903	0.163459	0.116444
Scenario 8	0.776667	-0.02346	0.307536	0.184306	0.12323	0.766667	-0.02027	0.291727	0.171846	0.119881
Scenario 9	0.650333	-0.04042	0.470162	0.220453	0.249709	0.688667	-0.03773	0.468729	0.239118	0.229611
Scenario 10	0.661	-0.03975	0.432387	0.207365	0.225022	0.695333	-0.03484	0.433484	0.223827	0.209657
Scenario 11	0.649	-0.05965	0.471088	0.221218	0.24987	0.684667	-0.02797	0.469655	0.238253	0.231403
Scenario 12	0.664	-0.05853	0.434616	0.207938	0.226678	0.684	-0.02412	0.434038	0.223162	0.210876
Scenario 13	0.648667	-0.06184	0.531446	0.243582	0.287864	0.685	-0.06422	0.528986	0.262509	0.266476
Scenario 14	0.678667	-0.04539	0.462315	0.227182	0.235134	0.722667	-0.0461	0.457898	0.244487	0.213411
Scenario 15	0.653333	-0.07019	0.531369	0.243797	0.287572	0.690667	-0.05931	0.528962	0.262241	0.266722
Scenario 16	0.677	-0.05123	0.461339	0.226964	0.234375	0.725667	-0.03867	0.457642	0.244286	0.213355

Table 4.10: *Performance measure estimates of a larger disconnected network of evidence for sample size 500 (fixed effect)*

DGM	COVERAGE d1	BIAS d1	EMP SE d1	MODEL SE d1	DIFFERENCE d1	COVERAGE d2	BIAS d2	EMP SE d2	MODEL SE d2	DIFFERENCE d2
Scenario 1	0.762	-0.00671	0.130585	0.077404	0.053181	0.76	-0.00387	0.125772	0.075511	0.050262
Scenario 2	0.763	-0.00623	0.139405	0.081622	0.057783	0.77	-0.00363	0.134556	0.08074	0.053817
Scenario 3	0.767	-0.01203	0.134247	0.080757	0.05349	0.747667	0.001143	0.126848	0.074918	0.051929
Scenario 4	0.766	-0.00921	0.142092	0.084571	0.057522	0.740333	0.002817	0.135615	0.080198	0.055416
Scenario 5	0.768667	-0.00643	0.155973	0.093952	0.062021	0.754333	-0.00548	0.146924	0.086531	0.060394
Scenario 6	0.771	-0.0069	0.160952	0.096328	0.064624	0.757333	-0.00803	0.152628	0.091432	0.061196
Scenario 7	0.780333	-0.00813	0.156575	0.096955	0.059619	0.755667	-0.00209	0.147219	0.086958	0.060261
Scenario 8	0.772	-0.00814	0.161683	0.098443	0.063241	0.763333	-0.00257	0.152748	0.091822	0.060927
Scenario 9	0.646333	-0.02251	0.23957	0.115184	0.124386	0.686	-0.01171	0.238209	0.124847	0.113362
Scenario 10	0.673333	-0.01868	0.217331	0.108282	0.10905	0.708	-0.00703	0.215313	0.117009	0.098304
Scenario 11	0.639	-0.04248	0.241245	0.115516	0.125729	0.689667	-0.00498	0.238839	0.124489	0.11435
Scenario 12	0.671333	-0.03762	0.218042	0.108618	0.109423	0.703333	0.001582	0.215483	0.116542	0.098941
Scenario 13	0.671333	-0.01748	0.260027	0.129804	0.130223	0.705667	-0.01148	0.257862	0.140076	0.117785
Scenario 14	0.704333	-0.01561	0.228365	0.119629	0.108736	0.736	-0.00961	0.226461	0.128844	0.097617
Scenario 15	0.67	-0.02761	0.260677	0.129756	0.13092	0.701667	-0.00704	0.258831	0.139882	0.11895
Scenario 16	0.710667	-0.02328	0.227501	0.119493	0.108008	0.729667	-0.00363	0.226366	0.128608	0.097758

4.7.4.5 Coverage by simulation scenarios

Figure 4.52 and 4.53 illustrate the impact of MAIC-adjusted NMA on treatments 2 and 3 respectively. The confidence intervals for the high-overlap scenarios are represented in red colour. Figure 4.52 and 4.53 show that when the overlap between studies was high, the highest coverage was 78% for both sample sizes 150 and 500. For the 8 scenarios where the coverage was high, the trend was similar. When the overlap was low, the coverage for each of the 8 scenarios decreased more. For sample size 150, in low overlap level, the highest coverage was 67% and the lowest was 65% and those for sample size 500 were 71% and 65% respectively. Within low overlap, for the scenarios where the correlation was high, a slight increase was seen for the coverage and this was true for both sample sizes.

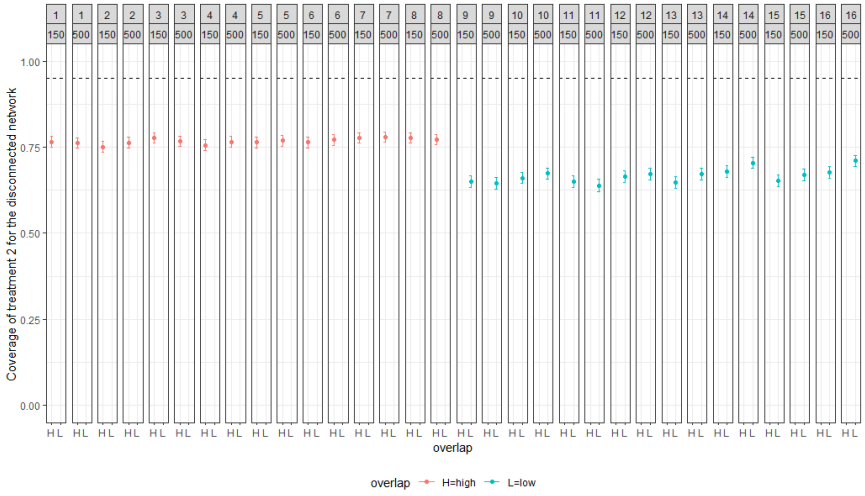


Figure 4.52: *Coverage by overlap for treatment 2*

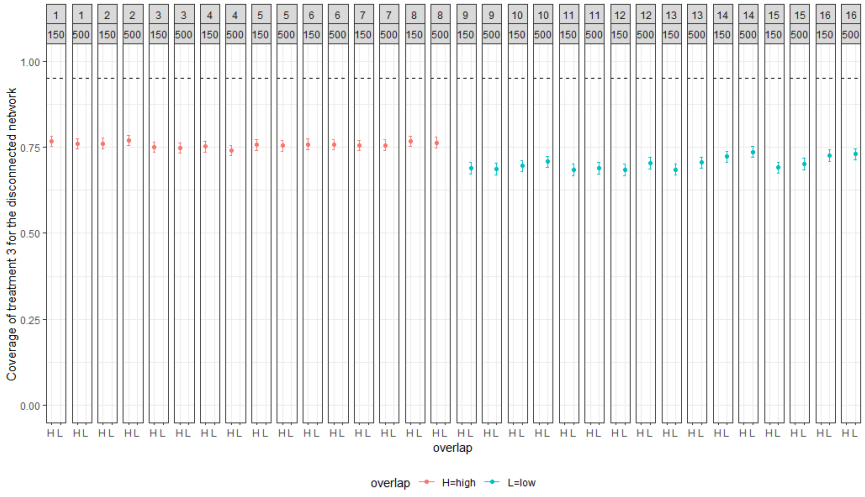
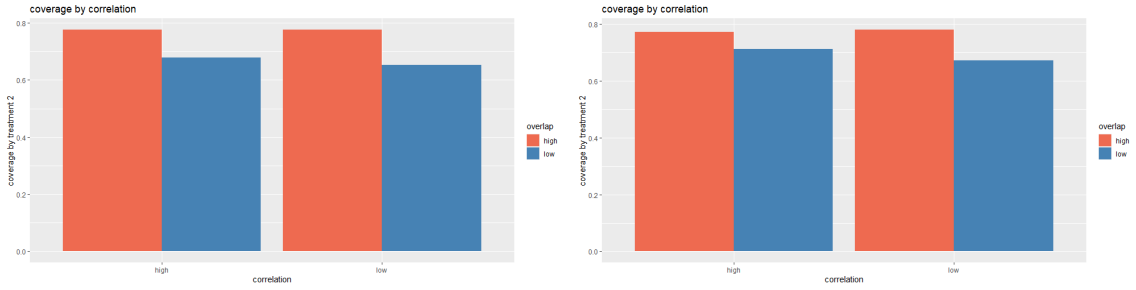


Figure 4.53: *Coverage by overlap for treatment 3*

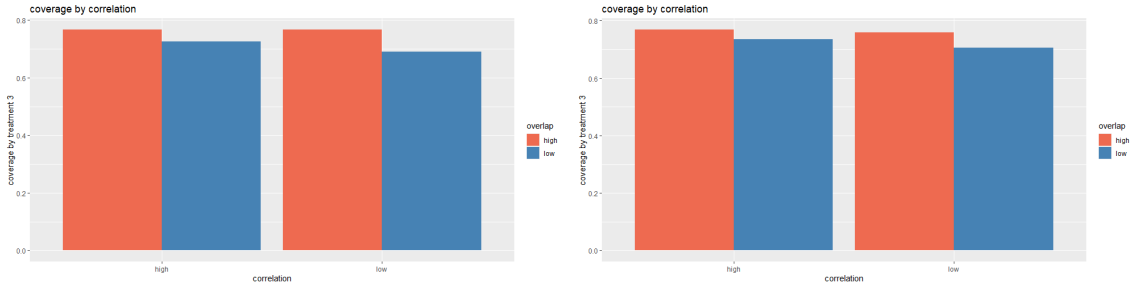
4.7.4.6 Coverage by correlation with overlap

Figure 4.54 and Figure 4.55 show the coverage with different levels of overlap and correlation for treatments 2 and 3 respectively. Figure 4.54 and 4.55 show that the coverage decreases slightly with a decrease in the overlap. However, the decrease in coverage was more for the small sample size compared to the big sample size. Additionally, when the overlap was low, the coverage was slightly higher for high correlation.



(a) *coverage by correlation for sample size 150* **(b)** *coverage by correlation for sample size 500*

Figure 4.54: *Coverage by correlation of treatment 2 for different sample sizes (disconnected larger network of evidence)*

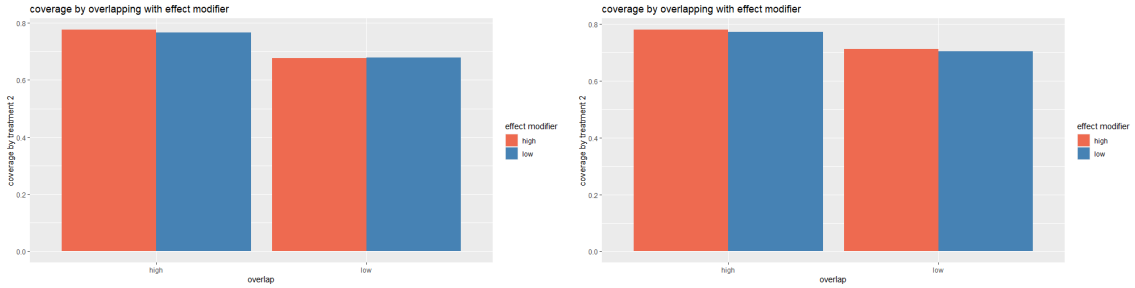


(a) *coverage by correlation for sample size 150* **(b)** *coverage by correlation for sample size 500*

Figure 4.55: *Coverage by correlation of treatment 3 for different sample sizes (disconnected larger network of evidence)*

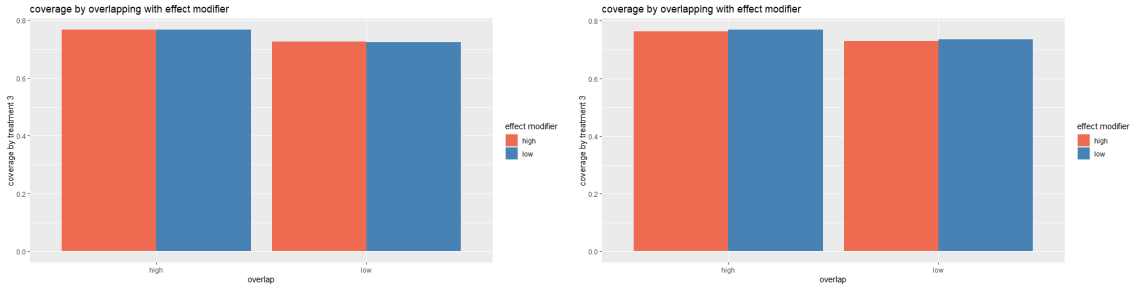
4.7.4.7 Coverage by overlap with effect-modifiers

Figure 4.56 and 4.57 show the coverage by overlap with effect-modifiers for treatments 2 and 3 respectively. The coverage is low with low overlap, but effect-modifiers show no impact on coverage irrespective of sample size and treatments.



(a) *coverage with EM for sample size 150* **(b)** *coverage with EM for sample size 500*

Figure 4.56: *Coverage by overlap with EM levels for treatment 2 (disconnected larger network of evidence)*

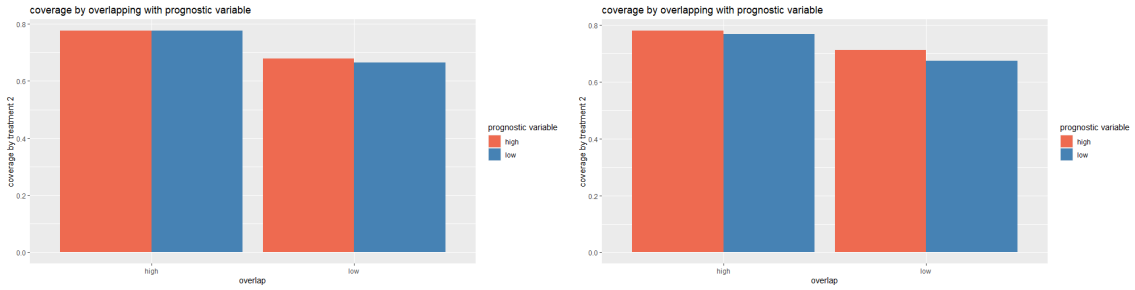


(a) *coverage with EM for sample size 150* **(b)** *coverage with EM for sample size 500*

Figure 4.57: *Coverage by overlap with EM levels for treatment 3 (disconnected larger network of evidence)*

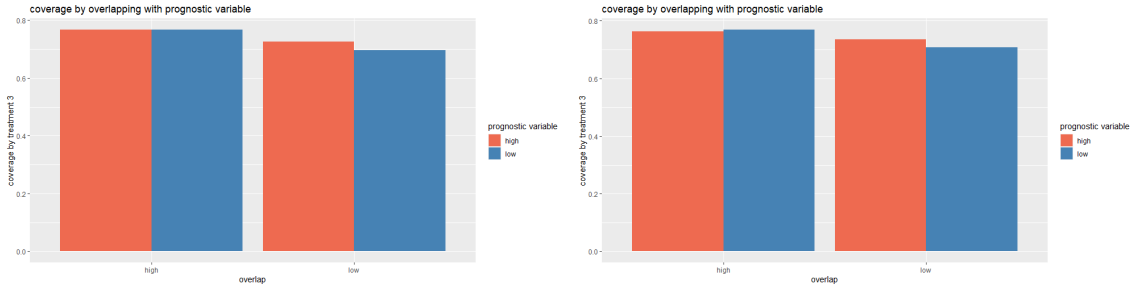
4.7.4.8 Coverage by overlap with prognostic variable

Figure 4.58 and 4.59 show the amount of coverage by different levels of overlap with respect to prognostic variables for treatments 2 and 3 respectively. For both sample sizes and treatments, the coverage was low with low overlap but the prognostic variable seems to have no effect on this.



(a) *coverage by PV for sample size 150* **(b)** *coverage by PV for sample size 500*

Figure 4.58: *Coverage by overlap with PV levels of treatment 2 (disconnected larger network of evidence)*



(a) *coverage by PV for sample size 150* **(b)** *coverage by PV for sample size 500*

Figure 4.59: *Coverage by overlap with PV levels of treatment 3 (disconnected larger network of evidence)*

4.7.5 Coverage by connected NMA and MAIC-adjusted NMA for the smaller network of evidence

Figure 4.60 and Figure 4.61 show the coverage for NMA estimates from the connected network of evidence and estimates from MAIC-adjusted NMA for the smaller network of evidence with sample sizes 150. Figure 4.62 and 4.63 show the coverage for NMA estimates from the connected network of evidence and estimates from MAIC-adjusted NMA for the smaller network of evidence with sample sizes 500. For each DGM, the coverage for the connected and MAIC-adjusted NMA was at a nominal level except for the DGM 9,10,12,13 in MAIC-adjusted NMA. In these DGMs, the overlap was slightly low and this was true for both treatments. For sample size 500 and for both treatments, a slight reduction in coverage of MAIC-adjusted NMA was seen for DGM 11 and 12.

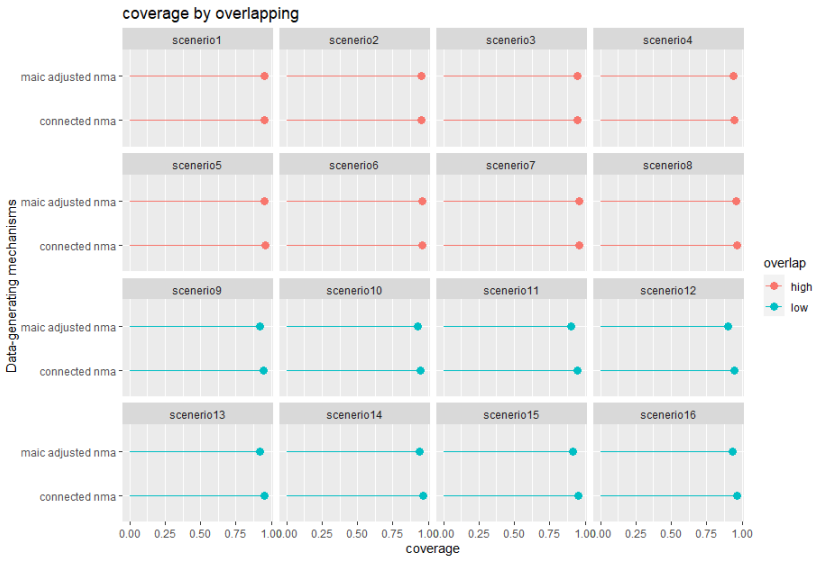


Figure 4.60: *Coverage by overlapping with treatment 2 for sample size 150*

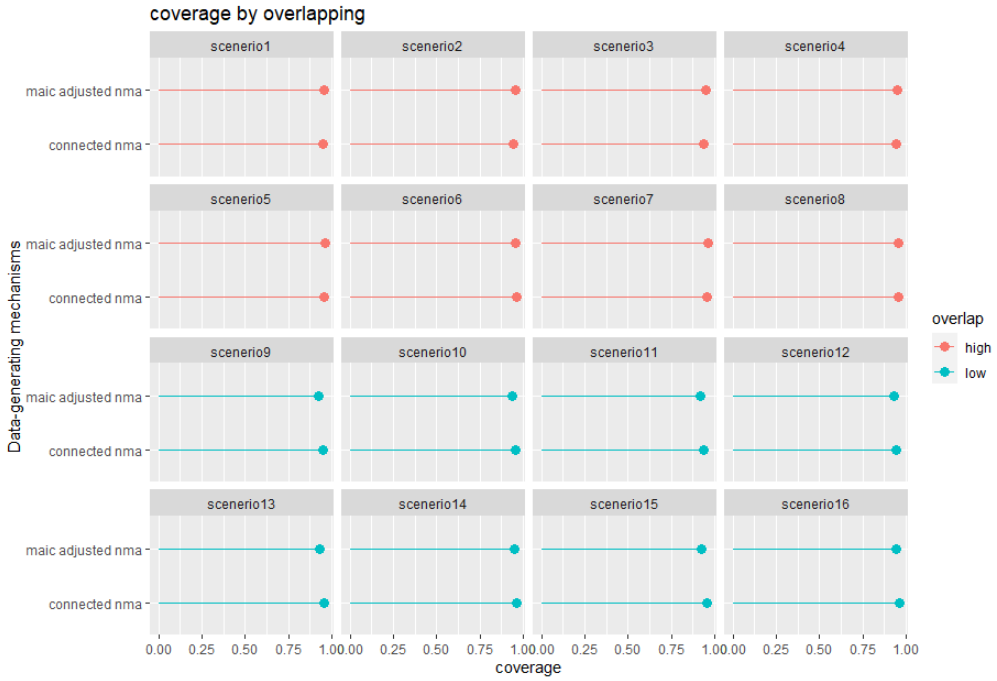


Figure 4.61: *Coverage by overlapping with treatment 3 for sample size 150*

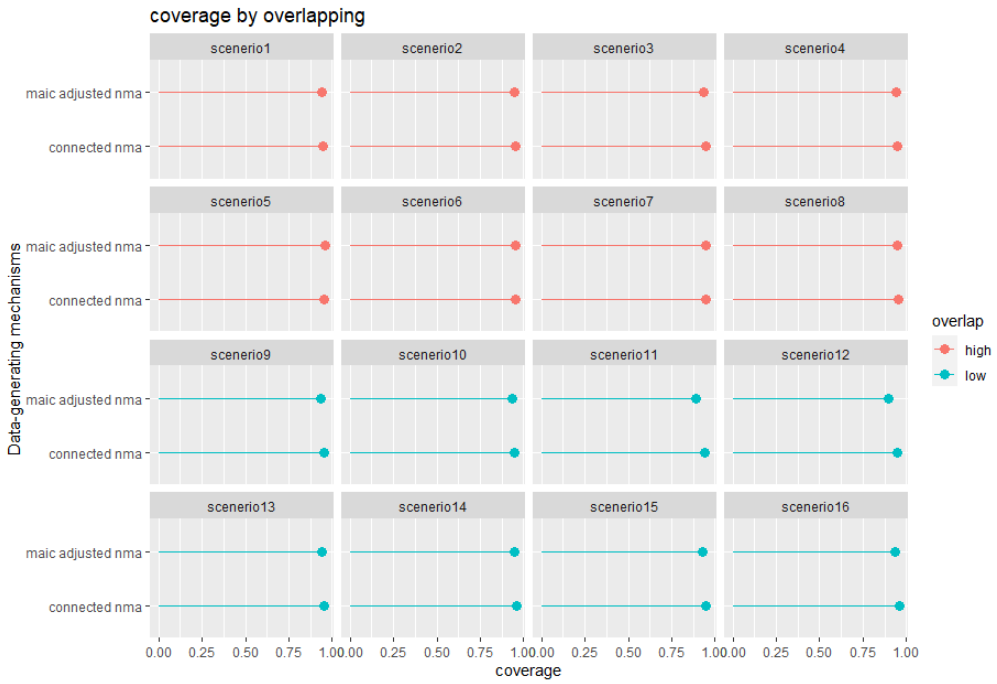


Figure 4.62: *Coverage by overlapping with treatment 2 for sample size 500*

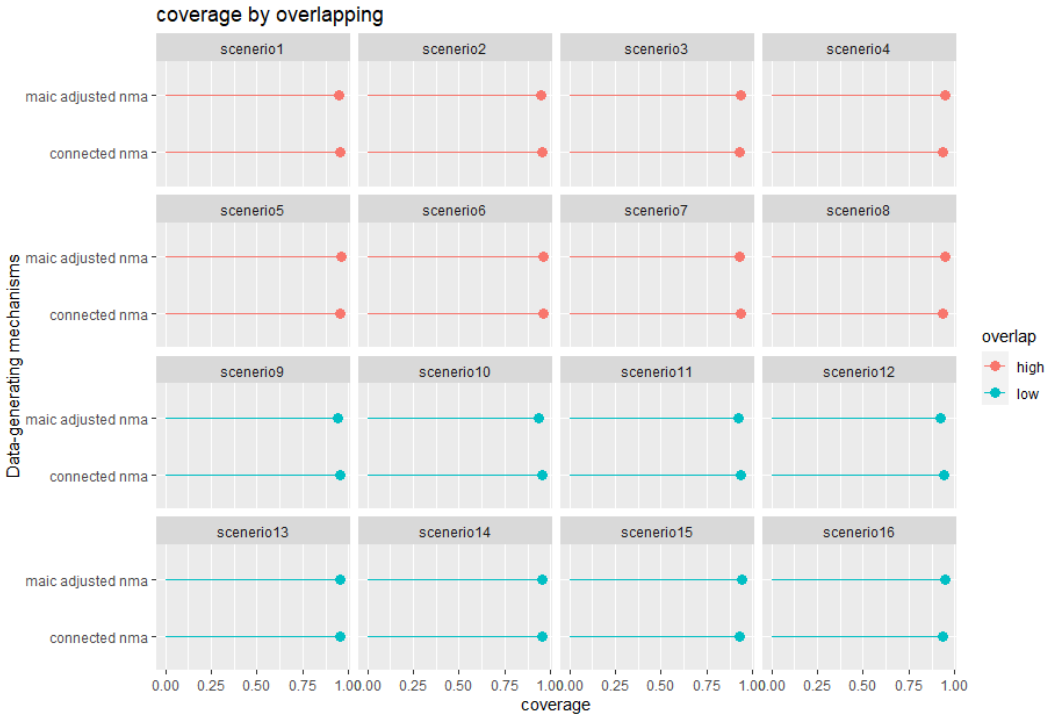
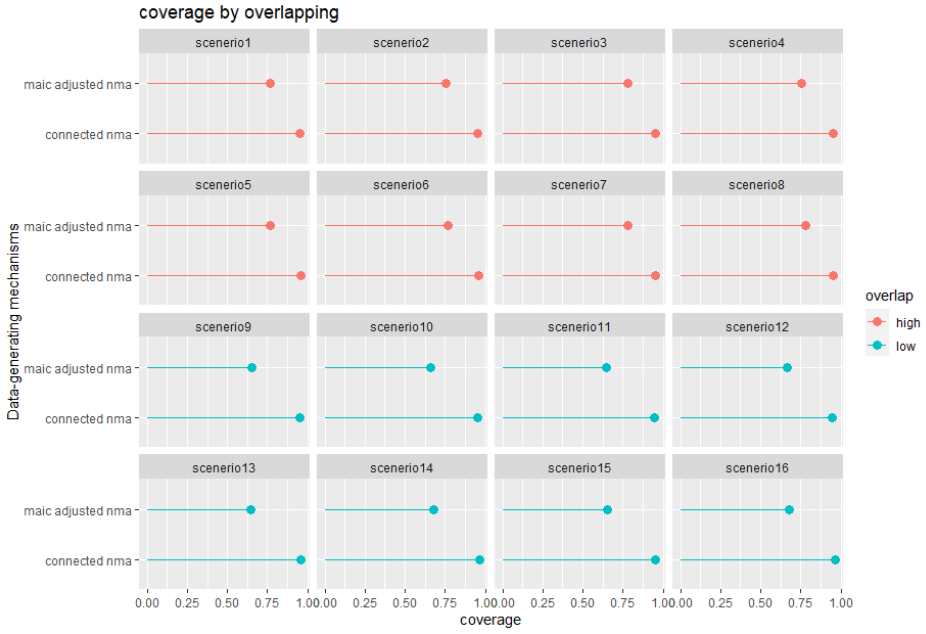


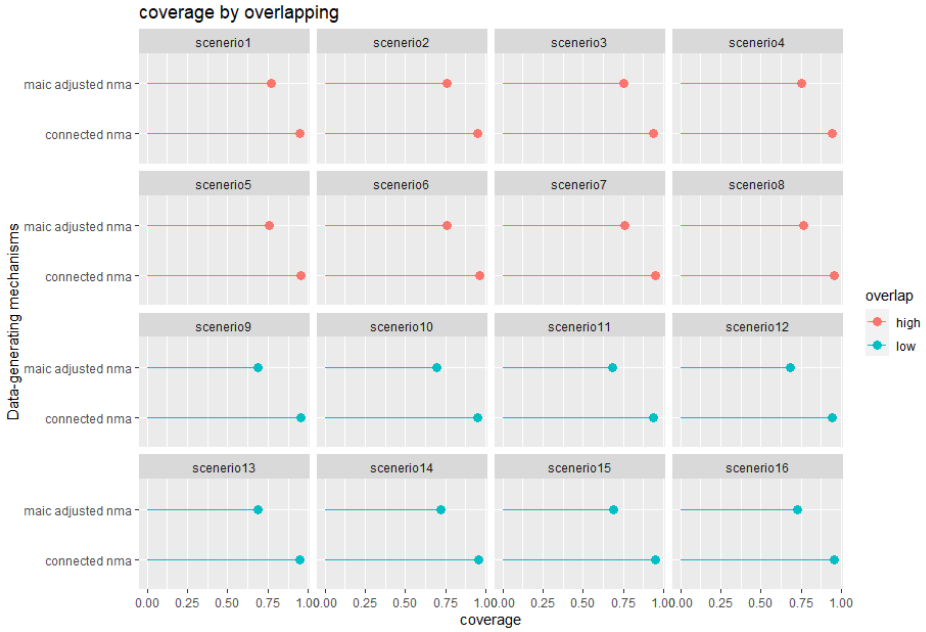
Figure 4.63: *Coverage by overlapping with treatment 3 for sample size 500*

4.7.6 Coverage by connected NMA and MAIC-adjusted NMA for the larger network of evidence

Figure 4.64 and Figure 4.65 show the coverage for NMA estimates from the connected network of evidence and estimates from MAIC-adjusted NMA for the larger network of evidence. Figure 4.64 and Figure 4.65 show the coverage for these two NMAs for sample sizes 150 and 500 respectively. For each DGM, the coverage for the connected NMA was at a nominal level whereas the coverage for the MAIC-adjusted NMA was below the nominal level. This was true irrespective of overlap and sample sizes.

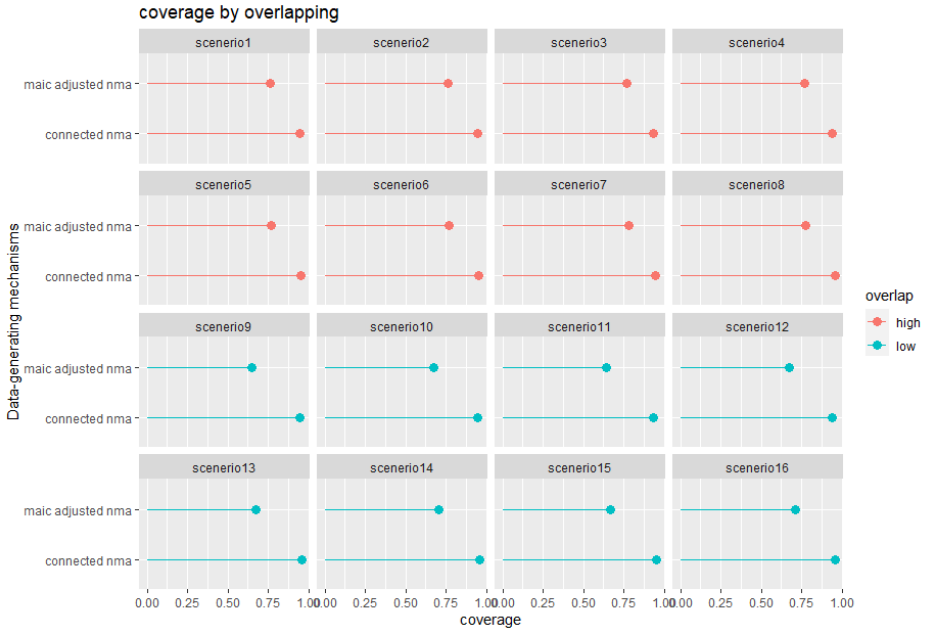


(a) *coverage by overlapping with treatment 2*

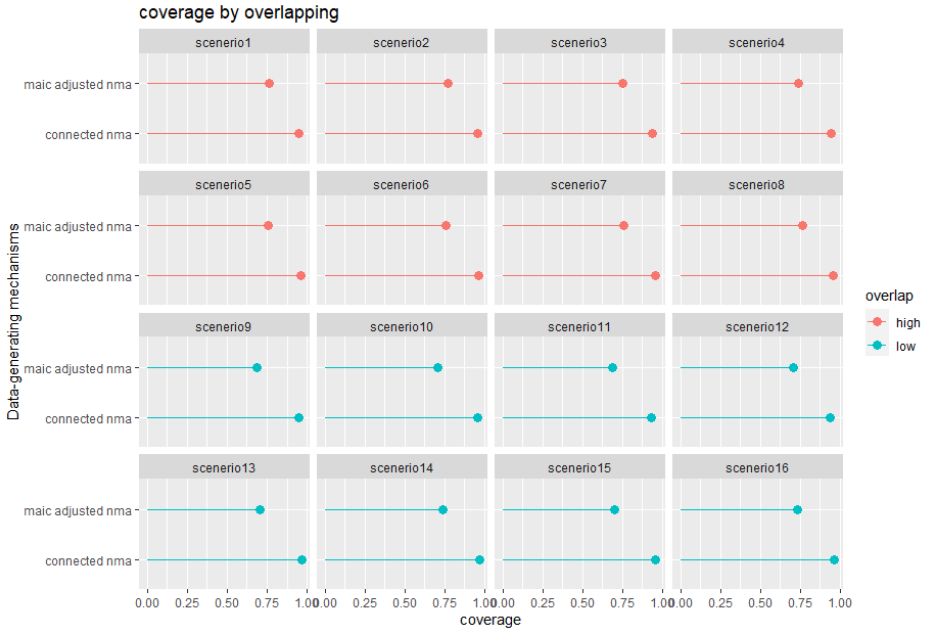


(b) *coverage by overlapping with treatment 3*

Figure 4.64: *Coverage by different NMA methods with sample size 150 (larger network of evidence)*



(a) *coverage by overlapping with treatment 2*



(b) *coverage by overlapping with treatment 3*

Figure 4.65: *Coverage by different NMA methods with sample size 500 (larger network of evidence)*

4.8 Discussion

In this simulation study, data were generated in four settings termed as connected smaller network, connected larger network, disconnected smaller network, and disconnected larger network. In connected networks, smaller and larger imply the number of studies simulated for each comparison. In the smaller network of evidence, 3 studies were generated whereas for the larger network of evidence, 10 studies were generated. Connected networks were formed by generating data with two arm RCTs. Fixed effect NMA were then estimated with both networks. Both the connected networks of evidence were then made into disconnected networks by dropping one arm from each study. Multiple unanchored MAICs were applied in the disconnected networks so that a connected network could be built to perform an NMA. Eventually, fixed effect NMA was applied with the MAIC estimates.

The first objective of conducting the connected NMAs was to demonstrate that the R code were correct and working well. The second objective was to assess where the results of the NMAs from disconnected networks deviate from the connected NMAs. In four data settings of the simulation study, the absolute size of biases appeared to be below one-half of the estimates SE, therefore, none of them can be termed as problematic following Schafer and Graham (2002). For both the smaller and larger connected networks of evidence, biases were found to be unrelated to overlap. Bigger biases can be seen with high overlap scenarios and vice-versa. However, the magnitude of biases was smaller for the larger network of evidence. The general trend in biases for the connected network of evidence was found to be higher biases associated with high effect-modifier levels for most of the cases. With prognostic variables, the trend was similar where most of the time, low biases were found with low prognostic variable levels. The coverage for both the smaller and larger connected network of evidence was at the nominal (95%) level as expected. The coverage seems to have no relationship with correlation as coverage was similar for different levels of correlation. Coverage was found to be unaffected with different levels of effect-modifier and prognostic variable levels for both connected networks of evidence.

For the smaller disconnected network of evidence, bias was found to be related to overlap. Lower biases were found with high overlap levels and vice-versa. This was expected as when MAIC is conducted to estimate relative treatment effect in a disconnected network, the estimate would be better with high overlap between studies and the bias will reduce. Additionally, bias was found to be high with high effect-modifier levels and low prognostic variable levels. However, when the MAIC-adjusted NMA was conducted for a smaller network of evidence with only 3 studies, the low coverage issue was not that severe in comparison to a larger disconnected network of evidence with 10 studies. In a smaller disconnected network of 3 studies, 1 MAIC-adjusted estimate was available per comparison whereas for a larger disconnected network of 10 studies, the number of MAIC-adjusted estimates was 5 and 4 for treatment 2 and 3 respectively. With the smaller disconnected network of evidence, in high overlap scenarios, the coverage was at the nominal (95%) level. Coverage starts to decrease slightly with low overlap scenarios. The coverage seems to be unaffected by correlation with high overlap level but with low overlap level, coverage slightly increases with high correlation. Similar to the smaller and larger connected network of evidence, coverage was found to be unaffected with different levels of effect-modifier and prognostic variable levels.

Analogous to the smaller disconnected network of evidence, in the larger disconnected network of evidence, biases were found to be related to overlap. Lower biases were found with high overlap levels and vice-versa. The general trend was found to be higher bias with high effect-modifier levels and higher prognostic variable levels. For the larger disconnected network of evidence, undercoverage was seen for all DGMs irrespective of overlap. When the overlap between study covariates was low, the reduction in coverage was higher than that of high overlap. Within the low overlap scenarios, coverage was found to be better with high correlation. The difference in coverage with the two levels of overlap did not change drastically: this is because though two levels of overlap were included, none of the overlaps were really low. Additionally, though coverage was affected mainly due to overlap between study covariates, prognostic and effect-modifying variables seem to have no effect on the coverage.

Prognostic and effect-modifying variables were found to have no effect on the undercoverage issue. However, they are related to the bias. The general trend was low bias with low effect-modifying variables and low prognostic variable level. Moreover, the magnitude of biases in the two MAIC-adjusted NMA was found to be lower for the larger disconnected network of evidence. Bias and overlap were found to be inversely related for both the smaller and larger disconnected networks of evidence. Throughout the simulation for the four networks of evidence, the highest biases were seen both for the connected and disconnected smaller network of evidence.

The major impact of performing an MAIC-adjusted NMA was seen in the coverage for each DGM. When the IPD from one study was used several times to conduct multiple MAICs, an unaccounted correlation emerged between studies. The independence between studies was violated which had a repercussion on the coverage. The deviation from the nominal (95%) level of coverage was more pronounced for the larger disconnected network of evidence. This was because during the estimation of disconnected NMA, the more times the IPD was used to get a relative effect estimate for an NMA, the more the assumption of between study independence was violated and the violation of this assumption becomes more visible through the reduction of coverage. Correlation between study covariates was found to be more crucial for the larger disconnected network of studies. The bigger in size of a disconnected network of evidence, the more coverage can be found with high correlation.

Low coverage/undercoverage issue for all DGM was found to be the major consequence of performing a fixed effect NMA with MAIC estimates for the larger disconnect network of evidence. The impact of MAIC estimates needs to be addressed for random effects NMA also as in a fixed effect NMA the true treatment effect is assumed to be fixed but shared by all the included studies. In reality, this is often not the case, and heterogeneity of treatment effect is expected. The next chapter will discuss the simulation of random effects NMA with MAIC estimates and its findings.

Chapter 5

Simulation Study with a MAIC-Adjusted Random Effects NMA

5.1 Introduction

In the previous chapter, a simulation study was performed in order to evaluate the consequences of conducting MAIC-adjusted NMA for a smaller and larger disconnected network of evidence for binary outcomes. This was done by simulating a smaller and a larger connected network of evidence based on the fixed effect of treatments. The network of evidence was then made disconnected by dropping arms from each study to turn each study into a single-arm study. Relative treatment estimates were then estimated by conducting multiple unanchored MAIC. Fixed effect NMA was then performed with the MAIC estimates. The focus of the current chapter is to explore the consequence of conducting MAIC-adjusted NMA for a disconnected network of evidence with a random effects model.

In a fixed effect NMA, the true treatment effect is assumed to be shared among the included studies, i.e. every study in the network of evidence aims to estimate the same parameter value. The studies in a fixed effect NMA are only subject to sampling error. On the contrary, in a random effects NMA, the true treatment effect is not bound to be fixed due to the heterogeneity among studies. Unlike the fixed effect NMA, the variance estimation in a random effects NMA comes from two sources: variance within each study and variance that comes from between studies. The assumption in a random effects NMA is that the true treatment effect can differ between studies and every study is evaluating a non-identical effect size. This non-identical effect size is exchangeable which means the true treatment effect comprises a distribution and the aim is to estimate the mean and dispersion parameters of this distribution. Usually, this distribution is considered to be a normal distribution. The fixed effect NMA model is considered to be a special case of a random effects NMA model where the variance parameter is considered to be zero. The assumption of a random effects NMA is more realistic as in practice, studies included in an NMA often differ on various aspects, therefore, fitting a random effects NMA is usually more pragmatic.

Section 5.2 of this chapter describes the aim of the simulation study whereas Section 5.3 gives details of how the simulation was designed to generate data to conduct a random effects NMA both for a smaller and larger network of evidence. Sections 5.4 to 5.6 describe what were the estimands of the study, and what methods and performance measures were evaluated respectively. Section 5.7 describes the findings from the simulation study. The chapter concludes with a discussion of the simulation results in Section 5.8.

5.2 Aims

Similar to the previous chapter, the goal of the simulation study in this chapter was to assess the appropriateness of MAIC estimates in a NMA setting, however, assuming random effects rather than fixed effects.

5.3 Data generating mechanism (DGM)

Similar to the previous chapter, data were generated in four settings termed as "connected smaller network of evidence", "connected larger network of evidence", "disconnected smaller network of evidence", and "disconnected larger network of evidence" with binary outcomes. The DGM for these networks of evidence in order to perform a random effects NMA was the same as that described in the previous chapter in Section 4.3. The main difference between fixed and random effects data generation is that, for the random effects model, a heterogeneity parameter tau (τ) was added during the data generation. The value of the τ parameter was fixed at 0.3 which indicates a moderate heterogeneity of treatment effect between studies was considered (Ren et al., 2018). The R code for data generation are given in D.1 of Appendix D.

5.4 Estimands

Analogous to the previous chapter, first a random effects NMA was performed after generating data from a connected network of evidence. The network of evidence was then disconnected by dropping one arm from each of the included studies. This makes the connected network of evidence into a disconnected network of evidence for single-arm studies as described in Figure 4.5 in the previous chapter. The relative treatment effect estimates were then estimated using multiple unanchored MAICs. The unanchored MAICs turn the disconnected network of evidence into a connected network of evidence again. A second random effects NMA was then performed using these MAIC-adjusted treatment effects. The estimands of interest were the overall treatment effect estimates from the connected NMA and the treatment effect estimates from the MAIC-adjusted NMA.

The simulation study was designed in such a way that it satisfies the shared effect modifier assumption as during the data generation process, all the treatments in a network of evidence had the same effect-modifier coefficient β_2.

5.5 Methods

Akin to the previous chapter, the following methods were applied to the data generated during the simulation exercise:

- A random effects NMA was applied to the simulated data of a connected network of evidence. The computation of the NMA was done using the Bayesian approach and using R package `multinma`. For the prior distributions of the treatment effects and study-specific intercepts, a normal distribution was used with $N(0, 100^2)$. An informative prior was used for the heterogeneity parameter τ which was Turner's prior as log-Normal (-2.56, 0.33). In a Bayesian analysis with a limited number of studies, the posterior distribution of between-study SD can be very broad or indefinite when a vague or weakly informative prior distribution is used. Turner et al. (2015) have developed a distribution on the log odds ratio scale for binary outcomes that can be used as prior distributions for heterogeneity parameter τ. Therefore, as Turner's prior can estimate more precise estimates for a log odds ratio, it was used as an informative prior here.

- Random effects MAIC-adjusted NMA: After generating data for a connected network of evidence (both smaller and larger), the next step was to transform this connected network of evidence into a disconnected network of evidence. To do this, the first study in the connected network was considered as the IPD study and the rest of the network as AgD studies. Therefore, after generating IPD for all the studies in the network, all the studies were converted to AgD except study 1. The connected network of evidence was transformed into single-arm studies by dropping one arm from each study. Treatment arm 1 was kept from the first study and for the rest of the studies, all arms were dropped except arms 2 and 3 (Figure 4.3 and Figure 4.5). Then multiple MAICs were conducted (depicted by an arrow line in Figure 4.3 and Figure 4.5) using the IPD from the first study. These MAIC estimates gave relative effects of intervention treatment (treatment 1) with treatments 2 and 3. Moreover, these multiple MAICs transform the disconnected network into a connected network to perform an NMA.

  MAIC uses weights calculated by the MoM/EB which gives more importance/weight to individuals in the IPD study who are more alike to the AgD study and less importance if they differ between studies. A logistic regression was used to assign the weights which in turn make the study individuals as similar as possible. As the MAICs was applied in unanchored form, all effect-modifiers and prognostic variables need to be included in the model. MAICs were applied upto the second moment i.e. balancing of covariate was done both for mean and standard deviation. After applying MAIC in a smaller and larger disconnected network of evidence, a MAIC-adjusted random effects NMA was performed and robust SE were estimated for the NMA estimates. The R package `MAIC` was used to perform the MAIC.

5.6 Performance Measures

In order to compare the performance of MAIC in a disconnected network of evidence, performance measures bias, model SE, empirical SE, and coverage probability were used.

- Bias: Statistical bias in simulation study gives an estimate of the systematic discrepancy between the true parameter and expected values of the results obtained from each simulated dataset. It can be defined as: $Bias = E[\hat{\theta}] - \theta$. The true parameter value was estimated by performing an NMA with a big sample size (1 million) and then using the estimates from this NMA as the true parameter value.
- Empirical SE: Empirical SE is the dispersion measure of the estimator in a simulation study. It represents the precision of an estimator as well as its true variability. An estimator is expected to be with low variance when it is applied to multiple datasets. It can be defined as: $EmpSE = \sqrt{Var(\hat{\theta})}$.
- Model SE: In a simulation study, when a method is applied to multiple datasets, the measure of the average of the SE reported by the method is known as model SE. It can be defined as: $ModelSE = \sqrt{E[\hat{se}(\hat{\theta})^2]}$. It is desired that the empirical SE is small which shows that the estimator is precise and the model SE is equal to empirical SE.
- Coverage probability: In a simulation study, coverage probability refers to the statistical technique where a percentage/proportion is calculated which shows how many confidence intervals include the true parameter value which is expected to be at (100×(1-α))% nominal level. It is common to fix the value of α at 0.05 i.e. at 95%.

The simulation study simulated 3000 repetitions/ datasets of each simulation scenario.

5.7 Results

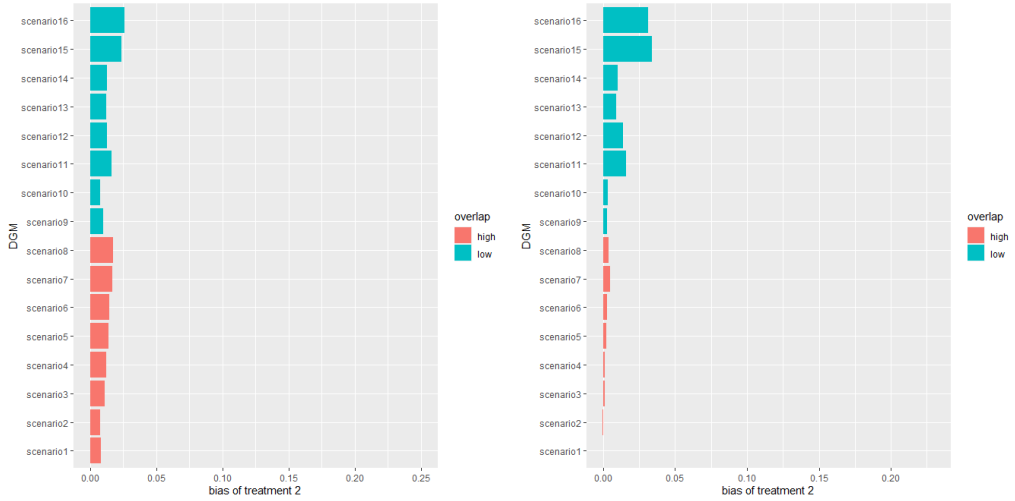
The results section describes the findings from both the connected and disconnected network of evidence. The results for both smaller and larger connected networks of evidence are described in subsections 5.7.1.1 to 5.7.2.8. Subsections 5.7.3.1 to 5.7.4.8 describe results for the disconnected network of evidence.

5.7.1 Results of simulation scenarios for the smaller connected network of evidence

Data were generated for the smaller connected network of evidence as described in Section 4.3.1 and in Figure 4.2 in the previous chapter where at first 3 RCTs were generated and then a random effects NMA was performed.

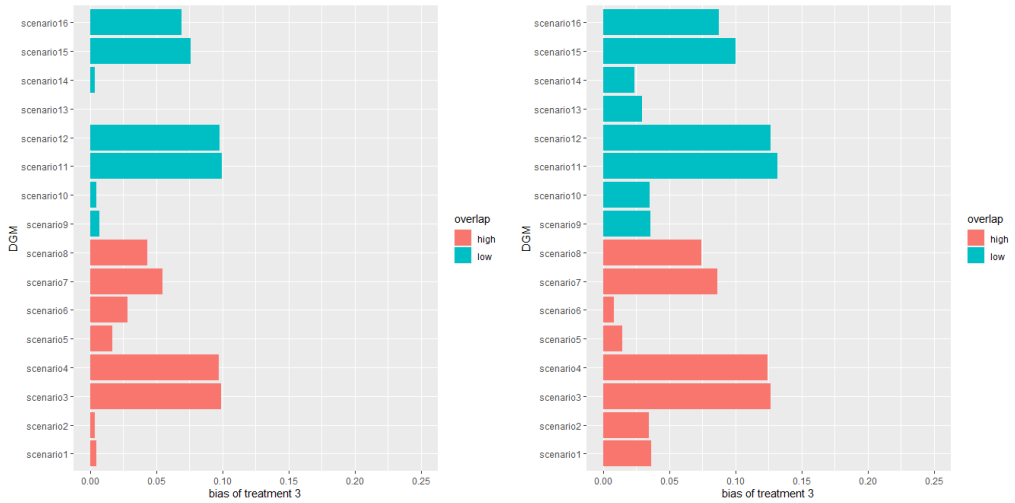
5.7.1.1 Bias by overlap

Figure 5.1 and 5.2 show the amount of bias for each DGM for treatment 2 and 3 respectively. The amount of bias was not found to be related to the amount of overlap. Overall, higher biases were found for scenarios 3, 4, 7, 8, 11, 12, 15, and in scenario 16 for treatment 3. One reason could be that the effect-modifier variable was high in these scenarios.



(a) *bias by overlap for sample size 150* **(b)** *bias by overlap for sample size 500*

Figure 5.1: *Bias of treatment 2 for different sample sizes (connected smaller network of evidence)*

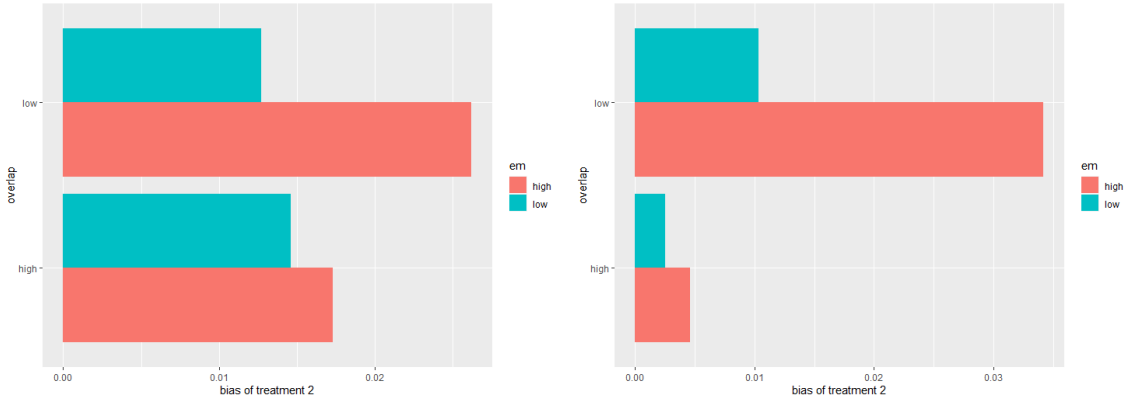


(a) *bias by overlap for sample size 150* **(b)** *bias by overlap for sample size 500*

Figure 5.2: *Bias of treatment 3 for different sample sizes (connected smaller network of evidence)*

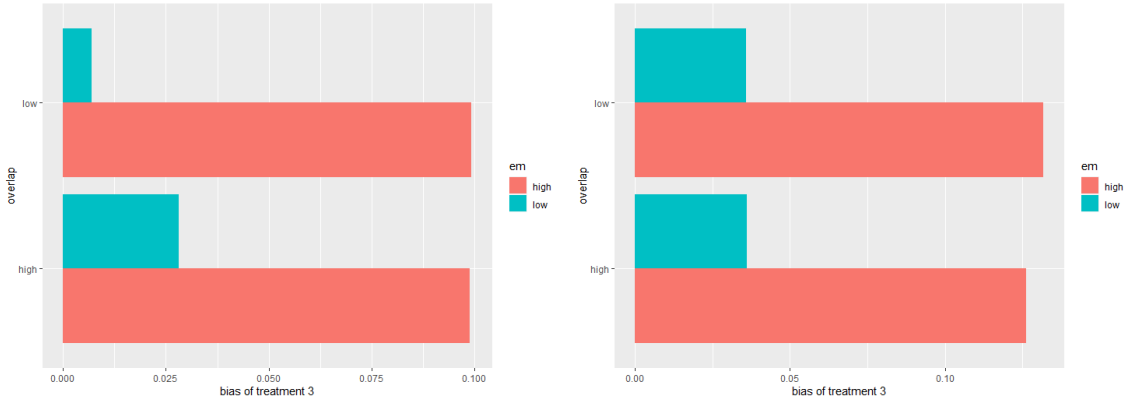
5.7.1.2 Bias by overlap with effect-modifiers

Figure 5.3 and 5.4 show the amount of bias by overlap for different levels of effect-modifier for treatments 2 and 3 respectively. Biases do not reduce with high overlap scenarios, but within each level of overlap, low biases were found with low effect-modifier levels.



(a) *bias by EM for sample size 150* **(b)** *bias by EM for sample size 500*

Figure 5.3: *Bias by overlap with EM levels for treatment 2 (connected smaller network of evidence)*

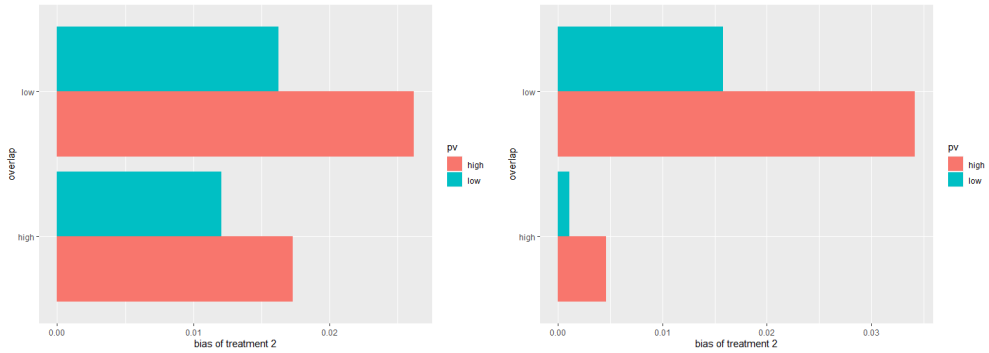


(a) *bias by EM for sample size 150* **(b)** *bias by EM for sample size 500*

Figure 5.4: *Bias by overlap with EM levels for treatment 3 (connected smaller network of evidence)*

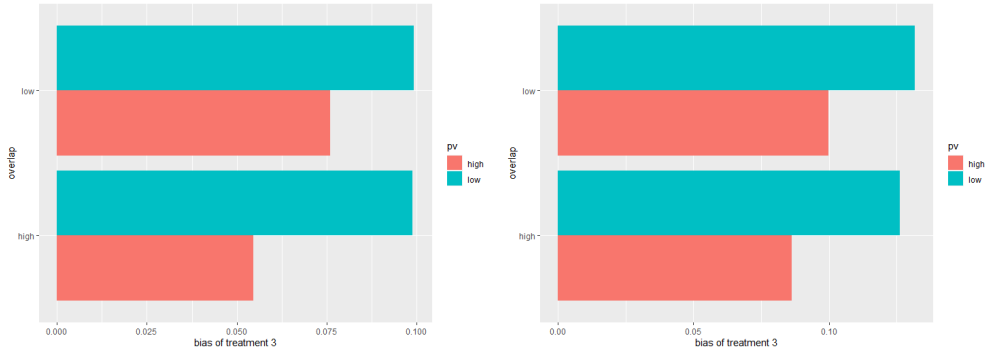
5.7.1.3 Bias by overlap with prognostic variable

Figure 5.5 and 5.6 show the amount of bias by overlap for different levels of prognostic variables for treatments 2 and 3 respectively. Biases do not reduce with high overlap but within the level of overlap, prognostic variables seem to be related to bias. In Figure 5.5 (a),(b), within each level of overlap, bias reduces with low prognostic variable level but the opposite can be seen for treatment 3 in Figure 5.6 (a),(b).



(a) *bias by PV for sample size 150* **(b)** *bias by PV for sample size 500*

Figure 5.5: *Bias by overlap with PV levels for treatment 2 (connected smaller network of evidence)*



(a) *bias by PV for sample size 150* **(b)** *bias by PV for sample size 500*

Figure 5.6: *Bias by overlap with PV levels for treatment 3 (connected smaller network of evidence)*

5.7.1.4 Empirical SE and model SE for the simulation scenarios

The simulation results of performance measures for the smaller connected network of evidence are summarised in Table 5.1 and Table 5.2 for sample sizes 150 and 500 respectively. Table 5.1 and Table 5.2 show the relative estimates of treatments 2 and 3 with treatment 1 from connected NMA for both sample sizes. The overall relative estimates of treatment 2 with the new intervention treatment 1 is termed as d1 and that of treatment 3 with treatment 1 is termed as d2. The empirical SE and model SE are similar or close to each other for each DGM. A quantity was estimated that calculates the difference between the empirical SE and model SE. The highest value of this difference was found to be 0.03 for the smaller connected network of evidence. Similar to the previous chapter, Tables are colour-coded to understand low-coverage and different magnitudes of biases. From the colour-coding it is evident that none of the coverage deviates from the nominal level. However, the biases were high for those scenarios where the effect-modifying variable was high.

Table 5.1: *Performance measure estimates of a smaller connected network of evidence for sample size 150 (random effects)*

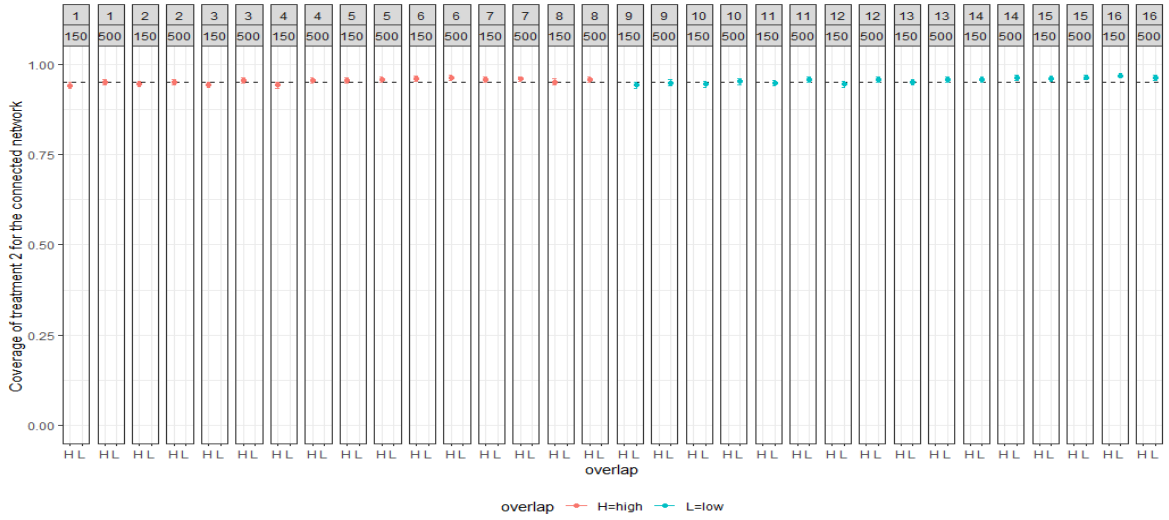
DGM	COVERAGE d1	BIAS d1	EMP SE d1	MODEL SE d1	DIFFERENCE d1	COVERAGE d2	BIAS d2	EMP SE d2	MODEL SE d2	DIFFERENCE d2
Scenario 1	0.941333	0.007951	0.312738	0.297306	0.015431	0.945	0.004784	0.399163	0.386282	0.01288
Scenario 2	0.945333	0.007441	0.308178	0.296448	0.01173	0.946333	0.003384	0.395782	0.386248	0.009534
Scenario 3	0.943	0.011223	0.320228	0.307662	0.012567	0.940667	0.0988	0.39084	0.384965	0.005875
Scenario 4	0.942667	0.012078	0.314083	0.306074	0.008009	0.946	0.097196	0.381498	0.384663	-0.00316
Scenario 5	0.955333	0.014197	0.339508	0.335143	0.004365	0.962	-0.01688	0.396616	0.404904	-0.00829
Scenario 6	0.960333	0.014625	0.32393	0.328001	-0.00407	0.968667	-0.02819	0.376562	0.401705	-0.02514
Scenario 7	0.958	0.016977	0.342934	0.346113	-0.00318	0.956	0.054651	0.38294	0.385981	-0.00304
Scenario 8	0.951333	0.01731	0.325112	0.336157	-0.01105	0.951333	0.043251	0.36251	0.362682	-0.00017
Scenario 9	0.942333	0.00985	0.303993	0.290753	0.01324	0.949333	0.006926	0.395337	0.384374	0.010963
Scenario 10	0.944333	0.007863	0.299791	0.290223	0.009568	0.951667	0.004878	0.392516	0.384266	0.00825
Scenario 11	0.947667	0.016254	0.306044	0.297105	0.008938	0.945	0.09929	0.384438	0.383512	0.000926
Scenario 12	0.945	0.01283	0.300849	0.295874	0.004975	0.948	0.097529	0.378428	0.383209	-0.00478
Scenario 13	0.949667	0.011926	0.315981	0.31171	0.004272	0.961667	0.000203	0.371985	0.384121	-0.01214
Scenario 14	0.957	0.012686	0.30205	0.307162	-0.00511	0.956667	0.003523	0.359153	0.352597	0.006557
Scenario 15	0.959667	0.023963	0.310433	0.316381	-0.00595	0.96	0.076045	0.35881	0.353643	0.005167
Scenario 16	0.968	0.026198	0.291165	0.310362	-0.0192	0.957667	0.069134	0.343606	0.341962	0.001643

Table 5.2: *Performance measure estimates of a smaller connected network of evidence for sample size 500 (random effects)*

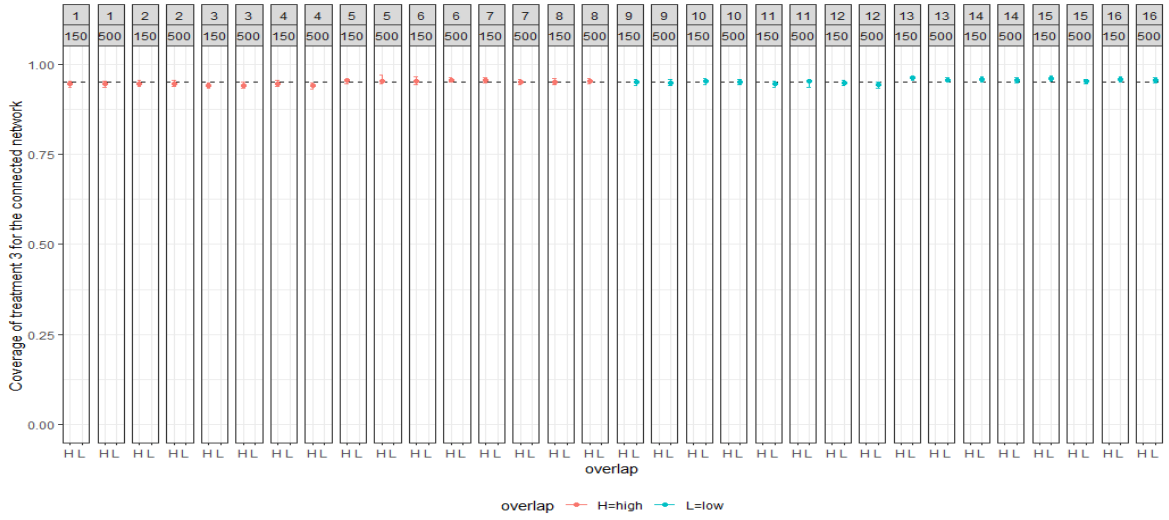
DGM	COVERAGE d1	BIAS d1	EMP SE d1	MODEL SE d1	DIFFERENCE d1	COVERAGE d2	BIAS d2	EMP SE d2	MODEL SE d2	DIFFERENCE d2
Scenario 1	0.951	-0.00013	0.236185	0.234268	0.001917	0.945	0.036187	0.319771	0.318559	0.001212
Scenario 2	0.949667	-0.00028	0.235321	0.233735	0.001586	0.946667	0.034755	0.316268	0.31857	-0.0023
Scenario 3	0.954333	-0.0011	0.237088	0.238192	-0.0011	0.941	0.126316	0.309048	0.318314	-0.00927
Scenario 4	0.955333	-0.00107	0.234147	0.237488	-0.00334	0.94	0.124426	0.302929	0.31801	-0.01508
Scenario 5	0.958	0.002195	0.242878	0.248239	-0.00536	0.962	0.014524	0.302236	0.314911	-0.01267
Scenario 6	0.962667	0.002551	0.232565	0.245299	-0.01273	0.966333	0.008047	0.288394	0.293555	-0.00516
Scenario 7	0.959	0.004615	0.241829	0.25264	-0.01081	0.951	0.086308	0.290458	0.305506	-0.01505
Scenario 8	0.957333	0.004012	0.229409	0.248475	-0.01907	0.952667	0.074179	0.272916	0.30409	-0.03117
Scenario 9	0.949	0.002753	0.234235	0.232082	0.002154	0.948667	0.035876	0.318898	0.318065	0.000833
Scenario 10	0.951667	0.003267	0.232298	0.231693	0.000604	0.949667	0.035105	0.317369	0.317959	-0.00059
Scenario 11	0.958	0.015795	0.230306	0.236471	-0.00617	0.934	0.131808	0.309768	0.320389	-0.01062
Scenario 12	0.957333	0.013754	0.227779	0.235742	-0.00796	0.942	0.126819	0.304623	0.32014	-0.01552
Scenario 13	0.958333	0.009074	0.232202	0.240367	-0.00817	0.966333	0.029551	0.29636	0.317842	-0.02148
Scenario 14	0.963	0.010337	0.223421	0.23828	-0.01486	0.956	0.023492	0.285381	0.287043	-0.00166
Scenario 15	0.963667	0.034194	0.227084	0.243954	-0.01687	0.962	0.099874	0.282488	0.29918	-0.01669
Scenario 16	0.963	0.031561	0.214734	0.240831	-0.0261	0.956	0.08754	0.268147	0.278403	-0.01026

5.7.1.5 Coverage by the simulation scenarios

The coverage of d1 and d2 from the connected NMA is at the nominal (95%) level for both sample sizes. Figure 5.7 (a),(b) shows the confidence intervals of the coverage estimates for varying levels of sample sizes for treatments 2 and 3 respectively. Red-coloured figures represent confidence intervals for the high overlap scenarios and blue-coloured figures represent confidence intervals for the low overlap scenarios. In the figures, the black horizontal line represents the nominal level of coverage. From the figures, it can be seen that the confidence intervals touch the nominal level for all 16 scenarios irrespective of sample sizes and overlap.



(a) coverage by overlap for treatment 2

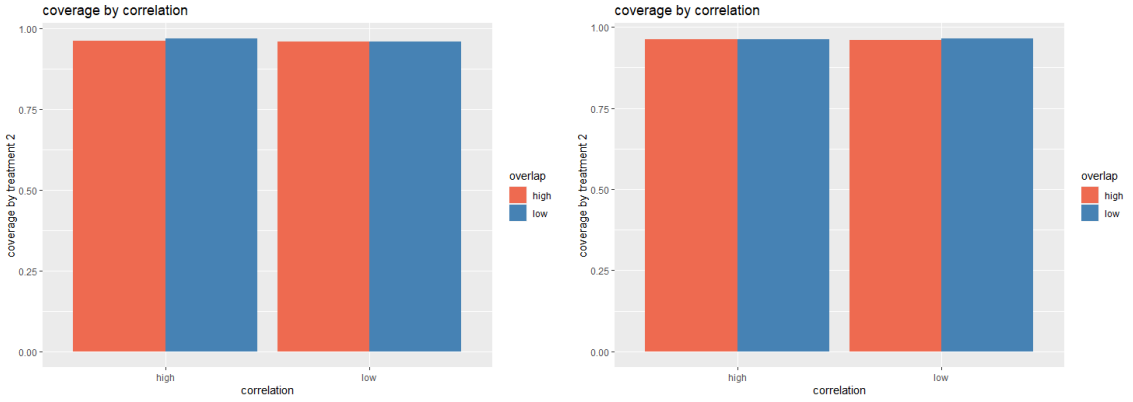


(b) coverage by overlap for treatment 3

Figure 5.7: *Coverage of treatments 2 and 3 for different sample sizes and overlap*

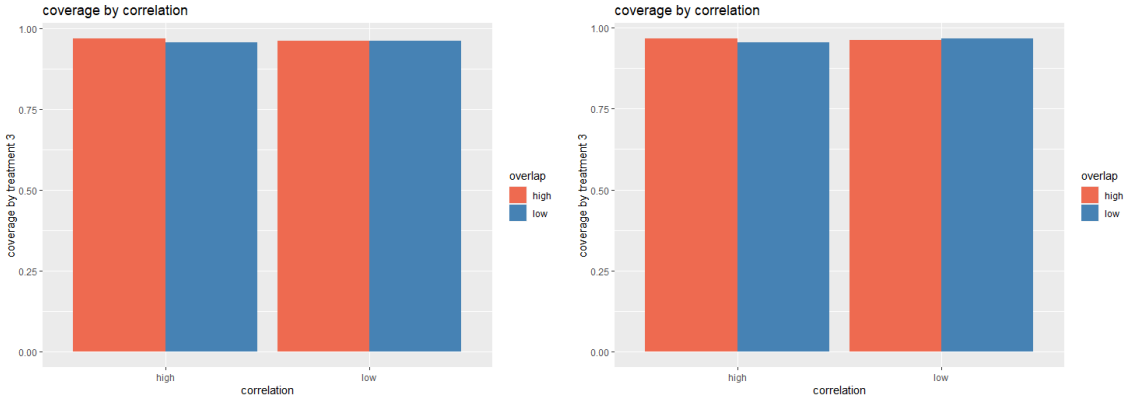
5.7.1.6 Coverage by correlation with overlap

Figure 5.8 and Figure 5.9 show the coverage with different levels of overlap and correlation for treatments 2 and 3 respectively. From the figures, it can be seen that the coverage was not found to be affected by different levels of correlation and overlap.



(a) *coverage by correlation for sample size 150* **(b)** *coverage by correlation for sample size 500*

Figure 5.8: *Coverage of treatment 2 for different sample sizes (connected smaller network of evidence)*

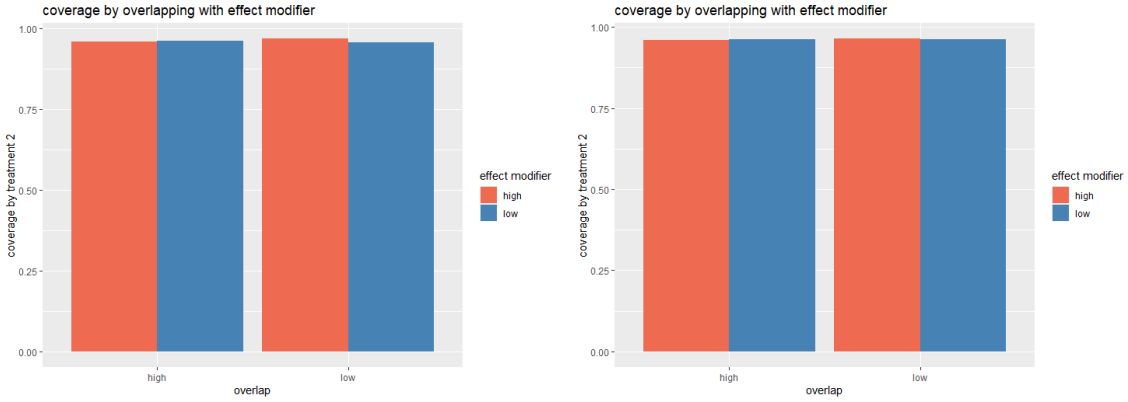


(a) *coverage by correlation for sample size 150* **(b)** *coverage by correlation for sample size 500*

Figure 5.9: *Coverage of treatment 3 for different sample sizes (connected smaller network of evidence)*

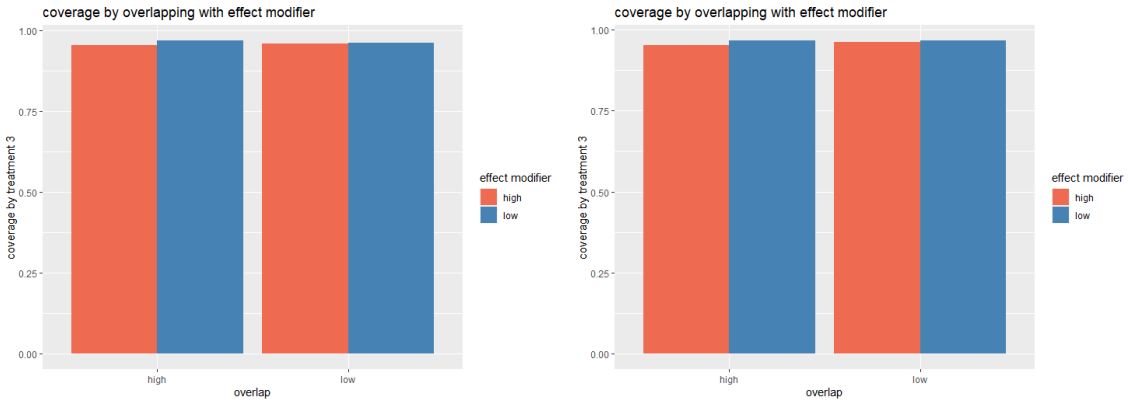
5.7.1.7 Coverage by overlap with effect-modifiers

Figure 5.10 and 5.11 show the coverage by different levels of overlap and effect-modifier for treatments 2 and 3 respectively. For both sample sizes and treatments, the coverage was similar with varying levels of overlap and effect-modifier variable.



(a) *coverage by EM sample size 150* **(b)** *coverage by EM sample size 500*

Figure 5.10: *Coverage by overlap with EM levels for treatment 2 (connected smaller network of evidence)*

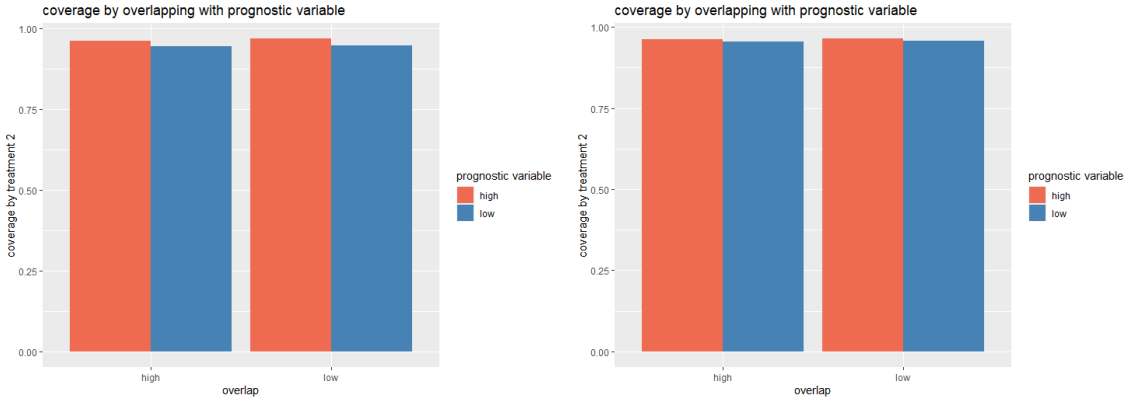


(a) *coverage by EM sample size 150* **(b)** *coverage by EM sample size 500*

Figure 5.11: *Coverage by overlap with EM levels for treatment 3 (connected smaller network of evidence)*

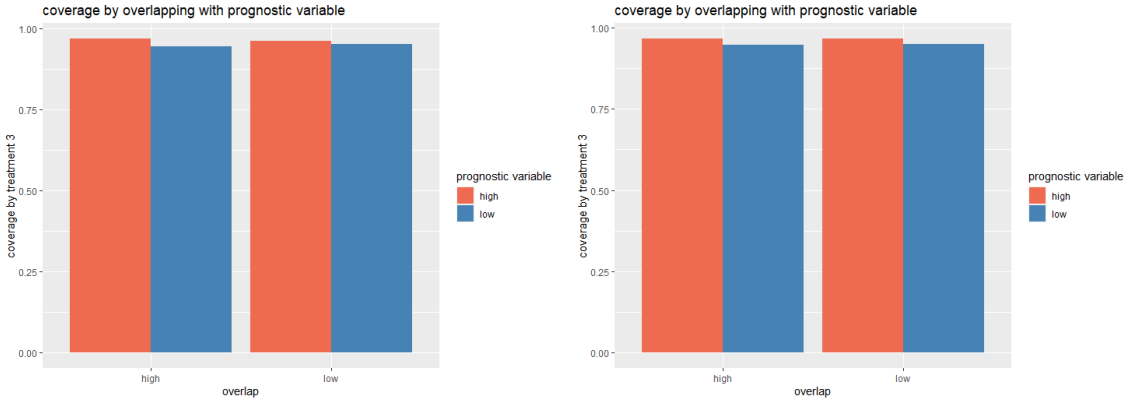
5.7.1.8 Coverage by overlap with prognostic variable

Figure 5.12 and 5.13 show the amount of coverage by different levels of overlap and prognostic variables for treatment 2 and 3 respectively. For both sample sizes and treatments, the coverage was found to be similar with varying levels of overlap and prognostic variables.



(a) *coverage by PV sample size 150* **(b)** *coverage by PV sample size 500*

Figure 5.12: *Coverage by overlap with PV levels for treatment 2 (connected smaller network of evidence)*



(a) *coverage by PV sample size 150* **(b)** *coverage by PV sample size 500*

Figure 5.13: *Coverage by overlap with PV levels for treatment 3 (connected smaller network of evidence)*

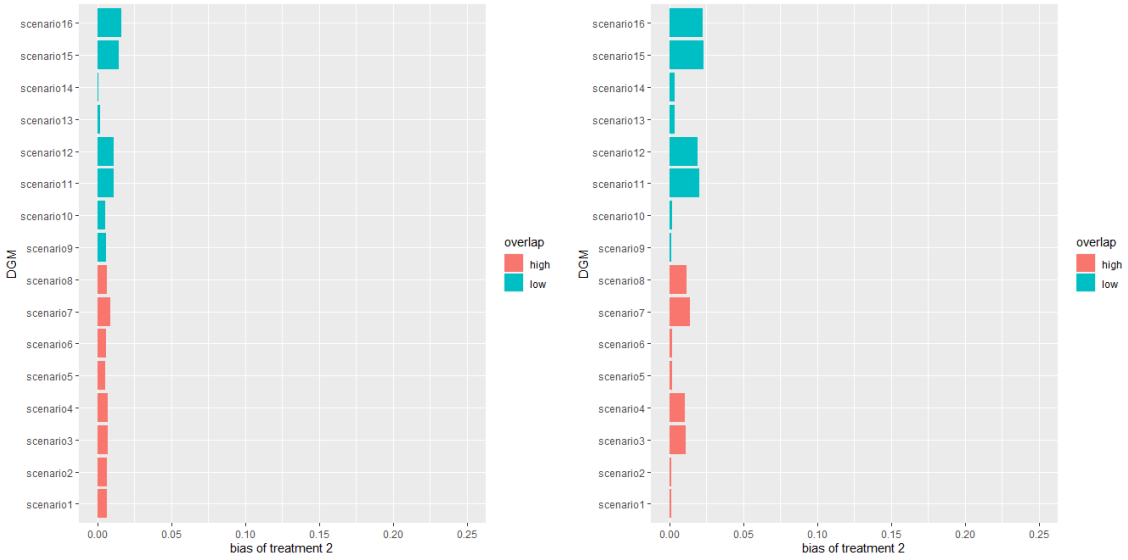
5.7.2 Results of simulation scenarios for the larger connected network of evidence

Data were generated for the larger connected network of evidence as described in Section 4.3.2 and in Figure 4.4 in the previous chapter where at first 10 RCTs were generated and then a random effects NMA was performed.

5.7.2.1 Bias by overlap

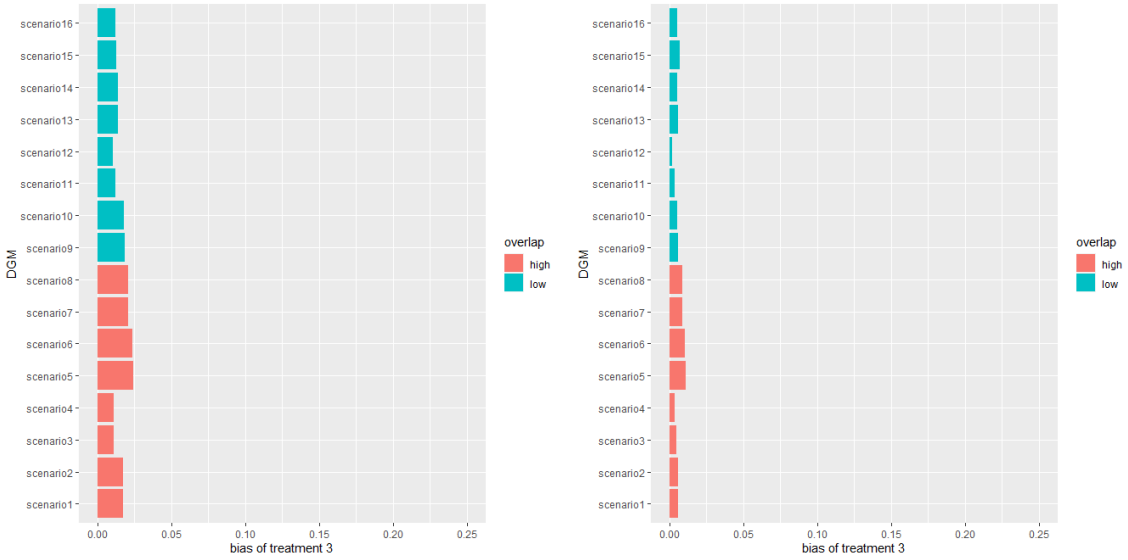
Figure 5.14 and 5.15 show the amount of bias for each DGM for treatment 2 and 3 respectively. Bias and overlap were not found to be related. In Figure 5.14 (a),(b), large biases can be seen for low overlap but the opposite can be seen in Figure 5.15. Although the findings are similar to the findings from the smaller connected network of evidence, the magnitude of

biases for the larger connected network of evidence is small compared to the smaller network of evidence.



(a) *bias by overlap for sample size 150* **(b)** *bias by overlap for sample size 500*

Figure 5.14: *Bias of treatment 2 for different sample sizes (connected larger network of evidence)*

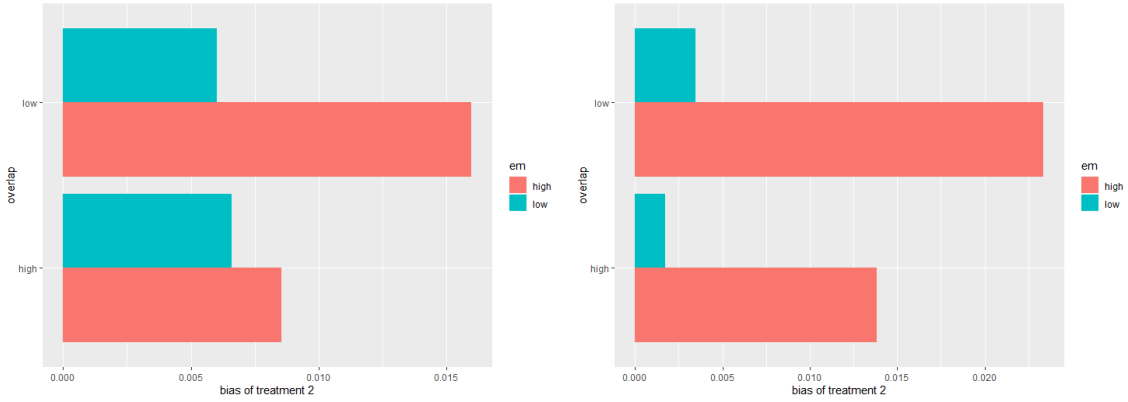


(a) *bias by overlap for sample size 150* **(b)** *bias by overlap for sample size 500*

Figure 5.15: *Bias of treatment 3 for different sample sizes (connected larger network of evidence)*

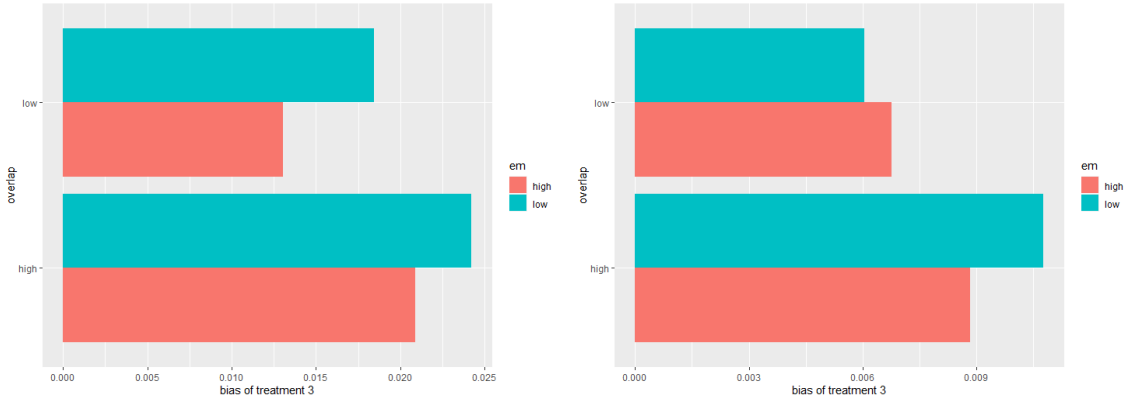
5.7.2.2 Bias by overlap with effect-modifiers

Figure 5.16 and 5.17 show the amount of bias by overlap for different levels of effect-modifier for treatments 2 and 3 respectively. Though biases were found to be unrelated with overlap, overall, lower biases were found with low effect-modifier levels.



(a) *bias by EM for sample size 150* **(b)** *bias by EM for sample size 500*

Figure 5.16: *Bias by overlap with EM levels for treatment 2 (connected larger network of evidence)*

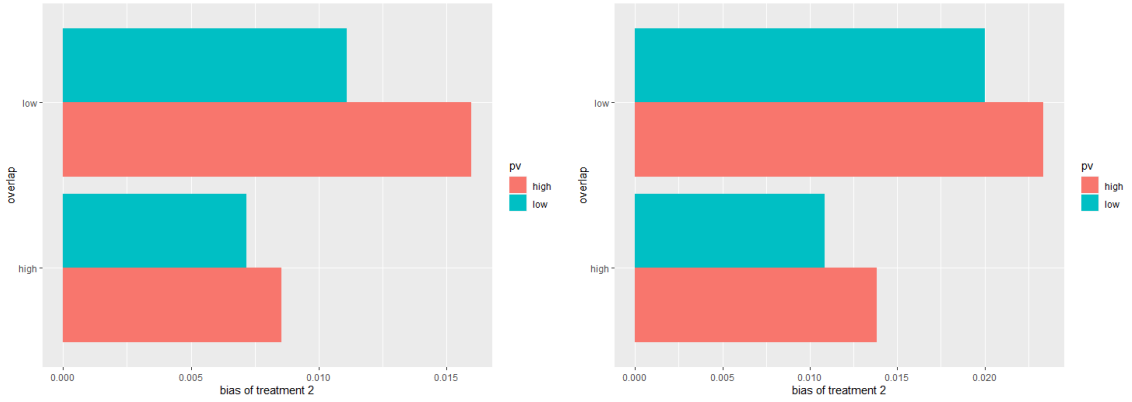


(a) *bias by EM for sample size 150* **(b)** *bias by EM for sample size 500*

Figure 5.17: *Bias by overlap with EM levels for treatment 3 (connected larger network of evidence)*

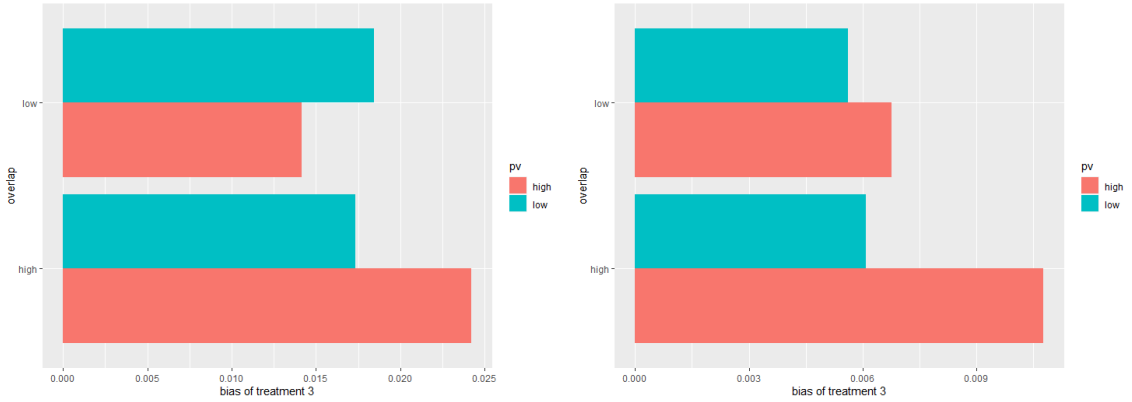
5.7.2.3 Bias by overlap with prognostic variable

Figure 5.18 and 5.19 show the amount of bias by overlap for different levels of prognostic variables for treatments 2 and 3 respectively. Overall, biases were found to be higher with high prognostic variable levels.



(a) *bias by PV for sample size 150* **(b)** *bias by PV for sample size 500*

Figure 5.18: *Bias by overlap with PV levels for treatment 2 (connected larger network of evidence)*



(a) *bias by PV for sample size 150* **(b)** *bias by PV for sample size 500*

Figure 5.19: *Bias by overlap with PV levels for treatment 3 (connected larger network of evidence)*

5.7.2.4 Empirical SE and model SE for the simulation scenarios

The simulation results of performance measures for the larger connected network of evidence are summarised in Table 5.3 and Table 5.4 for sample sizes 150 and 500 respectively. Table 5.3 and Table 5.4 show the relative estimates of treatments 2 and 3 with treatment 1 from connected NMA for both sample sizes. The empirical SE and model SE are similar or close to each other for each DGM. The highest value of the difference between the empirical SE and model SE was found to be 0.02 for the larger connected network of evidence. From the colour-coding, it can be seen that all were blue which means none of the scenarios shows undercoverage and the amount of biases was below 0.03.

Table 5.3: *Performance measure estimates of a larger connected network of evidence for sample size 150 (random effects)*

DGM	COVERAGE d1	BIAS d1	EMP SE d1	MODEL SE d1	DIFFERENCE d1	COVERAGE d2	BIAS d2	EMP SE d2	MODEL SE d2	DIFFERENCE d2
Scenario 1	0.944333	-0.00659	0.173408	0.170051	0.003357	0.943333	-0.01732	0.196758	0.192852	0.003906
Scenario 2	0.942333	-0.00658	0.17175	0.169526	0.002224	0.942	-0.01726	0.194364	0.192708	0.001656
Scenario 3	0.947667	0.007175	0.177104	0.174716	0.002388	0.946333	-0.01126	0.19228	0.191797	0.000483
Scenario 4	0.949	0.007016	0.174968	0.173649	0.001318	0.954	-0.01089	0.186323	0.19149	-0.00517
Scenario 5	0.946	-0.0051	0.189968	0.188188	0.001779	0.953	-0.02422	0.193171	0.201206	-0.00804
Scenario 6	0.95	-0.00596	0.182337	0.184263	-0.00193	0.966333	-0.0235	0.183931	0.199104	-0.01517
Scenario 7	0.95	0.008544	0.192544	0.192677	-0.00013	0.964	-0.02057	0.18859	0.191464	-0.00287
Scenario 8	0.955667	0.006402	0.181152	0.187321	-0.00617	0.969667	-0.02088	0.177861	0.199269	-0.02141
Scenario 9	0.944667	-0.00601	0.167136	0.164174	0.002962	0.939333	-0.01845	0.19578	0.191729	0.004052
Scenario 10	0.945	-0.00523	0.165585	0.163837	0.001748	0.944	-0.01787	0.194566	0.191493	0.003073
Scenario 11	0.953333	0.010833	0.165026	0.165739	-0.00071	0.947333	-0.01244	0.190709	0.191329	-0.00062
Scenario 12	0.952333	0.011102	0.162864	0.165097	-0.00223	0.955333	-0.01059	0.186832	0.190899	-0.00407
Scenario 13	0.955333	-0.00188	0.16352	0.168896	-0.00538	0.958667	-0.01391	0.181956	0.190724	-0.00877
Scenario 14	0.96	0.000553	0.1571	0.166778	-0.00968	0.954333	-0.01416	0.175735	0.179456	-0.00372
Scenario 15	0.958667	0.014432	0.160119	0.169432	-0.00931	0.956	-0.01303	0.176216	0.170222	0.005995
Scenario 16	0.965333	0.01597	0.151775	0.1667	-0.01493	0.953	-0.01242	0.166446	0.168826	-0.00238

Table 5.4: *Performance measure estimates of a larger connected network of evidence for sample size 500 (random effects)*

DGM	COVERAGE d1	BIAS d1	EMP SE d1	MODEL SE d1	DIFFERENCE d1	COVERAGE d2	BIAS d2	EMP SE d2	MODEL SE d2	DIFFERENCE d2
Scenario 1	0.941333	-0.00134	0.138512	0.135344	0.003169	00.945667	-0.00609	0.159789	0.159967	-0.00018
Scenario 2	0.945	-0.00156	0.13708	0.135004	0.002077	0.948	-0.0057	0.158549	0.159682	-0.00113
Scenario 3	0.942667	0.010851	0.138535	0.136925	0.00161	0.951333	-0.00496	0.155763	0.159271	-0.00351
Scenario 4	0.943	0.010633	0.136895	0.136216	0.000679	0.958333	-0.0034	0.152982	0.15879	-0.00581
Scenario 5	0.947667	0.001716	0.141753	0.141255	0.000498	0.96	-0.01077	0.153123	0.161543	-0.00842
Scenario 6	0.952	0.001648	0.135874	0.139138	-0.00326	0.967667	-0.01056	0.14549	0.15007	-0.00458
Scenario 7	0.95	0.013833	0.14069	0.142719	-0.00203	0.967333	-0.00883	0.146619	0.161195	-0.01458
Scenario 8	0.955	0.011516	0.132725	0.139676	-0.00695	0.974667	-0.00868	0.137604	0.159386	-0.02178
Scenario 9	0.938667	0.001256	0.135646	0.133312	0.002334	0.944667	-0.00562	0.158358	0.15975	-0.00139
Scenario 10	0.943667	0.001738	0.134317	0.133019	0.001298	0.947333	-0.00527	0.157405	0.159413	-0.00201
Scenario 11	0.945333	0.020015	0.13329	0.134782	-0.00149	0.953333	-0.00371	0.153318	0.160571	-0.00725
Scenario 12	0.955333	0.018984	0.13119	0.134164	-0.00297	0.955333	-0.00189	0.15084	0.160075	-0.00923
Scenario 13	0.951	0.003359	0.130692	0.133863	-0.00317	0.962667	-0.00605	0.14681	0.157646	-0.01084
Scenario 14	0.958	0.003473	0.125422	0.132283	-0.00686	0.969667	-0.00553	0.140658	0.156276	-0.01562
Scenario 15	0.956667	0.023372	0.126042	0.134251	-0.00821	0.970333	-0.00676	0.140276	0.157469	-0.01719
Scenario 16	0.964333	0.022524	0.119374	0.112106	0.007268	0.973	-0.00517	0.152552	0.155686	-0.00313

5.7.2.5 Coverage by the simulation scenarios

Figure 5.20 and 5.21 illustrates the coverage of the relative treatment effects d1 and d2 from the connected random effects NMA for varying levels of sample size and overlap. Figure 5.20 and 5.21 show the coverage for treatments 2 and 3 respectively. From the figure, it can be seen that the confidence interval of coverage for all the DGM was at the nominal level for both sample sizes.

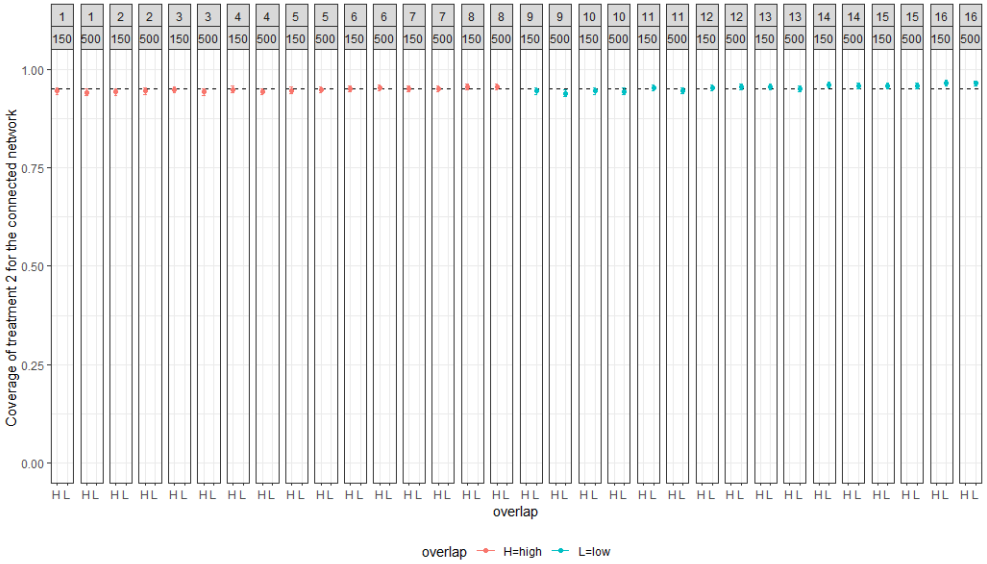


Figure 5.20: *Coverage by overlap for treatment 2*

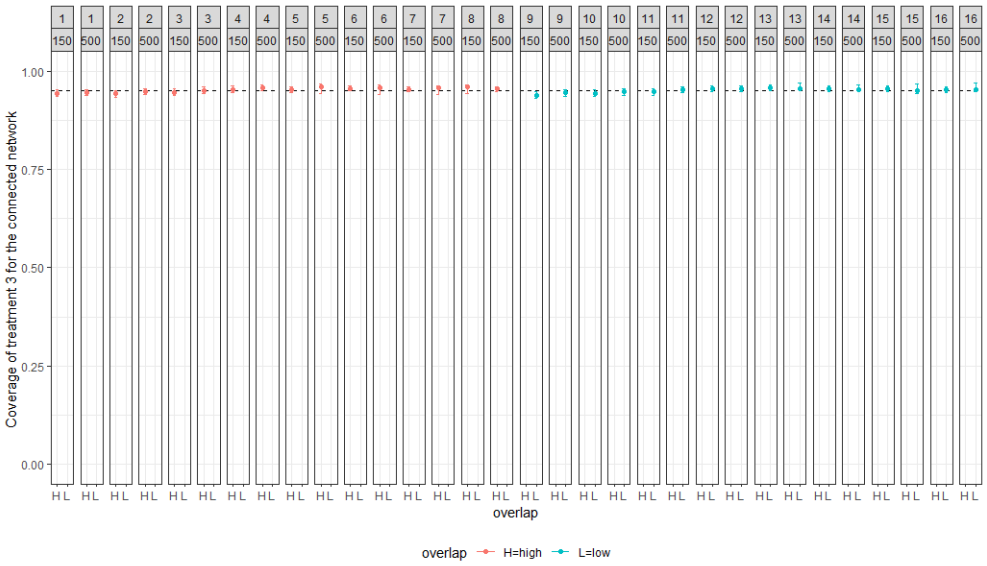
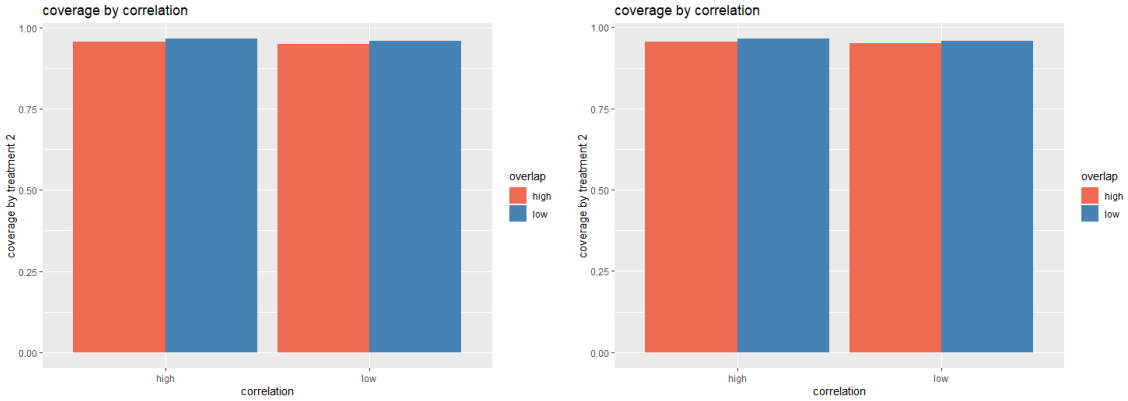


Figure 5.21: *Coverage by overlap for treatment 3*

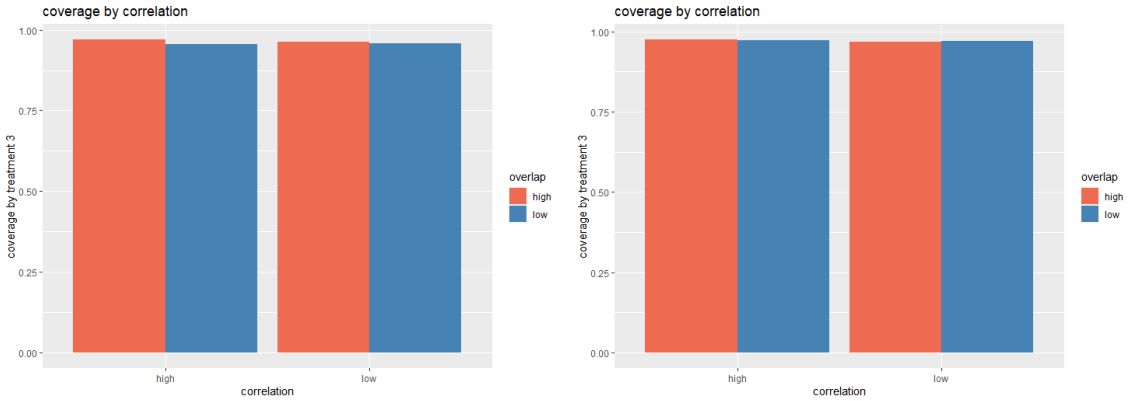
5.7.2.6 Coverage by correlation with overlap

Figure 5.22 and Figure 5.23 show the coverage with different levels of overlap and correlation for treatments 2 and 3 respectively. From the figures, it can be seen that the coverage is found to be unaffected by different levels of correlation and overlap.



(a) *coverage by correlation for sample size 150* **(b)** *coverage by correlation for sample size 500*

Figure 5.22: *Coverage of treatment 2 for different sample sizes (connected larger network of evidence)*

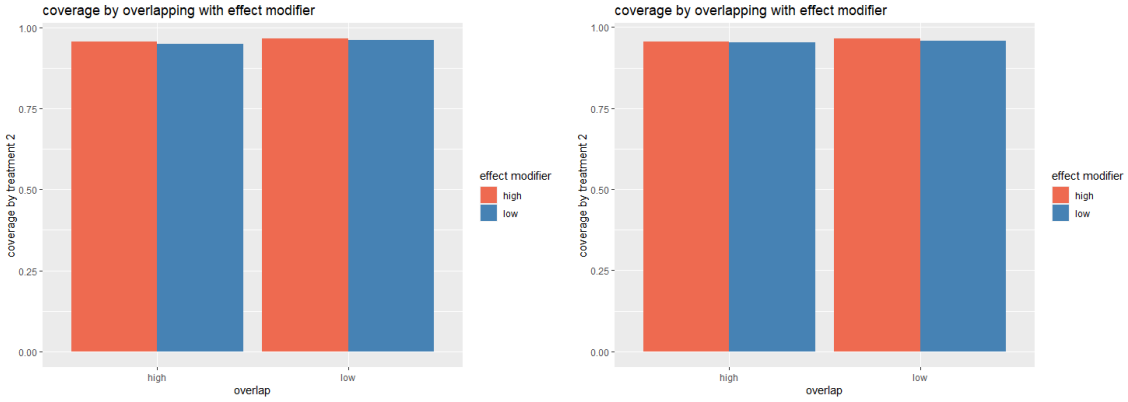


(a) *coverage by correlation for sample size 150* **(b)** *coverage by correlation for sample size 500*

Figure 5.23: *Coverage of treatment 3 for different sample sizes (connected larger network of evidence)*

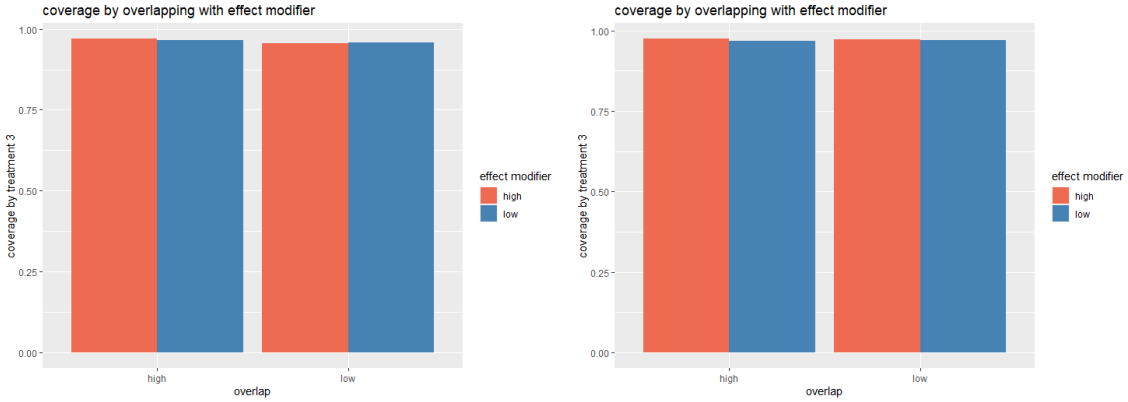
5.7.2.7 Coverage by overlap with effect-modifiers

Figure 5.24 and 5.25 show the coverage by different levels of overlap and effect-modifier for treatments 2 and 3 respectively. For both sample sizes, the coverage was found to be similar with varying levels of overlap and effect-modifier variables.



(a) *coverage by EM sample size 150* **(b)** *coverage by EM sample size 500*

Figure 5.24: *Coverage by overlap with EM levels for treatment 2 (connected larger network of evidence)*

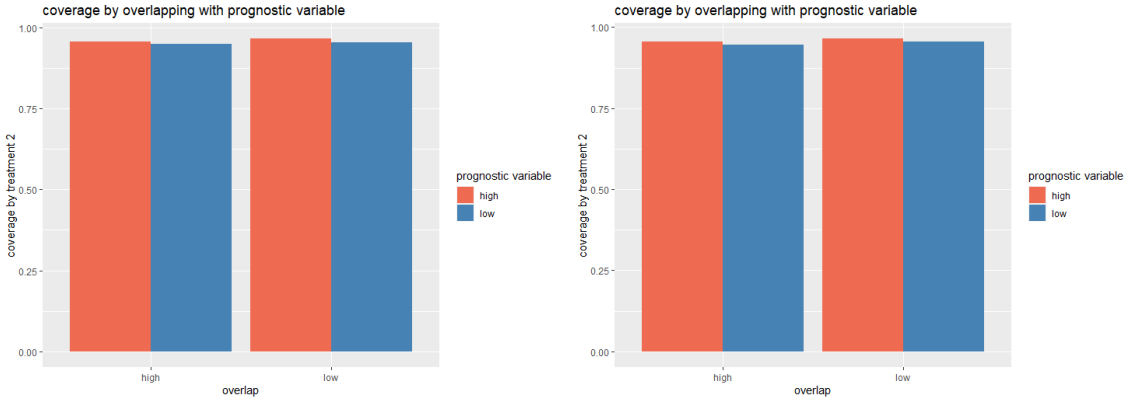


(a) *coverage by EM sample size 150* **(b)** *coverage by EM sample size 500*

Figure 5.25: *Coverage by overlap with EM levels for treatment 3 (connected larger network of evidence)*

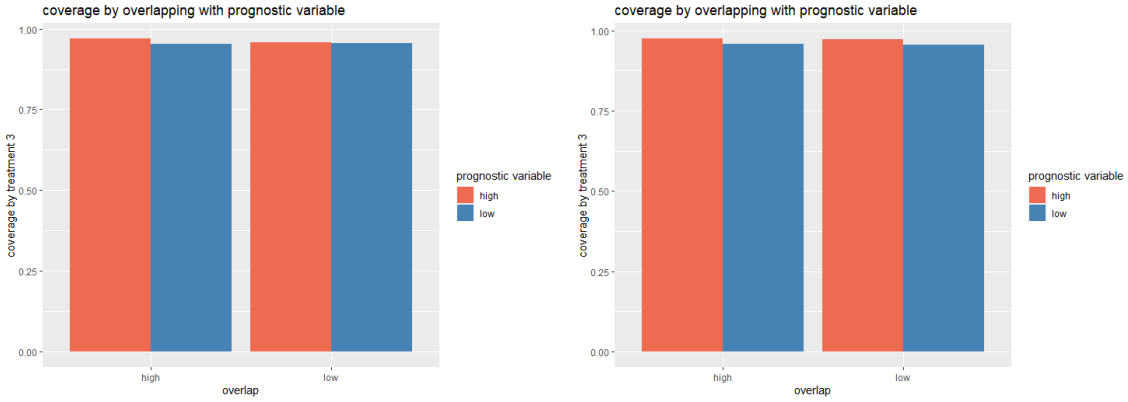
5.7.2.8 Coverage by overlap with prognostic variable

Figure 5.26 and 5.27 show the amount of coverage by different levels of overlap and prognostic variables for treatment 2 and 3 respectively. For both sample sizes, the coverage was similar with varying levels of overlap and prognostic variables.



(a) *coverage by PV sample size 150* **(b)** *coverage by PV sample size 500*

Figure 5.26: *Coverage by overlap with PV levels for treatment 2 (connected larger network of evidence)*



(a) *coverage by PV sample size 150* **(b)** *coverage by PV sample size 500*

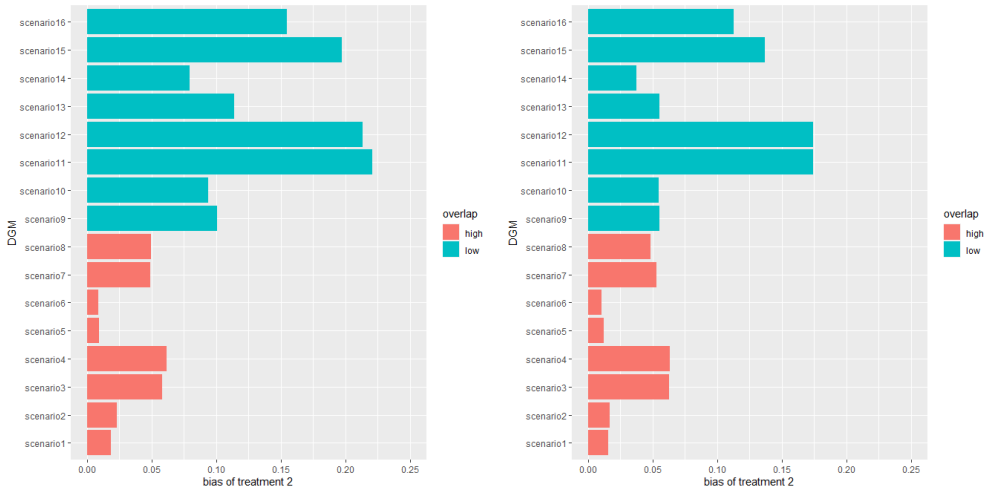
Figure 5.27: *Coverage by overlap with PV levels for treatment 3 (connected larger network of evidence)*

5.7.3 Results of simulation scenarios for the smaller disconnected network of evidence

Data were generated as described in the previous chapter in Section 4.3.1 and Figure 4.3 where first 3 RCTs were generated and then a random effects NMA was performed. These RCTs were then made disconnected artificially to implement MAIC. The MAIC estimates are then used to perform a NMA. Data were generated for 16 scenarios described in Table 4.1 from the previous chapter. The following sections describe the simulation results from the MAIC-adjusted random effects NMA.

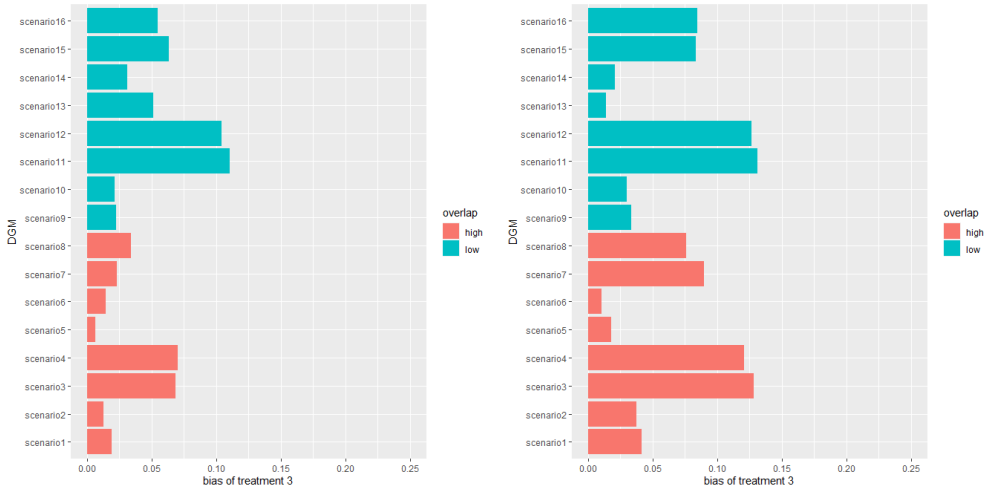
5.7.3.1 Bias by overlap

Figure 5.28 and 5.29 show the amount of bias for each DGM for treatment 2 and 3 respectively. The amount of bias decreases with the amount of overlap. The biases were low for high-overlap scenarios and high for low-overlap scenarios and it was found true for both sample sizes. Overall, for treatments 2 and 3, for both sample sizes, with high overlap, the amount of biases were high for scenarios 3, 4, 7, 8, and that was high for scenarios 11, 12, 15, 16 with low overlap. The effect-modifier level was high in these scenarios.



(a) *bias by overlap for sample size 150* **(b)** *bias by overlap for sample size 500*

Figure 5.28: *Bias of treatment 2 for different sample sizes (disconnected smaller network of evidence)*

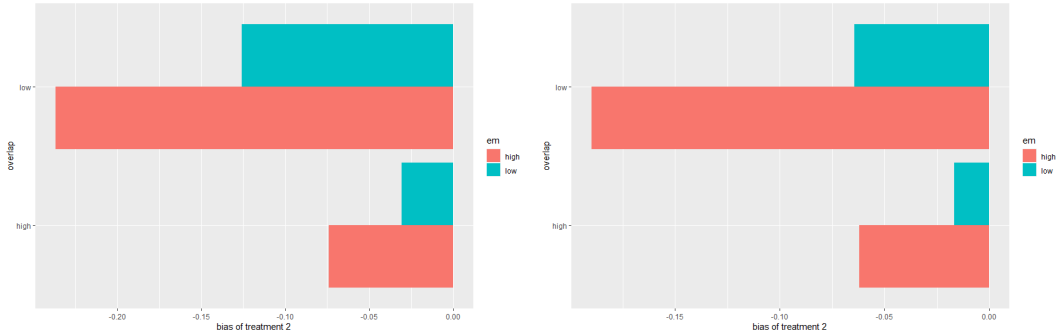


(a) *bias by overlap for sample size 150* **(b)** *bias by overlap for sample size 500*

Figure 5.29: *Bias of treatment 3 for different sample sizes (disconnected smaller network of evidence)*

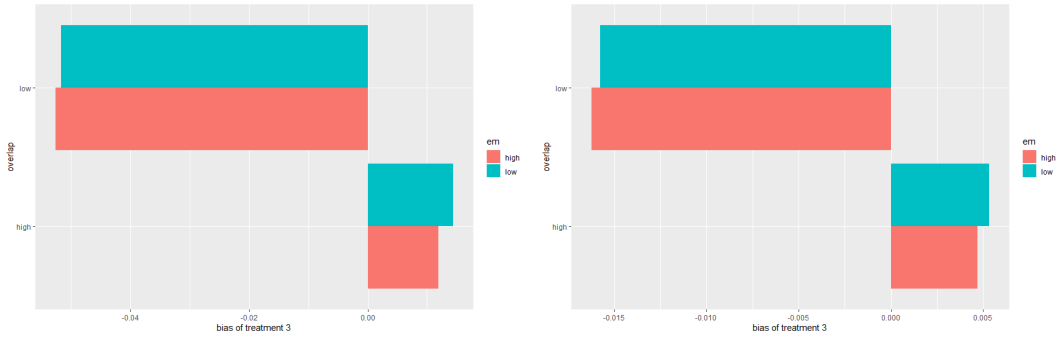
5.7.3.2 Bias by overlap with effect-modifiers

Figure 5.30 and 5.31 show the amount of bias by overlap for different levels of effect-modifier for treatments 2 and 3 respectively. In Figure 5.30 (a),(b), the bias reduces with high overlap level, and the reduction seems to depend on effect-modifier levels also. For treatment 2, the bias is low with the low level of effect-modifier. In Figure 5.31 (a),(b), for treatment 3, the bias reduces with high overlap, and within levels of overlap, biases were high with high effect-modifier level.



(a) *bias by EM for sample size 150* **(b)** *bias by EM for sample size 500*

Figure 5.30: *Bias by overlap with EM levels for treatment 2 (disconnected smaller network of evidence)*

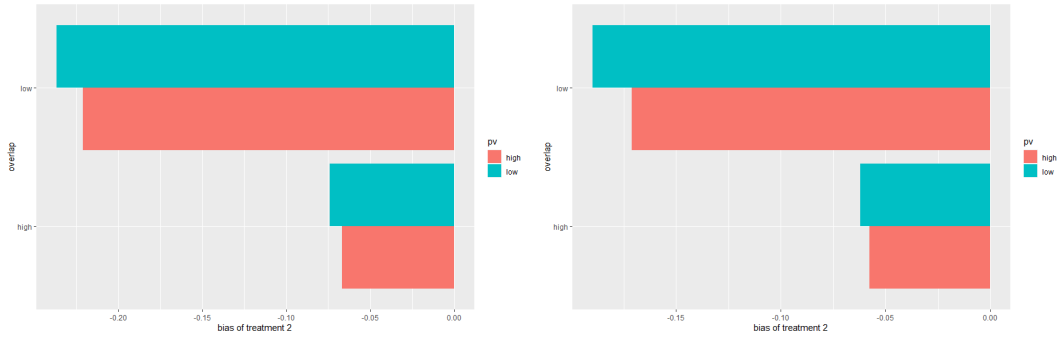


(a) *bias by EM for sample size 150* **(b)** *bias by EM for sample size 500*

Figure 5.31: *Bias by overlap with EM levels for treatment 3 (disconnected smaller network of evidence)*

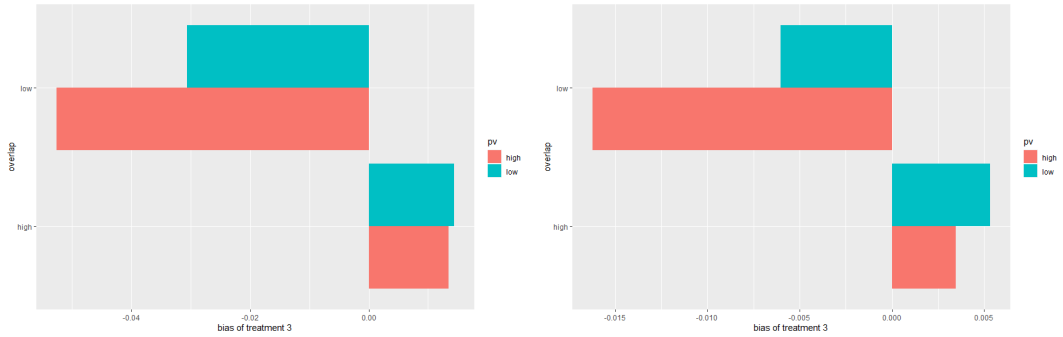
5.7.3.3 Bias by overlap with prognostic variable

Figure 5.32 and 5.33 show the amount of bias by overlap for different levels of prognostic variables for treatment 2 and 3 respectively. In Figure 5.32 (a),(b), the bias reduces with high overlap but within the levels of overlap, biases increase with the low level of the prognostic variable. In Figure 5.33 (a),(b) for treatment 3, for low levels of overlap, a bigger bias was found for higher levels of the prognostic variable, and the opposite was seen for high overlap level.



(a) *bias by PV for sample size 150* **(b)** *bias by PV for sample size 500*

Figure 5.32: *Bias by overlap with PV levels for treatment 2 (disconnected smaller network of evidence)*



(a) *bias by PV for sample size 150* **(b)** *bias by PV for sample size 500*

Figure 5.33: *Bias by overlap with PV levels for treatment 3 (disconnected smaller network of evidence)*

5.7.3.4 Empirical SE and model SE for the simulation scenarios

The simulation results of performance measures for the smaller disconnected network of evidence are summarised in Table 5.5 and Table 5.6 for sample sizes 150 and 500 respectively. Table 5.5 and Table 5.5 show the relative estimates of treatments 2 and 3 with treatment 1 from disconnected NMA for both sample sizes. The empirical SE and model SE are similar or close to each other for each DGM. The highest value of the difference between the empirical SE and model SE was found to be 0.05 for the smaller disconnected network of evidence. From colour-coding it can be seen that for high-overlap scenarios, coverage was at the nominal level, however, for low-overlap scenarios coverage dropped slightly for some scenarios. Overall, higher biases were seen for some scenarios where the effect-modifier level was high.

Table 5.5: *Performance measure estimates of a smaller disconnected network of evidence for sample size 150 (random effects)*

DGM	COVERAGE d1	BIAS d1	EMP SE d1	MODEL SE d1	DIFFERENCE d1	COVERAGE d2	BIAS d2	EMP SE d2	MODEL SE d2	DIFFERENCE d2
Scenario 1	0.939	-0.01824	0.45755	0.434619	0.022931	0.950667	0.019191	0.411141	0.403928	0.007213
Scenario 2	0.942	-0.02312	0.472778	0.450075	0.022703	0.946	0.01253	0.428948	0.422086	0.006861
Scenario 3	0.938667	-0.05844	0.467204	0.446218	0.020986	0.944	0.06868	0.403236	0.402183	0.001053
Scenario 4	0.938333	-0.06188	0.480128	0.460611	0.019517	0.945	0.070128	0.418169	0.420685	-0.00252
Scenario 5	0.949333	-0.00903	0.507259	0.492208	0.015051	0.958333	-0.0067	0.418388	0.435024	-0.01664
Scenario 6	0.955333	-0.00887	0.507327	0.500977	0.00635	0.960333	-0.01469	0.434867	0.452958	-0.01809
Scenario 7	0.951667	-0.0487	0.511446	0.50346	0.007986	0.964	0.023377	0.407426	0.43607	-0.02864
Scenario 8	0.953667	-0.04962	0.504539	0.508196	-0.00366	0.966	0.033977	0.422952	0.453997	-0.03104
Scenario 9	0.922	-0.1008	0.619539	0.583244	0.036295	0.933667	-0.02255	0.599245	0.576794	0.022452
Scenario 10	0.924667	-0.09391	0.598617	0.562299	0.036318	0.936667	-0.02116	0.578618	0.55636	0.022258
Scenario 11	0.907	-0.22065	0.617969	0.58431	0.033659	0.938667	0.110695	0.592457	0.575888	0.016569
Scenario 12	0.905333	-0.21339	0.593826	0.562969	0.030857	0.938333	0.104276	0.570603	0.554865	0.015738
Scenario 13	0.928333	-0.1139	0.678743	0.628803	0.04994	0.935667	-0.05154	0.633692	0.617731	0.015961
Scenario 14	0.937333	-0.07935	0.615603	0.600642	0.014961	0.945666	-0.03103	0.602755	0.59153	0.011226
Scenario 15	0.921333	-0.19712	0.672718	0.628158	0.04456	0.942333	0.063503	0.627207	0.617329	0.009878
Scenario 16	0.938	-0.15464	0.599771	0.600183	-0.00041	0.950333	0.054893	0.592307	0.590586	0.00172

Table 5.6: *Performance measure estimates of a smaller disconnected network of evidence for sample size 500 (random effects)*

DGM	COVERAGE d1	BIAS d1	EMP SE d1	MODEL SE d1	DIFFERENCE d1	COVERAGE d2	BIAS d2	EMP SE d2	MODEL SE d2	DIFFERENCE d2
Scenario 1	0.941667	-0.01571	0.338124	0.334469	0.003656	0.943667	0.041515	0.326339	0.323139	0.0032
Scenario 2	0.947333	-0.01693	0.342495	0.340306	0.002188	0.944667	0.037568	0.331148	0.329883	0.001265
Scenario 3	0.944	-0.06262	0.340091	0.33861	0.001481	0.941	0.128365	0.316198	0.322622	-0.00642
Scenario 4	0.944	-0.06321	0.343273	0.344068	-0.00079	0.948333	0.120803	0.318202	0.329342	-0.01114
Scenario 5	0.954667	-0.01233	0.348146	0.355413	-0.00727	0.961667	0.018001	0.314522	0.334229	-0.01971
Scenario 6	0.953	-0.01062	0.342801	0.359582	-0.01678	0.956	0.010263	0.309906	0.341095	-0.03119
Scenario 7	0.959	-0.05297	0.345465	0.359525	-0.01406	0.956667	0.089482	0.303551	0.334597	-0.03105
Scenario 8	0.960667	-0.04843	0.339651	0.36217	-0.02252	0.956	0.075963	0.295048	0.341639	-0.04659
Scenario 9	0.944333	-0.05538	0.390983	0.389555	0.001428	0.952333	0.033255	0.380915	0.38371	-0.0028
Scenario 10	0.947333	-0.0545	0.381993	0.380236	0.001757	0.946333	0.029761	0.374222	0.375523	-0.0013
Scenario 11	0.948	-0.17418	0.385467	0.389721	-0.00425	0.947667	0.131007	0.37351	0.383498	-0.00999
Scenario 12	0.949	-0.17398	0.375544	0.380527	-0.00498	0.948333	0.126393	0.365018	0.374893	-0.00988
Scenario 13	0.951667	-0.05531	0.398699	0.410822	-0.01212	0.963333	0.013801	0.380281	0.403304	-0.02302
Scenario 14	0.960667	-0.03731	0.372713	0.397649	-0.02494	0.961	0.020822	0.363791	0.392035	-0.02824
Scenario 15	0.944333	-0.13694	0.391108	0.410777	-0.01967	0.953333	0.083633	0.371091	0.402984	-0.03189
Scenario 16	0.958667	-0.11266	0.362348	0.39735	-0.035	0.957	0.08454	0.350094	0.391691	-0.0416

5.7.3.5 Coverage by simulation scenarios

Figure 5.34 and 5.35 illustrate the impact of MAIC adjusted random effects NMA on relative treatment effects for treatments 2 and 3 respectively. Figure 5.34, 5.35 show that when the overlap between studies was high, the coverage of treatment effects was at the nominal level for both sample sizes 150 and 500. When the overlap was low, the coverage started to decrease slightly, which was more visible for sample size 150. For treatment 2, the lowest coverage was 90% and 92% for sample sizes 150 and 500 respectively. For treatment 3, the lowest coverage was 93% for sample size 150 and for sample size 500, the rest of the coverages were at the nominal level.

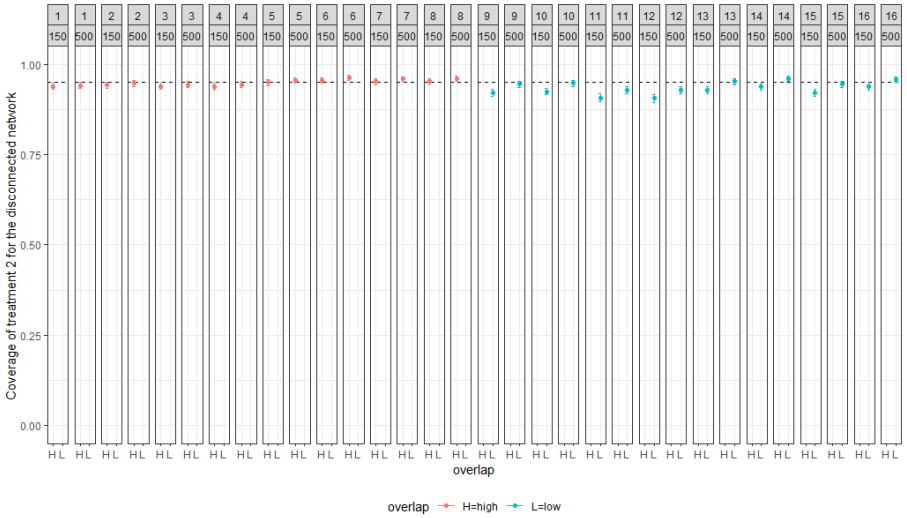


Figure 5.34: *Coverage by overlap for treatment 2*

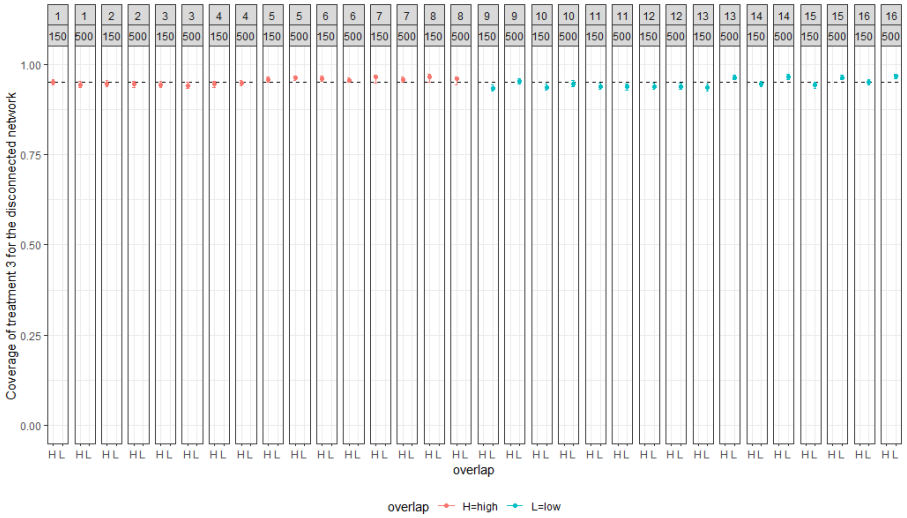
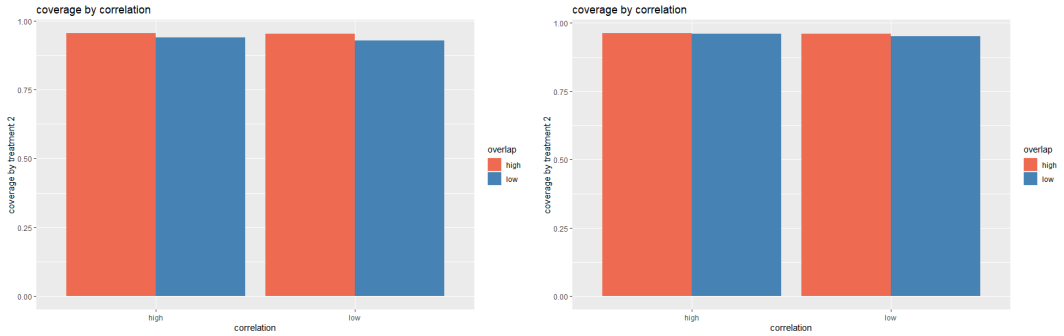


Figure 5.35: *Coverage by overlap for treatment 3*

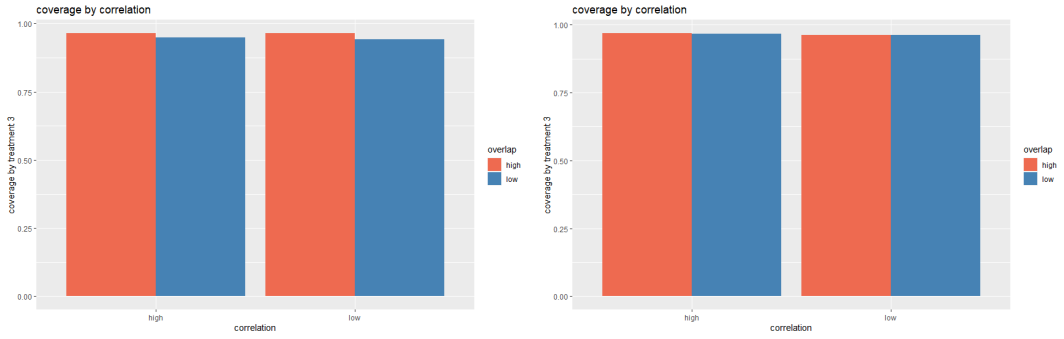
5.7.3.6 Coverage by correlation with overlap

Figure 5.36 and Figure 5.37 show the coverage with different levels of overlap and correlation for treatments 2 and 3 respectively. In Figure 5.36 (b) and Figure 5.37 (b), for sample size 500, the coverage is similar with different levels of overlap and correlation for both treatments. However, with the sample size of 150 for both treatments, within the low level of correlation, the coverage was slightly higher with high overlap.



(a) *coverage by correlation for sample size 150* **(b)** *coverage by correlation for sample size 500*

Figure 5.36: *Coverage by correlation of treatment 2 for different sample sizes (disconnected smaller network of evidence)*

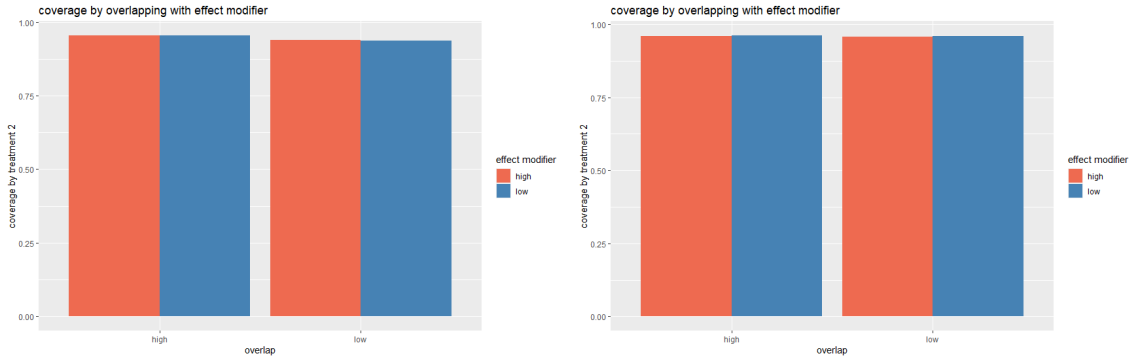


(a) *coverage by correlation for sample size 150* **(b)** *coverage by correlation for sample size 1000*

Figure 5.37: *Coverage by correlation of treatment 3 for different sample sizes (disconnected smaller network of evidence)*

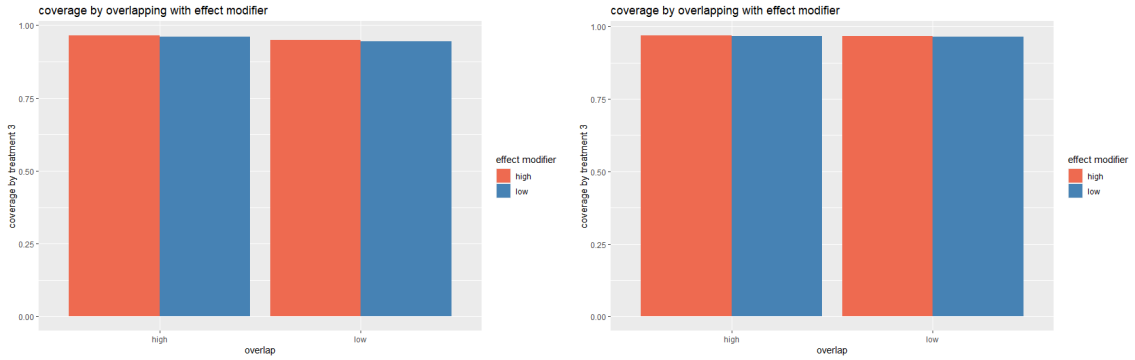
5.7.3.7 Coverage by overlap with effect-modifiers

Figure 5.38 and 5.39 show the coverage by different levels of overlap and effect-modifiers for treatments 2 and 3 respectively. In Figure 5.38 (a) and 5.39 (a), for both treatments 2 and 3, with sample size 150, the coverage decreases slightly with low overlap, but effect-modifiers show no impact on coverage irrespective of sample size and treatments.



(a) *coverage with EM for sample size 150* **(b)** *coverage with EM for sample size 500*

Figure 5.38: *Coverage by overlap with EM levels for treatment 2 (disconnected smaller network of evidence)*

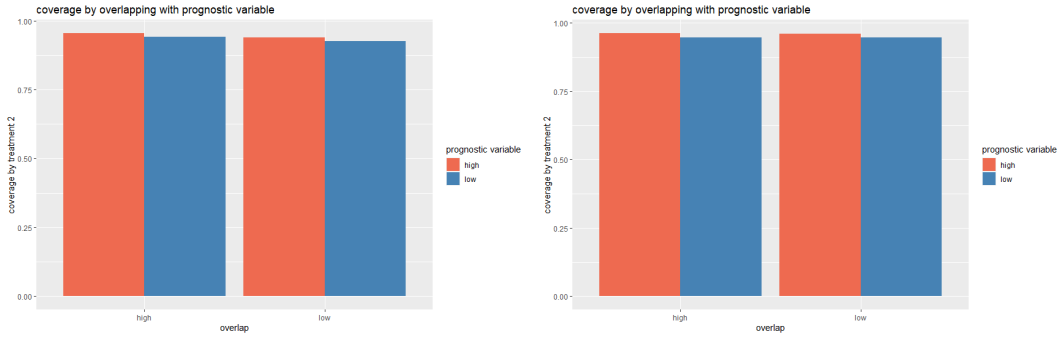


(a) *coverage with EM for sample size 150* **(b)** *coverage with EM for sample size 500*

Figure 5.39: *Coverage by overlap with EM levels for treatment 3 (disconnected smaller network of evidence)*

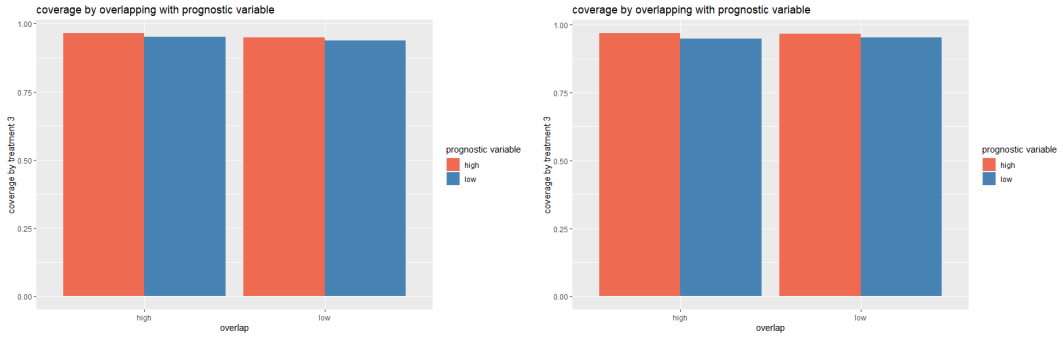
5.7.3.8 Coverage by overlap with prognostic variable

Figure 5.40 and 5.41 show the amount of coverage by different levels of overlap and prognostic variables for treatment 2 and 3 respectively. For both sample sizes, the coverage was similar with varying levels of overlap and prognostic variables.



(a) *coverage by PV for sample size 150* **(b)** *coverage by PV for sample size 500*

Figure 5.40: *Coverage by overlap with PV levels of treatment 2 (disconnected smaller network of evidence)*



(a) *coverage by PV for sample size 150* **(b)** *coverage by PV for sample size 500*

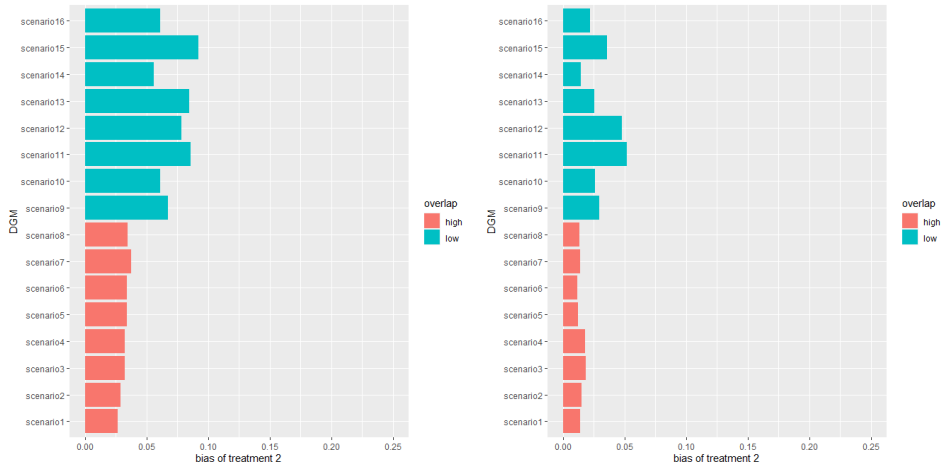
Figure 5.41: *Coverage by overlap with PV levels of treatment 3 (disconnected smaller network of evidence)*

5.7.4 Results of simulation scenarios for the larger disconnected network of evidence

Data were generated as described in Section 4.3.2 and in Figure 4.5 in the previous chapter where at first 10 RCTs were generated and then an NMA was performed. These RCTs were then made disconnected artificially to implement MAIC. The MAIC estimates were then used to perform a NMA again. Data were generated for 16 scenarios as described in Table 4.1 in Section 4.3.2. The following sections describe the simulation results from the MAIC-adjusted NMA.

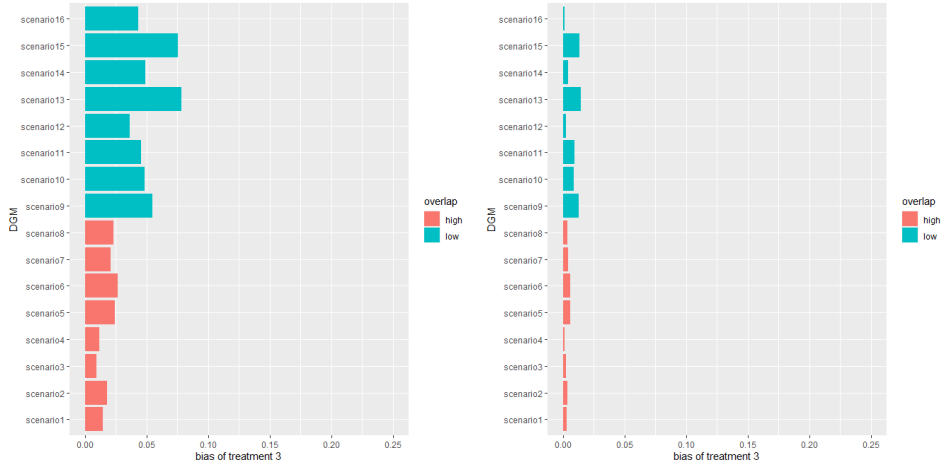
5.7.4.1 Bias by overlap

Figure 5.42 and 5.43 show the amount of bias for each DGM. The amount of bias seems to decrease with the amount of overlap. The biases were low for high-overlap scenarios and high for low-overlap scenarios and it was found true for both sample sizes. Scenarios 11, 12, 13, 15, 16 show more biases throughout different sample sizes and treatments. One reason could be that the coefficient of the effect-modifier was high in those scenarios.



(a) *bias by overlap for sample size 150* **(b)** *bias by overlap for sample size 500*

Figure 5.42: *Bias of treatment 2 for different sample sizes (disconnected larger network of evidence)*

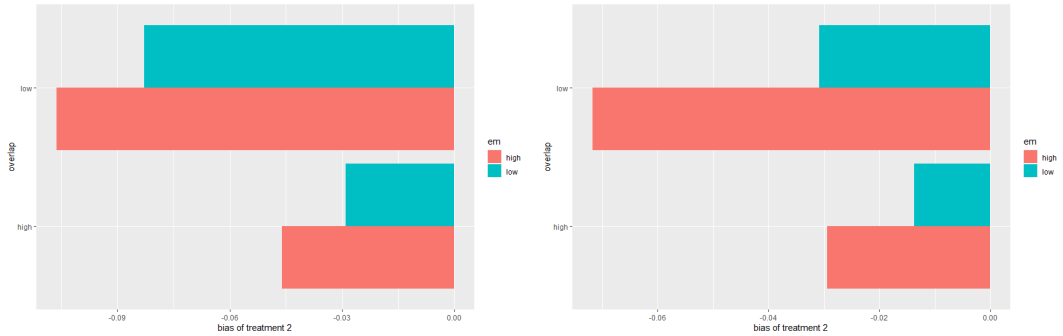


(a) *bias by overlap for sample size 150* **(b)** *bias by overlap for sample size 500*

Figure 5.43: *Bias of treatment 3 for different sample sizes (disconnected larger network of evidence)*

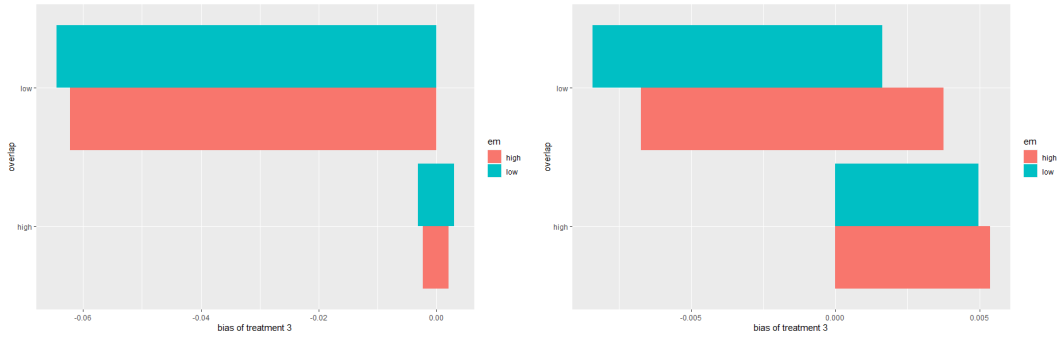
5.7.4.2 Bias by overlap with effect-modifiers

Figure 5.44 and 5.45 show the amount of bias by overlap for different levels of effect-modifier for treatments 2 and 3 respectively. In Figure 5.44 (a),(b), the bias reduces with high overlap, and the reduction seems to depend on effect-modifier levels also. For treatment 2, the bias is low with the low level of effect-modifier and this reduction is more prominent for sample size 500. In Figure 5.45 (a), for treatment 3, the bias reduces with high overlap but within levels of overlap, more bias is found with a low level of effect-modifier. However, in Figure 5.45 (b), within the levels of overlap, a similar amount of bias was seen for varying levels of effect-modifier.



(a) *bias by EM for sample size 150* **(b)** *bias by EM for sample size 500*

Figure 5.44: *Bias by overlap with EM levels for treatment 2 (disconnected larger network of evidence)*

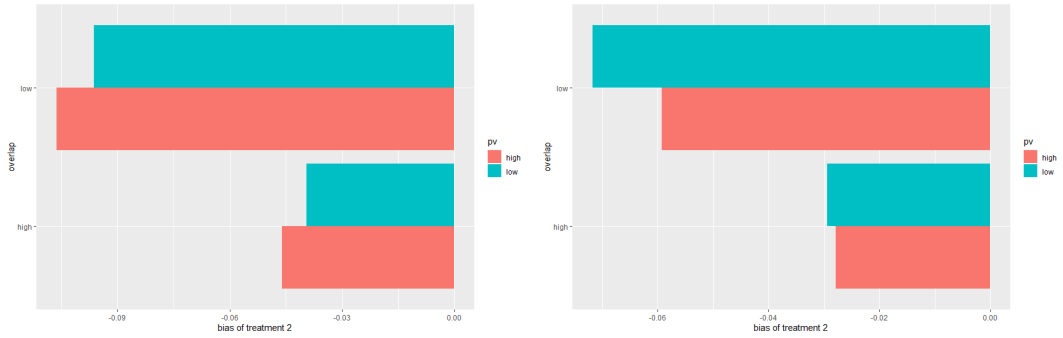


(a) *bias by EM for sample size 150* **(b)** *bias by EM for sample size 500*

Figure 5.45: *Bias by overlap with EM levels for treatment 3 (disconnected larger network of evidence)*

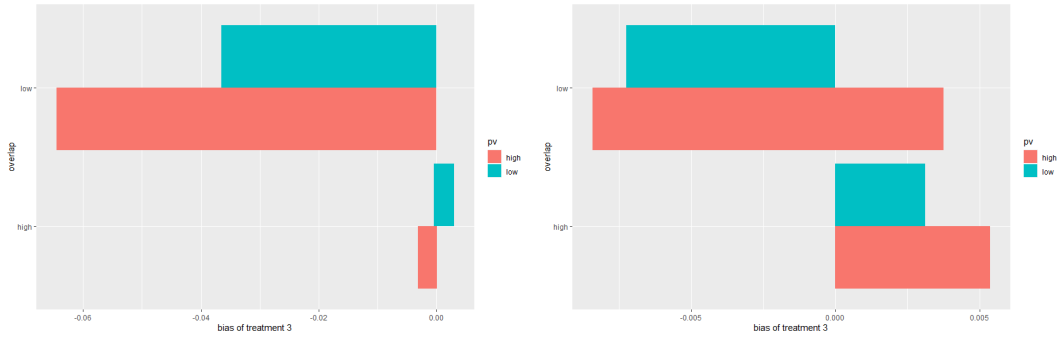
5.7.4.3 Bias by overlap with prognostic variable

Figure 5.46 and 5.47 show the amount of bias by overlap for different levels of prognostic variables for treatments 2 and 3 respectively. In Figure 5.46 (a), the bias reduces with high overlap but increases with the high level of prognostic variable coefficient. However, the opposite is seen for sample size 500 in Figure 5.46 (b). In Figure 5.46 (b), more bias is found for the low level of the prognostic variable. In Figure 5.47 (a),(b) for treatment 3, a bigger bias is seen for low overlap and high level of the prognostic variable.



(a) *bias by PV for sample size 150* **(b)** *bias by PV for sample size 500*

Figure 5.46: *Bias by overlap with PV levels for treatment 2 (disconnected larger network of evidence)*



(a) *bias by PV for sample size 150* **(b)** *bias by PV for sample size 500*

Figure 5.47: *Bias by overlap with PV levels for treatment 3 (disconnected larger network of evidence)*

5.7.4.4 Empirical SE and model SE for the simulation scenarios

The simulation results of performance measures for the larger disconnected network of evidence are summarised in Table 5.7 and Table 5.8 for sample sizes 150 and 500 respectively. Table 5.7 and Table 5.8 show the relative estimates of treatments 2 and 3 with treatment 1 from disconnected NMA for both sample sizes. The empirical SE was bigger than the model SE for each DGM which represents undercoverage. The highest value of the difference between the empirical SE and model SE was found to be 0.27 for the larger disconnected network of evidence. Additionally, all the simulation scenarios show deviation from the nominal level of coverage which is depicted by the red colour. Both moderate and higher magnitudes of biases were seen with low overlap scenarios.

Table 5.7: *Performance measure estimates of a larger disconnected network of evidence for sample size 150 (random effects)*

DGM	COVERAGE d1	BIAS d1	EMP SE d1	MODEL SE d1	DIFFERENCE d1	COVERAGE d2	BIAS d2	EMP SE d2	MODEL SE d2	DIFFERENCE d2
Scenario 1	0.829667	-0.02645	0.272367	0.191887	0.08048	0.826	-0.01434	0.283594	0.199352	0.084242
Scenario 2	0.818667	-0.02867	0.289902	0.198345	0.091557	0.819333	-0.01774	0.299896	0.207157	0.092739
Scenario 3	0.840333	-0.03231	0.273832	0.196775	0.077056	0.833333	-0.00926	0.278742	0.198378	0.080364
Scenario 4	0.830667	-0.03248	0.290843	0.202567	0.088275	0.827333	-0.01135	0.294561	0.206094	0.088467
Scenario 5	0.824	-0.03415	0.315096	0.216742	0.098354	0.820333	-0.02421	0.313692	0.214311	0.099381
Scenario 6	0.816333	-0.03387	0.327891	0.220117	0.107774	0.819333	-0.02667	0.324883	0.221671	0.103212
Scenario 7	0.834	-0.03743	0.31653	0.221471	0.095059	0.819667	-0.02109	0.31135	0.214804	0.096546
Scenario 8	0.825333	-0.03438	0.326813	0.223286	0.103527	0.826	-0.02317	0.322502	0.221914	0.100588
Scenario 9	0.692	-0.06712	0.497048	0.255591	0.241457	0.722333	-0.05494	0.508089	0.280105	0.227985
Scenario 10	0.718	-0.06092	0.449998	0.244247	0.20575	0.737667	-0.04823	0.464193	0.267129	0.197063
Scenario 11	0.692333	-0.08549	0.498047	0.256122	0.241925	0.723	-0.04553	0.505272	0.27953	0.225741
Scenario 12	0.717	-0.07801	0.450462	0.244622	0.20584	0.743333	-0.03614	0.460721	0.266335	0.194386
Scenario 13	0.695	-0.08474	0.550635	0.275978	0.274658	0.722333	-0.07844	0.550639	0.300755	0.249884
Scenario 14	0.732667	-0.05557	0.47943	0.261615	0.217815	0.762667	-0.04898	0.48681	0.285156	0.201654
Scenario 15	0.694	-0.09194	0.548814	0.27594	0.272874	0.727667	-0.07519	0.548471	0.30045	0.248022
Scenario 16	0.735667	-0.06126	0.478799	0.26138	0.217419	0.775333	-0.04343	0.482765	0.284732	0.198033

Table 5.8: *Performance measure estimates of a larger disconnected network of evidence for sample size 500 (random effects)*

DGM	COVERAGE d1	BIAS d1	EMP SE d1	MODEL SE d1	DIFFERENCE d1	COVERAGE d2	BIAS d2	EMP SE d2	MODEL SE d2	DIFFERENCE d2
Scenario 1	0.875333	-0.01409	0.18688	0.148451	0.038429	0.9	-0.00299	0.191902	0.160375	0.031527
Scenario 2	0.873667	-0.0149	0.192973	0.150598	0.042375	0.892667	-0.00377	0.19933	0.162997	0.036333
Scenario 3	0.882667	-0.01857	0.186151	0.149748	0.036403	0.901	-0.00253	0.187222	0.159451	0.027771
Scenario 4	0.874333	-0.01776	0.192181	0.151728	0.040453	0.897	-0.00125	0.19359	0.162058	0.031531
Scenario 5	0.877333	-0.01206	0.200688	0.156335	0.044353	0.894	-0.00581	0.198674	0.163872	0.034802
Scenario 6	0.874667	-0.01162	0.203268	0.157402	0.045867	0.9	-0.00597	0.199252	0.166509	0.032743
Scenario 7	0.880667	-0.01402	0.199721	0.157568	0.042153	0.904333	-0.00389	0.193847	0.163426	0.030421
Scenario 8	0.879333	-0.0135	0.200831	0.158075	0.042756	0.902667	-0.00332	0.193894	0.165979	0.027915
Scenario 9	0.775667	-0.02963	0.270003	0.170724	0.09928	0.818	-0.01285	0.27574	0.188255	0.087485
Scenario 10	0.802667	-0.02572	0.252266	0.166647	0.085619	0.838	-0.00855	0.260505	0.18385	0.076655
Scenario 11	0.770667	-0.05179	0.269216	0.170603	0.098614	0.817667	-0.00924	0.272658	0.187652	0.085006
Scenario 12	0.802333	-0.04779	0.250224	0.16631	0.083914	0.840333	-0.00268	0.256084	0.183083	0.073
Scenario 13	0.784667	-0.02541	0.287003	0.179934	0.107068	0.814	-0.01447	0.292566	0.197672	0.094894
Scenario 14	0.813333	-0.01451	0.260133	0.173257	0.086876	0.852	-0.00391	0.259878	0.190717	0.069161
Scenario 15	0.778	-0.0359	0.28444	0.17963	0.10481	0.816333	-0.0135	0.290157	0.197227	0.09293
Scenario 16	0.814333	-0.02212	0.256856	0.172732	0.084124	0.856333	-0.00141	0.256136	0.189939	0.066197

5.7.4.5 Coverage by simulation scenarios

Figure 5.48 and 5.49 illustrates the impact of MAIC-adjusted random effects NMA on treatments 2 and 3 respectively. Similar to the fixed effect NMA, the coverage is again not at a nominal (95%) level but the coverage of the random effects NMA model is greater than the fixed effect model. Figure 5.48 shows the coverage for treatment 2 for both sample sizes. When overlap between studies was high, the highest coverage was 84% for both sample sizes. For the 8 scenarios where the coverage was low, the highest coverage was 74% for sample size 150 and 81% for sample size 500. Figure 5.49 shows the coverage for treatment 3 for sample sizes 150 and 500 respectively. The highest coverage is 83% and 90% for sample sizes 150 and 500 respectively. When the overlap was low, the coverage for each of the 8 scenarios decreased more. For sample size 150, the highest coverage was 78% and the lowest was 72% and those for sample size 500 were 85% and 81% respectively. Within low overlap, for the scenarios where the correlation was high, a slight increase was seen for the coverage and this was true for both treatments.

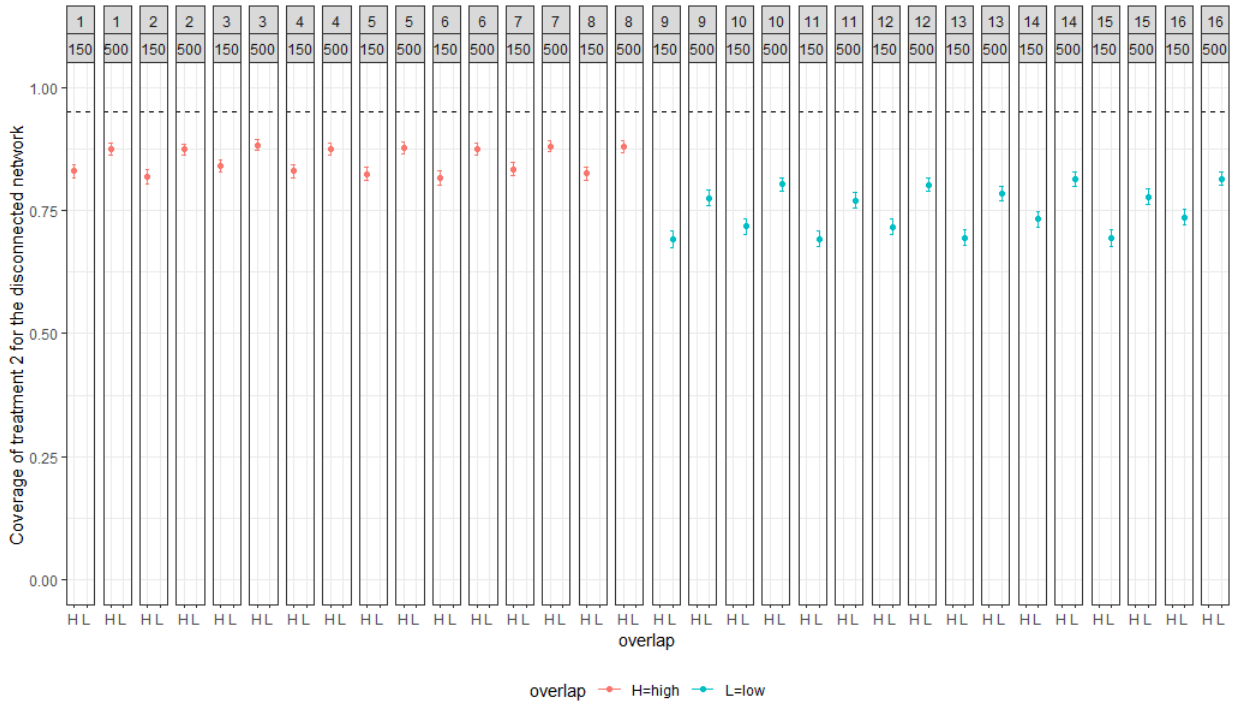


Figure 5.48: *Coverage by overlap for treatment 2*

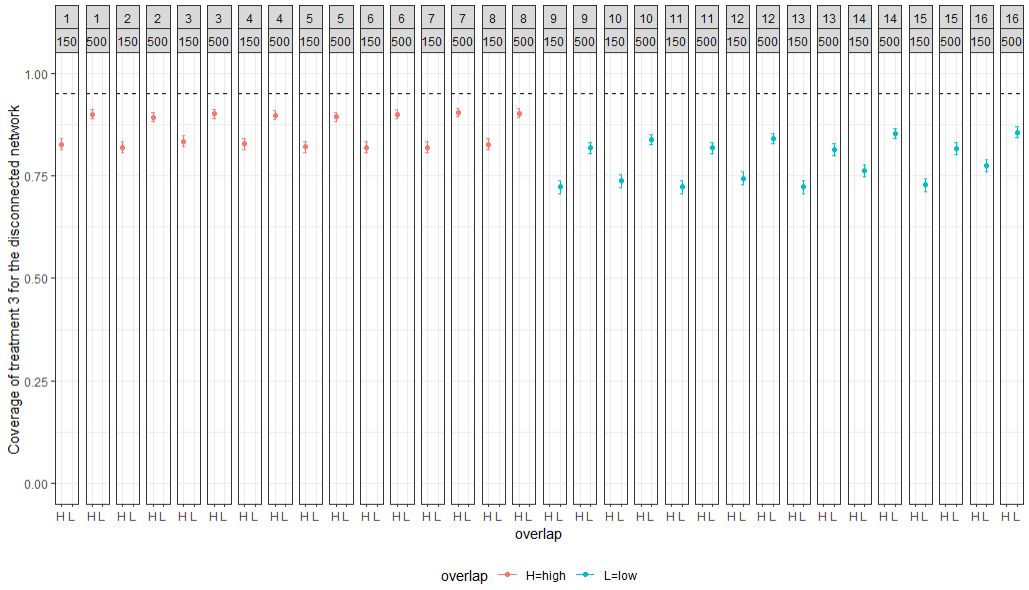
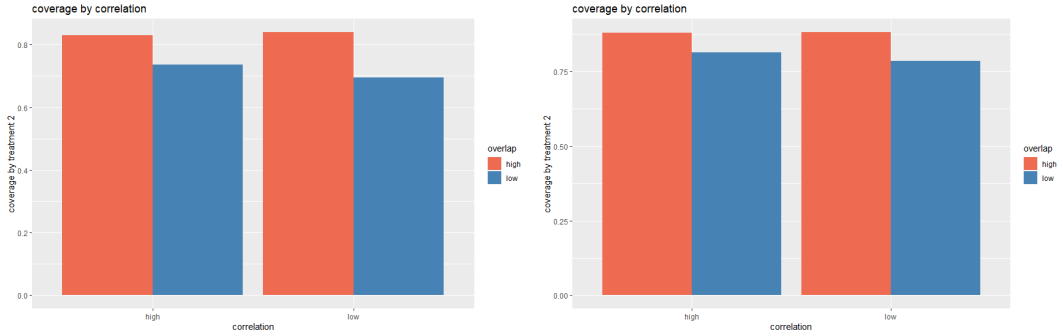


Figure 5.49: *Coverage by overlap for treatment 3*

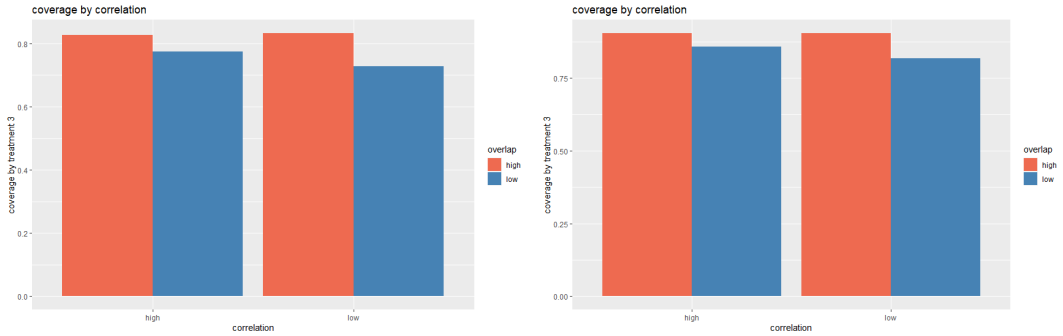
5.7.4.6 Coverage by correlation with overlap

Figure 5.50 and Figure 5.51 show the coverage for different levels of correlation and overlap for treatments 2 and 3 respectively. Figure 5.50 and Figure 5.51 show that within each level of overlap, the coverage decreases slightly with a decrease in the correlation.



(a) *coverage by correlation for sample size 150* **(b)** *coverage by correlation for sample size 500*

Figure 5.50: *Coverage by correlation of treatment 2 for different sample sizes (disconnected larger network of evidence)*

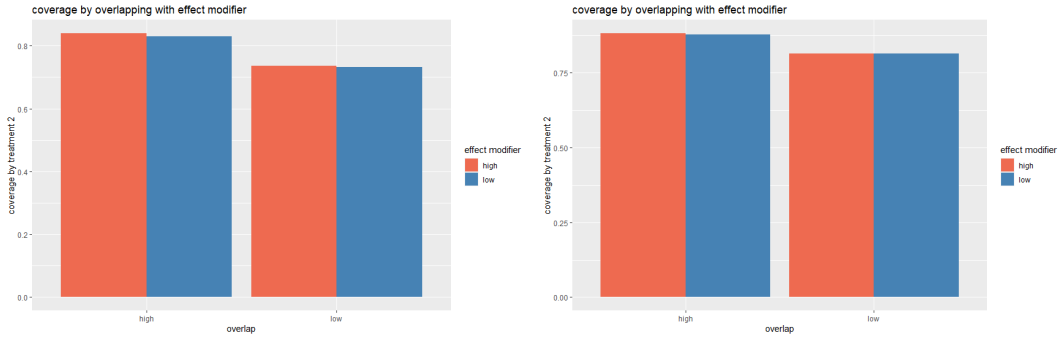


(a) *coverage by correlation for sample size 150* **(b)** *coverage by correlation for sample size 500*

Figure 5.51: *Coverage by correlation of treatment 3 for different sample sizes (disconnected larger network of evidence)*

5.7.4.7 Coverage by overlap with effect-modifiers

Figure 5.52 and 5.53 show the coverage by overlap with effect-modifiers for treatments 2 and 3 respectively. The coverage is low with low overlap, but effect-modifiers show no impact on coverage irrespective of sample size and treatments.

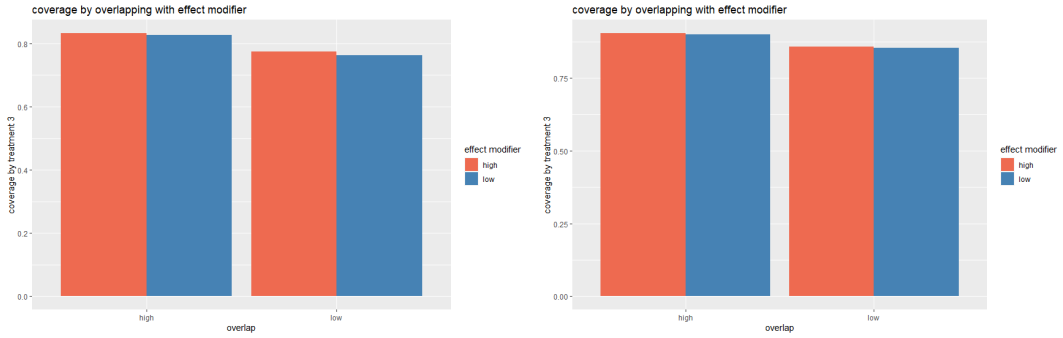


(a) *coverage with EM for sample size 150* **(b)** *coverage with EM for sample size 500*

Figure 5.52: *Coverage by overlap with EM levels for treatment 2 (disconnected larger network of evidence)*

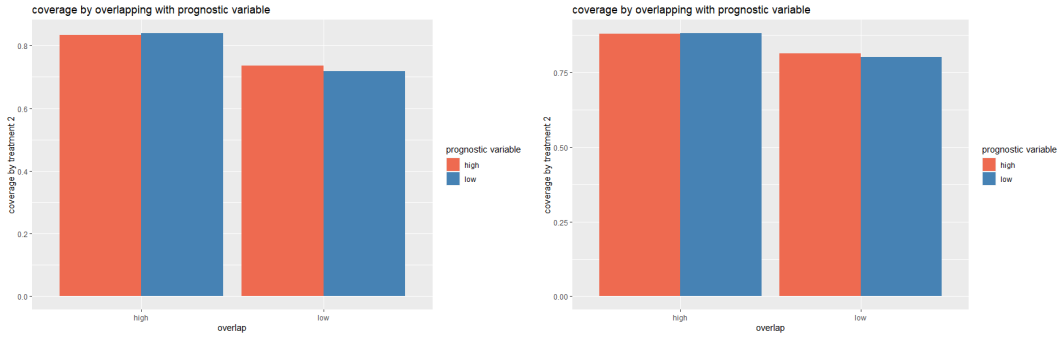
5.7.4.8 Coverage by overlap with prognostic variable

Figure 5.54 and 5.55 show the amount of coverage by overlap with respect to prognostic variables for treatments 2 and 3 respectively. For both sample sizes, the coverage was low with low overlapping but the prognostic variable seems to have no effect on this. Whether the overlap was high or low, the coverage seemed to be similar with different levels of the prognostic variable.



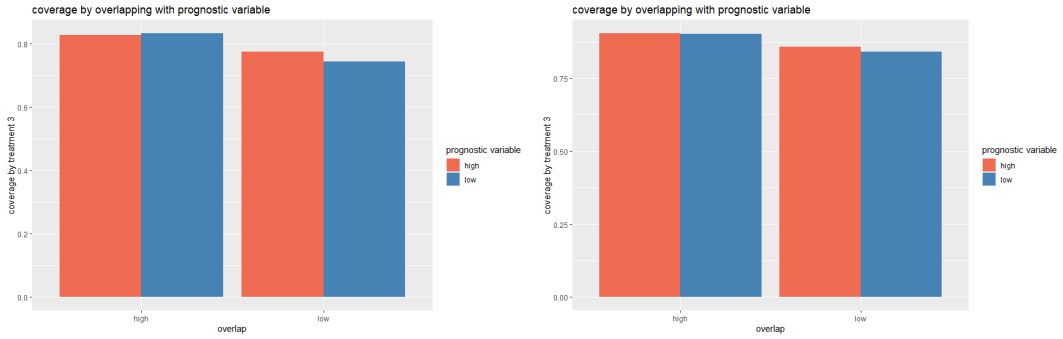
(a) *coverage with EM for sample size 150* **(b)** *coverage with EM for sample size 500*

Figure 5.53: *Coverage by overlap with EM levels for treatment 3 (disconnected larger network of evidence)*



(a) *coverage by PV for sample size 150* **(b)** *coverage by PV for sample size 500*

Figure 5.54: *Coverage by overlap with PV levels of treatment 2 for different sample sizes (disconnected larger network of evidence)*

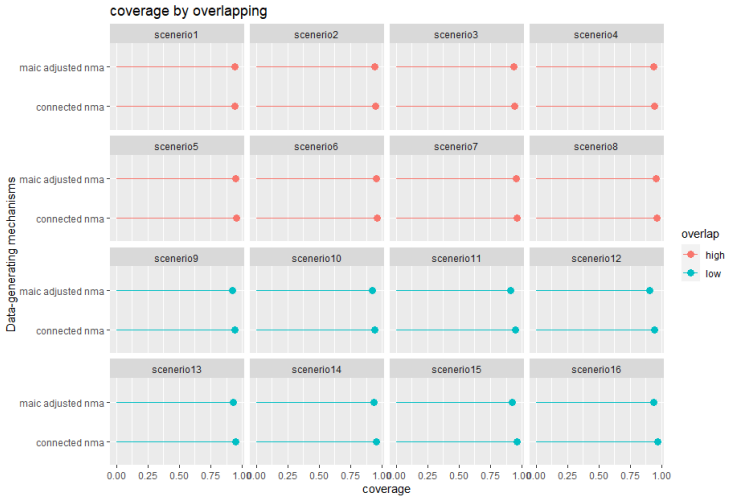


(a) *coverage by PV for sample size 150* **(b)** *coverage by PV for sample size 500*

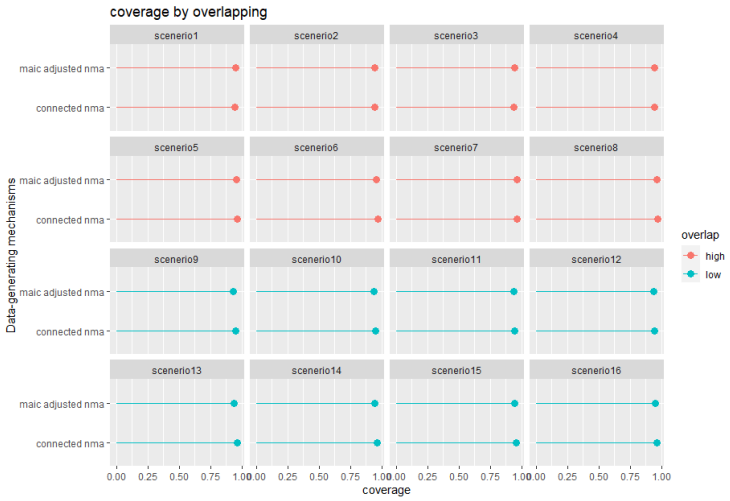
Figure 5.55: *Coverage by overlap with PV levels of treatment 3 for different sample sizes (disconnected larger network of evidence)*

5.7.5 Coverage by connected NMA and MAIC-adjusted NMA methods for the smaller network of evidence

Figure 5.56 and Figure 5.57 show the coverage for connected and disconnected NMA methods for the smaller network of evidence. Connected NMA refers to the NMA estimates from the connected network of evidence and disconnected NMA refers to the estimates from MAIC-adjusted NMA. Figure 5.56 and Figure 5.57 show the coverage for these two NMAs for sample sizes 150 and 500 respectively. In Figure 5.56, for each DGM, the coverage for the connected and disconnected NMA was at a nominal level except for the DGM 9,10,11,12,15 in MAIC-adjusted NMA. In those DGM, the coverage was slightly low and all these DGM were from low overlap. For sample size 500 and only for treatment 2, a slight reduction in coverage for MAIC-adjusted NMA was seen for DGM 11 and 12.

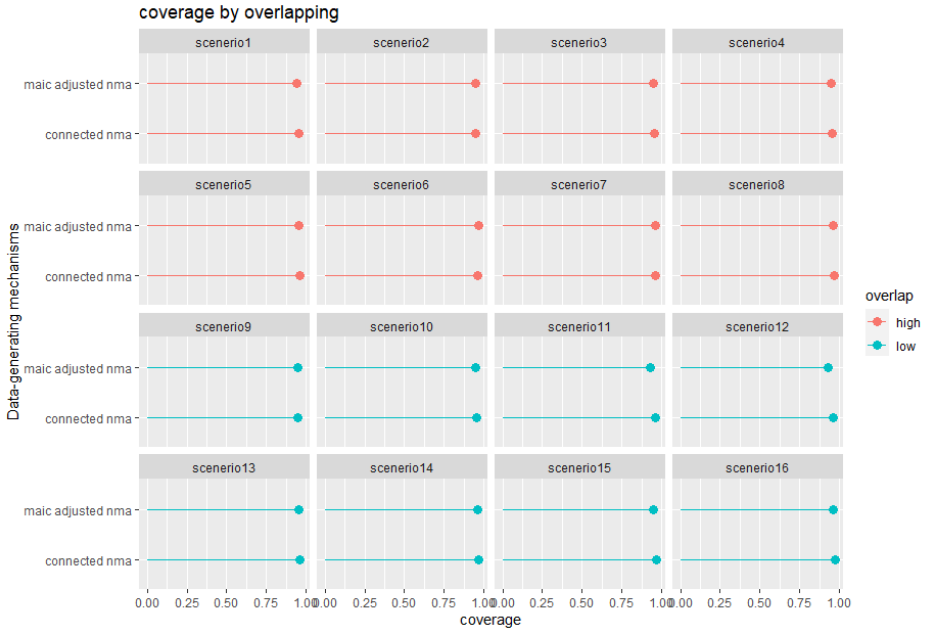


(a) *coverage by overlap with treatment 2*

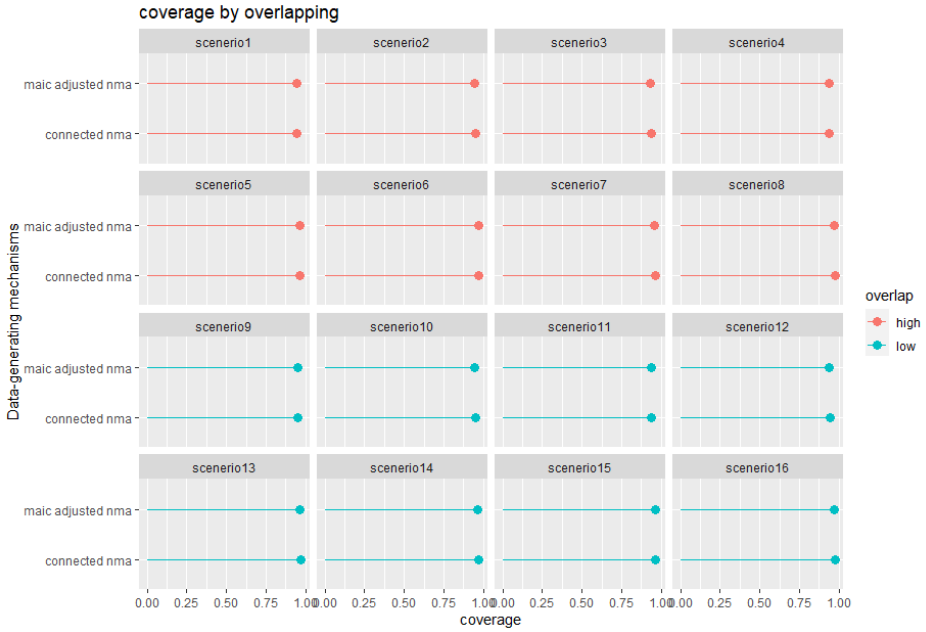


(b) *coverage by overlap with treatment 3*

Figure 5.56: *Coverage by different NMA methods with sample size 150 (smaller network of evidence)*



(a) *coverage by overlap with treatment 2*

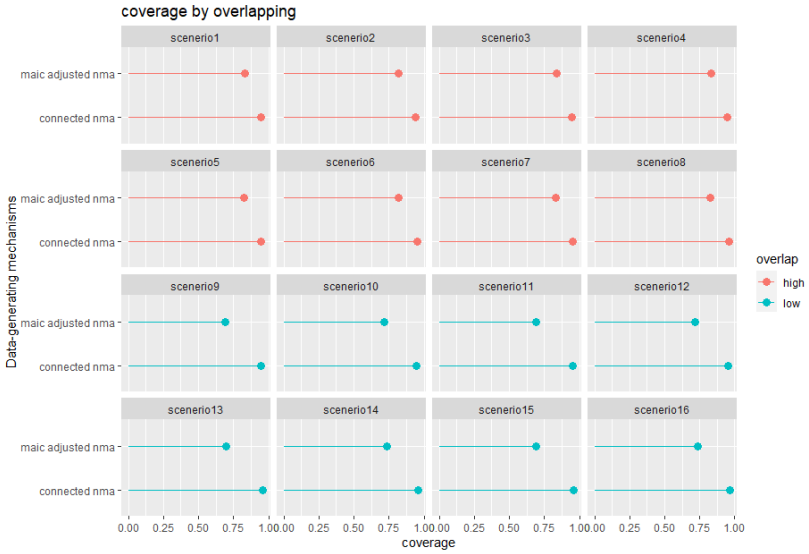


(b) *coverage by overlap with treatment 3*

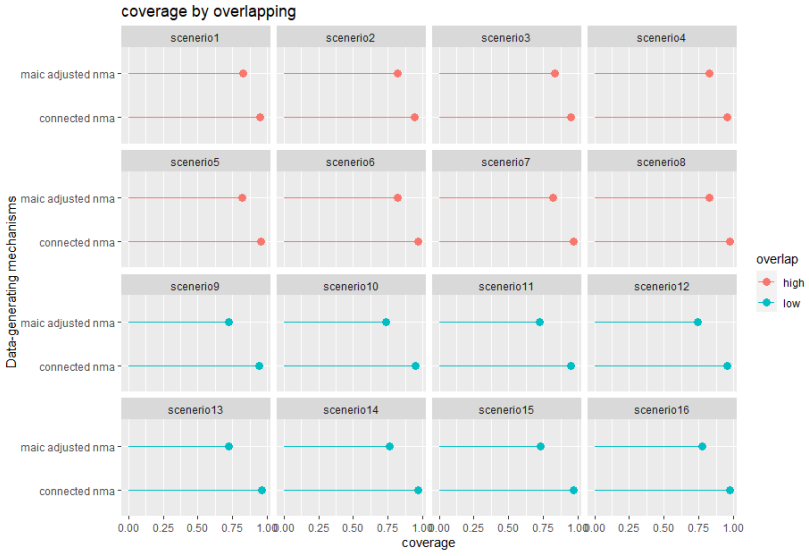
Figure 5.57: *Coverage by different NMA methods with sample size 500 (smaller network of evidence)*

5.7.6 Coverage by connected NMA and MAIC-adjusted NMA for the larger network of evidence

Figure 5.58 and Figure 5.59 show the coverage for connected NMA and MAIC-adjusted NMA for sample sizes 150 and 500 respectively. For each DGM, the coverage for the connected NMA was at a nominal level whereas the coverage for the disconnected NMA was below the nominal level. This was true irrespective of the overlap and sample sizes.

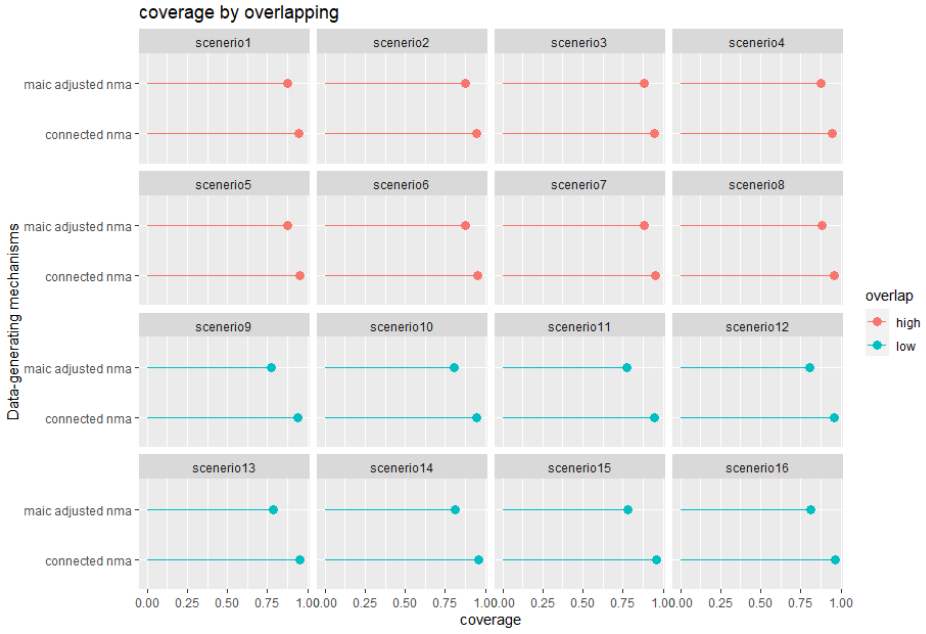


(a) *coverage by overlap with treatment 2*

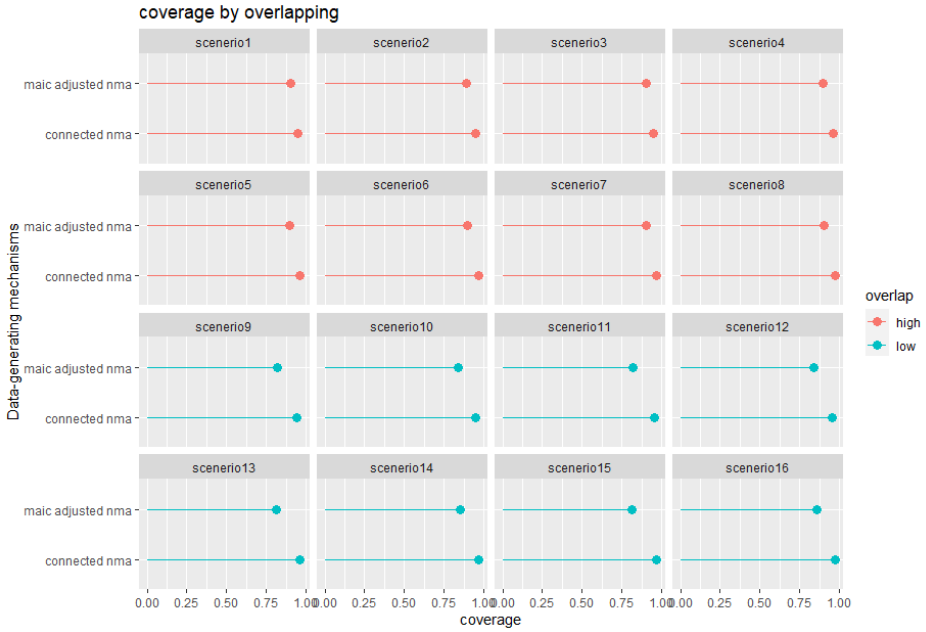


(b) *coverage by overlap with treatment 3*

Figure 5.58: *Coverage by different NMA methods with sample size 150 (larger network of evidence)*



(a) *coverage by overlap with treatment 2*



(b) *coverage by overlap with treatment 3*

Figure 5.59: *Coverage by different NMA methods with sample size 500 (larger network of evidence)*

5.8 Discussion

Biases were found to be independent of the two levels of overlap for both the smaller and larger connected network of evidence. Bigger biases can be seen with high overlap scenarios and vice-versa. In the connected network of evidence, in the scenarios where the biases were comparatively higher, the coefficient of the effect-modifying variable was also higher. This could be the reason for these big biases as the NMA method is not able to adjust for these variables. However, the magnitude of biases was smaller for the larger network of evidence in comparison to the smaller network of evidence. The general trend in biases for the two connected networks was found to be higher biases associated with high effect-modifier levels for most of the cases. With prognostic variables, the trend was similar where most of the time, bias was higher with high prognostic variable levels. The coverage for both the smaller and larger connected network of evidence was at the nominal level (95%) as expected. The coverage was found to be unassociated with correlation as well as different levels of effect-modifier and prognostic variable levels.

For the smaller disconnected network of evidence, bias was found to be inversely associated with overlap. Lower biases were found with high overlap levels and vice-versa. This was expected as when MAIC is conducted to estimate relative treatment effect in a disconnected network of evidence, the estimate would be better with high overlap between studies and the bias will reduce (Remiro-Azócar et al., 2021). Additionally, bias was found to be high with high effect-modifier levels and high prognostic variable levels. However, the undercoverage issue with the MAIC-adjusted NMA was less severe with the smaller network of evidence compared to the larger network of evidence. In high overlap scenarios, the coverage was at the nominal level in smaller disconnected network of evidence. Coverage starts to decrease slightly with low overlap scenarios and with a small sample size. The coverage seems to be unaffected by correlation with the high overlap level but with a low overlap level, coverage slightly increases with a high correlation. Similar to the smaller and larger connected network of evidence, coverage was found to be unaffected with different levels of effect-modifier and prognostic variable levels.

In the larger disconnected network of evidence, biases were again found to be inversely related to overlap which was similar to the smaller disconnected network of evidence. Lower biases were found with high overlap levels and vice-versa. The general trend was found to be higher bias with high effect-modifier levels and higher prognostic variable levels. For the larger disconnected network of evidence, undercoverage was found for all DGM irrespective of overlap. When the overlap between study covariates was low, the reduction in coverage was higher than that of high overlap. However, though undercoverage can be seen for all the DGM irrespective of overlap and sample sizes, the coverage was slightly better with the big sample size. Within the two levels of overlap, coverage was found to be slightly better with the high correlation level. Additionally, though coverage was affected mainly due to overlap between study covariates, prognostic and effect-modifying variables seem to have no effect on the coverage.

Overall, the findings from the random effects NMA model were similar to the fixed effect NMA model. In four data settings of the simulation study, the absolute size of biases ap-

peared to be below one-half of the estimates SE, therefore, none of them can be considered as problematic following Schafer and Graham (2002). As expected, similar to the previous chapter, throughout the simulation for the four networks of evidence, the highest biases were seen for the connected and disconnected smaller network of evidence. Similar to the previous chapter, undercoverage was also detected for the random effects NMA with MAIC estimates. This undercoverage of NMA estimates were more visible for the larger network of evidence. The difference between the coverage of fixed and random effects NMA was that the coverage was slightly better (closer to 95%) for the random effects NMA model compared to the fixed effect model. This was due to the fact the heterogeneity parameter τ added some extra level of randomness which inflates the SE. The SE is slightly bigger than the fixed effect model which causes more coverage for the random effects NMA model.

In both fixed effect and random effects NMA simulation chapters, data were generated for two disconnected networks of evidence termed as smaller and larger disconnected networks of evidence. The idea behind conducting the simulation for the two disconnected networks of evidence was to observe the impact of repeated use of the IPD in a MAIC-adjusted NMA for a smaller and a larger network of evidence. It was found that, for both fixed and random effects NMA, for a smaller network of evidence, undercoverage was not as severe as compared to the larger network of evidence. Nevertheless, for a smaller network of evidence, undercoverage was only observed for the low-overlap scenarios and coverage was at the nominal level for the high-overlap scenarios. With the larger disconnected network of evidence, for both fixed and random effects NMA, undercoverage was seen for all the DGM. This is due to the increased use of IPD. The more the IPD was used, the more ramification can be seen in the coverage. In spite of the fact that undercoverage was frequent in both fixed and random effects NMA model, the coverage of the random effects NMA model was better than the fixed effect model. Moreover, the undercoverage issue was found to be slightly better with the bigger sample size for the random effects NMA model which was not found in the fixed effect NMA.

The undercoverage issue was found to be more severe in a larger MAIC-adjusted NMA model for both fixed and random effects models. The MAIC-adjusted NMA when applied to a smaller network of evidence, the undercoverage issue was found to be increased with the decrease of overlap of covariates between studies. In a larger MAIC-adjusted NMA model, undercoverage was associated with both overlap and correlation between study covariates. Undercoverage was found to be better with a high overlap between study covariates and a high correlation of within-study covariates. In addition, the sample size was found to be related to undercoverage for the random effects model where the coverage was found to be slightly better with a bigger sample size.

It is expected that the empirical coverage probability conforms to 0.95% which is essential to get a proper type I error rate. An appropriate type I error rate ensures valid testing of a "no effect" null hypothesis. It is essential to get a correct SE in order to achieve the nominal level of coverage. The robust sandwich SE was estimated both for the fixed and random effects MAIC-adjusted NMA estimates, however, it was not able to reach the nominal level of coverage. A previous simulation study shows that MAIC can underestimate the empirical SE with small sample sizes and low overlap (Remiro-Azócar et al., 2021). However, the simulation

study in this thesis did not compute coverage for individual MAICs. The simulation study found that in addition to the sandwich estimator, repeated use of IPD has an impact on the coverage probability of the NMA estimates. This study shows that the more IPD from the same study was used for multiple unanchored MAICs, the more departure in the coverage probability can be seen.

Use of the same IPD for multiple MAICs breaks the independence of unit of analysis assumption. The violation of this assumption creates a correlation between studies which aggravates the poor coverage issue with the sandwich estimator. A previous simulation study by Belger et al. (2015b) suggests that when MAIC need to be conducted in a connected network of evidence with multiple IPD and AgD study, then the IPD can be matched against a) just one AgD study, b) the average patient characteristics from the AgD studies, c) the average mean and variances from the AgD studies or d) the distribution of patient characteristics using MCMC from the AgD studies. The study by Belger et al. (2015b) shows the importance of preserving the independence of the unit of analysis assumption. The correlation that emerges due to repeated use of IPD needs to be taken into account when estimating the uncertainty of estimates. In general, bootstrapping is a another option to capture uncertainty correctly. Therefore, in the next chapter, a new bootstrapping approach is developed and proposed to overcome this undercoverage issue for NMA estimates.

Chapter 6

Double-Bootstrapping: A Novel Method to Estimate Confidence Intervals

6.1 Introduction

In the previous two chapters, a simulation study has been conducted for a fixed effect and a random effects NMA. The motivation for the simulation study came from the NICE STA review chapter. In the review, for the appraisals carried out between 2018 to 2021 on single-arm studies, it was found that more than half of the appraisals (55%) had a larger disconnected network of evidence to estimate relative treatment estimates of a new intervention to comparators. In addition to that, it was found that a larger disconnected network of evidence was built either by comparing the new intervention with multiple comparators from various studies or by comparing the new intervention with a single comparator that comes from multiple studies. The reason for the network being disconnected was that no connected network of evidence can be found for the comparator treatments because either most of the comparator studies were also single-arm studies or no common comparator was found from the RCTs. For the purpose of estimating relative treatment estimates between the single-arm study with relevant comparators, either unanchored MAICs or STCs were implemented.

The aim of the simulation study was to assess the consequences of applying several unanchored MAICs for binary data using IPD from the same study to get a consistent synthesis of evidence in an NMA using a sandwich estimator ignoring the fact that MAIC is designed to evaluate relative treatment effects for a pair of studies. Moreover, when IPD from a single study was used repeatedly, the assumption of independence between studies was compromised. The simulation was designed to assess the MAIC-adjusted NMA with 3 treatments both for a smaller and a larger disconnected network of evidence. In the simulation, a smaller disconnected network evaluated 3 treatments consisting of 3 studies whereas a larger disconnected network was made of 10 studies with 3 treatments without any common treatment arm. The size of the network is a crucial issue as it is often the case that the number of studies per comparison is very small in a HTA submission. Therefore, one of the objectives of the simulation study was to evaluate MAIC-adjusted NMA when the number of studies

per comparison varies. The MAIC-adjusted NMA methods were assessed both for a fixed and a random effects NMA.

The significant findings from the simulation for both fixed and random effects NMA were similar. When MAIC was applied with a smaller and a larger disconnected network of evidence, a deviation from the nominal (95%) level of coverage was found for each of the relative treatment effect estimates. With a smaller disconnected network of evidence, coverage was found to be at a nominal level when the overlap between study covariates was high. The coverage starts to decrease with low-overlap scenarios. The undercoverage issue was found to be more problematic with a larger disconnected network of evidence where it can be seen for each DGM irrespective of overlap level. Prognostic and effect-modifying variables were found to have no effect on the undercoverage issue, however, they were found to have an effect on the bias. Though undercoverage was found for both the larger disconnected network of evidence in a fixed and random effects NMA, the magnitude of undercoverage was higher for the fixed effect NMA. Coverage was comparatively better for the random effects NMA due to the heterogeneity parameter τ which takes into account variation between included studies. The τ parameter inflates the SE which results in a higher coverage for the random effects NMA.

Coverage of an estimate to nominal level is a crucial issue as the confidence interval (CI) is a measure of accuracy for the estimates as well as a measure of unreliability. In order to make inferences with respect to a population and its parameters, the CI is used as an important instrument. A conventional CI uses the sample data to calculate a likely range of estimates for an unknown parameter that helps to make judgments about the population. The width of a CI determines how precise or imprecise the estimate is, therefore, a CI with a larger width expresses more uncertainty about the estimate. Though a CI can be estimated with different levels of confidence, e.g. 90%, 99%, estimating it with a confidence level of 95% is the most common approach. The percentage in a CI describes the confidence that the interval includes the true population parameter. A bootstrap CI is an advancement in the estimation of CI that helps to estimate a CI without following assumptions which are often unrealistic.

Section 6.2 of this chapter describes the bootstrap methods that can be used to get an accurate CI. Section 6.3 describes the simulation study that was conducted to overcome the problem of undercoverage and reports the results of the double-bootstrapping for the fixed and random effects NMA. The chapter concludes with a discussion of the double-bootstrapping results in Section 6.4.

6.2 Methods

6.2.1 Bootstrapping for estimating confidence intervals (CI)

In a frequentist/conventional way, the construction of a CI mainly depends on the normality of the sampling distribution and its SD. The idea of a sampling distribution is a collection of all sample estimates when the population is resampled repeatedly. In reality, only a single sample is available to make an inference on the sampling distribution, therefore, certain assumptions need to be made. The "central limit theorem" allows that for a large sample size,

the sampling distribution will approach a normal distribution, and the SE of the population can be approximated by the SD of the sampling distribution. A problem arises if the sample size is not fairly large to implement the central limit theorem, which in turn makes it unreasonable to assume that the sampling distribution is normal.

The bootstrap CI has gained popularity over the conventional CI technique due to its simplistic approach. In a bootstrap approach, a representative sample of the population is resampled with replacement multiple times to calculate the quantity of interest on each of these resamples. The resamples then can be used to determine the sampling distribution of the quantity of interest. The SD of the sampling distribution can be used to construct a CI. Unlike the conventional CI, a bootstrap CI can be constructed without making assumptions about the underlying distribution of the data. A bootstrap CI can work well even when the sampling distribution for the quantity of interest is asymmetric. Bootstrap does not need to depend on normality assumption as the sampling distribution can simply be observed by iterating the original sample multiple times. It is straightforward to estimate the SE and CI using bootstrapping as bootstrapping is asymptotically consistent and precise than the conventional CI (Cline, 2019).

A bootstrap CI can be classified as parametric and non-parametric on the basis of how the samples have been extracted. In the case of parametric bootstrapping, each bootstrap sample of a specific size is extracted from a parametric distribution, whereas, a non-parametric bootstrap sample is extracted from the original sample with replacement without specifying any distributional form. In a non-parametric bootstrap, the sample observations that are included in a bootstrap sample, need to be generated with a uniform probability of being chosen from the original sample and also with the assumption of independence. There are various kinds of bootstrap CI namely percentile, percentile-t, and bias-corrected percentile intervals (Tibshirani and Efron, 1993).

6.2.2 Double-bootstrapping for estimating confidence intervals

Improvement in the coverage accuracy and bias correction are the two main sources of problems where the double-bootstrap is mainly used (Chang and Hall, 2015). A double-bootstrap CI is an extension of a single bootstrap CI where after extracting the first bootstrap sample, it is again used to get the second bootstrap sample. Therefore, the bootstrapping in a double-bootstrap can be described as a two-step process, where, the original sample is used to generate bootstrap samples at the first step, and then the bootstrap samples that are generated at the first step are bootstrapped again to generate bootstrap samples for the second step. The usage of the double-bootstrapping technique regularly appears in literature (Martin, 1990; Shi, 1992; Booth and Hall, 1994; Vinod, 1995; Booth and Presnell, 1998; Letson and McCullough, 1998; McCullough and Vinod, 1998; Lee and Young, 1999; Balcombe et al., 2008; Arasan and Adam, 2014; Chronopoulos et al., 2015; Chang and Hall, 2015; Ishwaran and Lu, 2019). Double-bootstrapping in the construction of a CI can reduce the discrepancy between nominal and empirical coverage probability and is capable of more accurate coverage than single bootstrapped CI (Arasan and Adam, 2014; Shang, 2021).

When a bootstrap sample is generated from a given sample, a dependency is developed be-

tween the probability distribution of the bootstrapped sample and the process that generates the data. In a double-bootstrap procedure, reiterating the bootstrap sample again can bring down this dependence (Chang and Hall, 2015; Shang, 2021). Bootstrapping can be used to approximate the unknown distribution of a statistic. During the bootstrap procedure, the original sample is resampled with replacement repeatedly which in turn creates an empirical distribution of the unknown distribution. The confidence level of this empirical distribution converges to 1-α (α=0.05) when the sample size n is big (Beran, 1987). Nevertheless, there is a discrepancy in confidence level error when the sample is finite. It has been proved that, for a one-sided CI, the double-bootstrap can eliminate coverage error by a factor O($n^{-1/2}$) whereas for a two-sided CI, the quantity is O(n^{-1}) (Chang and Hall, 2015).

A measure of distance that serves the part of error correction for coverage can be calculated for a statistic using a single bootstrap and a double-bootstrap sample (Shang, 2021). Let t be a statistic whose distribution is unknown. When t is estimated from a single bootstrap, the empirical coverage probability of t can be calculated as $\frac{L(\hat{t^*},\hat{t})>L(\hat{t},t)}{B_1}$, where $L(\hat{t},t)$ estimates the distance between the sample and population level of a statistic, L($\hat{t^*},\hat{t}$) estimates the distance between the real sample and the single bootstrap sample, and B_1 denotes the total number of samples that are drawn at the single-bootstrap level. In opposition to this, for t, the empirical coverage probability for the double-bootstrap can be estimated as $\frac{L(\hat{t^{**}},\hat{t^*})>L(\hat{t},t)}{B_1*B_2}$, where $L(\hat{t^{**}},\hat{t^*})$ estimates the distance between the single bootstrap sample and double-bootstrap sample and B_2 denotes the total number of samples that are drawn at the second level of a double-bootstrap. Chang and Hall (2015) showed that instead of taking a large no of samples at the second level in double-bootstrap, B_2 can be 1 which benefits in lowering the computation time as well as confidence level error.

6.3 A simulation study of double-bootstrapping in fixed and random effects NMA

In the previous two chapters, a MAIC-adjusted NMA was estimated both for a fixed and a random effects model. A robust sandwich estimator was used to estimate the SE for NMA estimates. Undercoverage was found to be one of the major consequences in MAIC-adjusted NMA estimates. Undercoverage of the NMA estimates was detected as more problematic for the larger disconnected network of evidence compared to a smaller disconnected network of evidence. The repeated use of the same IPD in multiple MAICs creates a dependency between studies. To fix these issues, during the estimation of MAIC-adjusted NMA, double-bootstrapping was used to estimate the coverage probability to take into account both the uncertainty and dependence between studies.

6.3.1 DGM for a larger connected and a larger disconnected network

In order to implement the double-bootstrapping, a simulation study was designed. During the simulation process, first, binary data were generated for a larger connected network of 10 RCTs with 3 treatments to conduct a fixed and a random effects NMA. Six studies compared treatments 2 and 1 and four studies compared treatments 3 and 1. Data was generated with two continuous covariates for each RCT where one of them was prognostic and the other one

was effect-modifying variable. The simulation was conducted by setting the following data properties: the number of nodes in the network, the sample size per study, network density, and the nature of prognostic and effect-modification in the network. Each of the RCT was equal in size. A sparse network i.e. a triangular network was built with no closed loops as described in Chapter 4 in Figure 4.4.

After generating data for a larger connected network of evidence, the next step was to transform this connected network of evidence into a disconnected network of evidence so that MAIC-adjusted NMA can be conducted. The first study in the network was considered as the study with IPD and the rest of the network as studies with AgD. Therefore, after generating IPD for all the studies in the network, they were converted to AgD except for study one. The connected network of evidence was transformed into single-arm studies by dropping one arm from each study. Figure 4.5 in Chapter 4 illustrates the process where each oval shape node represents the study arm with treatment number and an RCT was depicted by joining two nodes with a solid line. The treatment arm that was dropped from each study was depicted by striking through the treatment number. Treatment arm 1 was kept from the first study and for the rest of the studies, all arms were dropped except arms 2 and 3.

After the transformation, the disconnected network of evidence consisted of 1 IPD and 9 AgD studies. The next step was to apply 9 MAICs to generate relative treatment effect estimates for a NMA. The double-bootstrapping was used in this stage which is described in Figure 6.1. During this step, when the IPD was used for each MAIC, instead of taking the original IPD, double-bootstrapping was applied. In this process, first, a bootstrap sample was extracted from the original simulated IPD, and the bootstrapped sample was bootstrapped again. The second bootstrapped sample was then used for the computation of MAIC weights. The process of double-bootstrapping was repeated for the 9 MAICs, and then an NMA was estimated with the MAIC estimates. Additionally, this whole process was bootstrapped multiple times to get the bootstrapped NMA estimates. The mean of bootstrapped estimates was then taken as the overall treatment effect estimate and the SD of the bootstrapped estimates was used for the estimation of coverage probability. All the performance measures were also calculated as described in the previous chapters (Chapters 4 & 5).

The simulation study evaluates the change in five factors in a full factorial design. Taking two values from each factor results in a 2x2x2x2x2=32 scenario. The factors were sample size, the correlation coefficient between covariates, the strength of effect-modification, the strength of the prognostic factor, and the overlap of covariates between studies. The values of the different levels of factors are depicted in Chapter 4 in Table 4.1. Table 4.1 shows different combinations of all the factors except sample size. When these 16 scenarios are again combined with sample sizes 150 and 500 per arm, it results in 32 scenarios. The R codes for the double-bootstrapping are given in E.1 and E.2 of Appendix E.

6.3.2 Estimands

The estimands of interest in this study were the overall treatment effect estimates of double-bootstrapped MAIC-adjusted fixed and random effects NMA for treatments 2 and 3.

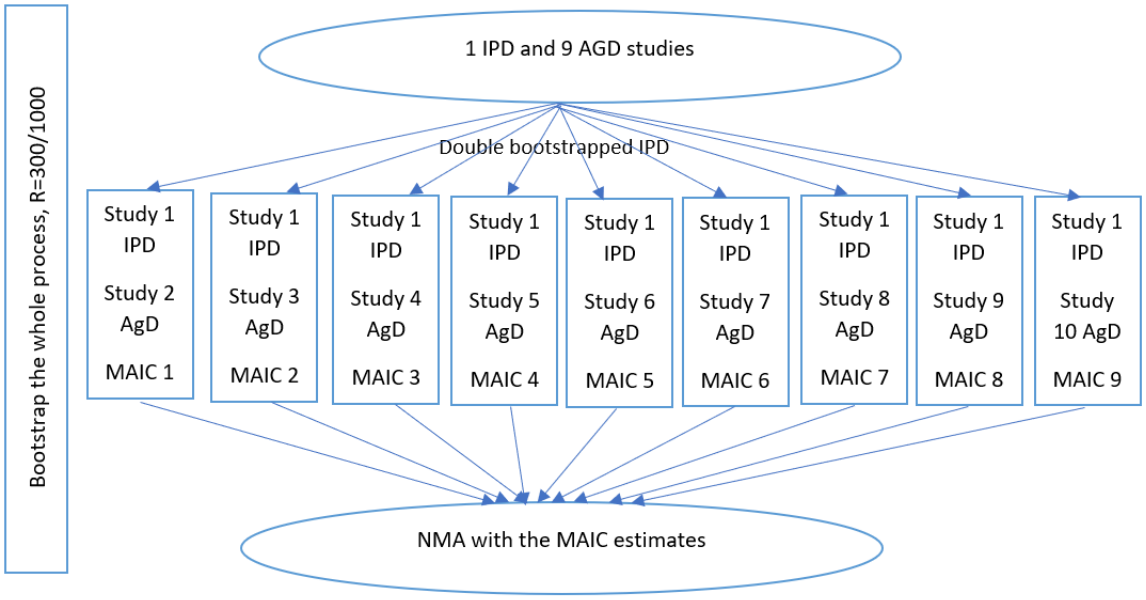


Figure 6.1: *Double-bootstrapping in the estimation of MAIC-adjusted NMA*

6.3.3 Methods

The following methods were applied to the data generated during the simulation exercise:

- A double-bootstrapped MAIC-adjusted fixed effect NMA was performed and double-bootstrapped SE were estimated for the NMA estimates. The calculation of the NMA was done using the Bayesian approach and using R package `multinma`. The R package `MAIC` was used to perform the MAIC. For the prior distributions of the treatment effects and study-specific intercepts, a normal distribution was used with $N(0, 100^2)$. MAIC was applied to match discrepancy between individuals from different studies with respect to patient demographics or covariate values (Signorovitch et al., 2010). MAIC uses weights which are calculated by the MoM/EB that gives more importance/weight to individuals in the IPD study who are more alike to the AgD study and less importance if they differ between studies. MAICs were applied upto the second moment i.e. balancing of covariate was done both for mean and standard deviation.

- A double-bootstrapped MAIC-adjusted random effects NMA was performed and double-bootstrapped standard errors were estimated for the NMA estimates. For the prior distributions of the treatment effects and study-specific intercepts, a normal distribution was used with $N(0, 100^2)$. An informative prior was used for the heterogeneity parameter τ which was Turner's prior as log-Normal(-2.56, 0.33).

6.3.4 Performance Measures

In order to compare the performance of double-bootstrapped MAIC adjusted fixed and random effects NMA in a larger disconnected network of evidence, performance measures bias, model SE, empirical SE, and coverage probability were used.

- Bias: Statistical bias in simulation study gives an estimate of the systematic discrepancy between the true parameter and expected values of the results obtained from each simulated dataset. It can be defined as: $Bias = E[\hat{\theta}] - \theta$. The true parameter value was estimated by performing an NMA with a big sample size (1 million) and then using the estimates from this NMA as the true parameter value.

- Empirical SE: Empirical SE is the dispersion measure of the estimator in a simulation study. It represents the precision of an estimator as well as its true variability. An estimator is expected to be with low variance when it is applied to multiple datasets. It can be defined as: $EmpSE = \sqrt{Var(\hat{\theta})}$.

- Model SE: In a simulation study, when a method is applied to multiple datasets, the measure of the average of the SE reported by the method is known as model SE. It can be defined as: $ModelSE = \sqrt{E[\hat{se}(\hat{\theta})^2]}$. It is desired that the empirical SE is small which shows that the estimator is precise and the model SE is equal to empirical SE.

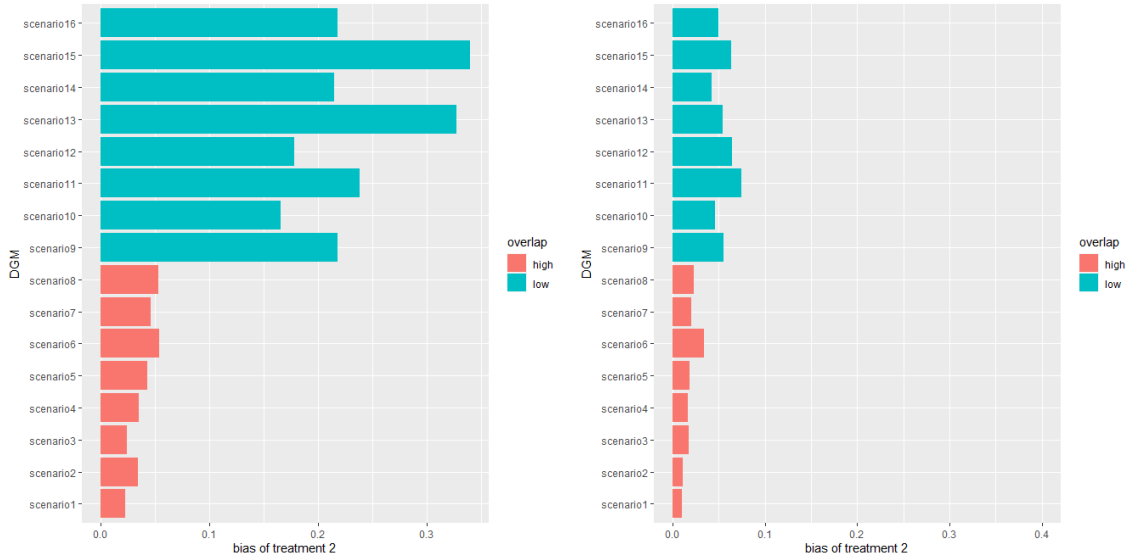
- Coverage probability: In a simulation study, coverage probability refers to the statistical technique where a percentage/proportion is calculated which shows how many confidence intervals include the true parameter value which is expected to be at (100×(1-α))% nominal level. It is common to fix the value of α at 0.05 i.e. at 95%.

6.3.5 Results for the simulation study of fixed effect NMA with double-bootstrapping

The simulation study simulated 1000 MCMC samples for each DGM. In each MCMC sample, double-bootstrapping was implemented to calculate MAIC estimates which were then used to perform a fixed effect NMA. This whole process was again bootstrapped 300 times to get bootstrapped NMA estimates.

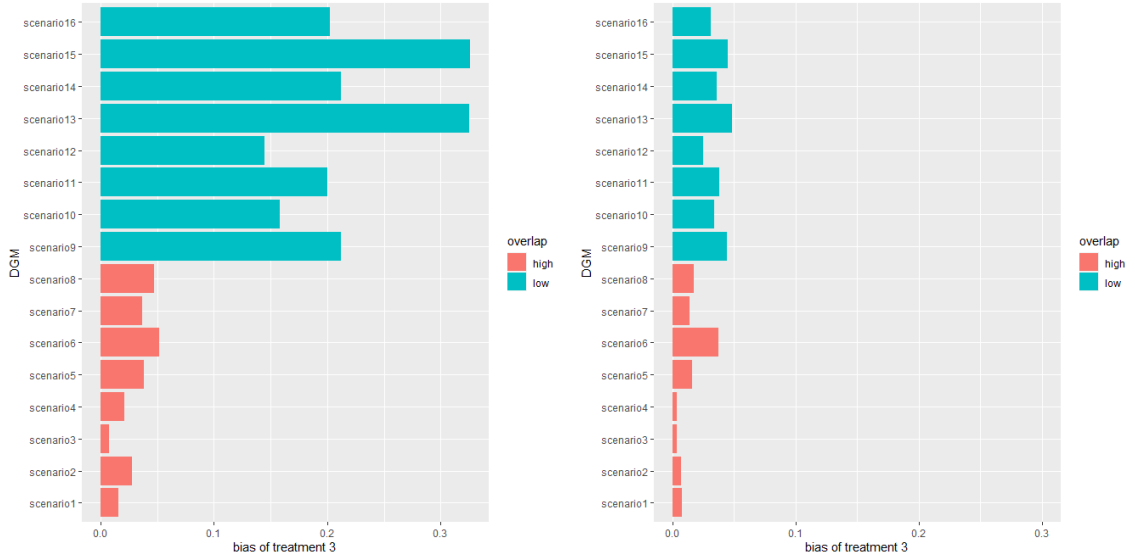
6.3.5.1 Bias by overlap

Figure 6.2 and 6.3 show the amount of bias in each DGM for treatments 2 and 3 respectively. The amount of bias was found to be related to the amount of overlap. In Figure 6.2, 6.3, large biases can be seen for low overlap scenarios and vice-versa. The magnitude of biases was higher for the sample size of 150. Overall, higher biases were found for scenarios 9, 11, 13 and 15. Other than the common fact that overlap was low in these scenarios, the correlation coefficient for study covariates was also low.



(a) *bias by overlap for sample size 150* **(b)** *bias by overlap for sample size 500*

Figure 6.2: *Bias of treatment 2 for different sample sizes (disconnected larger network of evidence)*



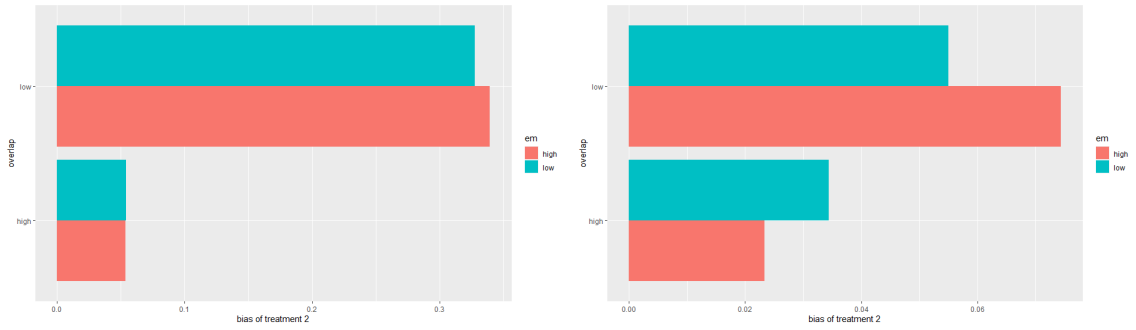
(a) *bias by overlap for sample size 150* **(b)** *bias by overlap for sample size 500*

Figure 6.3: *Bias of treatment 3 for different sample sizes (disconnected larger network of evidence)*

6.3.5.2 Bias by overlap with effect-modifiers

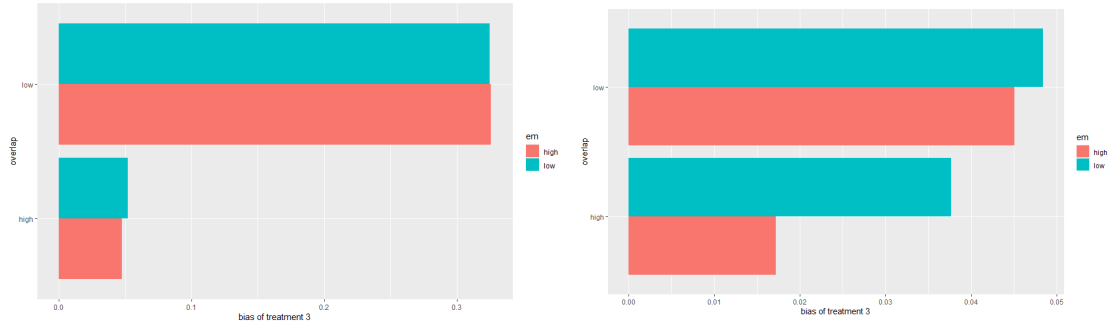
Figure 6.4 and 6.5 show the amount of bias by overlap for different levels of effect-modifier for treatments 2 and 3 respectively. Biases were found low with high overlap scenarios, but

within each level of overlap, there seems to be no effect of the effect-modifier. In Figure 6.4 (a) and 6.5 (a), for the sample size 150, bias was found to be unaffected by effect-modifier for both treatments 2 and 3. However, for the sample size of 500, most of the time, bias was found to be higher with a low effect-modifier level.



(a) *bias by EM for sample size 150* **(b)** *bias by EM for sample size 500*

Figure 6.4: *Bias by overlap with EM levels for treatment 2 (disconnected larger network of evidence)*

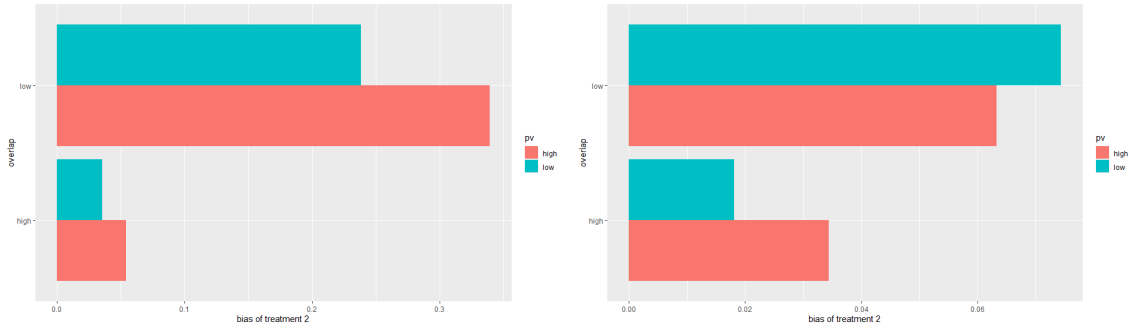


(a) *bias by EM for sample size 150* **(b)** *bias by EM for sample size 500*

Figure 6.5: *Bias by overlap with EM levels for treatment 3 (disconnected larger network of evidence)*

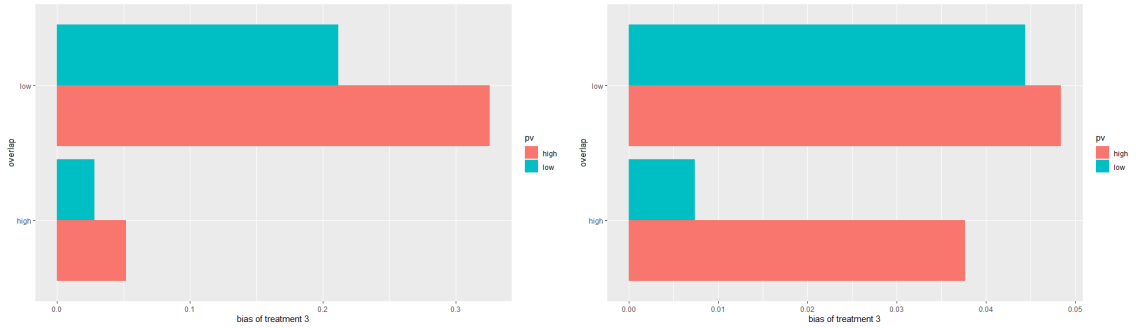
6.3.5.3 Bias by overlap with prognostic variable

Figure 6.6 and 6.7 show the amount of bias by overlap for different levels of prognostic variables for treatments 2 and 3 respectively. Mostly, bias was found to be higher with a high prognostic variable level.



(a) *bias by PV for sample size 150* **(b)** *bias by PV for sample size 500*

Figure 6.6: *Bias by overlap with PV levels for treatment 2 (disconnected larger network of evidence)*



(a) *bias by PV for sample size 150* **(b)** *bias by PV for sample size 500*

Figure 6.7: *Bias by overlap with PV levels for treatment 3 (disconnected larger network of evidence)*

6.3.5.4 Empirical SE and model SE by simulation scenarios

The simulation results of performance measures for the larger disconnected network of evidence are summarised in Table 6.1 and Table 6.2 for sample sizes 150 and 500 respectively. Table 6.1 and Table 6.2 show the relative estimates of treatments 2 and 3 with treatment 1 from MAIC-adjusted NMA with double-bootstrapping. The overall relative estimates of treatment 2 with the new intervention treatment 1 is termed as d1 and that of treatment 3 with treatment 1 is termed as d2. The empirical SE and model SE were similar or close to each other for each DGM. A quantity was estimated that calculates the difference between the empirical SE and model SE. The highest value of this difference was found to be 0.16 for the sample size of 150. From the colour-coding, it is evident that none of the coverage deviates from the nominal level. However, the moderate and higher biases were seen mainly for low-overlap scenarios.

Table 6.1: *Performance measure estimates of a larger disconnected network of evidence with double-bootstrapping for sample size 150 (fixed effect)*

DGM	COVERAGE d1	BIAS d1	EMP SE d1	MODEL SE d1	DIFFERENCE d1	COVERAGE d2	BIAS d2	EMP SE d2	MODEL SE d2	DIFFERENCE d2
Scenario 1	0.951	-0.02268	0.243313	0.245023	-0.00171	0.961	-0.01591	0.238351	0.250357	-0.01201
Scenario 2	0.956	-0.03452	0.262177	0.268789	-0.00661	0.961	-0.02814	0.259935	0.27516	-0.01522
Scenario 3	0.949	-0.02465	0.248301	0.246097	0.002204	0.952	-0.00784	0.241879	0.251124	-0.00924
Scenario 4	0.952	-0.03523	0.266669	0.269228	-0.00256	0.961	-0.02109	0.259375	0.275305	-0.01593
Scenario 5	0.955	-0.04307	0.297756	0.317185	-0.01943	0.953	-0.03852	0.284051	0.326017	-0.04197
Scenario 6	0.96	-0.05389	0.311153	0.328236	-0.01708	0.961	-0.05172	0.297674	0.334869	-0.03719
Scenario 7	0.961	-0.04656	0.30115	0.31224	-0.01109	0.961	-0.03704	0.287514	0.323619	-0.03611
Scenario 8	0.954	-0.05352	0.313554	0.325549	-0.012	0.96	-0.04756	0.300974	0.334618	-0.03364
Scenario 9	0.96	-0.21777	0.569565	0.689003	-0.11944	0.954	-0.2119	0.568027	0.714053	-0.14603
Scenario 10	0.955	-0.16587	0.49609	0.56744	-0.07135	0.955	-0.15804	0.504685	0.586434	-0.08175
Scenario 11	0.954	-0.23829	0.572838	0.69369	-0.12085	0.957	-0.20038	0.566344	0.707056	-0.14071
Scenario 12	0.953	-0.17831	0.500164	0.564827	-0.06466	0.954	-0.14492	0.506054	0.588992	-0.08294
Scenario 13	0.959	-0.32754	0.707031	0.85726	-0.15023	0.959	-0.32508	0.713614	0.878731	-0.16512
Scenario 14	0.961	-0.21481	0.575214	0.681715	-0.1065	0.962	-0.21225	0.582929	0.709655	-0.12673
Scenario 15	0.953	-0.33957	0.70754	0.86036	-0.15282	0.957	-0.32591	0.72548	0.883803	-0.15832
Scenario 16	0.957	-0.21813	0.579594	0.677165	-0.09757	0.952	-0.20239	0.581086	0.702473	-0.12139

Table 6.2: *Performance measure estimates of a larger disconnected network of evidence with double-bootstrapping for sample size 500 (fixed effect)*

DGM	COVERAGE d1	BIAS d1	EMP SE d1	MODEL SE d1	DIFFERENCE d1	COVERAGE d2	BIAS d2	EMP SE d2	MODEL SE d2	DIFFERENCE d2
Scenario 1	0.94	-0.01051	0.136679	0.128732	0.007947	0.957	-0.00739	0.127531	0.13157	-0.00404
Scenario 2	0.94	-0.01163	0.143482	0.13951	0.003972	0.956	-0.00716	0.135659	0.142875	-0.00722
Scenario 3	0.94	-0.0181	0.138908	0.128906	0.010003	0.954	-0.00345	0.129766	0.131644	-0.00188
Scenario 4	0.94	-0.01652	0.145927	0.139526	0.006401	0.95	-0.0036	0.136332	0.142571	-0.00624
Scenario 5	0.941	-0.01879	0.159996	0.155447	0.004549	0.95	-0.01633	0.151193	0.158742	-0.00755
Scenario 6	0.95	-0.03442	0.252392	0.243951	0.008442	0.96	-0.03765	0.240952	0.251243	-0.01029
Scenario 7	0.94	-0.02084	0.160996	0.155085	0.00591	0.95	-0.01391	0.152975	0.158684	-0.00571
Scenario 8	0.943	-0.02338	0.164215	0.161706	0.00251	0.958	-0.01719	0.158237	0.165179	-0.00694
Scenario 9	0.94	-0.05501	0.239335	0.247703	-0.00837	0.95	-0.0444	0.235402	0.252132	-0.01673
Scenario 10	0.951	-0.04613	0.215126	0.225748	-0.01062	0.954	-0.03391	0.21283	0.23128	-0.01845
Scenario 11	0.94	-0.07441	0.240825	0.247749	-0.00692	0.954	-0.03797	0.237974	0.25233	-0.01436
Scenario 12	0.94	-0.06433	0.216142	0.225822	-0.00968	0.955	-0.02485	0.213546	0.231353	-0.01781
Scenario 13	0.958	-0.054	0.254727	0.277181	-0.02245	0.956	-0.04844	0.254893	0.282743	-0.02785
Scenario 14	0.945	-0.04229	0.234118	0.242962	-0.00884	0.953	-0.03611	0.235193	0.249144	-0.01395
Scenario 15	0.953	-0.06328	0.256217	0.277078	-0.02086	0.953	-0.04507	0.254523	0.282822	-0.0283
Scenario 16	0.953	-0.04975	0.233985	0.243402	-0.00942	0.953	-0.031	0.235903	0.248672	-0.01277

6.3.5.5 Coverage by simulation scenarios

Figure 6.8 and 6.9 show the CIs of the coverage estimates for varying levels of overlap and sample sizes for treatments 2 and 3 respectively. The coverage of d1 and d2 from the double-bootstrapped MAIC-adjusted NMA was at the nominal level for both sample sizes. Red-coloured figures represent CIs for the high overlap scenarios and blue-coloured figures represent CIs for the low overlap scenarios. In the figures, the black horizontal line represents the nominal level of coverage. From the figure, it can be seen that the CIs touch the nominal level for all 16 scenarios irrespective of sample sizes and overlap.

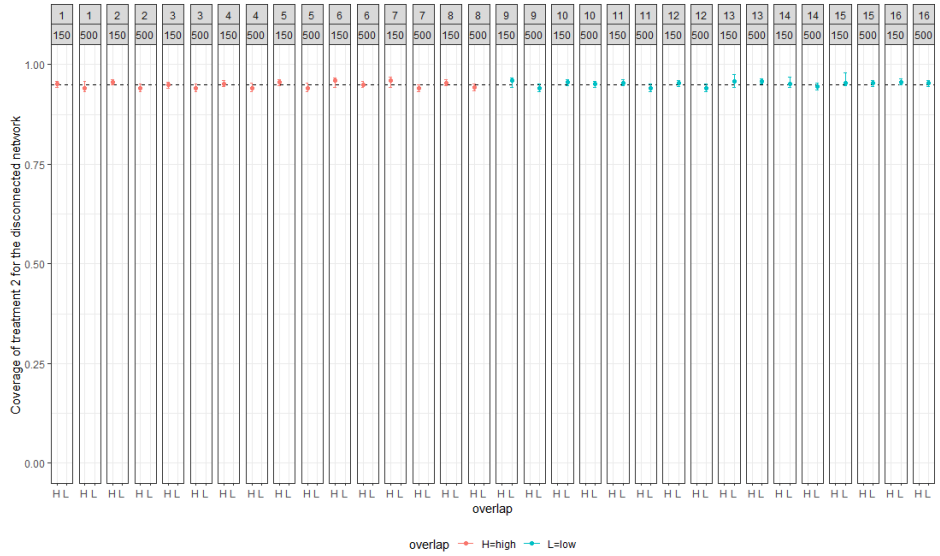


Figure 6.8: *Coverage by overlap for treatment 2*

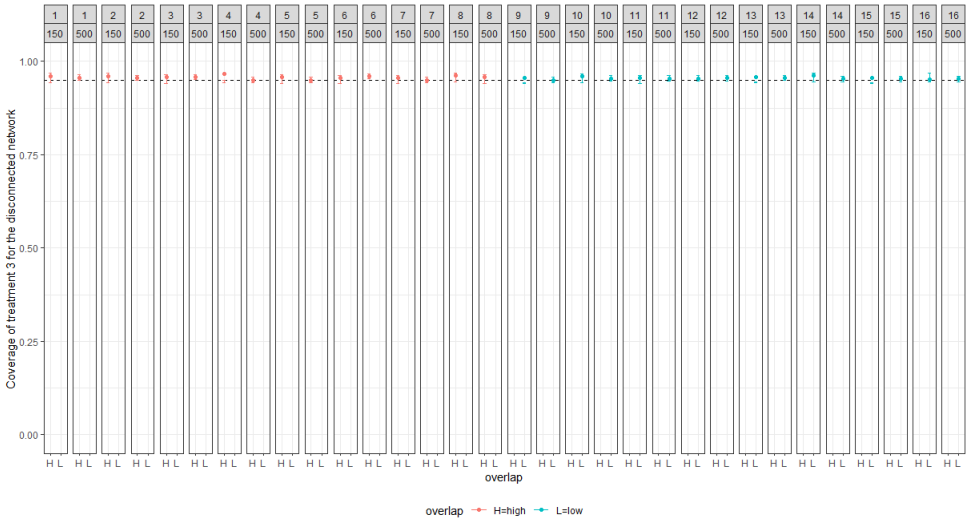
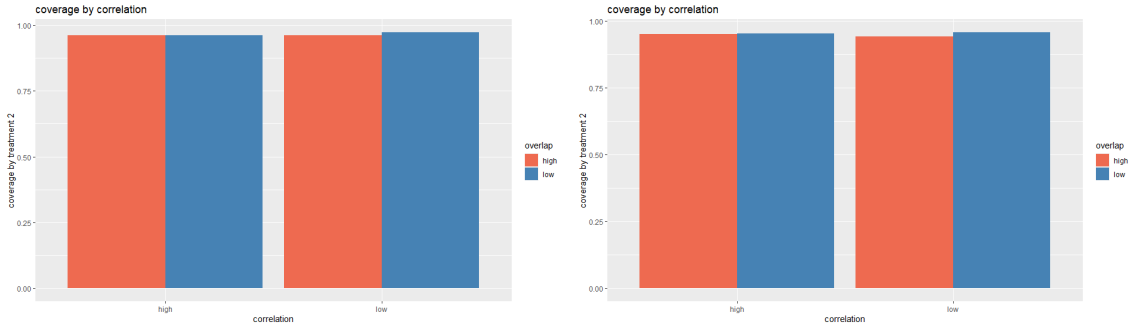


Figure 6.9: *Coverage by overlap for treatment 3*

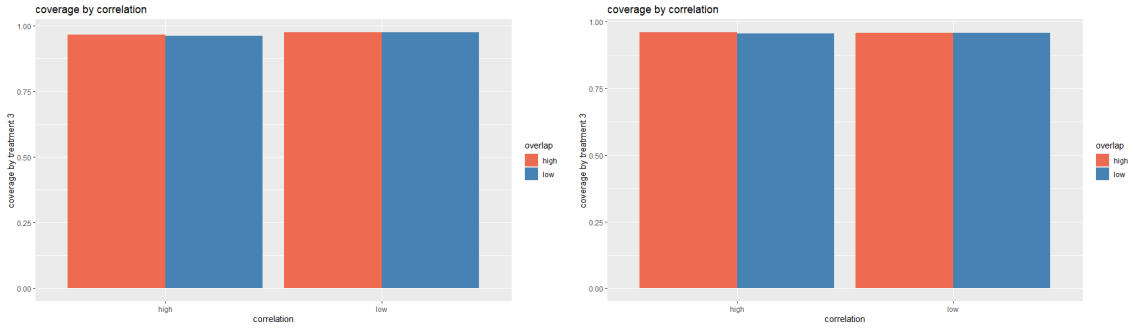
6.3.5.6 Coverage by correlation with overlap

Figure 6.10 and Figure 6.11 show the coverage with different levels of overlap and correlation for treatments 2 and 3 respectively. From the figures, it can be seen that the coverage is not found to be affected by different levels of correlation and overlap.



(a) *coverage by correlation for sample size 150* **(b)** *coverage by correlation for sample size 500*

Figure 6.10: *Coverage by correlation of treatment 2 for different sample sizes (disconnected larger network of evidence)*

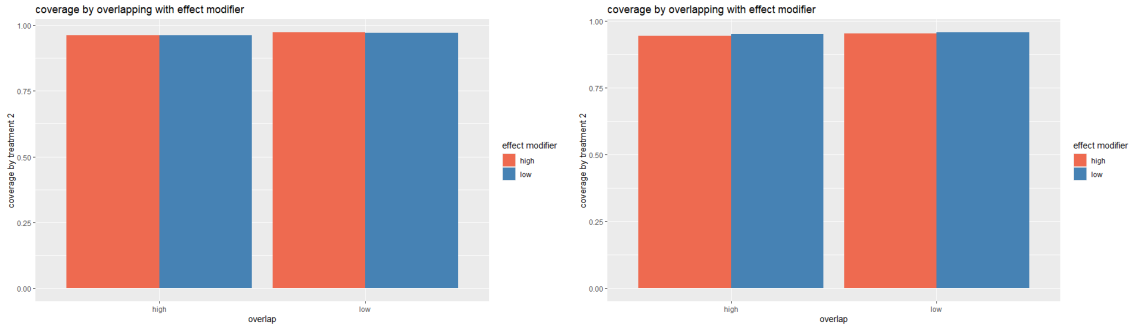


(a) *coverage by correlation for sample size 150* **(b)** *coverage by correlation for sample size 500*

Figure 6.11: *Coverage by correlation of treatment 3 for different sample sizes (disconnected larger network of evidence)*

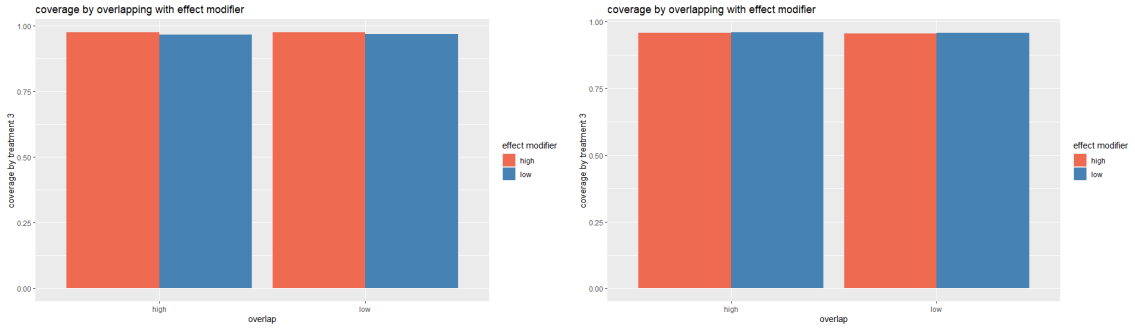
6.3.5.7 Coverage by overlap with effect-modifiers

Figure 6.12 and 6.13 show the coverage by different levels of overlap and effect-modifier for treatments 2 and 3 respectively. For both sample sizes and treatments, the coverage was similar with varying levels of overlap and effect-modifier variable.



(a) *coverage with EM for sample size 150* **(b)** *coverage with EM for sample size 500*

Figure 6.12: *Coverage by overlap with EM levels for treatment 2 (disconnected larger network of evidence)*

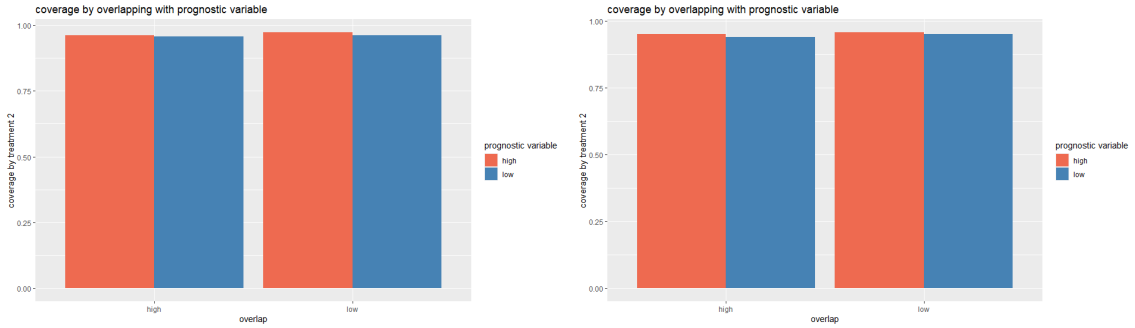


(a) *coverage with EM for sample size 150* **(b)** *coverage with EM for sample size 500*

Figure 6.13: *Coverage by overlap with EM levels for treatment 3 (disconnected larger network of evidence)*

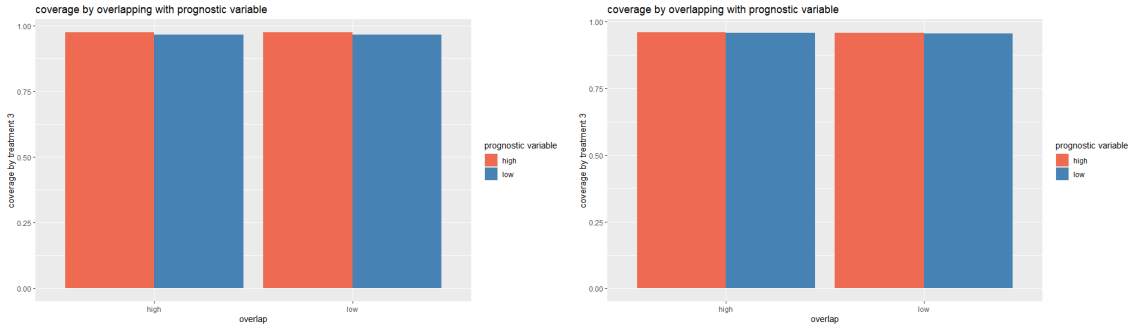
6.3.5.8 Coverage by overlap with prognostic variable

Figure 6.14 and 6.15 show the amount of coverage by different levels of overlap and prognostic variables for treatments 2 and 3 respectively. For both sample sizes and treatments, the coverage was found to be similar with varying levels of overlap and prognostic variables.



(a) *coverage by PV for sample size 150* **(b)** *coverage by PV for sample size 500*

Figure 6.14: *Coverage by overlap with PV levels of treatment 2 (disconnected larger network of evidence)*



(a) *coverage by PV for sample size 150* **(b)** *coverage by PV for sample size 500*

Figure 6.15: *Coverage by overlap with PV levels of treatment 3 (disconnected larger network of evidence)*

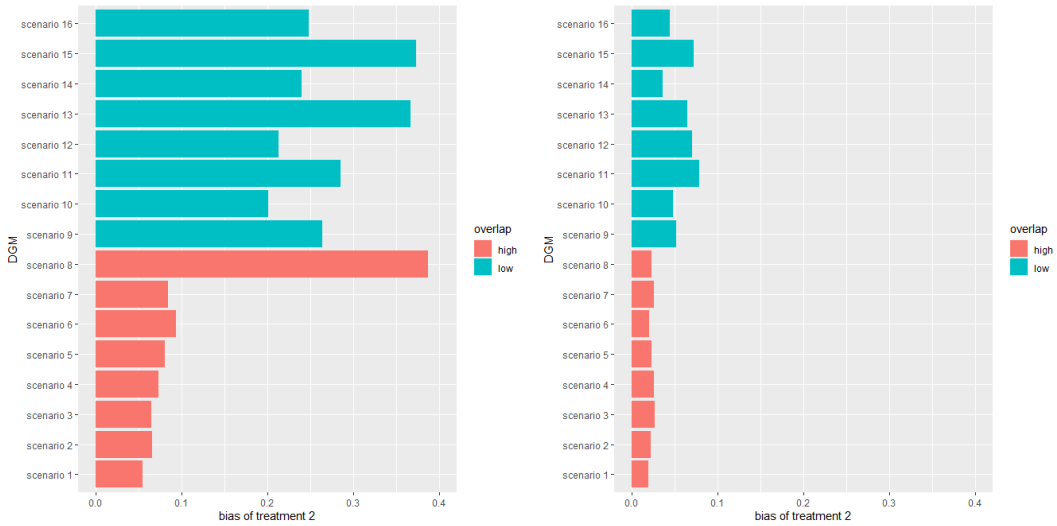
6.3.6 Results for the simulation study of random effects NMA with double-bootstrapping

The simulation study simulated 1000 MCMC samples for each DGM. In each MCMC sample, double-bootstrapping was implemented to calculate MAIC estimates. The MAIC estimates were then used to perform random effects NMA. This whole process was again bootstrapped 1000 times to get bootstrapped NMA estimates. The number of bootstraps for the whole process was bigger for random effects NMA (1000) compared to the fixed effect (300) as in fixed effect NMA, the estimates quickly stabilise to nominal level with fewer number of bootstraps.

6.3.6.1 Bias by overlap

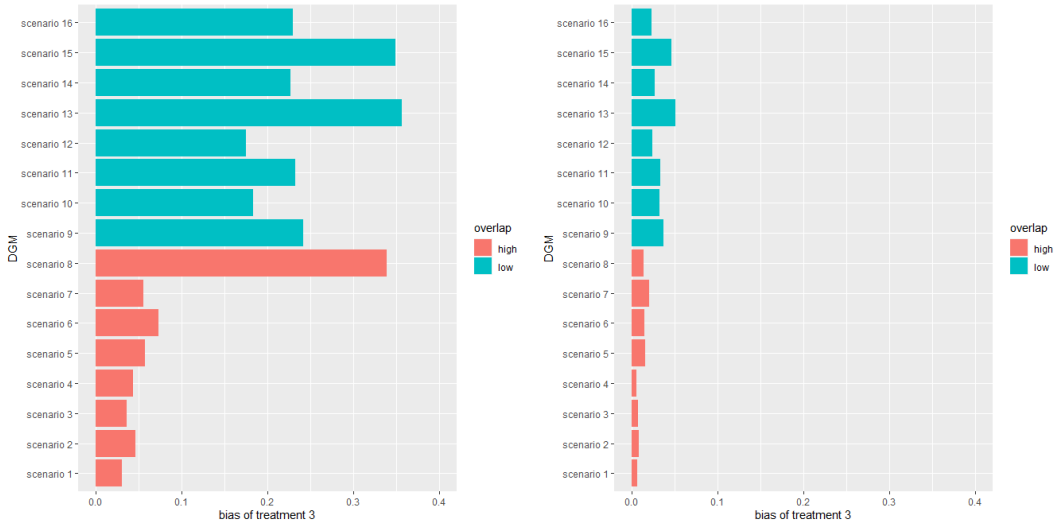
Figure 6.16 and 6.17 show the amount of bias in each DGM for treatments 2 and 3 respectively. Scenarios 1 to 8 are from the high-overlap scenarios whereas scenarios 9 to 15 are from low-overlap. The amount of bias was found to be related to the amount of overlap. In Figure 6.16, 6.17, large biases can be seen for low overlap scenarios and vice-versa. Though bias was

low with high overlap, a bigger bias was found for scenario 8 in sample size 150. Overall, higher biases were found for scenarios 8, 12 and 15. In all these scenarios, the effect-modifier coefficient for study covariates was high. However, the magnitude of biases was low for the sample size of 500.



(a) *bias by overlap for sample size 150* **(b)** *bias by overlap for sample size 500*

Figure 6.16: *Bias of treatment 2 for different sample sizes (disconnected larger network of evidence)*

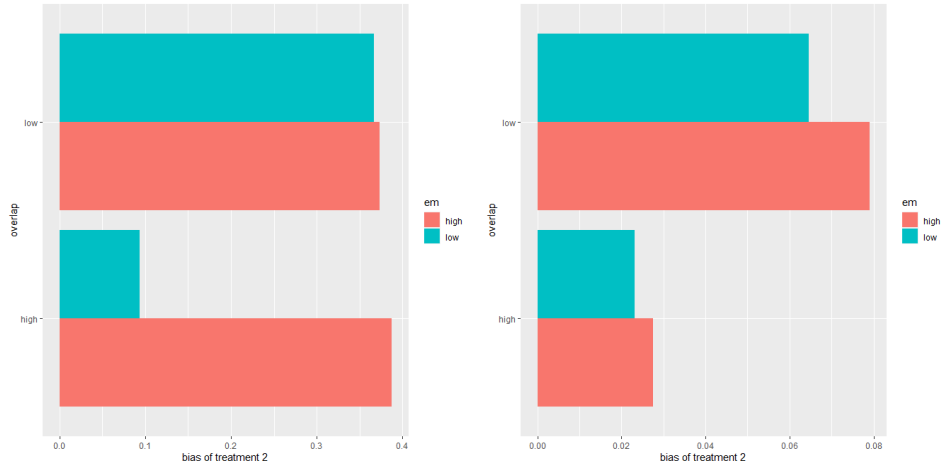


(a) *bias by overlap for sample size 150* **(b)** *bias by overlap for sample size 500*

Figure 6.17: *Bias of treatment 3 for different sample sizes (disconnected larger network of evidence)*

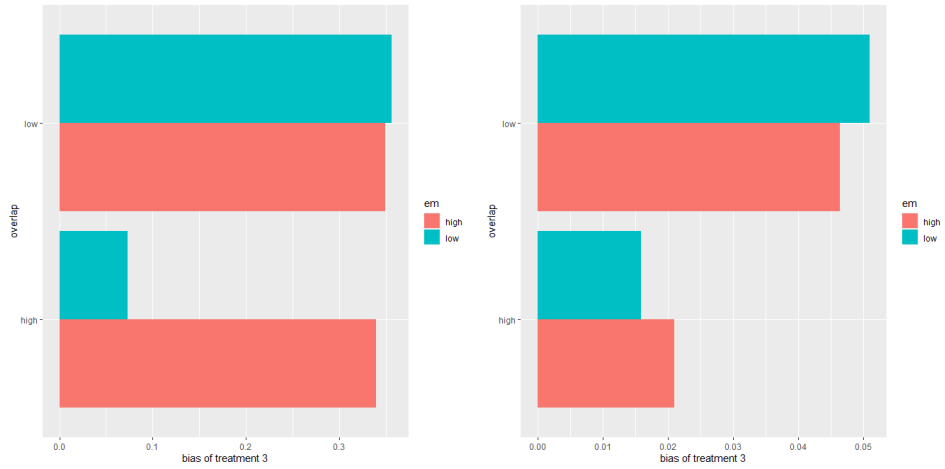
6.3.6.2 Bias by overlap with effect-modifiers

Figure 6.18 and 6.19 show the amount of bias by overlap for different levels of effect-modifier for treatments 2 and 3 respectively. Overall, low biases were mainly found to be associated with high-overlap scenarios. However, for low overlap scenarios, biases were low with a low effect-modifier level only for the sample size of 500. In Figure 6.18 and 6.19, with high overlap scenarios biases were comparatively low with low effect-modifier level.



(a) *bias by EM for sample size 150* **(b)** *bias by EM for sample size 500*

Figure 6.18: *Bias by overlap with EM levels for treatment 2 (disconnected larger network of evidence)*

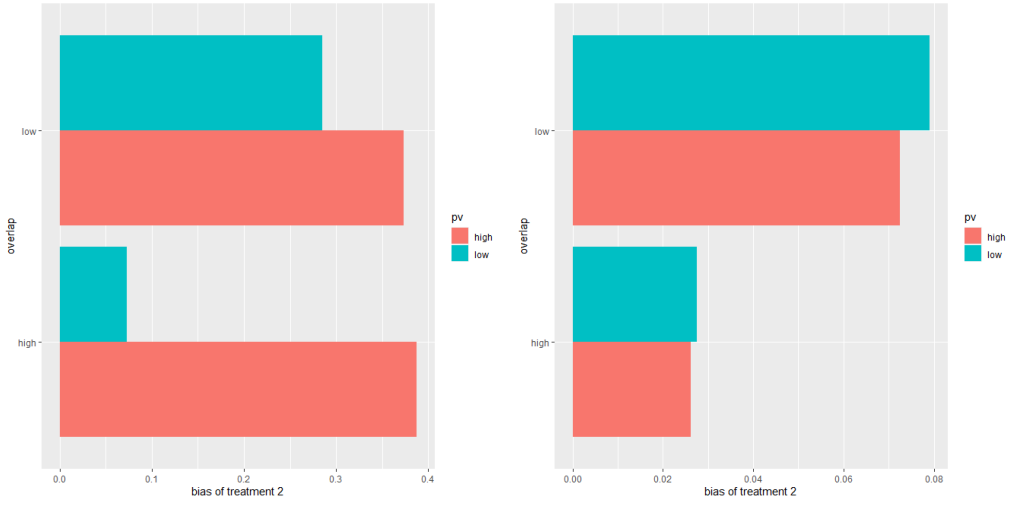


(a) *bias by EM for sample size 150* **(b)** *bias by EM for sample size 500*

Figure 6.19: *Bias by overlap with EM levels for treatment 3 (disconnected larger network of evidence)*

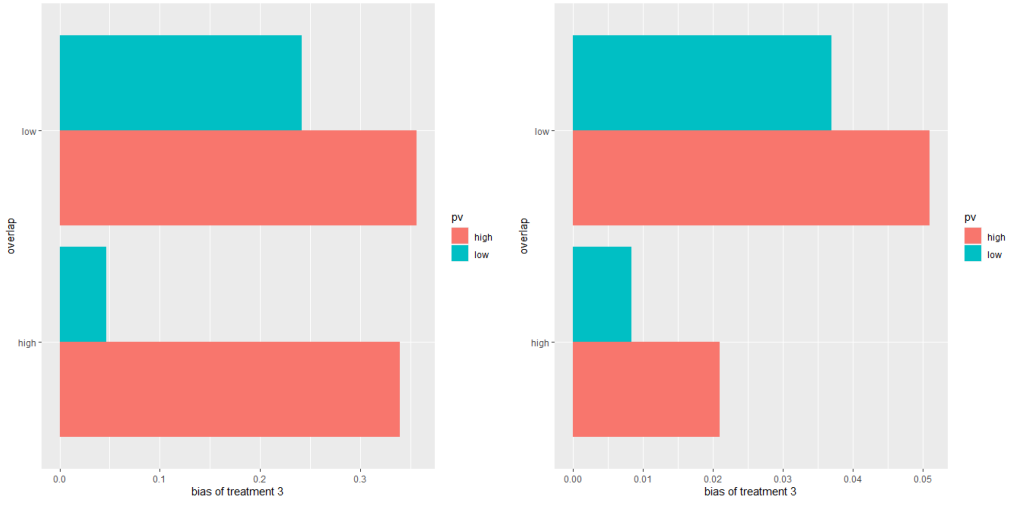
6.3.6.3 Bias by overlap with prognostic variable

Figure 6.20 and 6.21 show the amount of bias by overlap for different levels of prognostic variables for treatments 2 and 3 respectively. Mostly, bias was found to be higher with a high prognostic variable level.



(a) *bias by PV for sample size 150* **(b)** *bias by PV for sample size 500*

Figure 6.20: *Bias by overlap with PV levels for treatment 2 (disconnected larger network of evidence)*



(a) *bias by PV for sample size 150* **(b)** *bias by PV for sample size 500*

Figure 6.21: *Bias by overlap with PV levels for treatment 3 (disconnected larger network of evidence)*

6.3.6.4 Empirical SE and model SE by simulation scenarios

The simulation results of performance measures for the larger disconnected network of evidence are summarised in Table 6.3 and Table 6.4 for sample sizes 150 and 500 respectively. Both the tables show the relative estimates of treatments 2 and 3 with treatment 1 from MAIC-adjusted NMA with double-bootstrapping. From the tables, it can be seen that the coverage of d1 and d2 from the double-bootstrapped MAIC-adjusted NMA was at the nominal level. The empirical SE and model SE were similar or close to each other for each DGM. Similar to fixed effect model, in the random effect model, none of the coverage deviates from the nominal level. However, moderate and large biases were seen mainly for the sample size 159 (small sample). The magnitude of biases was low for the sample size of 500.

Table 6.3: *Performance measure estimates of a larger disconnected network of evidence with double-bootstrapping for sample size 150 (random effects)*

DGM	COVERAGE d1	BIAS d1	EMP SE d1	MODEL SE d1	DIFFERENCE d1	COVERAGE d2	BIAS d2	EMP SE d2	MODEL SE d2	DIFFERENC d2
Scenario 1	0.95043	-0.05532	0.277479	0.277187	0.030292	0.952527	-0.0307	0.287718	0.283661	0.004057
Scenario 2	0.954141	-0.0662	0.298244	0.272099	0.026145	0.959293	-0.04687	0.305833	0.290614	0.015219
Scenario 3	0.950714	-0.06531	0.280203	0.287822	0.032381	0.952619	-0.03668	0.285273	0.28406	0.001213
Scenario 4	0.950833	-0.07336	0.295896	0.272119	0.023781	0.951792	-0.0439	0.303624	0.280528	0.023095
Scenario 5	0.953535	-0.08103	0.325266	0.310585	0.014681	0.951576	-0.05732	0.323738	0.318348	0.00539
Scenario 6	0.945	-0.09353	0.345591	0.32803	0.017561	0.953	-0.07301	0.345345	0.33829	0.007056
Scenario 7	0.952083	-0.08486	0.327492	0.312392	0.0151	0.95625	-0.05594	0.323445	0.319766	0.003679
Scenario 8	0.958	-0.38736	0.694798	0.70114	-0.00634	0.956634	-0.33937	0.684558	0.70231	-0.01775
Scenario 9	0.955451	-0.2639	0.598786	0.692965	-0.09418	0.952549	-0.24156	0.596517	0.598211	-0.00169
Scenario 10	0.955882	-0.20071	0.504254	0.568101	-0.06385	0.94902	-0.18341	0.519282	0.59311	-0.07383
Scenario 11	0.953333	-0.28484	0.599552	0.587045	0.012507	0.952222	-0.23241	0.595399	0.589138	0.006261
Scenario 12	0.950516	-0.21344	0.504969	0.513626	-0.05866	0.952577	-0.17482	0.522596	0.589071	-0.06648
Scenario 13	0.953737	-0.36702	0.697248	0.700998	-0.00375	0.951717	-0.35611	0.693865	0.700892	-0.00703
Scenario 14	0.950417	-0.23968	0.557497	0.5572	0.000297	0.9525	-0.22736	0.560309	0.586297	-0.02599
Scenario 15	0.95	-0.37286	0.697489	0.70324	-0.00575	0.951	-0.34914	0.682808	0.72311	-0.0403
Scenario 16	0.953792	-0.248	0.568649	0.570551	-0.0019	0.95275	-0.22931	0.571899	0.58117	-0.00927

Table 6.4: *Performance measure estimates of a larger disconnected network of evidence with double-bootstrapping for sample size 500 (random effects)*

DGM	COVERAGE d1	BIAS d1	EMP SE d1	MODEL SE d1	DIFFERENCE d1	COVERAGE d2	BIAS d2	EMP SE d2	MODEL SE d2	DIFFERENC d2
Scenario 1	0.950333	-0.02009	0.180259	0.189234	-0.00898	0.948	-0.00696	0.182814	0.181894	0.00092
Scenario 2	0.941667	-0.02282	0.189098	0.180937	0.008162	0.941667	-0.00842	0.191422	0.19413	-0.00271
Scenario 3	0.953667	-0.02756	0.179436	0.179371	6.50E-05	0.948333	-0.00768	0.181125	0.182066	-0.00094
Scenario 4	0.949667	-0.02638	0.184733	0.181226	0.003507	0.952667	-0.00545	0.189409	0.184585	0.004824
Scenario 5	0.952	-0.0231	0.196022	0.195219	0.000803	0.944	-0.01593	0.193258	0.198595	-0.00534
Scenario 6	0.956	-0.02092	0.202938	0.202434	0.000504	0.955333	-0.01539	0.203227	0.206336	-0.00311
Scenario 7	0.954444	-0.02622	0.201538	0.201714	-0.00018	0.953333	-0.021	0.19643	0.19483	0.0016
Scenario 8	0.954667	-0.02309	0.199738	0.192516	0.007222	0.961333	-0.01424	0.197032	0.196254	0.000778
Scenario 9	0.953333	-0.05177	0.268748	0.268398	0.00035	0.951333	-0.03697	0.272229	0.272803	-0.00057
Scenario 10	0.954333	-0.04889	0.243401	0.245853	-0.00245	0.956333	-0.03296	0.251959	0.250714	0.001244
Scenario 11	0.951889	-0.07908	0.25918	0.246188	0.012993	0.955556	-0.03349	0.258812	0.250779	0.008033
Scenario 12	0.95	-0.07036	0.240656	0.246379	-0.00572	0.951333	-0.02437	0.245495	0.231239	0.014256
Scenario 13	0.955714	-0.06464	0.277209	0.274999	0.002209	0.951667	-0.05099	0.280269	0.280273	-4.40E-06
Scenario 14	0.95954	-0.03624	0.260138	0.262226	-0.00209	0.957241	-0.0276	0.262932	0.26761	-0.00468
Scenario 15	0.954444	-0.07251	0.277605	0.274501	0.003104	0.956333	-0.04645	0.279523	0.279641	-0.00012
Scenario 16	0.955556	-0.04432	0.258297	0.243028	0.015269	0.95536	-0.02333	0.258534	0.248523	0.010012

6.3.6.5 Coverage by simulation scenarios

Figure 6.22 and 6.23 show the CI of the coverage estimates for varying levels of overlap and sample sizes for treatments 2 and 3 respectively. The coverage of d1 and d2 from the double-bootstrapped MAIC-adjusted random effects NMA was at the nominal level for both sample sizes. Red-coloured figures represent CIs for the high overlap scenarios and blue-coloured figures represent CIs for the low overlap scenarios. The black horizontal line represents the nominal level of coverage. From the figure, it can be seen that the CI touches the nominal level for all DGM irrespective of sample sizes and overlap.

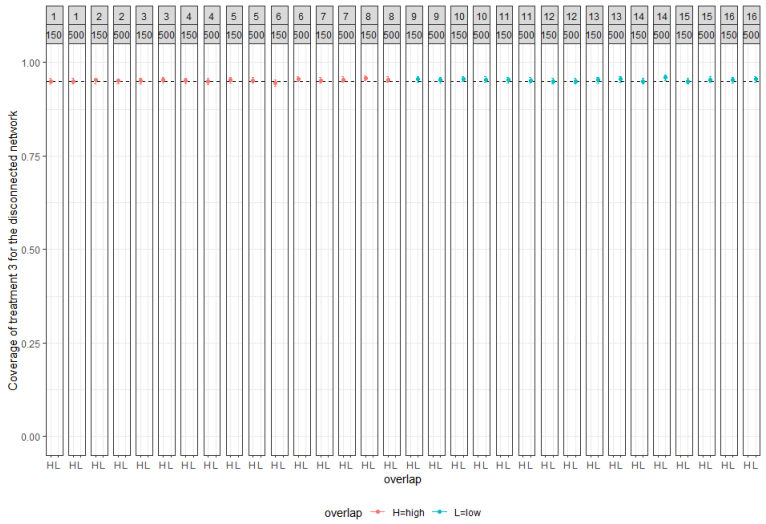


Figure 6.22: *Coverage by overlap for treatment 2*

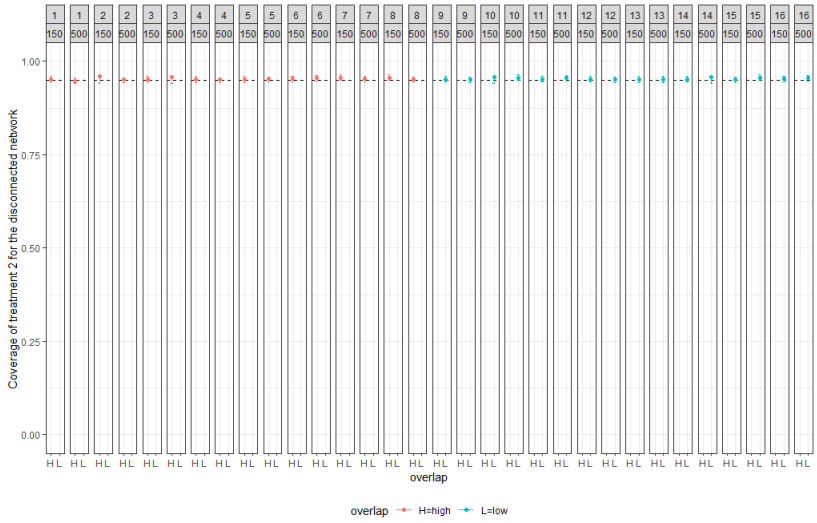
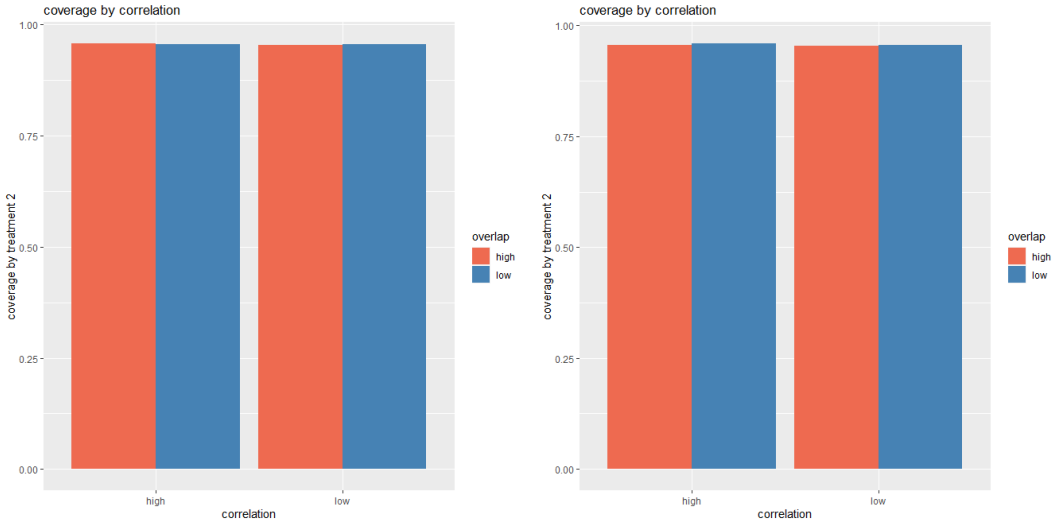


Figure 6.23: *Coverage by overlap for treatment 3*

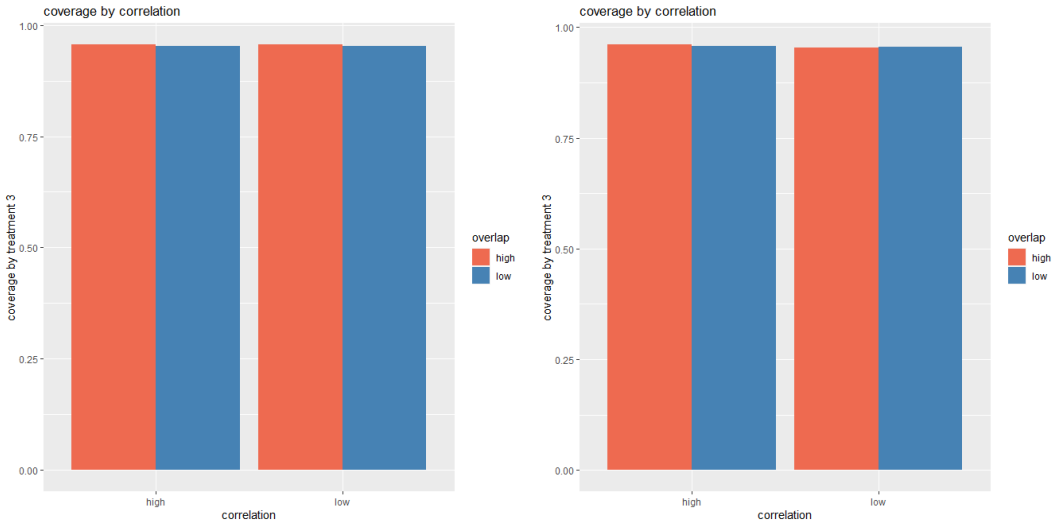
6.3.6.6 Coverage by correlation with overlap

Figure 6.24 and Figure 6.25 show the coverage with different levels of overlap and correlation for treatments 2 and 3 respectively. From the figures, it can be seen that the coverage was not found to be affected by different levels of correlation and overlap.



(a) *coverage by correlation for sample size 150* **(b)** *coverage by correlation for sample size 500*

Figure 6.24: *Coverage by correlation of treatment 2 for different sample sizes (disconnected larger network of evidence)*

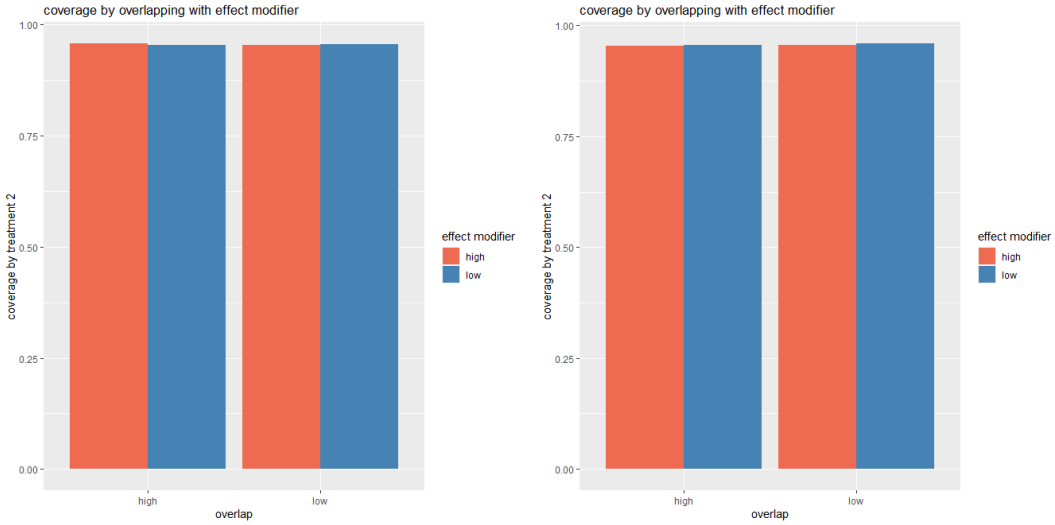


(a) *coverage by correlation for sample size 150* **(b)** *coverage by correlation for sample size 500*

Figure 6.25: *Coverage by correlation of treatment 3 for different sample sizes (disconnected larger network of evidence)*

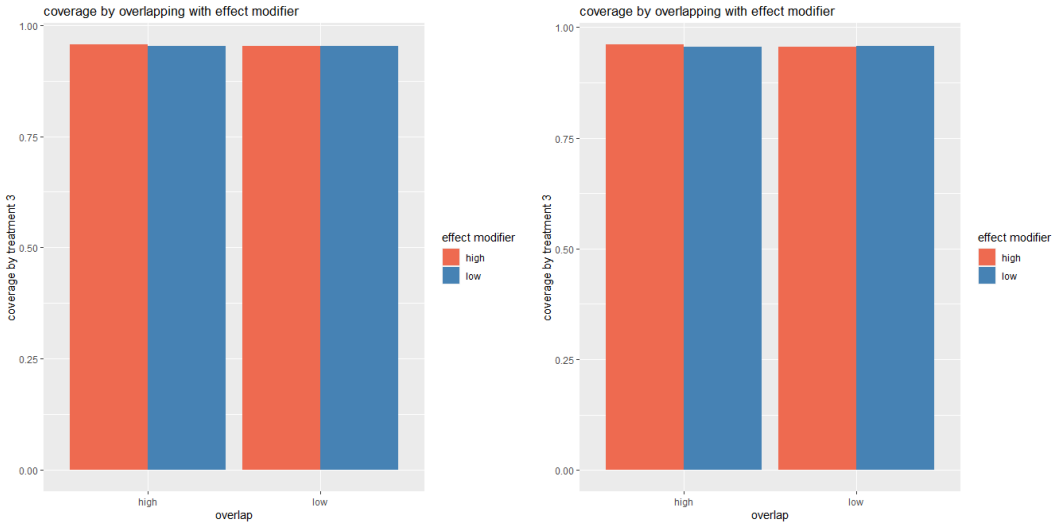
6.3.6.7 Coverage by overlap with effect-modifiers

Figure 6.26 and 6.27 show the coverage by different levels of overlap and effect-modifier for treatments 2 and 3 respectively. For both sample sizes and treatments, the coverage was similar with varying levels of overlap and effect-modifier variable.



(a) *coverage with EM for sample size 150* **(b)** *coverage with EM for sample size 500*

Figure 6.26: *Coverage by overlap with EM levels for treatment 2 (disconnected larger network of evidence)*

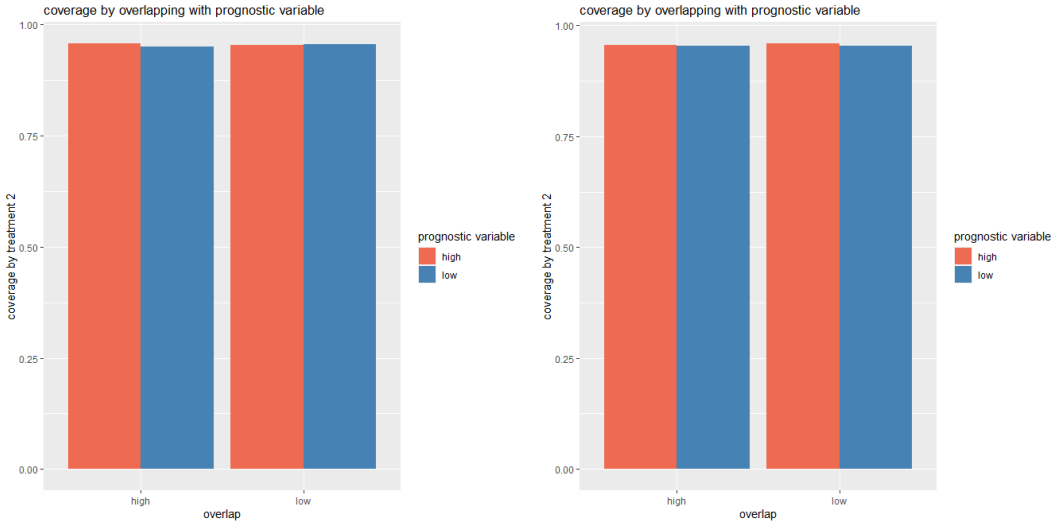


(a) *coverage with EM for sample size 150* **(b)** *coverage with EM for sample size 500*

Figure 6.27: *Coverage by overlap with EM levels for treatment 3 (disconnected larger network of evidence)*

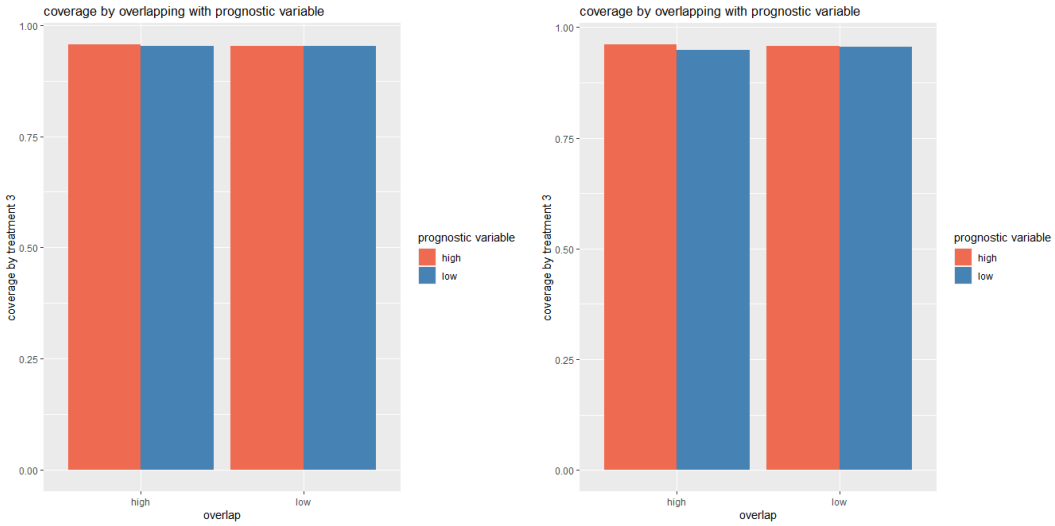
6.3.6.8 Coverage by overlap with prognostic variable

Figure 6.28 and 6.29 show the amount of coverage by different levels of overlap and prognostic variables for treatments 2 and 3 respectively. For both sample sizes and treatments, the coverage was found to be similar with varying levels of overlap and prognostic variables.



(a) *coverage by PV for sample size 150* **(b)** *coverage by PV for sample size 500*

Figure 6.28: *Coverage by overlap with PV levels of treatment 2 (disconnected larger network of evidence)*



(a) *coverage by PV for sample size 150* **(b)** *coverage by PV for sample size 500*

Figure 6.29: *Coverage by overlap with PV levels of treatment 3 (disconnected larger network of evidence)*

6.4 Discussion

Previous simulation chapters showed that when MAIC was implemented in a larger disconnected network of evidence to conduct a NMA, the most significant effect was seen in the coverage probability. Undercoverage issue was found for each DGM with a larger network of evidence. The repeated use of the same IPD in an NMA along with the sandwich estimator was causing this low coverage. In order to overcome this issue, instead of using the sandwich estimator, double-bootstrapping was used for the estimation of uncertainty in a fixed and random effects NMA.

Both for the double-bootstrapped MAIC-adjusted fixed and random effects NMA of the larger disconnected network of evidence, bias was found to be inversely related to overlap. Lower biases were found with high overlap levels and vice-versa. The general trend was higher bias with higher prognostic and effect-modifying variable levels. Moreover, the magnitude of biases was found to be lower for the sample size of 500 compared to the sample size 150. Double-bootstrapping was found to increase the coverage to the nominal (95%) level both for fixed and random effects NMA. Additionally, though bias was affected mainly due to overlap between study covariates as well as prognostic and effect-modifying variables, coverage was found to be unaffected by these factors.

The highest biases were found specifically for the sample size 150 with low overlap scenarios both for the fixed and random effects NMA. Nevertheless, the magnitude of bias with small sample size and low overlap was found to be higher for the random effects NMA model in comparison to the fixed effect NMA. Therefore, for a bigger sample size, low biases can be found with double-bootstrapping but with a smaller sample size and low overlap, biases can increase significantly.

Double-bootstrapping was found to solve the problem of undercoverage. However, for the disconnected network of fixed and random effects NMA, the biases were found to be comparatively high with low-overlap scenarios with a smaller sample size. In MAIC, the adjustment of covariates is made for the AgD study population which means the covariate distribution in the AgD study needs to be covered by the IPD study. When overlap is poor between IPD and AgD study, MAIC will generate large weights and may fail to produce a valid estimate. Therefore, in the double-bootstrap, the amount of bias increases with low overlap scenarios. For the random effects NMA, biases were found to be high in comparison to the fixed effect NMA. Moreover, though double-bootstrapping solves the undercoverage issue, the main difficulty found to be computational issue. Double-bootstrapping was very resource-intensive as well as time-consuming.

Despite the fact that the population-adjustment method MAIC is able to mitigate the difference between studies when prognostic and effect-modifier variables are imbalanced, it is not generalisable to larger networks of treatments without additional assumptions. Moreover, in MAIC, the adjustment is made for the AgD study population which is often not the population of interest. Additionally, in practice the correlation information of the covariates in the AgD study is unavailable, therefore, MAIC assumes that the joint covariate distribution in the AgD study is the product of the published marginal distributions. Furthermore,

the propensity score model that is used to estimate weights, is also sensitive to model misspecification bias even though the propensity score model does not make reference to the outcome. A newly developed method called ML-NMR is a recent addition to the population-adjustment method that is able to handle a larger network of evidence. ML-NMR is an extension of the standard NMA framework to combine IPD with AgD. It fits a regression model using the studies that have IPD available and then integrates the regression model over the AgD studies. ML-NMR is able to overcome the aggregation bias which is typically found in conventional NMA as well as can produce treatment estimates in any target population. Although ML-NMR adjusts for differences in the covariates between populations in estimating the relative treatment effects of multiple treatments simultaneously, at present, this method is suitable for a connected network of evidence. Currently, ML-NMR is not able to include single-arm studies with a larger disconnected network of evidence.

In the next chapter, a dataset from a case study will be used in order to show the practical application of MAIC-adjusted NMA with and without double-bootstrapping.

Chapter 7

Double-Bootstrapping with a Case Study

7.1 Introduction

In the simulation chapters (Chapters 4 and 5), it was found that undercoverage of the relative treatment effect estimates was one of the substantial consequences of a MAIC-adjusted NMA. The problem of undercoverage gets more serious for the larger disconnected network of evidence compared to a smaller disconnected network of evidence as undercoverage was seen for all DGM. Moreover, the magnitude of undercoverage was also found to be higher for the larger disconnected network of evidence. The reason for this lies in the repeated use of IPD. In a larger disconnected network of evidence, to apply a MAIC-adjusted NMA, IPD from the single-arm study was used more times compared to a smaller disconnected network of evidence. The assumption of independence between studies was ignored which results in an unappraised correlation between studies.

A double-bootstrapped approach was suggested in the previous chapter to amend this undercoverage issue. A double-bootstrapped approach which is an augmentation of a single bootstrap CI was used to increase the coverage probability of NMA estimates to the nominal level. The use of double-bootstrapping breaks the dependency between the studies. For the fixed effect model, though the coverage reach the nominal level (95%) for all DGM, however, there was an increase in bias found for the low overlap scenarios with the sample size 150. Nevertheless, with the sample size of 500, the bias was considerably lower compared to the sample size of 150.

In this chapter, double-bootstrapping has been implemented with a case study. The data for the case study comes from Dose Ranging Efficacy And Safety with Mepolizumab (DREAM) study on severe eosinophilic asthma exacerbation that is an RCT conducted by GlaxoSmithKline (GSK) Pharmaceuticals (Pavord et al., 2012). The objective of this chapter is to show how double-bootstrapping can be executed in a real-data setting. For this, first, a connected NMA was conducted with all the available biologics for severe asthma, and then the connected network of evidence was made disconnected artificially so that the unanchored MAICs could be carried out. The IPD from the DREAM study was used to implement multiple MAICs.

A MAIC-adjusted NMA as well as a double-bootstrapped MAIC-adjusted NMA were then performed.

Sections 7.2 and 7.3 of this chapter describe the case study and aim of the current study respectively. Section 7.4 illustrates the methods that were implemented and Section 7.5 summarize the results. Section 7.6 compares the results whereas Section 7.7 makes a discussion of the results. The chapter concludes by stating the strengths and limitations of the study in Section 7.8.

7.2 Description of the case study

7.2.1 Population (severe eosinophilic asthma)

Asthma is a common, non-communicable chronic disorder. It differs from other chronic diseases as it causes a health condition that needs prolonged treatment. An asthma patient suffers from sensitive airways that contract and become narrow when they got irritated. The irritated airways then cause coughing or wheezing (Henriksen et al., 2020). Asthma is a non-curable but manageable disease with the help of precaution and proper treatment. Although there have been huge advances in controlling chronic childhood asthma, acute exacerbations remain a common occurrence during viral seasons. Even though acute asthma has a low mortality rate, the condition causes a large healthcare burden because of its frequency. Among asthma patients, around 5%-10% suffer from severe asthma which can require both high doses of inhaled corticosteroids and oral corticosteroids (Charles et al., 2022).

A severe asthma patient is defined as someone who requires treatment with high-dose inhaled corticosteroids (ICS) plus a second controller and/or systemic corticosteroids to preclude uncontrolled asthma (Chung et al., 2014). Severe asthma patients suffer from frequent exacerbations even with proper medication. Additionally, frequent exacerbations with treatment side effects cause asthma patients diminishing quality of life. Eosinophilic asthma is a type of asthma that is severe in nature and usually comes on in adults. Eosinophilic asthma is caused by eosinophils which are a type of white blood cell. One of the roles of eosinophils is to fight disease by swelling but in eosinophilic asthma, these cells cause severe inflammation in the respiratory airways system by overreacting and causing asthma. People with severe asthma are found to have a 50% prevalence of eosinophilic asthma (Cohort, 2021).

7.2.2 DREAM study

The data used in this study was from an RCT called DREAM study. The DREAM clinical study compared the efficacy of three doses of mepolizumab for severe eosinophilic asthma patients. Mepolizumab is a humanized monoclonal antibody, that works against interleukin-5, works on both sputum and blood eosinophil counts by reducing them, and helps in the need for treatment with systemic glucocorticoids. Interleukin-5 releases eosinophils in the bone marrow, however, too much eosinophil causes airway inflammation. Mepolizumab helps to relieve asthma symptoms by controlling inflammation in the airways of the lungs. The DREAM study was a multicentre, double-blind, placebo-controlled study undertaken at 81 centers in 13 countries (Pavord et al., 2012). The criteria of patients eligibility in the study

were aged 12–74 years who had signs of eosinophilic inflammation with a history of recurrent severe asthma exacerbations. Present smokers, past history of more than 10 packs of smoking, significant comorbidity, pregnancy, and inferior record on treatment adherence were considered as the exclusion criteria.

Patients in the DREAM study were randomly assigned (in a 1:1:1:1 ratio) to receive one of three doses of intravenous mepolizumab (75 mg, 250 mg, or 750 mg) or matched placebo. The hypothesis that had been tested in DREAM study was that mepolizumab reduces the frequency of asthma exacerbations. As secondary objectives, the study assessed blood and sputum levels of eosinophils, asthma control, asthma-related quality of life, and forced expiratory volume in one second (FEV1). FEV1 is the maximum amount of air one can forcefully exhale in one second which describes the degree of airway obstruction caused by asthma. The treatment mepolizumab was found to be effective and well tolerated in patients with severe eosinophilic asthma. Demographic information, spirometry measurements, blood eosinophil counts, and scores on the asthma control questionnaire (ACQ) were obtained every 4 weeks for 52 weeks. The primary outcome was the rate of clinically significant asthma exacerbations. The baseline covariates which were also considered predictive of the outcome are given in Table 7.1. No variable in the study was identified as effect-modifying variable.

Table 7.1: *Prognostic covariates in DREAM study*

Prognostic covariates	Region	Sex	Age	Weight	Number of exacerbations in the year before the study
Prognostic covariates	Use of maintenance oral corticosteroids	Percentage of predicted FEV1	Airway reversibility	Blood eosinophil count	IgE concentration

7.3 Aim of the study

The aim of the study in this chapter was to demonstrate the use of double-bootstrapping in a MAIC-adjusted NMA for estimating relative treatment effects using a real dataset. The outcome measure of the study was the rate of exacerbations. First, a connected NMA was performed for severe eosinophilic asthma. Then the connected network of evidence was made disconnected by transforming each of the RCTs into a single-arm study. The objective of making the network disconnected was to imitate a situation that is often found in NICE STAs where an intervention from a single-arm study needs to be compared in a larger disconnected network of evidence. The IPD from the DREAM study was used to perform multiple unanchored MAICs. Performing several unanchored MAIC converts the disconnected network into a connected network. With the estimates from the unanchored

MAICs, an MAIC-adjusted NMA was then conducted. Eventually, the MAIC-adjusted NMA was then re-estimated using double-bootstrapping.

7.4 Methods

In this section, three methods are illustrated: a standard NMA for a connected network of evidence on eosinophilic asthma, a MAIC-adjusted NMA and a double-bootstrapped MAIC-adjusted NMA for the disconnected network of evidence.

7.4.1 A connected NMA for severe eosinophilic asthma

A connected NMA was conducted so that its estimates can be compared with MAIC-adjusted NMA (with and without double-bootstrapping). As in the DREAM study, no variable was identified as an effect-modifier, therefore, a standard connected NMA was performed that is capable of giving relative treatment effect estimates by adjusting prognostic variables.

7.4.1.1 Data contributing to the NMA

To perform a properly connected NMA, the first and foremost step is to conduct a systematic literature review (SLR) in multiple databases to identify the relevant studies. A proper SLR was not carried out for this study as the aim was to explain the steps that are required for executing double-bootstrapping using a case study, not estimating an updated relative treatment effect of mepolizumab compared to existing biologics. Therefore, in order to conduct a connected NMA, a connected network of evidence was formed using the available biologics for severe eosinophilic asthma. To find out the connected network of evidence for eosinophilic asthma, an existing systematic review for severe eosinophilic asthma was used (Agache et al., 2020). Agache et al. (2020) have identified five biologics that are used for severe eosinophilic asthma: these are benralizumab, dupilumab, mepolizumab, omalizumab and reslizumab. Among these, omalizumab was the first approved antibody for severe allergic asthma. Although all of the drugs are used for eosinophilic asthma, their application differs according to asthma phenotype and endotype (Agache, 2019). Phenotype involves grouping individuals with similar observable characteristics while endotype focuses on groupings based on underlying molecular mechanisms or treatment responses (Lin et al., 2013). Table 7.2 shows the studies that were included in the standard connected NMA.

Table 7.2: *Studies included in the connected NMA*

	Biologics			
	Mepolizumab	**Benralizumab**	**Dupilumab**	**Omalizumab**
Study 1	Pavord et al., 2012	Bleecker et al., 2016	Wenzel et al., 2016	Hanania et al., 2013
Study 2	Bel et al., 2014	FitzGerald et al., 2016	Castro et al., 2018	Busse et al., 2013
Study 3		Nair et al., 2017	Rabe et al., 2018	

Though Agache et al. (2020) have identified all the available drugs for eosinophilic asthma treatment, no NMA can be found that have taken into account all the existing drugs due to the unavailability of a common placebo/standard of care (SOC). The review by Agache et al. (2020) has compared each aforementioned asthma drug to a particular SOC. One study was found that compared mepolizumab and omalizumab in an indirect comparison where mepolizumab was found to be better in reducing exacerbations, however, the study emphasized that there was a discrepancy in the study populations in terms of severity, and also the results from the study were clinically not relevant due to heterogeneous pathways and population for the drugs (NICE, 2020). A study was found with observational data that uses target trial emulation (TTE) for assessing the comparative effectiveness between omalizumab, mepolizumab, and dupilumab in severe asthma where dupilumab was found to be most efficacious followed by omalizumab and then mepolizumab (Akenroye et al., 2023).

7.4.1.2 Statistical methods for the NMA

In order to perform a NMA with a connected network of evidence, the assumptions of transitivity and consistency need to be satisfied that assume that among studies prognostic and effect-modifiers are evenly distributed. In this study, the difference between the identified studies was ignored and the placebo/ SOC arm was assumed to be the same for each study. Moreover, reslizumab was excluded from the NMA as no published articles can be accessed for the drug. Figure 7.1 illustrates the process where 10 studies were included for the connected NMA. Each oval-shaped node represents a study arm with a treatment name and an RCT is depicted by joining two nodes with a solid line. Treatment SOC was used as the reference treatment. In the connected NMA, 10 studies were included where 2 studies were of mepolizumab, 3 studies of benralizumab, 3 of dupilumab, and 2 of omalizumab.

The NMA was undertaken using the rate of exacerbation as the outcome. Therefore the NMA analyses arm-level rate data using the number of exacerbations and the person-years at risk in each arm using Poisson likelihood and a log link function. For the NMA estimation R package `multinma` was used that estimates an NMA in a Bayesian framework using Stan. Both fixed effects and random effects models were estimated. For the fixed and random effects NMA, $N(0, 100^2)$ was used as prior distributions for the treatment effects and study-specific intercepts. Additionally, for the random effects NMA, $half\text{-}Normal(0.15)$ prior was used for the heterogeneity parameter τ.

7.4.2 A MAIC-adjusted NMA for severe eosinophilic asthma

After conducting a connected NMA, a MAIC-adjusted NMA was performed using multiple unanchored MAICs.

7.4.2.1 Data contributing to the NMA

After getting a connected network of evidence with severe eosinophilic asthma, the next step was to convert the connected network of evidence into a disconnected network of evidence. The mepolizumab arm with 250 mg from DREAM study was considered as a single-arm

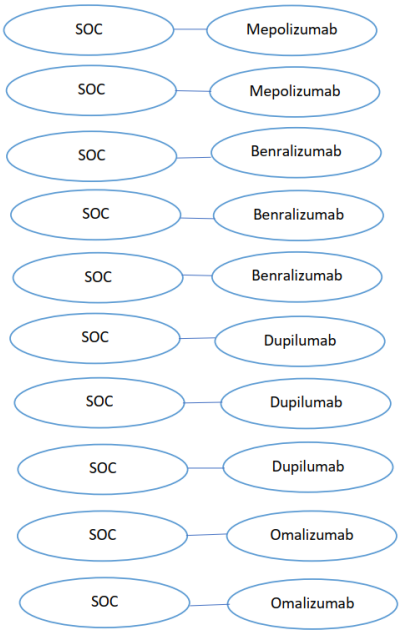


Figure 7.1: *Making of a connected network of evidence for severe eosinophilic asthma*

study with a sample size of 152 patients. The IPD from this arm was used to perform multiple unanchored MAICs. Figure 7.2 illustrates the process where the first node is the 250 mg mepolizumab arm from the DREAM study. The connected network of evidence that was formed earlier was made disconnected by dropping one arm from each study which has been depicted by striking through the treatment name, and each MAIC has been depicted by an arrow line. Other than the mepolizumab arm of the DREAM study, all other arms only had AgD in the form of number of exacerbations and the person-years at risk. Performing multiple MAICs and estimating relative treatment effects again made the disconnected network of evidence into a connected network of evidence where treatment mepolizumab was used as the reference treatment.

7.4.2.2 Statistical methods for the MAIC-adjusted NMA

As the MAICs were applied in unanchored form, all effect-modifiers and prognostic variables need to be included in the MAIC logistic regression model for the weight calculation. Due to the data availability issue from the AgD studies, only four covariates sex, age, baseline peripheral blood eosinophil count, and exacerbation frequency in the previous year were considered for adjustment between the single-arm and each AgD study. All of these variables were defined as prognostic variables, and none of the variables were identified as effect-modifiers in the DREAM study.

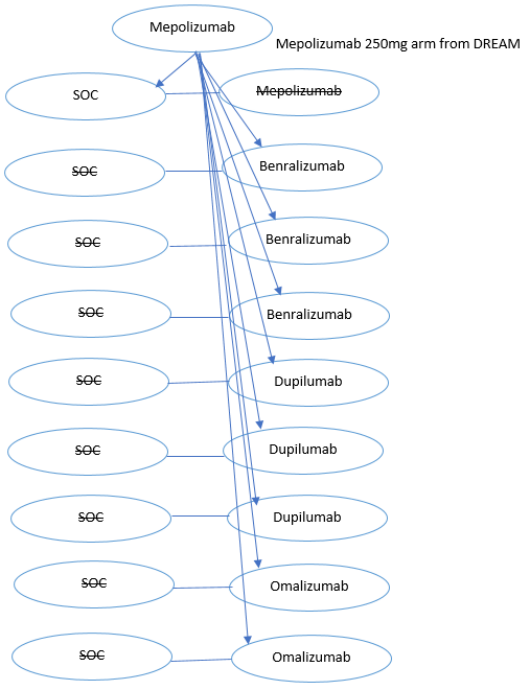


Figure 7.2: *Making of a disconnected network of evidence for severe eosinophilic asthma*

At the beginning of the analysis, 9 unanchored MAIC were performed. In each of these MAICs, the IPD from the DREAM study was reweighted adjusting for 4 covariates to match the baseline covariates in AgD study. Three out of the 9 MAICs, could not produce any weights due to overlap issues of covariates sex and age between studies. In these studies, the sex was either only male or female. The overlap between age variables was quite small which causes the non-convergence issue. Therefore, in the final analysis, only the covariates baseline peripheral blood eosinophil count and exacerbation frequency in the previous year were included. Additionally, only these two covariates were found to be associated with the efficacy of mepolizumab in reducing asthma exacerbation in DREAM study. MAICs were applied up to the first moment i.e. balancing of covariates was done only for the mean as the logistic model was not able to produce meaningful weights when the second moment was included in the model.

In order to perform an NMA with rate data, three quantities are required which are the number of events, number of participants, and total number of exposure time. For MAIC-adjusted NMA, these quantities need to be estimated involving MAIC weights. Therefore, in the MAIC-adjusted NMA, the weighted number of events, weighted number of participants, and weighted total number of exposure time were estimated as follows:

The R codes for MAIC-adjusted NMA are given in F.2 in Appendix F.

Quantity	Estimation procedure
Weighted number of events	$\sum$(individual patient weight*event indicator for each patient)
Weighted exposure time	$\sum$(individual patient weight*exposure time for each patient)
Weighted number of participants	$\sum$ (individual weights)

7.4.3 A double-bootstrapped MAIC-adjusted NMA for severe eosinophilic asthma

The process for a double-bootstrapped MAIC-adjusted NMA can be described with the following steps that was implemented with each unanchored MAIC:

Step 1: Let $\mathcal{X}$ denote the original IPD from DREAM study. Instead of using the original data $\mathcal{X}$ directly for the calculation of MAIC weights, $\mathcal{X}$ was bootstrapped once which is denoted by $\mathcal{X}^*$.

Step 2: At the second step the bootstrap sample generated earlier as $\mathcal{X}^*$ was again bootstrapped which can be denoted by $\mathcal{X}^{**}$. Therefore, the original IPD from DREAM study was bootstrapped twice, and the bootstrapped sample $\mathcal{X}^{**}$ was included for the estimation of MAIC weights. Using the MAIC weights, a weighted number of events, weighted exposure time, and a weighted number of participants were estimated in each MAIC.

Step 3: The weighted quantities that were estimated in step 2 from each MAIC, were then used as input in an NMA setting and a MAIC-adjusted NMA was performed with the MAIC estimates. The whole process was again bootstrapped 2000 times which produced 2000 NMA estimates. The mean and SD of these bootstrapped estimates were considered to be the final estimate of relative treatment effects and SE respectively. The R codes for double-bootstrapped MAIC-adjusted NMA are given in F.3 in Appendix F.

7.5 Results

In this section, results from the connected NMA, MAIC-adjusted NMA, and double-bootstrapped MAIC-adjusted NMA have been presented.

7.5.1 Results from connected NMA for severe eosinophilic asthma

Figure 7.3 shows the network diagram for the connected network of evidence. In the network, SOC was the common treatment that was connected with mepolizumab, benralizumab, dupilumab, and omalizumab. Table 7.3 shows the results from the connected NMA for severe eosinophilic asthma. The connected NMA was fitted for both fixed and random effects NMA.

The log rate ratio (LRR) together with 95% CrI of SOC, benralizumab, dupilumab, and omalizumab compared to mepolizumab in the fixed effect model was 0.24 (0.006, 0.48), -0.29 (-0.55, -0.03), -0.43 (-0.72, -0.14) and 0.15 (-0.13, 0.44)) respectively. The LRR in the random effects model was 0.27 (0.038, 0.57), -0.29 (-0.67, -0.08), -0.45 (-0.86, -0.04) and 0.16 (-0.27, 0.59) respectively. The LRR for the treatments benralizumab and dupilumab were negative which indicates a lower rate of exacerbation in comparison to mepolizumab, however, the LRR for the treatments SOC and omalizumab was positive that indicates in comparison to mepolizumab, these treatments were not able to reduce the rate of exacerbations. The CrI

for omalizumab includes 0, therefore, the result was statistically not significant.

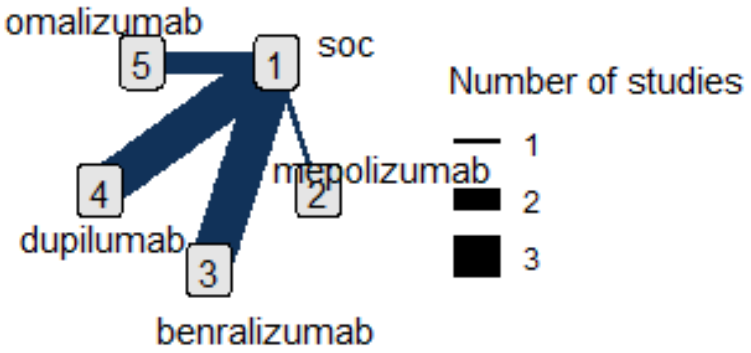


Figure 7.3: *Network diagram for the connected network of evidence*

Table 7.3: *Parameter estimates from the connected NMA for fixed and random effects model*

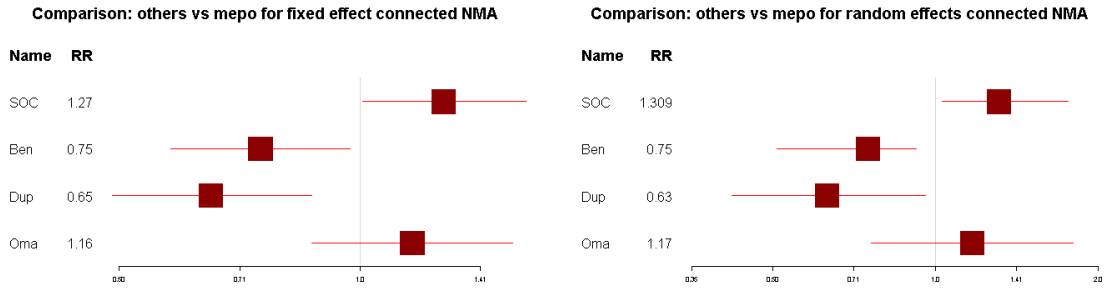
	Fixed effect model			**Random effect model**		
	MEAN (LRR)	**SE**	**CrI**	**MEAN (LRR)**	**SE**	**CrI**
d1 (SOC vs Mepo)	0.24	0.12	(0.006, 0.48)	0.27	0.16	(0.038, 0.57)
d3 (Ben vs Mepo)	-0.29	0.13	(-0.55, -0.03)	-0.29	0.19	(-0.67, -0.08)
d4 (Dup vs Mepo)	-0.43	0.14	(-0.72, -0.14)	-0.45	0.21	(-0.86, -0.04)
d5 (Oma vs Mepo)	0.15	0.15	(-0.13, 0.44)	0.16	0.22	(-0.27, 0.59)
tau				0.12		(0.005, 0.33)
DIC	31.4			32.2		
Residual deviance	18.7			19.7		
pD	13.1			15.3		

[a] Mepo=Mepolizumab, Ben=Benralizumab, Dup=Dupilumab, Oma=Omalizumab
[b] LRR=log rate ratio

The residual deviance was close to 20 data points for both models which shows that both models fit well. Moreover, Deviance Information Criterion (DIC) and pD were also close for both models. The DIC is used to compare models with the same likelihood and data such as between fixed and random effects models. It also penalises for model complexity, the more complex a model is, the more penalties will be implemented. Lower values of the DIC suggest a more parsimonious model. pD indicates the effective number of parameters and a lower

value of pD is desired.

Figure 7.4 shows the forest plot both for the fixed and random effects connected NMA. As in the estimation of NMA, mepolizumab was used as reference treatment, therefore, in the forest plot comparisons have been made to mepolizumab. The forest plot was built using the table of parameter estimates (Table 7.3). In the forest plot, the x-axis follows a logarithmic scale, therefore, all argument values for the plot i.e. mean estimates and upper and lower confidence values from the table were transformed into exponential form. The forest plots show that dupilumab and benralizumab were more effective than mepolizumab and the results were statistically significant; omalizumab and SOC were less effective than mepolizumab, which was statistically significant for SOC. However, omalizumab touches the line of null effect which means it shows statistically insignificant results when compared with mepolizumab.



(a) *forest plot for fixed effect connected NMA* **(b)** *forest plot for random effects connected NMA*

Figure 7.4: *Forest plot for fixed and random effects connected NMA model*

Figure 7.5 and Figure 7.6 show the ranking of the treatments for fixed and random effects models respectively. Dupilumab shows the highest efficacy in reducing asthma exacerbation with respect to mepolizumab followed by benralizumab. Mepolizumab was found to be third in ranking in reducing asthma exacerbations followed by omalizumab and SOC. Furthermore, the heterogeneity parameter τ was found to be 0.12 in the random effects model which indicated a moderate level of heterogeneity existed between studies.

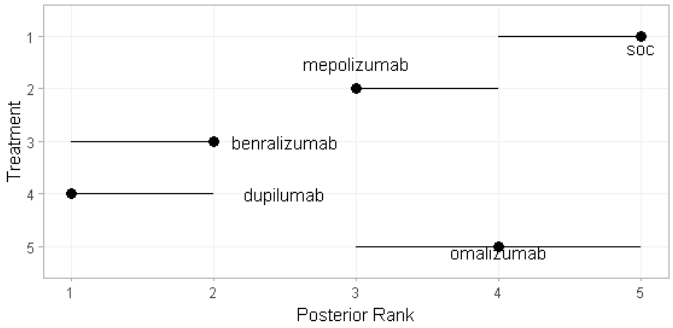


Figure 7.5: *Ranks of the treatments in a FE connected NMA*

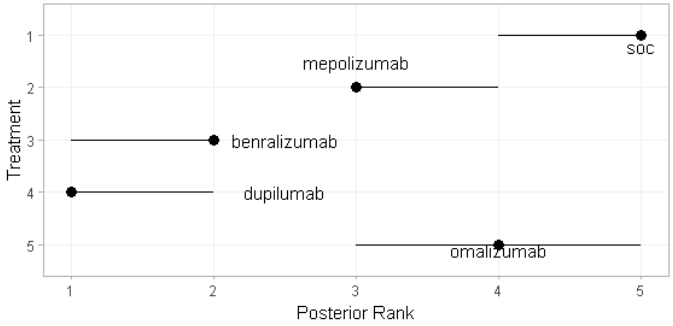


Figure 7.6: *Ranks of the treatments in a RE connected NMA*

7.5.2 Results from MAIC-adjusted NMA for severe eosinophilic asthma

Table 7.4 shows the mean estimates of the baseline covariates in MAICs before and after weighting the intervention and the comparator studies. From the table, it can be seen that the MoM that was used to estimate MAIC weights, matches the mean baseline patients characteristics of the comparator study to the weighted intervention study. Figures of the weight distribution for 9 MAICs are given in Figures F.1, F.2, F.3, F.4, F.5, F.6, F.7, F.8, F.9 in Appendix F. In the figures, weights are shown both in their original scale and after rescaling them which makes it easier to examine. A rescaled weight > 1 corresponds to a person with a higher weight in the reweighted population than the original data and a rescaled weight $<$ 1 corresponds to a person with a lower weight in the reweighted population than the original data. In Table 7.4, the smallest ESS was found for MAIC 8 which can also be seen in Figure F.8 where most of the patients were found to have weights of 0.

Figure 7.7 shows the connected network of evidence for the MAIC-adjusted NMA. In the network, mepolizumab was the common treatment that was connected with SOC, benralizumab, dupilumab, and omalizumab. Table 7.5 shows the results from the MAIC-adjusted fixed and random effects NMA for severe eosinophilic asthma. Similar to the connected NMA, both

Table 7.4: *Baseline variables before and after weighting*

Mepolizumab vs Benralizumab		**ESS/N**	**Blood eosinophil count**	**Exacerbation frequency in the previous year**
MAIC 1	Before MAIC	152	1.78	389.80
	After MAIC	47.44	2.90	390
	Comparator study	400	2.90	390
MAIC 2	Before MAIC	152	1.78	389.80
	After MAIC	149.80	1.70	349
	Comparator study	425	1.70	349
MAIC 3	Before MAIC	152	1.78	389.80
	After MAIC	56.63	2.80	462
	Comparator study	73	2.80	462
Mepolizumab vs Dupilumab		**ESS/N**	**Blood eosinophil count**	**Exacerbation frequency in the previous year**
MAIC 4	Before MAIC	152	1.78	389.80
	After MAIC	120.87	2.17	347.40
	Comparator study	154	2.17	347.40
MAIC 5	Before MAIC	152	1.78	389.80
	After MAIC	132.40	2.07	348.98
	Comparator study	631	2.07	349
MAIC 6	Before MAIC	152	1.78	389.80
	After MAIC	139.97	2.01	370
	Comparator study	103	2.01	370
Mepolizumab vs Omalizumab		**ESS/N**	**Blood eosinophil count**	**Exacerbation frequency in the previous year**
MAIC 7	Before MAIC	152	1.78	389.80
	After MAIC	129.19	2.0	535.01
	Comparator study	427	2.0	535.01
MAIC 8	Before MAIC	152	1.78	389.80
	After MAIC	14.47	1.0	50.96
	Comparator study	157	1.0	51
Mepolizumab vs SOC		**ESS/N**	**Blood eosinophil count**	**Exacerbation frequency in the previous year**
MAIC 9	Before MAIC	152	1.78	389.80
	After MAIC	36.88	2.90	230
	Comparator study	66	2.90	230

fixed and random effects models were fitted.

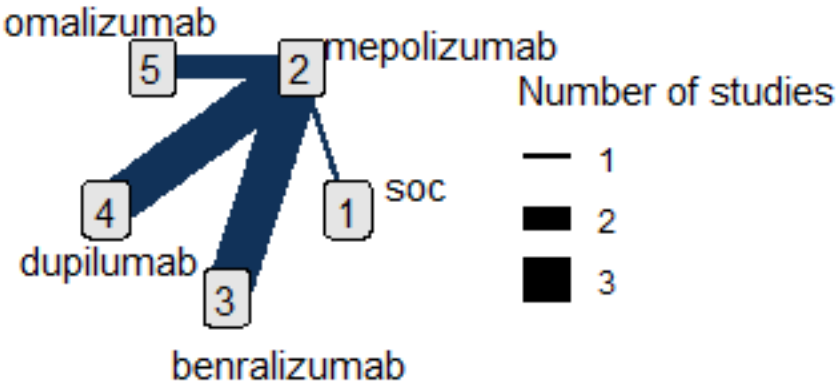


Figure 7.7: *Network diagram for the MAIC-adjusted network of evidence*

The LRR together with 95% CrI of SOC, benralizumab, dupilumab, and omalizumab compared to mepolizumab in the fixed effect model was 1.06 (0.62, 1.52), 0.06 (-0.11, 0.25), -0.23 (-0.44, -0.019) and 0.19 (-0.06, 0.45) respectively. The LRR in the random effects model was 1.06 (0.54, 1.59), 0.07 (-0.16, 0.34), -0.21 (-0.49, -0.07), and 0.23 (-0.11, 0.64) respectively. The LRR for all the treatments was positive except treatment 4, which was dupilumab. Therefore, in comparison to mepolizumab, only dupilumab showed a lower rate of exacerbation. Mepolizumab scored second in reducing asthma exacerbation followed by benralizumab, omalizumab and then SOC. The results for benralizumab and omalizumab were statistically not significant as the CrI includes 0.

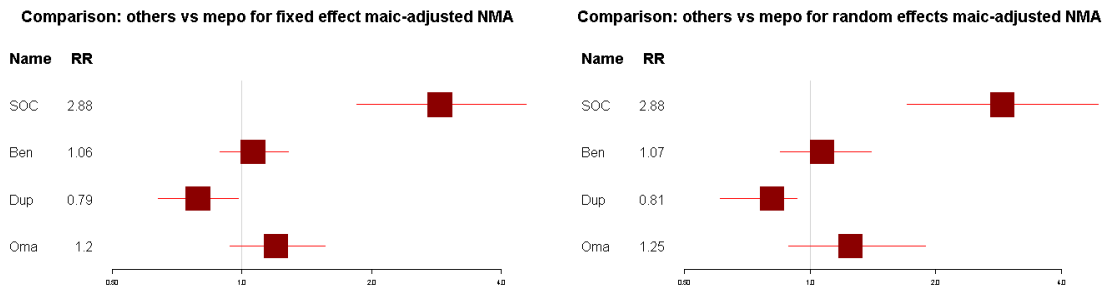
The residual deviance was close to 18 data points for both models. Moreover, DIC and pD were also close for both models. The τ parameter was close to the connected random effects NMA model which indicates a moderate level of heterogeneity between studies. Although, MAIC can adjust for treatment effect-modifiers, however, the variables that were adjusted in the analysis were prognostic variables only. Moreover, due to data availability issues, only two variables were adjusted in the unanchored MAICs. These could be the reasons that the τ parameter was similar to the connected NMA.

Table 7.5: *Parameter estimates from the MAIC-adjusted NMA for fixed and random effects model*

	Fixed effect model			Random effects model		
	MEAN (LRR)	**SE**	**CrI**	**MEAN (LRR)**	**SE**	**CrI**
d1 (SOC vs Mepo)	1.06	0.18	(0.62,1.52)	1.06	0.17	(0.54, 1.59)
d3 (Ben vs Mepo)	0.06	0.09	(-0.11,0.25)	0.07	0.10	(-0.16,0.34)
d4 (Dup vs Mepo)	-0.23	0.09	(-0.44,-0.019)	-0.21	0.09	(-0.49, -0.07)
d5 (Oma vs Mepo)	0.19	0.11	(-0.06,0.45)	0.23	0.12	(-0.11,0.64)
tau				0.118		(0.003, 0.27)
DIC	30.9			30.6		
Residual deviance	17.3			16.6		
pD	13.2			14.1		

[a] Mepo=Mepolizumab, Ben=Benralizumab, Dup=Dupilumab, Oma=Omalizumab
[b] LRR=log rate ratio

Figure 7.8 shows the forest plot both for fixed and random effects MAIC-adjusted NMA model. Similar to connected NMA, the forest plot of MAIC-adjusted NMA showed that dupilumab was more effective than mepolizumab and the results were statistically significant. Omalizumab, benralizumab and SOC were found to be less effective than mepolizumab. However, both omalizumab and benralizumab touch the line of null effect which indicates statistically insignificant results.



(a) *forest plot for fixed effect maic-adjusted NMA* **(b)** *forest plot for random effects maic-adjusted NMA*

Figure 7.8: *Forest plot for fixed and random effects maic-adjusted NMA model*

Figure 7.9 and Figure 7.10 show the ranking of the treatments for fixed and random effects models respectively. Similar to the connected NMA, dupilumab shows the highest efficacy in reducing asthma exacerbation followed by mepolizumab. After dupilumab, benralizumab was found to be third in efficacy followed by omalizumab and then SOC. Though in the connected NMA, benralizumab was found to be more efficacious than mepolizumab, the opposite was seen here.

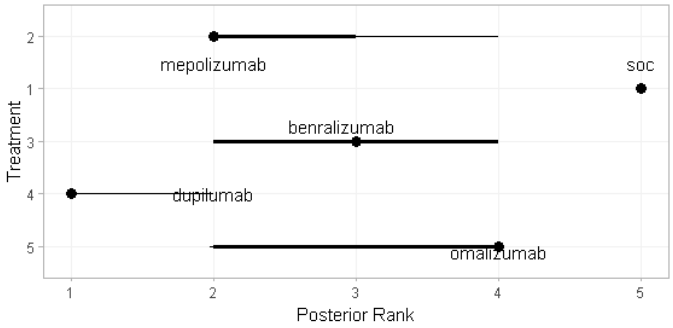


Figure 7.9: *Ranks of the treatments in a MAIC-adjusted FE NMA*

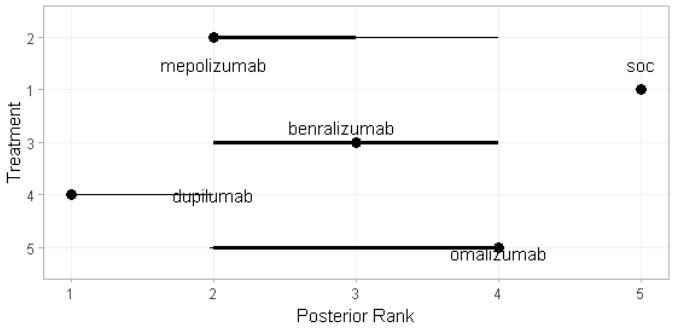


Figure 7.10: *Ranks of the treatments in a MAIC-adjusted RE NMA*

7.5.3 Results from double-bootstrapped MAIC-adjusted NMA for severe eosinophilic asthma

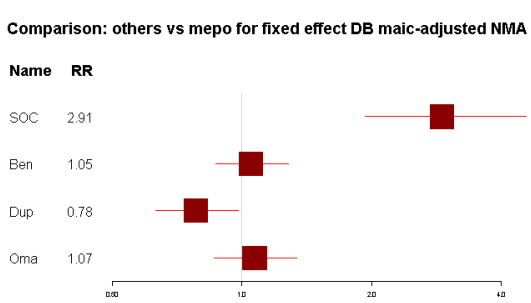
Table 7.6 shows the parameter estimates of a double-bootstrapped MAIC-adjusted NMA for a fixed and random effects model. In the double-bootstrapped MAIC-adjusted NMA, the original IPD from DREAM study was bootstrapped 2000 times. During each iteration, an MAIC-adjusted NMA was estimated where instead of using the original IPD in the estimation of each MAIC, a double-bootstrapped sample of the original IPD was used. After the bootstrapped iterations, 2000 MAIC-adjusted NMA results were estimated. The mean and SD of the 2000 estimates were taken as the final estimates. The relative treatment effect estimates were similar to MAIC-adjusted NMA from the previous section.

Figure 7.11 shows the forest plot both for fixed and random effects double-bootstrapped MAIC-adjusted NMA model. Similar to MAIC-adjusted NMA, the forest plot of double-bootstrapped MAIC-adjusted NMA showed that dupilumab was more effective than mepolizumab and the results were statistically significant. Omalizumab, benralizumab and SOC were found to be less effective than mepolizumab. However, both omalizumab and benralizumab touch the line of null effect which indicates statistically insignificant results.

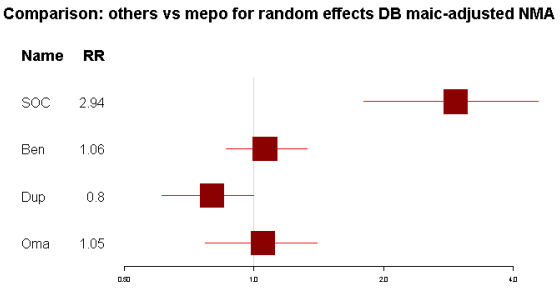
Table 7.6: *Parameter estimates from the double-bootstrapped MAIC adjusted fixed and random effects model*

	Fixed effect model			Random effects model		
	Mean (LRR)	**SE**	**CrI**	**Mean (LRR)**	**SE**	**CrI**
d1 (SOC vs Mepo)	1.07	0.23	(0.66,1.52)	1.08	0.26	(0.59,1.52)
d3 (Ben vs Mepo)	0.05	0.10	(-0.13,0.25)	0.06	0.12	(-0.15,0.29)
d4 (Dup vs Mepo)	-0.24	0.11	(-0.45,-0.02)	-0.22	0.14	(-0.49,0.00)
d5 (Oma vs Mepo)	0.07	0.12	(-0.15,0.30)	0.05	0.16	(-0.26,0.34)
tau				0.121	(0.005, 0.30)	
DIC			36.63		36.61	
Residual deviance			23.22		21.94	
pD			13.11		14.20	

[a] Mepo=Mepolizumab, Ben=Benralizumab, Dup=Dupilumab, Oma=Omalizumab
[b] LRR=log rate ratio



(a) *forest plot for fixed effect DB maic-adjusted NMA*



(b) *forest plot for random effects DB maic-adjusted NMA*

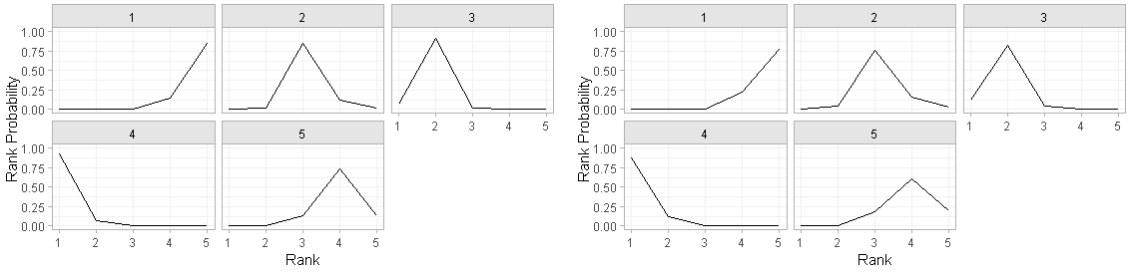
Figure 7.11: *Forest plot for fixed and random effects DB maic-adjusted NMA model*

7.6 Result comparison

Figure 7.12, and 7.13 show the posterior rank probability for the connected and MAIC-adjusted NMA. For a specific outcome, ranking probability evaluates the probable ranking of a treatment with respect to several comparator treatments. In Figure 7.12, it can be seen that the highest ranking was achieved by treatment 4 (dupilumab) followed by 3 (benralizumab), 2 (mepolizumab), 5 (omalizumab) and 1(SOC). For the MAIC-adjusted NMA (with or without double-bootstrapping), the most efficacious treatment was again treatment 4 (dupilumab), followed by treatment 2 (mepolizumab). The second and third position was swapped in the MAIC-adjusted NMA in comparison to the connected NMA, the second position was taken by treatment 2 (mepolizumab) and then treatment 3 (benralizumab). Omalizumab and SOC scored fourth and fifth rank respectively in all the models.

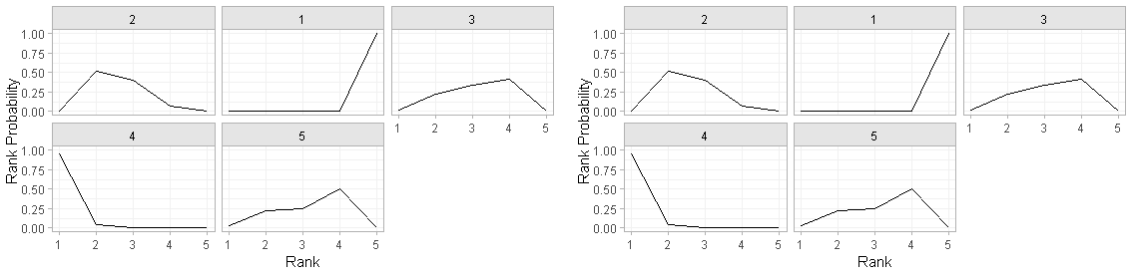
A negative value in the point estimate means the corresponding treatment is capable of

reducing asthma exacerbations compared to the reference. The point estimate of SOC in the standard connected and MAIC-adjusted NMA were positive which means SOC failed to reduce asthma exacerbation in comparison to mepolizumab. However, the magnitude of point estimate for SOC was more further from 0 in MAIC-adjusted NMA which means SOC was proved to be much less efficacious in MAIC-adjusted NMA compared to connected NMA. The point estimate of benralizumab showed opposite results. It was found to be more efficacious than mepolizumab in connected NMA but less efficacious in the MAIC-adjusted NMA. However, the CrI in the MAIC-adjusted NMA includes 0 which means the results were statistically not significant. Moreover, the point estimate of dupilumab showed similar results in both NMAs, nonetheless, the magnitude of the estimate of dupilumab was found more efficacious in the connected NMA. Furthermore, the point estimate of omalizumab was similar, even so, the CrI included 0 in both NMAs which makes the results statistically insignificant. In addition to the point estimates, the width of most of the CrI in MAIC-adjusted NMA was narrower than the connected NMA.



(a) *posterior rank probability for fixed effect* **(b)** *posterior rank probability for random effects*

Figure 7.12: *Posterior rank probability for connected fixed and random effects NMA model*



(a) *posterior rank probability for fixed effect* **(b)** *posterior rank probability for random effects*

Figure 7.13: *Posterior rank probability for both fixed and random effects MAIC-adjusted NMA model*

Table 7.7 shows the consensus between different models on the ranking of the treatments. A clear consensus was found on the ranking for SOC, omalizumab, and dupilumab where dupilumab scored the highest rank, omalizumab in fourth rank, and SOC scored the lowest rank. Benralizumab scored second in connected NMA which was replaced by mepolizumab in MAIC-adjusted NMA.

Table 7.7: *Ranking comparison between different NMA models (both in FE and RE model)*

	Connected NMA	**MAIC adjusted NMA**	**MAIC adjusted NMA (DB)**
SOC	5	5	5
Mepolizumab	3	2	2
Benralizumab	2	3	3
Dupilumab	1	1	1
Omalizumab	4	4	4

[a] DB=Double-bootstrapped

7.7 Discussion

The aim of this chapter was to demonstrate the implementation of the double-bootstrapped MAIC-adjusted NMA approach. First, a connected NMA was carried out to estimate the relative treatment effect for severe eosinophilic asthma. The connected network of evidence was then made disconnected artificially by dropping one arm from each RCT which makes each RCT into a single-arm study with AgD. The 250 mg mepolizumab arm from the DREAM study was considered as a single-arm study with IPD. Then multiple unanchored MAICs were estimated by repeatedly using the IPD from the DREAM study. A MAIC-adjusted NMA as well as a double-bootstrapped MAIC-adjusted NMA was then estimated.

The standard connected NMA was performed to estimate relative treatment effects instead of a ML-NMR. ML-NMR introduced by Phillippo et al. (2020a) is a recent developement which incorporates both IPD and AgD in an NMA setting to estimate relative treatment effects. It is multi-level as it embeds a probabilistic approach to population adjustment by allowing both aggregate data and individual data into a single probabilistic model and as it uses a linear model using available IPD, hence it is termed as network meta-regression. Similar to STC, this method uses a linear model using IPD to perform meta-regression, however, ML-NMR differs from STC as ML-NMR is capable of generalising population-adjustment to a larger network of evidence where the relative effects can be estimated in any specified target population. ML-NMR aims to make an unbiased comparison of the outcome variable by taking into account imbalance in effect-modifiers. In the DREAM study all the variables were prognostic and no variables were identified as effect-modifiers, therefore, it was not necessary to balance out these variables with the use of ML-NMR. A standard NMA was considered to be enough for estimating relative treatment effects assuming that effect-modifier variables are evenly distributed. For the connected NMA, the estimands were unadjusted relative effects of treatments, however, for MAIC-adjusted NMA, the estimands were adjusted population-average relative effects for the AgD study.

The results from the connected NMA show that dupilumab has the highest efficacy in comparison to mepolizumab followed by benralizumab, omalizumab and then SOC. The results

from the MAIC-adjusted NMA and double-bootstrapped MAIC-adjusted NMA was similar where dupilumab showed the highest efficacy with respect to mepolizumab, followed by benralizumab, and omalizumab. The results from the standard NMA and MAIC-adjusted NMA were similar except the ranking for benralizumab which was scored third in MAIC-adjusted NMAs but second in standard NMA. One reason could be that MAIC-adjusted NMA was able to adjust for only two variables due to availability and convergence issues. In both the standard and MAIC-adjusted NMA, mepolizumab was found to be better than omalizumab. However, few studies have evaluated mepolizumab with omalizumab; hence no consensus can be found on the efficacy of these drugs. The results from those studies were in opposite directions in terms of the efficacy of mepolizumab with omalizumab.

Between the fixed and random effects double-bootstrapped MAIC-adjusted NMA, the DIC and residual deviance both were lower for the random effects NMA compared to the fixed effect. Even so, the difference in DIC between the fixed and random effects models was less than 3 points, which means both models were similar. Moreover, the point estimates and CrI for MAIC-adjusted NMA and double-bootstrapped MAIC-adjusted NMA were similar. The SE of the double-bootstrapped MAIC-adjusted NMA was higher than MAIC-adjusted NMA due to the use of double-bootstrapping. Nevertheless, the results that are presented in Table 7.6 are only for one run with the IPD. The whole process needs to run a sufficient number of times (such as 1000) to get the bootstrapped SE. The SE for the standard NMA and MAIC-adjusted NMA(with or without double-bootstrap) was not comparable. This was because the results in 7.3 were presented for the mepolizumab arm instead of the reference arm SOC. The relative treatment effects with mepolizumab were estimated via the consistency equation from the connected NMA. As there were only two studies available with mepolizumab, therefore the SE of the estimates increased. Consequently, the SE from the standard and MAIC-adjusted NMA should not be compared.

In all models, the heterogeneity parameter τ was similar which showed moderate heterogeneity between studies. Due to a limited number of studies per comparison a weakly informative prior for τ was used for all the analysis. This could be the reason for the stationary τ in all models.

7.8 Strength and limitations of the study

In the absence of head-to-head comparison between multiple comparator treatments of a particular health condition, this study illustrates how to make multiple comparisons simultaneously using double-bootstrapped MAIC-adjusted NMA where a correct level of coverage for the NMA estimates can be preserved with the use of double-bootstrapping.

The main limitation of the study was that the unanchored MAICs were not able to adjust for all the prognostic and effect-modifying variables. This was because, in the DREAM study although the prognostic variables were defined, there was no information on whether any variables were effect-modifiers or not. Additionally, during the estimation of MAICs, all prognostic variables were not included due to availability and convergence issues which could have retained residual bias between studies. The inability to adjust for all prognostic

variables could be the reason that the ranking of treatments 2 and 3 was swapped between standard and MAIC-adjusted NMA. Moreover, as no effect-modifying variables were identified in DREAM study, it was assumed that ML-NMR was not imperative which is developed to adjust effect-modifier variables in a connected NMA setting. The standard NMA was considered to be capable of adjusting all prognostic variables and there was no imbalance between studies due to effect-modifiers. Additionally, during the estimation of the standard NMA all the placebo/SOC arm were considered similar which may not be true. Furthermore, all the models showed a moderate level of heterogeneity between studies which indicates the possibility of the presence of unobserved effect-modifiers which none of the models were able to adjust for.

The results of the NMAs from this case study chapter should be interpreted with caution. This is because in order to conduct NMA with the real dataset of asthma, the formation of a connected network was required. For this, to find out the connected network of evidence for eosinophilic asthma, an existing systematic review for severe eosinophilic asthma was used. Although the review identified all the available treatments for eosinophilic asthma, no NMA could be conducted with the treatments as there was no common placebo/standard of care (SOC) exists. Additionally, the assumptions of transitivity and consistency need to be satisfied for estimating a credible NMA. In this case study chapter, the difference between the identified studies was completely ignored to satisfy the assumptions, and the placebo/ SOC arm was assumed to be the same for each study to perform a standard, MAIC-adjusted and double-bootstrapped MAIC-adjusted NMA. The target of this chapter was to demonstrate how double-bootstrapping can be executed with a real dataset as well as to identify practical challenges for the application of double-bootstrapping, not to make a valid comparison among comparator treatments.

Chapter 8

Discussion and Conclusion

A single-arm study is exceptional in its design where all the patients receive the same treatment and there is no control arm to estimate the relative treatment effect. This characteristic of the single-arm study does not allow a direct estimate of treatment effects like an RCT. Therefore, undertaking treatment comparisons using the single-arm study has to be done in an indirect way. An indirect comparison can either be in anchored or in unanchored form where an anchored indirect comparison refers to the relative treatment comparison via a common comparator which takes into account randomisation within studies. Therefore, in anchored comparison it is not necessary to adjust for prognostic variables; only effect-modifying variables are adjusted for. An unanchored indirect comparison needs to be done when common comparators do not exist between studies, therefore, it does not consider randomisation within studies. An unanchored indirect comparison is more problematic as it relies on the conditional constancy of absolute effects assumption. The assumption presumes that all prognostic and effect-modifiers have been adjusted which is demanding to meet in practice. With a single-arm study, the comparison is always in an unanchored form where the relative treatment estimate needs to be made with an external comparator. Therefore, these variables need to be adjusted to obtain a valid comparison.

In HTA, estimating relative treatment effects comparison with a single-arm study is not only difficult as the external control/comparator arm needs to be balanced with respect to prognostic and effect-modifying variables but also access to IPD is often restricted. In HTA, when a pharmaceutical company wants to estimate the relative treatment effect with a single-arm study by adjusting prognostic and effect-modifying covariates, most of the time the IPD is available for their own study but not for the comparator study. Due to this limited availability of IPD, conventional adjustment methods like propensity score matching or regression adjustment are not possible, and population-adjustment methods that can adjust study arms using both IPD and AgD need to be applied.

This thesis provides an exploration of indirect comparisons with population adjustment methods using single-arm studies in HTA. The thesis, as a whole, can be described in five key points. First, in the NICE STA review chapter, it was noticed that unanchored MAIC and STC were used with single-arm studies in a larger disconnected network of evidence despite the fact that they were not being developed for multiple comparisons simultaneously. Second,

in the review of methods chapter, two NMA-based methods namely random baseline NMA and NMA with matching were identified that could be used to estimate relative treatment effects in a larger disconnected network of evidence using single-arm study, however, they were found to be difficult to implement as they must satisfy a lot of assumptions in order to get reliable estimates. Third, the subsequent two chapters (Chapters 4 & 5) design a simulation study to explore the consequence of applying MAIC in a larger disconnected network of evidence with repeated use of IPD from a single-arm study, both for a fixed effect and random effects NMA. Undercoverage of MAIC-adjusted NMA estimates was found to be the prime concern in the simulation study. Fourth, in Chapter 6, a double-bootstrapping approach was developed to overcome the problem of undercoverage issue that was found during the simulation study with the NMA estimates. Finally, in Chapter 7, to illustrate the practical use of the double-bootstrapping approach, it was applied to a real dataset that was collected in an RCT.

The objective of this chapter is to provide a discussion of the thesis in a structured and systematic way. Sections 8.1 to 8.5 narrate the main finding of the thesis in a coherent way. Section 8.6 explains the strengths and limitations of the research whereas Section 8.7 proposes ideas for future investigation. Section 8.8 sums up the thesis with a conclusion.

8.1 Review of methods using single-arm studies in NICE single technology appraisals

Estimating relative treatment effects with single-arm studies comes with additional difficulties as single-arm studies lack a comparator arm which turns the relative comparison into an unanchored form. In Chapter 2, a review was done to assess how comparisons against relevant comparators were obtained from single-arm studies in NICE STAs when IPD was partially available . The focus of the review was NICE because worldwide it is appraised as one of the most influential HTA bodies.

Earlier a review was done by Phillippo et al. (2019a) on NICE STAs where the objective was to explore how the population-adjustment methods were used for any outcome. The timeframe of Phillipo's review was 2010 to 2018 and it was done for both anchored and unanchored cases. However, the review on NICE STAs in this thesis was restricted only to single-arm studies for unanchored comparison with a time frame from 2018 to 2021. This time frame was chosen to include more recent STAs information. In Phillipo's review, 83% of the identified applications were found to be in oncology which shows the frequent use of population-adjustment methods in this field.

Out of 260 TAs that were identified in the NICE STA review (Chapter 2), 20 TAs were found to be eligible for inclusion as the pivotal study was a single-arm study with partial availability of IPD (Section 2.3). Of the 20 included TAs, 80% used population-adjustment methods, and all the TAs were found to be in oncology. Unanchored MAIC was found to be the most applied population-adjustment method, followed by unanchored STC (Section 2.3.2). MAIC (Signorovitch et al., 2010; Ishak et al., 2015b) was developed based on a reweighting technique whereas STC (Caro and Ishak, 2010; Ishak et al., 2015b) depends on regression adjustment. Both methods were developed to adjust the baseline covariates for a pair of study populations

and cannot be extended for multiple studies i.e. for a larger network of evidence.

It was found that more than half of the TAs in the review (55%) have to make treatment comparisons for a larger disconnected network of evidence which includes not just multiple comparisons but multiple sources of evidence per comparison as well (Section 2.3.3). The larger disconnected network of evidence was found to be formed when a treatment from the single-arm study needs to be compared to more than one comparator treatment from different studies or the single-arm study treatment needs to be compared to a single comparator but from multiple studies. Multiple MAICs were applied in order to deal with multiple comparators. Additionally, when MAIC was used in a disconnected network of evidence for estimating relative treatment effect, the IPD from the single-arm study was used several times which breaks the independence between the studies. Some TAs have used STC to construct a predicted treatment arm for each single-arm study. This predicted arm forms a newly-connected network and then analyses were made with FP NMA. Although MAIC and STC are not developed to make multiple comparisons, they can still be used if an additional assumption called shared effect modifier can be assumed. None of the STAs have tried to justify this assumption.

Although in the review the use of both MAIC and STC were found, the frequency of using MAIC (65%) was more than STC. Reduction in ESS was found to be one of the major problems in applying MAIC. In order to overcome the problem of small ESS, some companies often discarded adjusting all unbalanced baseline covariates. Very small ESS is quite alarming as it suggests that there is a serious paucity of overlap between study covariates. When there exists a scarcity of overlap, weighting methods like MAIC cannot extrapolate beyond the data observed in the IPD, in which case MAIC estimates become biased. Most of the TAs did not adjust all the identified prognostic and effect-modifier variables due to availability issues. Therefore, unmeasured prognostic and effect-modifier variables increase the possibility of residual bias.

Description of the uncertainty estimates of the MAIC method was also very vague for most of the TAs included in the review. For MAIC, robust sandwich estimator and bootstrapping were mentioned for a few TAs. In STC, bootstrapping was used as the method to capture uncertainty. In MAIC, bootstrapping and sandwich estimators both estimate uncertainty from the data and discard overly strong assumptions about the weights by not treating the weights as fixed and known. However, compared to the sandwich estimator, bootstrapping is computationally very challenging.

The review of NICE STAs have both similarities, as well as dissimilarities with the review by Phillippo et al. (2019a). Similar to Phillipo's review, the applications of population-adjustment methods also turned out to be in oncology where survival outcomes were found to be the most common outcome type. A substantial decrease in ESS has been found in both of the reviews which in turn made the comparisons dependent on very few numbers of individuals in the IPD study. Additionally, the application of population-adjustment methods turned out to be very prevalent for a larger network of evidence in Phillipo's review which was found to be true for the review in this thesis also. On the contrary, in Phillipo's review,

no TAs were proved to be attempted to adjust residual bias whereas in this review, one TA was found that had used "out sample" method to quantify residual bias.

The main contribution of the NICE review chapter was that it pointed out two important facts that need to be considered. First, Chapter 2 sheds light on the issue that there is a lack of appropriate population-adjustment methods when there are multiple comparators. When an intervention from the single-arm study needs to be compared in a larger disconnected network of evidence where the network includes not just multiple comparisons but multiple sources of evidence per comparison, population-adjustment methods were used recurrently using the same IPD from the single-arm study. Moreover, in some TAs, the population-adjustment methods were used multiple times to use the estimates in a NMA setting so that relevant comparators can be compared simultaneously ignoring the fact that these methods are not capable of doing so. Second, this chapter also draws attention that it was crucial to investigate what methods are available in the literature that can estimate relative treatment effects with single-arm studies and also if the methods are able to handle multiple comparisons with multiple sources of evidence per comparison simultaneously. All these issues set the scene to conduct a second review to identify methods for unanchored indirect comparisons with single-arm studies.

8.2 A systematic review of methods for unanchored indirect comparisons with single-arm study

After conducting the review on NICE STAs, the next objective was to explore all the available methods in the literature that could be used to estimate relative treatment effects with single-arm studies. In order to do so, a second review was carried out (Chapter 3). The review aimed to understand how and under what situations the methods can be used. At the beginning of the review, 682 references were identified which comes down to 13 articles after screening (Section 3.3). Of these 13 articles, 4 methods were identified which could be used with single-arm study to estimate the relative treatment effect. Random baseline NMA method was discussed by Thom et al. (2015); Goring et al. (2016) and Béliveau et al. (2017) and NMA with matching was identified in Leahy et al. (2019). MAIC was identified from several articles including Signorovitch et al. (2010); Cucherat et al. (2020); Jiang and Ni (2020); Remiro-Azócar et al. (2020). STC was identified from Caro and Ishak (2010); Remiro-Azócar et al. (2020).

All four methods identified in the review of methods chapter were different from each other in various aspects (Section 3.4.5). The methods can be classified under two categories. Methods that can estimate a relative treatment effect from a single-arm study with only one comparator treatment or those that can estimate relative treatment effects with multiple comparators with multiple sources of evidence per comparison simultaneously.

8.2.1 Comparison of a treatment with a single comparator

Both matching adjusted indirect comparison (MAIC) and simulated treatment comparison (STC) are suitable for estimating the relative treatment effect of an intervention treatment

with a single comparator. When a treatment from a single-arm study needs to be compared with a comparator treatment after balancing out the baseline covariates, an unanchored MAIC/STC can be applied where the comparator treatment can come from another single-arm study or from one arm of an RCT for which only AgD information is available (Section 3.4.1 & 3.4.2).

MAIC uses the reweighting method in a propensity score logistic regression model whereas STC uses regression adjustment (linear, logistic, or any time-to-event regression model) to balance out all known effect-modifiers and prognostic variables between the IPD study with respect to the AgD study. An overlap between AgD and IPD study variables is needed to apply the logistic model in MAIC, in the absence of which MAIC cannot extrapolate beyond the observed data. However, unlike MAIC, STC is capable of extrapolating beyond the observed values. Nonetheless, MAIC targets a marginal treatment effect whereas the version of STC (anchored) that is described in Phillippo et al. (2016), targets a conditional treatment effect due to the non-collapsibility issue. The version of STC that uses parametric G-computation or model-based standardisation can produce a covariate-adjusted marginal effect estimate (Remiro-Azócar, 2021). Misspecifying the relationship between covariates in the MAIC propensity score regression model or the outcome and covariates in STC regression model can produce biased estimates.

8.2.2 Comparison of a treatment with multiple comparators from multiple sources of evidence per comparison

Random baseline NMA and NMA with matching are suitable for estimating the relative treatment effects of a treatment from a single-arm study to multiple comparators with multiple sources of evidence per comparison (Section 3.4.3 & 3.4.4).

Random baseline NMA uses the assumption of exchangeability that assumes the similarity of a reference arm throughout a disconnected network of evidence and makes the inclusion of single-arm study treatment possible. In NMA with matching, a single-arm study treatment is matched to any arm of a connected network based on the similarity of patient characteristics including both effect-modifier and prognostic variables. The main criticism of random baseline NMA is the assumption of exchangeability that can interfere with randomisation. To address this interference, recently, a reformation has been done on random baseline NMA known as "reference prediction", introduced by Thom et al. (2022) where studies in the NMA are separated into groups by putting RCTs connected to the reference treatment into one group, and single-arm studies into another group. In NMA with matching, any arm of the connected network can be taken for a match as long as the comparator arm covariates are similar to the single-arm study. A reformation has been done on it known as ALM by Thom et al. (2022) where Euclidian distance has been used to assess the closeness of the single-arm study treatment with another arm of the network concerning patient characteristics.

8.2.3 Suitability of methods in different situations

Adjustment and identification of all prognostic and effect-modifier variables is a vital requirement for methods with single-arm studies. MAIC and STC compare an intervention with

a comparator treatment. Although MAIC and STC are not suitable for a larger network of evidence, they are capable of adjusting all known prognostic and effect-modifier variables using the propensity/regression model (Section 3.4.5.4). When this cannot be done, i.e. if unmeasured prognostic and effect-modifier variables exist, residual bias will affect relative treatment effects. MAIC/STC are not appropriate for larger networks as they focus on a different target population in every single comparison, which is the population of the comparator study. They can be extended for a larger network of evidence by including the shared effect modifier assumption. This assumes that the competing treatments belong to the same class with similar clinical properties and they share the same set of treatment effect-modifiers which allows relative treatment effects to be interpreted into any population. This assumption is hard to meet in practice and is untestable.

Random baseline NMA and NMA with matching can compare an intervention with a single-arm study to a larger network of multiple comparators with multiple sources of evidence per comparison (Section 3.4.5.5.2). The larger network can include both RCTs and single-arm studies. NMA-based methods can handle a larger network of evidence, however, to get a valid estimate, they need to satisfy a lot of assumptions such as transitivity, consistency along with the robust assumption of conditional constancy of absolute effects. Failure to satisfy these assumptions makes the methods invalid. Although random baseline NMA was found to be "safer" than NMA with matching as it is more conservative, its main criticism lies in the exchangeability assumption which is often impractical and generates a biased estimate. In the random baseline NMA, covariates can be adjusted by conducting a within-study and between-study covariate adjustment on treatment but it is often the case that an adequate number of studies are not available for each treatment effect for conducting the between-study covariate adjustment. Between-study covariate adjustment/interaction effect is harder to detect than within-study interaction as the former needs to differentiate the interaction effect from the random noise (Dias et al., 2011a). In NMA by matching and in ALM, adjustment of variables is done by finding an arm that matches with the single-arm study. However, this matching can be difficult if the information on important covariates is not present across the network. Additionally, there is no clear guideline on how much similarity can be considered sufficient for a valid comparison. Moreover, the reference prediction and ALM model cannot adjust for treatment effect-modifiers, they can only take into account prognostic variables.

A major issue with NMA-based methods is that though they can be used in a larger disconnected network of evidence, however, to get a trustworthy estimate, the disconnected network of evidence needs to consist of a sufficiently large connected and a disconnected part where several studies should be available per comparison for both parts. A larger disconnected network is common in HTA, however, the network usually consists of several single-arm studies. Therefore, a connected part with several studies per comparison is very rare in NICE STAs with single-arm studies. Furthermore, random baseline NMA is able to use both IPD and AgD but reference prediction and ALM are still not available for that. They can only be used with AgD.

None of the methods found in Chapter 3 can be declared as ideal for every situation as every method functions well when it satisfies certain conditions. The rarity of a connected part

with multiple studies per comparison in a larger disconnected network of evidence makes the NMA-based methods difficult to carry out. Moreover, reference prediction and ALM are not able to adjust for effect-modifier variables which is a key concern. As adjustment of variables is a vital issue with single-arm study, therefore, these NMA-based methods were not further considered. MAIC and STC are useful for estimating treatment effects in a pair of studies by adjusting imbalanced variables but they are not suitable for a larger network without making additional assumptions. However, both MAIC and STC were found to be used in a meta-analysis setting to handle a larger disconnected network of evidence in NICE STA review (Chapter 2) with the recurrent use of the IPD from the single-arm study. All these motivated the design of a simulation study to observe the consequences of the identified issues.

The main contribution of Chapter 3 is that it identifies methods other than MAIC and STC that could be used in estimating relative treatment effects with single-arm studies and discuss the conditions under which the methods are justifiable. This chapter provides a thorough discussion of the identified methods with their pros and cons and discusses the rationale of not using the identified NMA methods for the simulation study.

8.3 Simulation Study with a fixed and a random effects NMA

In Chapter 2 it was noticed that to get a coherent synthesis of overall treatment effect estimates in a larger disconnected network of evidence with a single-arm study, unanchored MAIC or STC were applied in an NMA setting without justifying the shared effect modifier assumption. Additionally, in this process, the IPD from the single-arm study was used multiple times which violates the independence among the studies. To understand the consequence of performing a population-adjusted NMA, a simulation study was designed for binary outcomes (Chapters 4 & 5). The rationale for using binary outcome comes from the discussion of single-arm study in drug development (Section 1.4).

It was perceived that oncology and haemato-oncology are the specific areas where the use of single-arm studies was very prevalent and where time-to-event outcomes are most prevalent for establishing clinical effectiveness followed by binary outcomes. Binary outcomes like overall response rate (ORR) and complete response rate (CRR) are frequently used for establishing clinical effectiveness. In comparison to binary data, simulation with time-to-event event data is quite complex. Therefore, to start with a simple setting, the simulation study in this thesis was done with binary data. However, it is surely possible to extend the simulation study for time-to-event data. A simulation study by Remiro-Azócar et al. (2021) with time-to-event data found that population adjustment method MAIC shows undercoverage where model standard errors underestimated the empirical standard errors with small sample size and low overlap scenarios. Although the simulation study by Remiro-Azócar et al. (2021) was done for a pair of studies with a common treatment, the undercoverage issue matches the results of the simulation in this thesis. This indicates the simulation of single-arm studies with binary data seems extendable for time-to-event data where undercoverage issues are also expected to be found.

To evaluate the population-adjustment method MAIC is of paramount importance as the fre-

quency of using MAIC in HTA submissions turned out to be much higher than STC (Chapter 2), therefore, MAIC was chosen for the simulation study. It was essential to understand the impact of performing MAIC-adjusted NMAs and formally evaluate the performance of this approach.

The goal of the data generation setting in the simulation study was to replicate a common scenario in HTA, where the treatment from a single-arm study needs to make relative treatment comparisons with multiple comparator treatments with multiple sources of evidence per comparison and the single-arm study has IPD on outcome and covariates but only AgD values for comparator treatments. Data generation was done for four settings with 3 treatments using binary outcomes (Section 4.3). The four settings were "connected smaller network of evidence", "connected larger network of evidence", "disconnected smaller network of evidence", and "disconnected larger network of evidence".

A connected network means a collection of RCTs with respect to a common treatment whereas a disconnected network of evidence refers to a collection of studies without any common treatment. Smaller and larger network refers to how many studies were available per comparison. After generating data for a connected network of evidence, it was transformed into a disconnected network of evidence by dropping one arm from each RCT. Both the connected and disconnected networks were generated as smaller (3 studies) and larger (10 studies) where a smaller network of evidence refers to one study per comparison whereas a larger network refers to multiple studies per comparison. To evaluate the MAIC-adjusted NMA with respect to the size of the network was an essential issue as in HTA submission the number of studies per comparison is often very limited.

The simulation study evaluates the change in five factors in a full factorial design (Section 4.3.2). The five factors were the overall sample size (150 and 500), the correlation coefficient between covariates (weak and strong), the strength of effect-modification and prognostic variable (weak and strong), and between study overlap (weak and strong). The estimands of interest were the overall treatment effect estimates from the MAIC-adjusted NMA. The simulation satisfies the shared effect modifier assumption as it was needed to apply the MAIC estimates in an NMA. For a smaller and larger disconnected network of evidence, MAIC was applied multiple times and then a MAIC-adjusted fixed and random effects NMA was performed. Robust SE were used for estimating the uncertainty of the NMA estimates.

All the coding was done using R programming language and I am claiming myself as the sole contributor for all the R codes. Developing all the R codes and running them to get simulation results were quite challenging. Before starting the R coding, it needed a clear understanding and perfectness of R syntax e.g. looping. Moreover, all the R codes were run using high performance computer (HPC). Running codes in HPC is not straightforward as one needs to learn the Linux operating system and also getting all the results can be quite time-consuming as one needs to wait in a queuing process to get his/her HPC results. It took me a solid month to learn how to use HPC and how I can adjust my R codes to run in HPC using R codes on parallelism.

Overall, the findings from the random effects MAIC-adjusted NMA were similar to the fixed effect NMA. The major impact of performing an MAIC-adjusted NMA was seen in the coverage of NMA estimates for each DGM. Deviation from nominal (95%) coverage was seen as a consequence of repeated use of IPD from the single-arm study. Additionally, the robust SE was not able to capture the true variability of the NMA estimates which resulted in low coverage. The undercoverage issue with the MAIC-adjusted NMA was less severe with the smaller network of evidence compared to the larger network of evidence. This was because the assumption of independence of the unit of analysis was broken more in a larger network of evidence which resulted in a more severe undercoverage issue with the larger network. Moreover, bias was found to be inversely associated with overlap between studies and higher biases were found with high effect-modifier and prognostic variable levels.

The difference between the undercoverage of fixed and random effects NMA was that the undercoverage issue was slightly better (close to 95%) for the random effects MAIC-adjusted NMA model compared to the fixed effect model. As the heterogeneity parameter τ added some extra level of randomness, it inflates the SE of the NMA estimates which in turn causes an improvement in coverage. Undercoverage issue was found to be better with a high overlap between study covariates and a high correlation of within-study covariates. Moreover, for the random effects MAIC-adjusted NMA model, the coverage was found to be slightly better with a bigger sample size.

Overall, the simulation study shows that when no prognostic and effect-modifiers variables were missing (conditional constancy of absolute effects), the shared effect-modifier assumptions were satisfied, and a high level of overlap exists between studies, a MAIC-adjusted NMA was able to estimate the true relative treatment effect of treatments. Although the MAIC-adjusted NMA was not free of biases, however, the magnitude of biases was not tremendously unacceptable following Schafer and Graham (2002). The main concern lies in the undercoverage issue which makes the NMA estimates unreliable. Therefore, it was of vital importance to find a solution to this issue. The main contribution of the simulation study is that it attempts to assess the consequence of repeated use of IPD in an NMA setting using estimates from a population-adjustment method with a robust sandwich estimator. Even so, the simulation study is not completely flawless as the study did not explore the consequences when MAIC-adjusted NMA is applied violating the aforementioned assumptions and the overlap issue.

8.3.1 Comparison of the simulation study with existing studies in literature

Since the release of TSD 18 by Phillippo et al. (2016), several simulation studies have been done to explore the appropriateness of the population-adjustment methods (Ishak et al., 2015b; Belger et al., 2015b; Kühnast et al., 2017; Leahy and Walsh, 2019; Hatswell et al., 2020; Phillippo et al., 2020b; Remiro-Azócar et al., 2020). In this section, first a discussion will be conducted on existing simulation studies in literature and then a comparison will be made with the simulation in this thesis.

In the simulation study by Jiang and Ni (2020), the bias and efficiency of unanchored MAIC

was assessed with two single-arm studies for time-to-event outcomes. Unanchored MAIC was found to have the potential to generate unbiased estimates if all effect-modifier as well as prognostic variables were balanced. In the absence of effect-modification, biases were still found due to the non-balancing of prognostic variables. Hatswell et al. (2020) assess the performance of unanchored MAIC where an intervention was compared to a control by matching on first or higher moments using a variety of conditions for time-to-event outcomes. A variety of scenarios were then tested by varying several factors including the survival model, type of variables for matching (binary as opposed to continuous), the relative importance of covariates, efficacy of treatment, matching also on nuisance parameters, degree of overlap, effects of the unobserved variables and so on. Under suitable conditions of higher overlap and matching on important variables, MAIC appeared as a competent method to address biases. However, with a small sample size, omission of imbalanced important variables, poor overlap or matching on variables with limited impact on the outcome causes biased estimates. When MAIC showed poor performance, it worsened if matching was also performed on higher moments.

A simulation study by Remiro-Azócar et al. (2020) compared standard unadjusted indirect comparisons (Bucher method), anchored MAIC and STC under a variety of scenarios for time-to-event outcomes with continuous covariates. MAIC was found to be the least biased method followed by conventional STC and the Bucher method. Standard errors and coverage rates are often valid in MAIC but with small sample sizes and poor covariate overlap, model SE underestimate the empirical SE due to the use of a robust sandwich estimator for estimating SE that results in empirical coverage rates significantly below the nominal coverage.

Phillippo et al. (2020b) have done an extensive simulation study to compare ML-NMR, anchored MAIC and STC for binary outcomes in a range of scenarios under various failures of assumptions. The factors that were varied were sample size, missing effect-modifiers, effect-modification strength, varying between-study overlap, validity of the shared effect modifier assumption, validity of extrapolation, different correlations and covariate distributions. ML-NMR and STC outperformed MAIC by eliminating bias when the requisite assumptions were met. Unlike other studies, the validity of MAIC was seriously questioned as it performed poorly in all simulation scenarios. In a few scenarios with low overlap and small sample size, MAIC even showed larger bias in comparison to standard indirect comparison i.e. to Bucher method. As a weighting method the inability of MAIC to extrapolate when the AgD study is not contained sufficiently within the population of the IPD study causes this poor performance. Justifications were also mentioned as to why other studies failed to discover this drawback of MAIC. Either in other studies, a good amount of overlap exists between studies with continuous covariates or they were focused on binary covariates, where issues only arise when covariate proportions are close to zero or one in the IPD study. Phillippo et al. (2020b) found that though MAIC SE was estimated using bootstrapping, however, deviation from nominal coverage was again found for some scenarios. Bias was found to be associated with deviation from the nominal coverage level. Bias and coverage were found to be inversely related. However, the opposite was seen for coverage and overlap. With high overlap between studies, MAIC was able to reduce more bias that results in higher coverage.

Leahy and Walsh (2019) did a simulation study to assess the impact of using anchored MAIC prior to running a Bayesian network meta-analysis with time-to-event outcomes for a connected network of evidence. In total 4 studies were simulated where 3 studies have IPD and 1 study has AgD. When the distribution of covariates in the IPD studies was very different from the AgD studies, i.e. with low overlap between studies the standard NMA model in general gives worse estimates (high MAE) in comparison to the estimates when covariates are similar between the studies. Posterior SD was used to measure uncertainty. With low overlap between IPD and AgD studies, the MAIC gives a posterior SD that was similar to the standard NMA model for both the direct and the indirect estimates of treatments. This means uncertainty in the estimates increases by running a MAIC. However, the MAE of the indirect estimate was quite similar for the standard NMA model and the MAIC model. Both fixed effect and random effects NMA models were fitted where the coverage of the random effects model was closer to the nominal 95% CrI than the fixed effect model. The posterior SD was much smaller for the fixed effect models compared with the random effects models, which explains the lower coverage. The coverage was found to be decrease when the covariate-treatment interaction increased. When the IPD studies had a relatively low overlap with the AgD study, the MAIC model generated a larger measure of heterogeneity in comparison to the standard NMA models. The large amount of reweighting probably caused this.

Most of the simulation studies that exist in the literature are for anchored cases with survival outcomes. Although studies by Jiang and Ni (2020) and Hatswell et al. (2020) were on unanchored MAIC, those were conducted for a pair of studies. Two studies were identified that have included MAIC in a NMA setting (Leahy and Walsh (2019); Belger et al. (2015b)), however, these were also for a connected network of evidence using anchored MAIC. The study by Belger et al. (2015b) was not available as a full article, therefore it was not discussed. No study has been found that has applied unanchored MAIC in a NMA setting. Therefore, it may be reasonable to claim that the simulation study carried out in this thesis is the first attempt to observe the consequence of unanchored MAIC in a NMA setting.

As in the study by Hatswell et al. (2020), small sample size and poor overlap were also found to be associated with biases in the simulation study in this thesis. Undercoverage was also found in the study by Remiro-Azócar et al. (2020) where it was claimed that the use of robust SE was the cause of the undercoverage of MAIC estimates. Undercoverage was also seen in the study by Phillippo et al. (2020b) where bootstrapping was used for estimating SE. Phillippo et al. (2020b) explains the deviation from nominal coverage rates as a result of both the bias and the SE whereas, Remiro-Azócar et al. (2020) suggests that for MAIC, the sandwich estimator is the sole cause for low coverage. The simulation study in this thesis also used a sandwich estimator for estimating uncertainty for MAIC weights. However, the simulation in this thesis did not test undercoverage within each MAIC. The undercoverage issue was found for each MAIC-adjusted NMA estimates. Higher biases were found with low overlap scenarios where the undercoverage issue was found to get more severe. However, the highest biases were seen for the smaller disconnected network of evidence where the undercoverage issue was mainly found in low-overlap cases. On the contrary, undercoverage worsened when the same IPD was used multiple times in a larger disconnected network of evidence compared to a smaller network of evidence. Therefore, the simulation study shows that the

use of the sandwich estimator along with the repeated use of IPD affected the coverage of NMA estimates.

Issues identified in the study by Leahy and Walsh (2019) between standard NMA and MAIC NMA cannot be compared with the simulation study in this thesis. Along with the shared effect modifier assumption in the simulation study, the factors were also not varied as they were in the study by Leahy and Walsh (2019). In this thesis, the objective of the inclusion of the standard NMA was to make sure that the R codes are working correctly and also to check whether the coverage of the standard NMA at the nominal level or not. Leahy and Walsh (2019) also found the undercoverage issue. Moreover, parallel to Leahy and Walsh (2019), the simulation study also found that the random effects model was closer to the nominal 95% CrI in comparison to the fixed effect model. When running an NMA model with MAIC, Leahy and Walsh (2019) preserve the independence assumption of the unit of analysis by either weighting covariates for separate IPD studies or pooling all IPD studies to conduct a MAIC. This was not the case for the simulation in this thesis as one of the objectives was to assess the impact of using the same IPD repeatedly for conducting multiple MAICs.

8.4 Development of double-bootstrapping with MAIC

For the purpose of overcoming the undercoverage problem, a new and novel method called "double-bootstrapping with MAIC" was proposed and applied to increase the coverage to nominal level (95%) (Chapter 6). To evaluate the MAIC with double-bootstrapping, a simulation study was designed. In the simulation study, data were generated for a larger connected and disconnected network of evidence to apply double-bootstrapping both for fixed and random effects MAIC-adjusted NMA. The data generation for the connected and disconnected network of evidence was the same as it was described in Chapters 4 and 5 (Section 4.3 & 5.3). To execute double-bootstrapping in a MAIC-adjusted NMA, when the IPD from the single-arm study needs to be used in MAIC, instead of taking the original IPD, double-bootstrapping should be applied. In this process, for each individual MAIC, a bootstrap sample should be extracted from the original IPD, and then the bootstrapped sample needs to be bootstrapped again. The second bootstrapped sample is then required to be used for the computation of MAIC weights.

The simulation study evaluated the change in five factors as described in Section 8.3. The estimands of interest were the overall treatment effect estimates from the double-bootstrapped MAIC-adjusted NMA. The simulation satisfied the shared effect modifier assumption as it was needed to apply the MAIC estimates in an NMA setting.

Both for the fixed and random effects MAIC-adjusted NMA, double-bootstrapping was found to increase the coverage of NMA estimates to the nominal (95%) level. Overlap between study covariates, prognostic, and effect-modifying variables was found to have no effect on the coverage. However, biases were found to be lower with high overlap, and low prognostic and effect-modifying variable levels. The highest biases were found for the combination of a small (150) sample size with low overlap scenarios. Therefore, low biases can be achieved with double-bootstrapping for a bigger sample size, but with a smaller sample size and low

overlap, biases can increase substantially.

Though double-bootstrapping was found to solve the problem of undercoverage, it was not completely flawless. The main difficulty of the execution of the double-bootstrapping was that it was computationally very intensive and time-consuming. The simulation study simulated 1000 MCMC samples for each DGM. In each MCMC sample, double-bootstrapping was implemented to calculate MAIC estimates. The MAIC estimates were then used to perform fixed/random effects NMA. This whole process was again bootstrapped 300 (for fixed effect NMA)/1000 (for random effects NMA) times to get bootstrapped NMA estimates. The R codes were run using HPC with R parallelism where each MCMC sample with double-bootstrapping was assigned to one computer processor or core to make the calculation faster. It would take approximately 2 days to get results for 1000 MCMC samples.

The usage of double-bootstrapping is quite frequent in the statistical literature. A double-bootstrap equation was derived by Shi (1992) for confidence limit estimation and double-bootstrapping was found to be better than the percentile method and equally good as a single bootstrap method such as accelerated bias-correction (BCa). Vinod (1995) applied the double-bootstrap approach in a ridge regression to achieve accurate CIs when the classical methods or the single bootstrap were unable to obtain it. Letson and McCullough (1998) proved with an example that CI with double-bootstrapping converges more quickly than a single-bootstrap. McCullough and Vinod (1998) showed with examples that the double-bootstrap has better convergence properties for consistent CI. Arasan and Adam (2014) states that with large sample sizes and low censoring, Wald can still be employed, however, when the sample size is small as well as it is censored and truncated, double-bootstrapped CI should be employed. The coverage probability can significantly deviate from the nominal level when Wald is applied to smaller data sets that are censored or truncated. Chang and Hall (2015) applied double-bootstrapping for bias corrections and CI. They showed that double-bootstrapping is insensitive to the number of resamples used in the second bootstrap stage, where a single second bootstrap sample can be enough for bias correction. In the paper by Chronopoulos et al. (2015) a double bootstrap algorithm was developed which proved to improve coverage probabilities for obtaining confidence intervals.

In the NICE STA review (Chapter 2) it was found that other than the sandwich estimator, bootstrapping was also used as a measure of uncertainty for MAIC weights. However, both methods are subject to undercoverage issues (Phillippo et al., 2020b; Remiro-Azócar et al., 2020). Chapter 6 adds value to the fact that for uncertainty estimates in a MAIC-adjusted NMA estimates, double-bootstrapping could be an alternate approach instead of a sandwich estimator or single bootstrapping. Additionally, several application were found in the literature on double-bootstrapped CI, however, none of them were applied for estimating the coverage of NMA estimates. Therefore, the application of double-bootstrapping with MAIC can be considered as a first attempt with a population-adjustment method in an NMA setting. Moreover, simulation with double-bootstrapping also sheds light on the fact that though double-bootstrapping is capable of giving the nominal level of coverage, however, caution should be made with the combination of low-overlap with a small sample size as the magnitude of biases can increase substantially.

8.5 Double-bootstrapping with a case study

The aim of Chapter 7 was to demonstrate how to implement double-bootstrapping with a case study and also to find practical problems associated with implementing the method. For this, a real dataset from the DREAM study was used (Pavord et al., 2012). In the DREAM study, data was collected on severe eosinophilic asthma exacerbation where the outcome measure was the rate of exacerbations. Prior to performing the double-bootstrapping with the DREAM study, first, a standard connected NMA was performed with 5 available biologics for asthma (Section 7.4.1). The rate of exacerbation was used as the outcome measure. Both fixed and random effects NMA models were estimated in a Bayesian framework with Poisson likelihood and a log link function. After performing the connected NMA with severe eosinophilic asthma, a MAIC-adjusted NMA (with and without double-bootstrapping) was performed (Section 7.4.2 and 7.4.3). For this, the connected network of evidence was converted into single-arm studies. The DREAM study arm 250 mg of mepolizumab was considered to be the single-arm study with IPD and all the RCTs were transformed into single-arm by dropping one arm from each study (Section 7.4.2.1). Multiple unanchored MAICs were performed using IPD from the DREAM study and then with the MAIC estimates a MAIC-adjusted NMA and a double-bootstrapped MAIC-adjusted NMA were performed.

The first problem that occurred during the estimation process of double-bootstrapping was that in the IPD (DREAM study) study no clear information was available on effect-modifying variables, only variables were mentioned which were predictive of the outcome. A data availability issue was found which restricted the adjustment in MAIC only for 4 variables. Eventually, in addition to availability issues, due to non-convergence issues, the MAIC model was able to adjust for only 2 variables. This may have caused residual bias to be present in unanchored MAICs. The next issue that was encountered was the difficulty of the estimation procedure. Double-bootstrapping is computationally very challenging as well as time-consuming. The results of the double-bootstrapping that were presented in Chapter 7 were only for one iteration where the number of bootstraps was 2000 (R=2000). This was run on a personal computer with 4GB RAM and it took almost 2 hours (fixed effect model) and 6 hours (random effects model) to run the codes for one iteration. This procedure needs to be repeated a moderately large number of times (for example 2000) where each time the bootstrapping should be done sufficiently with a big number (for example 2000). Moreover, the individual MAICs in a MAIC-adjusted NMA is capable of adjusting for covariates that are imbalanced between studies, it does not guarantee that when the estimates are included in an NMA setting, the distribution of effect-modifiers will still be same across the network. Therefore, in implementing the double-bootstrapping in a NMA setting, the shared effect modifier assumption was made which may not be true in practice. As a consequence, the estimates can show biased results. Furthermore, in MAIC, low-overlap between studies is a crucial issue as MAIC cannot extrapolate in low-overlap cases. In Chapter 6 it was found that double-bootstrapping shows high bias with a small sample size and low overlap.

The contribution of this chapter is that it shows the steps and identifies practical problems in the implementation of the double-bootstrapping approach with a real dataset.

8.6 Strengths and limitations of the research

The strengths of the thesis can be described with some key aspects. Firstly, the simulation study that has been undertaken in the thesis allows for exploring the consequences of recurrent use of IPD with a single-arm study by implementing population-adjustment method MAIC in a NMA for a disconnected network of evidence. Moreover, the simulation study also examines the impact of varying the size of the network of evidence which is an essential issue as in HTA a limited number of studies per treatment comparison is quite prevalent. Undercoverage of the NMA estimates was identified to be the main concern in MAIC-adjusted NMA. In addition, the thesis attempts to solve the undercoverage issue of NMA estimates by developing a novel method called "double-bootstrapping with MAIC". Furthermore, the practical issues that could be encountered in the execution of double-bootstrapping have been illustrated with an example using data from an RCT.

In addition to the strengths of the thesis, the limitations need to be considered also. It is worth noting that, the simulation study in this thesis has been done for binary outcomes. However, survival outcomes are the most prevalent outcome types in single-arm studies which is evident from Chapter 1. Further simulation studies are required with time-to-event outcomes. Furthermore, the simulation study includes only one prognostic and effect modifier variable. In practice, it is not uncommon to find more than 10 covariates being balanced during the adjustment process (Phillippo et al., 2019a). Moreover, during the simulation process, both a smaller and larger connected network of evidence was simulated where a network was defined as smaller or larger according to how many studies were available per comparison. A smaller network of evidence refers to a network of studies where there is only one study per comparison whereas a larger network of evidence refers to multiple studies per comparison. Nevertheless, in HTA, smaller and larger networks are commonly built according to how many treatments are evaluated in the network. The simulation study in this thesis did not evaluate the effect of MAIC-adjusted NMA for this particular setting.

By design the simulation study upholds all the assumptions that are required for indirect treatment comparisons and valid population adjustments. It does not tell us the consequences in case of failures in assumptions. During the data generation for the simulation study, a shared effect-modifier assumption was met which kept the effect-modifier variables balanced across the disconnected network of evidence. In practice, this can rarely be seen. Additionally, in the simulation study, the data-generating mechanism was known, therefore, prognostic and effect modifiers were also known. The simulation study satisfied no unmeasured effect-modifiers and prognostic variables assumption when conducting the unanchored MAICs which is again not very common in practice. Typically, one cannot make this determination in practice because full information on covariate data may not be measured or reported. Moreover, determining the effect modifier status of variables can be challenging, especially for novel treatments with minimal expertise in the clinical sector and earlier empirical data. The logistic regression model for estimating the MAIC weights was correct in the simulation study because all the covariates have been included in the model and the balancing property holds for the weights. The misspecification of the logistic regression model by the incorrect omission of covariates was not assessed in the simulation study.

The MAIC approach was predicated on the internal validity of included studies, which calls for suitable designs, the absence of non-compliance, and suitable sample sizes. This indicates that cross-trial changes that are perfectly confounded with treatment nature cannot be accounted for by MAIC. Moreover, overlap between study covariates is a major issue with MAIC as it can not function well in the case of low overlap between studies. Therefore, in the simulation study, the amount of overlap was kept sufficiently high. In practice, this is often not the case. In MAIC, the covariate correlations of the AgD study need to be included in the weighting model. The correlations are taken to be equal to the correlations between covariates in the pseudo population created by weighting the AgD population since it is common that no covariate correlation information from the AgD population can be obtained. Therefore, it cannot be balanced by the inclusion in the weighting model. Bias will arise from this assumption if higher-order interactions involving two or more factors are not present or omitted. the simulation study did not evaluate the consequence of violating this assumption. The simulation study with the double-bootstrapping approach also enlightens that caution should be made when double-bootstrapping is applied with small sample sizes and low overlap of covariates between studies as the magnitude of biases can turn out to be troublesome.

8.7 Scope for further research

The findings from the simulation study show the repercussions of implementing unanchored MAIC in a larger disconnected network of evidence satisfying the shared effect-modifier and conditional consistency of absolute effects i.e. no missing prognostic and effect-modifier assumptions. Future study needs to be done to explore the consequences when these assumptions are not met. Additionally, the simulation was done for binary outcomes, future studies should evaluate MAIC-adjusted NMA for continuous and time-to-event outcomes also. In the simulation study of this thesis, the larger disconnected network of evidence was designed with 10 studies (6 studies with treatment 2 and 4 studies with treatment 3) and double-bootstrapping was used for this larger disconnected network of evidence. Future studies should evaluate the effect of double-bootstrapping for a network of evidence larger than 10 studies. Other simulation studies need to be done where the smaller and larger networks are built according to the number of treatments instead of the number of studies per treatment comparison. Furthermore, two levels of covariate overlap were used in the simulation study termed as low (61%) and high (88%), however, none of the overlap was really low. Further studies should evaluate the consequences of MAIC-adjusted NMA with a further reduction of overlap. Moreover, the simulation was designed for unanchored MAIC. Other simulation studies should be designed to investigate the consequences of implementing unanchored STC in a larger disconnected network of evidence.

8.8 Conclusion

The main contribution of this thesis is that it has developed a method called “double-bootstrapping with MAIC” to estimate the SE of MAIC-adjusted NMA estimates. MAIC with double-bootstrapping can be used in the case when a treatment from a single-arm study

needs to be compared with multiple comparators with multiple sources of evidence per comparator with binary outcome. Double-bootstrapping breaks the correlation that emerged due to the use of same IPD in a MAIC-adjusted NMA. Instead of using the original IPD in conducting multiple MAICs, the IPD was bootstrapped twice for each MAIC, and then the final bootstrapped IPD was used in calculating MAIC estimates for a NMA. Both for the fixed and random effects NMA of a larger disconnected network of evidence, with double-bootstrapping, bias was found to be inversely related to overlap. Lower biases were found with high overlap levels and vice-versa. The general trend was higher bias with higher prognostic and effect-modifying variable levels. Moreover, the magnitude of biases was found to be lower for the sample size of 500 compared to the sample size of 150. Double-bootstrapping was found to increase the coverage to the nominal (95%) level both for fixed and random effects NMA. Additionally, though bias was affected mainly due to overlap between study covariates as well as prognostic and effect-modifying variables, coverage was found to be unaffected by these factors. The main challenge found with double-bootstrapping was that it is computationally very challenging and resource intensive.

The use of MAIC and STC are discouraged in a larger network of evidence (both connected and disconnected) due to the fact that they are developed to balance covariates between a pair of treatments. The anchored and unanchored population-adjusted methods satisfy conditional constancy of the relative effects and conditional constancy of the absolute effects respectively which are less strong assumptions than the constancy of relative effects made for standard NMA with only AgD. MAIC and STC make a comparison where the target population is the population of the AgD study which often does not match the population for the decision. A newly developed method called ML-NMR that is an extended version of meta-regression NMA, is a population-adjustment method that combines IPD and AgD from multiple studies on treatments of interest. In ML-NMR, the IPD is used to define an individual-level regression model like STC, and aggregate data are fitted by integrating over the covariate distribution to form the likelihood. ML-NMR differs from STC as the linear model in ML-NMR is embedded inside a probabilistic model. ML-NMR make use of all available information and is generalisable to treatment networks of any size in any target population. It avoids aggregation bias by design, unlike other meta-regression approaches. Despite all these advantages, the application of this approach still does not include single-arm studies as the method is currently only applicable for a connected network of evidence.

The biases of MAIC-adjusted NMA estimates (with robust SE or with double-bootstrapping) were found to be higher for low overlap between studies along with high prognostic and effect-modifier variables levels. This was found to be true both for the fixed and random effects NMA. This is because the validity of unanchored MAIC-adjusted NMA depends on how correctly the individual MAICs are able to adjust the imbalanced variables between studies. With MAIC, the logistic model starts to produce insensible weights when the overlap starts to decrease significantly. With low-overlap, large weights are assigned to few observations which makes the MAIC estimates unstable with larger SE. In the case of complete non-overlap between studies, MAIC is not able to produce any weight in which case it fails to produce any estimate. In earlier simulation studies it was found that unable to adjust for all effect-modifiers as well as prognostic variables, low overlap between studies, small sample

size, inclusion of variables with limited impact on the outcome, variables that are not linked to the outcome, and variables that are already well-matched between studies causes MAIC to give biases estimates with reduced precision (Ishak et al., 2015b; Belger et al., 2015b; Kühnast et al., 2017; Leahy and Walsh, 2019; Hatswell et al., 2020; Phillippo et al., 2020b; Remiro-Azócar et al., 2020).

This thesis mainly explores the practical consequences of applying unanchored MAICs in a NMA setting to estimate the relative treatment effect of a single-arm study treatment with multiple comparators. The simulation study in this thesis found that breaking the independence of the unit of analysis by using the same IPD multiple times along with robust SE contributes to the presence of undercoverage for NMA estimates. Double-bootstrapping is advised with MAIC-adjusted NMA as a solution to this problem. However, caution should be taken when low overlap exists between studies with a small sample size as the simulation showed that biases were comparatively high with a smaller sample size along with low overlap. Despite the fact that undercoverage was found to be a less serious problem for the smaller disconnected network of evidence, still double-bootstrapping will be advised to be on the safer side.

References

Agache, I. (2019). Severe asthma phenotypes and endotypes. In *Seminars in immunology*, volume 46, page 101301. Elsevier.

Agache, I., Song, Y., Rocha, C., Beltran, J., Posso, M., Steiner, C., Alonso-Coello, P., Akdis, C., Akdis, M., Canonica, G. W., et al. (2020). Efficacy and safety of treatment with dupilumab for severe asthma: A systematic review of the eaaci guidelines—recommendations on the use of biologicals in severe asthma. *Allergy*, 75(5):1058–1068.

Agrawal, S., Arora, S., Amiri-Kordestani, L., de Claro, R. A., Fashoyin-Aje, L., Gormley, N., Kim, T., Lemery, S., Mehta, G. U., Scott, E. C., et al. (2023). Use of single-arm trials for us food and drug administration drug approval in oncology, 2002-2021. *JAMA oncology*, 9(2):266–272.

Akenroye, A. T., Segal, J. B., Zhou, G., Foer, D., Li, L., Alexander, G. C., Keet, C. A., and Jackson, J. W. (2023). Comparative effectiveness of omalizumab, mepolizumab, and dupilumab in asthma: A target trial emulation. *Journal of Allergy and Clinical Immunology*, 151(5):1269–1276.

Arasan, J. and Adam, M. B. (2014). Double bootstrap confidence interval estimates with censored and truncated data. *Journal of Modern Applied Statistical Methods*, 13(2):22.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research*, 46(3):399–424.

Austin, P. C. (2014). The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. *Statistics in medicine*, 33(7):1242–1258.

Balcombe, K., Fraser, I., Latruffe, L., Rahman, M., and Smith, L. (2008). An application of the dea double bootstrap to examine sources of efficiency in bangladesh rice farming. *Applied Economics*, 40(15):1919–1925.

Belger, M., Brnabic, A., Kadziola, Z., Petto, H., and Faries, D. (2015a). Alternative weighting approaches for matching adjusted indirect comparisons (maic). *Value in Health*, 18(3):A31–A32.

Belger, M., Brnabic, A., Kadziola, Z., Petto, H., and Faries, D. (2015b). Inclusion of multiple studies in matching adjusted indirect comparisons (maic). *Value in Health*, 18(3):A33.

Béliveau, A., Goring, S., Platt, R. W., and Gustafson, P. (2017). Network meta-analysis of disconnected networks: How dangerous are random baseline treatment effects? *Research synthesis methods*, 8(4):465–474.

Beran, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika*, 74(3):457–468.

Booth, J. and Presnell, B. (1998). Allocation of monte carlo resources for the iterated bootstrap. *Journal of Computational and Graphical Statistics*, 7(1):92–112.

Booth, J. G. and Hall, P. (1994). Monte carlo approximation and the iterated bootstrap. *Biometrika*, 81(2):331–340.

Buhl, R., Korn, S., Menzies-Gow, A., Aubier, M., Chapman, K. R., Canonica, G. W., Picado, C., Donica, M., Kuhlbusch, K., Korom, S., et al. (2020). Prospective, single-arm, longitudinal study of biomarkers in real-world patients with severe asthma. *The Journal of Allergy and Clinical Immunology: In Practice*, 8(8):2630–2639.

CADTH (2014). Canadian agency for drugs and technologies in health. procedure for the cadth common drug review 2014. `https://www.cadth.ca/media/cdr/process/Procedure_for_CADTH_CDR.pdf`.

Caldwell, D. M., Dias, S., and Welton, N. J. (2015). Extending treatment networks in health technology assessment: how far should we go? *Value in Health*, 18(5):673–681.

Caro, J. J. and Ishak, K. J. (2010). No head-to-head trial? simulate the missing arms. *Pharmacoeconomics*, 28(10):957–967.

Chang, J. and Hall, P. (2015). Double-bootstrap methods that use a single double-bootstrap simulation. *Biometrika*, 102(1):203–214.

Charles, D., Shanley, J., Temple, S.-N., Rattu, A., Khaleva, E., and Roberts, G. (2022). Real-world efficacy of treatment with benralizumab, dupilumab, mepolizumab and reslizumab for severe asthma: A systematic review and meta-analysis. *Clinical & Experimental Allergy*, 52(5):616–627.

Chronopoulos, D. K., Girardone, C., and Nankervis, J. C. (2015). Double bootstrap confidence intervals in the two-stage dea approach. *Journal of Time Series Analysis*, 36(5):653–662.

Chung, K. F., Wenzel, S. E., Brozek, J. L., Bush, A., Castro, M., Sterk, P. J., Adcock, I. M., Bateman, E. D., Bel, E. H., Bleecker, E. R., et al. (2014). International ers/ats guidelines on definition, evaluation and treatment of severe asthma. *European respiratory journal*, 43(2):343–373.

Cline, G. (2019). *Nonparametric Statistical Methods Using R*. Scientific e-Resources.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

Cohort, G. R.-L. S. A. (2021). Eosinophilic and noneosinophilic asthma. *Chest*, 160(3):814–830.

Collignon, O., Schritz, A., Senn, S. J., and Spezia, R. (2020). Clustered allocation as a way of understanding historical controls: Components of variation and regulatory considerations. *Statistical methods in medical research*, 29(7):1960–1971.

Cooper, N. J., Peters, J., Lai, M. C., Juni, P., Wandel, S., Palmer, S., Paulden, M., Conti, S., Welton, N. J., Abrams, K. R., et al. (2011). How valuable are multiple treatment comparison methods in evidence-based health-care evaluation? *Value in Health*, 14(2):371–380.

Cucherat, M., Laporte, S., Delaitre, O., Behier, J.-M., d'Andon, A., Binlich, F., Bureau, S., Cornu, C., Fouret, C., Labouret, N. H., et al. (2020). From single-arm studies to externally controlled studies. methodological considerations and guidelines. *Therapies*, 75(1):21–27.

Dahabreh, I. J., Robertson, S. E., Petito, L. C., Hernán, M. A., and Steingrimsson, J. A. (2019). Efficient and robust methods for causally interpretable meta-analysis: Transporting inferences from multiple randomized trials to a target population. *Biometrics*.

Dewey, A. and Drahota, A. (2016). Introduction to systematic reviews: online learning module cochrane training. *Retrieved from*.

Dias, S. and Ades, A. (2016). Absolute or relative effects? arm-based synthesis of trial data. *Research synthesis methods*, 7(1):23.

Dias, S., Sutton, A. J., Ades, A., and Welton, N. J. (2013). Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Medical Decision Making*, 33(5):607–617.

Dias, S., Sutton, A. J., Welton, N. J., and Ades, A. (2011a). Nice dsu technical support document 3: heterogeneity: subgroups, meta-regression, bias and bias-adjustment.

Dias, S., Welton, N. J., Sutton, A. J., and Ades, A. (2011b). Nice dsu technical support document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials.

Djulbegovic, B., Glasziou, P., Klocksieben, F. A., Reljic, T., VanDenBergh, M., Mhaskar, R., Ioannidis, J. P., and Chalmers, I. (2018). Larger effect sizes in nonrandomized studies are associated with higher rates of ema licensing approval. *Journal of Clinical Epidemiology*, 98:24–32.

Donegan, S., Williamson, P., D'Alessandro, U., and Tudur Smith, C. (2013). Assessing key assumptions of network meta-analysis: a review of methods. *Research synthesis methods*, 4(4):291–323.

Eichler, H.-G., Bloechl-Daum, B., Abadie, E., Barnett, D., König, F., and Pearson, S. (2010). Relative efficacy of drugs: an emerging issue between regulatory agencies and third-party payers. *Nature reviews Drug discovery*, 9(4):277–291.

Evans, S. R. (2010). Clinical trial structures. *Journal of experimental stroke & translational medicine*, 3(1):8.

Faria, R., Alava, M. H., Manca, A., and Wailoo, A. J. (2015). Nice dsu technical support document 17: The use of observational data to inform estimates of.

Fleetwood, K. (2020). An introduction to network meta-analysis (mixed treatment comparison / indirect treatment comparison) - quantics biostatistics. `https://www.quantics.co.uk/blog/an-introduction-to-network-meta-analysis/`.

Goring, S., Gustafson, P., Liu, Y., Saab, S., Cline, S., and Platt, R. (2016). Disconnected by design: analytic approach in treatment networks having no common comparator. *Research synthesis methods*, 7(4):420–432.

Goring, S., Taylor, A., Müller, K., Li, T. J. J., Korol, E. E., Levy, A. R., and Freemantle, N. (2019). Characteristics of non-randomised studies using comparisons with external controls submitted for regulatory approval in the usa and europe: a systematic review. *BMJ open*, 9(2):e024895.

Greenland, S. (1987). Interpretation and choice of effect measures in epidemiologic analyses. *American journal of epidemiology*, 125(5):761–768.

Griffiths, E. A., Macaulay, R., Vadlamudi, N. K., Uddin, J., and Samuels, E. R. (2017). The role of noncomparative evidence in health technology assessment decisions. *Value in Health*, 20(10):1245–1251.

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political analysis*, pages 25–46.

Harari, O., Soltanifar, M., Cappelleri, J. C., Verhoek, A., Ouwens, M., Daly, C., and Heeg, B. (2022). Network meta-interpolation: Effect modification adjustment in network meta-analysis using subgroup analyses. *Research Synthesis Methods*.

Hartman, E., Grieve, R., Ramsahai, R., and Sekhon, J. S. (2015). From sample average treatment effect to population average treatment effect on the treated: combining experimental with observational studies to estimate population treatment effects. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pages 757–778.

Hatswell, A. J., Baio, G., Berlin, J. A., Irs, A., and Freemantle, N. (2016). Regulatory approval of pharmaceuticals without a randomised controlled study: analysis of ema and fda approvals 1999–2014. *BMJ open*, 6(6).

Hatswell, A. J., Freemantle, N., and Baio, G. (2020). The effects of model misspecification in unanchored matching-adjusted indirect comparison (maic): Results of a simulation study. *Value in Health*.

Henriksen, D. P., Bodtger, U., Sidenius, K., Maltbaek, N., Pedersen, L., Madsen, H., Andersson, E. A., Norgaard, O., Madsen, L. K., and Chawes, B. L. (2020). Efficacy of omalizumab in children, adolescents, and adults with severe allergic asthma: a systematic review, meta-analysis, and call for new trials using current guidelines for assessment of severe asthma. *Allergy, Asthma & Clinical Immunology*, 16(1):1–12.

Hoaglin, D. C., Hawkins, N., Jansen, J. P., Scott, D. A., Itzler, R., Cappelleri, J. C., Boersma, C., Thompson, D., Larholt, K. M., Diaz, M., et al. (2011). Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ispor task force on indirect treatment comparisons good research practices: part 2. *Value in health*, 14(4):429–437.

IQWiG (2016). Institute for quality and efficiency in health care. iqwig general methods version 4.2 2016. `https://www.iqwig.de/download/IQWiG_General_Methods_Version_%204-2.pdf`.

Ishak, K., Rael, M., Phatak, H., Masseria, C., and Lanitis, T. (2015a). Simulated treatment comparison of time-to-event (and other non-linear) outcomes. *Value in Health*, 18(7):A719.

Ishak, K. J., Proskorovsky, I., and Benedict, A. (2015b). Simulation and matching-based approaches for indirect comparison of treatments. *Pharmacoeconomics*, 33(6):537–549.

Ishwaran, H. and Lu, M. (2019). Standard errors and confidence intervals for variable importance in random forest regression, classification, and survival. *Statistics in medicine*, 38(4):558–582.

Jansen, J. P., Crawford, B., Bergman, G., and Stam, W. (2008). Bayesian meta-analysis of multiple treatment comparisons: an introduction to mixed treatment comparisons. *Value in Health*, 11(5):956–964.

Jiang, Y. and Ni, W. (2020). Performance of unanchored matching-adjusted indirect comparison (maic) for the evidence synthesis of single-arm trials with time-to-event outcomes. *BMC medical research methodology*, 20(1):1–9.

Kühnast, S., Schiffner-Rohe, J., Rahnenführer, J., and Leverkus, F. (2017). Evaluation of adjusted and unadjusted indirect comparison methods in benefit assessment. *Methods of information in medicine*, 56(03):261–267.

Lambert, P. C., Sutton, A. J., Abrams, K. R., and Jones, D. R. (2002). A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *Journal of clinical epidemiology*, 55(1):86–94.

Leahy, J., Thom, H., Jansen, J. P., Gray, E., O'Leary, A., White, A., and Walsh, C. (2019). Incorporating single-arm evidence into a network meta-analysis using aggregate level matching: Assessing the impact. *Statistics in medicine*, 38(14):2505–2523.

Leahy, J. and Walsh, C. (2019). Assessing the impact of a matching-adjusted indirect comparison in a bayesian network meta-analysis. *Research synthesis methods*, 10(4):546–568.

Lee, S. M. and Young, G. A. (1999). The effect of monte carlo approximation on coverage error of double-bootstrap confidence intervals. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(2):353–366.

Letson, D. and McCullough, B. (1998). Better confidence intervals: The double bootstrap with no pivot. *American journal of agricultural economics*, 80(3):552–559.

Lin, T.-Y., Poon, A. H., and Hamid, Q. (2013). Asthma phenotypes and endotypes. *Current opinion in pulmonary medicine*, 19(1):18–23.

López-López, J. A., Sterne, J. A., Thom, H. H., Higgins, J. P., Hingorani, A. D., Okoli, G. N., Davies, P. A., Bodalia, P. N., Bryden, P. A., Welton, N. J., et al. (2017). Oral anticoagulants for prevention of stroke in atrial fibrillation: systematic review, network meta-analysis, and cost effectiveness analysis. *bmj*, 359.

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

Lugogo, N. L., Akuthota, P., Mathur, S. K., Burnette, A., Sumino, K., Arya, N., Clarke, N., Lindsley, A. W., Ackert, J.-P. L., Ambrose, C., et al. (2022). Passage: A phase 4, single-arm, open-label study to evaluate the effectiveness and safety of tezepelumab in patients with severe asthma, including under-represented groups. *Chest*, 162(4):A37–A40.

Martin, M. A. (1990). On bootstrap iteration for coverage correction in confidence intervals. *Journal of the American Statistical Association*, 85(412):1105–1118.

McCullough, B. and Vinod, H. (1998). Implementing the double bootstrap. *Computational Economics*, 12:79–95.

Miguel, A., HERNAN, R., and JAMES, M. (2023). *Causal Inference: What If*. CRC PRESS.

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

NICE (2009). Guide to the single technology appraisal process. `https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/Guide-to-the-single-technology-appraisal-process.pdf`.

NICE (2013). Guide to the methods of technology appraisal 2013.

NICE (2020). National institute for health and care excellence. mepolizumab for treating severe refractory eosinophilic asthma. final appraisal determination. `https://www.nice.org.uk/guidance/ta431/documents/final-appraisal-determinationdocument`.

NICE (2022). Nice health technology evaluations: the manual.

O'Connor, A. M., Coetzee, J. F., da Silva, N., and Wang, C. (2013). A mixed treatment comparison meta-analysis of antibiotic treatments for bovine respiratory disease. *Preventive veterinary medicine*, 110(2):77–87.

Pavord, I. D., Korn, S., Howarth, P., Bleecker, E. R., Buhl, R., Keene, O. N., Ortega, H., and Chanez, P. (2012). Mepolizumab for severe eosinophilic asthma (dream): a multicentre, double-blind, placebo-controlled trial. *The Lancet*, 380(9842):651–659.

Petto, H., Kadziola, Z., Brnabic, A., Saure, D., and Belger, M. (2019). Alternative weighting approaches for anchored matching-adjusted indirect comparisons via a common comparator. *Value in Health*, 22(1):85–91.

Phillippo, D., Ades, T., Dias, S., Palmer, S., Abrams, K. R., and Welton, N. (2016). Nice dsu technical support document 18: methods for population-adjusted indirect comparisons in submissions to nice.

Phillippo, D. M. (2021). multinma: An r package for bayesian network meta-analysis of individual and aggregate data. In *Evidence Synthesis and Meta-Analysis in R Conference 2021*.

Phillippo, D. M., Ades, A. E., Dias, S., Palmer, S., Abrams, K. R., and Welton, N. J. (2018). Methods for population-adjusted indirect comparisons in health technology appraisal. *Medical Decision Making*, 38(2):200–211.

Phillippo, D. M., Dias, S., Ades, A., Belger, M., Brnabic, A., Schacht, A., Saure, D., Kadziola, Z., and Welton, N. J. (2020a). Multilevel network meta-regression for population-adjusted treatment comparisons. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 183(3):1189–1210.

Phillippo, D. M., Dias, S., Ades, A., and Welton, N. J. (2020b). Assessing the performance of population adjustment methods for anchored indirect comparisons: A simulation study. *Statistics in Medicine*, 39(30):4885–4911.

Phillippo, D. M., Dias, S., Ades, A., and Welton, N. J. (2020c). Equivalence of entropy balancing and the method of moments for matching-adjusted indirect comparison. *Research synthesis methods*, 11(4):568–572.

Phillippo, D. M., Dias, S., Elsada, A., Ades, A., and Welton, N. J. (2019a). Population adjustment methods for indirect comparisons: a review of national institute for health and care excellence technology appraisals. *International journal of technology assessment in health care*, 35(3):221–228.

Phillippo, D. M., Dias, S., Welton, N. J., Caldwell, D. M., Taske, N., and Ades, A. (2019b). Threshold analysis as an alternative to grade for assessing confidence in guideline recommendations based on network meta-analyses. *Annals of internal medicine*, 170(8):538–546.

Pilette, C., Chaudhuri, R., Lee, F., Lee, J., Sanchez, C., Jakes, R., MSc11, P. R., and Howarth, P. M. (2019). International, prospective real-world study of mepolizumab in patients with severe asthma at one year: Realiti-a.

Remiro-Azócar, A. (2021). Target estimands for population-adjusted indirect comparisons. *arXiv preprint arXiv:2112.08023*.

Remiro-Azócar, A. (2022a). Some considerations on target estimands for health technology assessment. *Statistics in Medicine*, 41(28):5592.

Remiro-Azócar, A. (2022b). Two-stage matching-adjusted indirect comparison. *BMC medical research methodology*, 22(1):1–16.

Remiro-Azócar, A., Heath, A., and Baio, G. (2020). Methods for population adjustment with limited access to individual patient data: A simulation study. *arXiv preprint arXiv:2004.14800*.

Remiro-Azócar, A., Heath, A., and Baio, G. (2021). Methods for population adjustment with limited access to individual patient data: a review and simulation study. *Research synthesis methods*, 12(6):750–775.

Remiro-Azócar, A., Heath, A., and Baio, G. (2022). Parametric g-computation for compatible indirect treatment comparisons with limited individual patient data. *Research Synthesis Methods*, 13(6):716–744.

Ren, S., Oakley, J. E., and Stevens, J. W. (2018). Incorporating genuine prior information about between-study heterogeneity in random effects pairwise and network meta-analyses. *Medical Decision Making*, 38(4):531–542.

Riley, R. D., Lambert, P. C., and Abo-Zaid, G. (2010). Meta-analysis of individual participant data: rationale, conduct, and reporting. *Bmj*, 340.

Russek-Cohen, E. (2022). Discussion of "target estimands for population-adjusted indirect comparisons" by antonio remiro-azocar. *Statistics in Medicine*, 41(28):5573–5576.

Schafer, J. L. and Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological methods*, 7(2):147.

Schmitz, S., Adams, R., and Walsh, C. (2013). Incorporating data from various trial designs into a mixed treatment comparison model. *Statistics in medicine*, 32(17):2935–2949.

Schmitz, S., Maguire, Á., Morris, J., Ruggeri, K., Haller, E., Kuhn, I., Leahy, J., Homer, N., Khan, A., Bowden, J., et al. (2018). The use of single armed observational data to closing the gap in otherwise disconnected evidence networks: a network meta-analysis in multiple myeloma. *BMC medical research methodology*, 18(1):1–18.

Scullard, P., Abdelhamid, A., Steel, N., and Qureshi, N. (2011). Does the evidence referenced in nice guidelines reflect a primary care population? *British journal of general practice*, 61(584):e112–e117.

Sealey, S., Edathodu, A., and Mukku, S. R. (2014). International hta referencing–a reality? *Value in Health*, 17(3):A28.

Serret-Larmande, A., Zenati, B., Dechartres, A., Lambert, J., and Hajage, D. (2023). A methodological review of population-adjusted indirect comparisons reveals inconsistent reporting and suggests publication bias. *Journal of Clinical Epidemiology*.

Shang, H. L. (2021). Double bootstrapping for visualizing the distribution of descriptive statistics of functional data. *Journal of Statistical Computation and Simulation*, 91(10):2116–2132.

Shi, S. G. (1992). Accurate and efficient double-bootstrap confidence limit method. *Computational statistics & data analysis*, 13(1):21–32.

Shim, S., Yoon, B.-H., Shin, I.-S., and Bae, J.-M. (2017). Network meta-analysis: application and practice using stata. *Epidemiology and health*, 39.

Signorovitch, J. E., Wu, E. Q., Andrew, P. Y., Gerrits, C. M., Kantor, E., Bao, Y., Gupta, S. R., and Mulani, P. M. (2010). Comparative effectiveness without head-to-head trials. *Pharmacoeconomics*, 28(10):935–945.

Stevens, J. W., Fletcher, C., Downey, G., and Sutton, A. (2018). A review of methods for comparing treatments evaluated in studies that form disconnected networks of evidence. *Research synthesis methods*, 9(2):148–162.

Stuart, E. A., Cole, S. R., Bradshaw, C. P., and Leaf, P. J. (2011). The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(2):369–386.

Sutton, A., Ades, A., Cooper, N., and Abrams, K. (2008). Use of indirect and mixed treatment comparisons for technology assessment. *Pharmacoeconomics*, 26(9):753–767.

TA510 (2018). Daratumumab monotherapy for treating relapsed and refractory multiple myeloma. `https://www.nice.org.uk/guidance/ta783/documents/committee-papers`.

TA522 (2018). Pembrolizumab for locally advanced or metastatic urothelial cancer where cisplatin is unsuitable. .

TA525 (2018). Atezolizumab for treating metastatic urothelial cancer after platinum-based chemotherapy. `https://www.nice.org.uk/guidance/ta525/documents/committee-papers-2`.

TA529 (2018). Crizotinib for treating ROS1-positive advanced non-small-cell lung cancer. `https://www.nice.org.uk/guidance/ta529/documents/committee-papers-2`.

TA530 (2018). Nivolumab for treating metastatic or unresectable urothelial cancer after platinum-based chemotherapy. `https://www.nice.org.uk/guidance/ta530/documents/committee-papers-3`.

TA540 (2018). Pembrolizumab for treating relapsed or refractory classical Hodgkin lymphoma. `https://www.nice.org.uk/guidance/ta540/documents/committee-papers-2`.

TA554 (2018). Tisagenlecleucel for treating relapsed or refractory B-cell acute lymphoblastic leukaemia in people aged up to 25 years. `https://www.nice.org.uk/guidance/ta554/documents/committee-papers`.

TA567 (2019). Tisagenlecleucel-T for treating relapsed or refractory diffuse large B-cell lymphoma. `https://www.nice.org.uk/guidance/ta567/documents/committee-papers`.

TA571 (2019). Brigatinib for treating ALK-positive nonsmall-cell lung cancer after crizotinib. `https://www.nice.org.uk/guidance/ta571/documents/committee-papers`.

TA592 (2019). Cemiplimab for treating metastatic or locally advanced cutaneous squamous cell carcinoma. `https://www.nice.org.uk/guidance/ta802/documents/committee-papers`.

TA604 (2019). Idelalisib for treating follicular lymphoma refractory to 2 treatments. `https://www.nice.org.uk/guidance/ta604/documents/committee-papers`.

TA628 (2020). Lorlatinib for previously treated ALKpositive advanced non-small-cell lung cancer. `https://www.nice.org.uk/guidance/ta628/documents/committee-papers-2`.

TA630 (2020). Larotrectinib for treating NTRK fusionpositive advanced solid tumours. https://www.nice.org.uk/guidance/ta630/documents/committee-papers.

TA643 (2020). Entrectinib for treating ROS1 fusionpositive locally advanced or metastatic non-small-cell lung cancer. https://www.nice.org.uk/guidance/ta643/documents/committee-papers.

TA644 (2020). Entrectinib for treating NTRK fusionpositive solid tumours. https://www.nice.org.uk/guidance/ta644/documents/committee-papers.

TA704 (2021). Trastuzumab deruxtecan for treating HER2-positive unresectable or metastatic breast cancer after 2 or more anti-HER2 therapies. https://www.nice.org.uk/guidance/ta704/documents/committee-papers.

TA716 (2021). Nivolumab with ipilimumab for previously treated metastatic colorectal cancer with high microsatellite instability or mismatch repair deficiency. https://www.nice.org.uk/guidance/ta716/documents/committee-papers.

TA722 (2021). Pemigatinib for treating relapsed or refractory advanced cholangiocarcinoma with FGFR2 alterations. .

TA742 (2021). Selpercatinib for treating advanced thyroid cancer with RET alterations. https://www.nice.org.uk/guidance/ta742/documents/committee-papers.

TA756 (2021). Fedratinib for disease-related splenomegaly and symptoms in myelofibrosis. https://www.nice.org.uk/guidance/ta756/documents/committee-papers.

Thom, H. (2020). Network meta-analysis on disconnected evidence networks. what can be done?

Thom, H., Leahy, J., and Jansen, J. P. (2022). Network meta-analysis on disconnected evidence networks when only aggregate data are available: Modified methods to include disconnected trials and single-arm studies while minimizing bias. *Medical Decision Making*, page 0272989X221097081.

Thom, H. H., Capkun, G., Cerulli, A., Nixon, R. M., and Howard, L. S. (2015). Network meta-analysis combining individual patient and aggregate data from a mixture of study designs with an application to pulmonary arterial hypertension. *BMC medical research methodology*, 15(1):1–16.

Tibshirani, R. J. and Efron, B. (1993). An introduction to the bootstrap. *Monographs on statistics and applied probability*, 57(1).

Truong, B., Tran, L.-A. T., Le, T. A., Pham, T. T., and Vo, T.-T. (2023). Population adjusted-indirect comparisons in health technology assessment: A methodological systematic review. *Research Synthesis Methods*, 14(5):660–670.

Tsimberidou, A.-M., Braiteh, F., Stewart, D. J., and Kurzrock, R. (2009). Ultimate fate of oncology drugs approved by the us food and drug administration without a randomized trial. *Journal of Clinical Oncology*, 27(36):6243–6250.

Turner, R. M., Jackson, D., Wei, Y., Thompson, S. G., and Higgins, J. P. (2015). Predictive distributions for between-study heterogeneity and simple methods for their application in bayesian meta-analysis. *Statistics in Medicine*, 34(6):984–998.

Turner, R. M., Lloyd-Jones, M., Anumba, D. O., Smith, G. C., Spiegelhalter, D. J., Squires, H., Stevens, J. W., Sweeting, M. J., Urbaniak, S. J., Webster, R., et al. (2012). Routine antenatal anti-d prophylaxis in women who are rh (d) negative: meta-analyses adjusted for differences in study design and quality. *PLoS One*, 7(2):e30711.

Vinod, H. D. (1995). Double bootstrap for shrinkage estimators. *Journal of Econometrics*, 68(2):287–302.

Wang, J. (2021). On matching-adjusted indirect comparison and calibration estimation. *arXiv preprint arXiv:2107.11687*.

WHO (2020). World health organization. `https://www.who.int/medical_devices/assessment/en/`.

Appendices

Appendix A

Data extraction for NICE STA review

A.1 Tables1

Table A.1: *NICE STA data extraction table*

TA number	**TA 525**	**TA 604**	**TA 510**
Name of the pivotal study	IMvigor210	study 101-09	MMY2002
Publication date of the appraisal	13/06/2018	02/10/2019	14/03/2018
Therapeutic Area	oncology	oncology	oncology
What types of outcome measures were indirectly compared?	Time-to-event	Time-to-event	Time-to-event
Which time-to-event was indirectly compared	OS	OS+PFS	OS+PFS
What type of network is being considered?	Multiple comparison (larger network)	Single comparison	Multiple comparison (larger network)

Continued on next page

Table A.1 – continued from previous page

TA number	TA 525	TA 604	TA 510
What methodology used for making unanchored indirect comparison	NMA by STC	MAIC	MAIC
How covariates were identified	predictive performance	availability of both study	literature review +clinical expert
How many variables were included in the model	4	not mentioned	MAIC 1= 11, MAIC 2= 5
Were all identified prognostic and effect modifier variables included in the model	yes	no	no
If no what was the reason	not mentioned	lack of availability	lack of availability
Other than prognostic and effect modifier variables, were other variables also included	no	no	no
Did the model only include main effects	yes	not mentioned	not mentioned
Did the model include second order terms	no	not mentioned	not mentioned
MAIC effective sample size(%)	not applicable	3.8(5.27%)	MAIC 1= 84(56.75%) MAIC 2=80(64%)

Continued on next page

Table A.1 – continued from previous page

TA number	TA 525	TA 604	TA 510
If NMA was conducted, any attempt made to check if any inconsistencies are found in the connected part of the network?	no	not applicable	not applicable
If NMA was conducted, was heterogeneity among studies was assesses?	yes	not applicable	not applicable
If yes, what was the amount of heterogeneity identified?	moderate	not applicable	not applicable
Was at least two studies were available on each contrast for the heterogeneity parameter	not mentioned	not applicable	not applicable
Along with the chosen method, were other methods also discussed?	yes	no	no
Was any justification given for the chosen method?	yes	not applicable	not applicable
How many events were available for time to event outcome(%)	not applicable	not available	not available

Continued on next page

Table A.1 – continued from previous page

TA number	TA 525	TA 604	TA 510
In cost effectiveness , what approach was used for extrapolation of time to event data	parametric model with unadjusted survival function	two stage	parametric model with unadjusted survival function
In clinical effectiveness, what adjustment was made for time to event data	adjusted time varying HR from NMA	weighted KM survival function	weighted Cox proportional hazard model
Is overlaping between weighed IPD and reconstructed IPD has been checked/ commented on	not applicable	yes	not applicable
If no, what procedure was taken to ensure overlapping between weighed IPD and reconstructed IPD	not applicable	sample inflation appraoach	not applicable
Was the population for the extrapolation clearly defined?	no	yes	no
Treatment effects are estimated for which population?	IPD	AgD	IPD
Had any justification given for transportable treatment effects if they are estimated for IPD population?	no	not applicable	no
If PH assumption was made, was it tested?	not applicable	not applicable	for unadjusted
What procedure has been taken to measure uncertainty	bootstrap	not mentioned	sandwich estimator

Continued on next page

Table A.1 – continued from previous page

TA number	TA 525	TA 604	TA 510
Any attempt made to estimate residual bias?	no	no	no

Table A.2: *NICE STA data extraction table*

TA number	TA 530	TA 628	TA 643
Name of the pivotal study	CheckMate 275 and CheckMate 032	Study 1001	ALKA, STARTRK-1 and STARTRK-2
Publication date of the appraisal	04/07/2018	13/05/2020	12/08/2020
Therapeutic Area	oncology	oncology	oncology
What types of outcome measures were indirectly compared?	Time-to-event	Time-to-event	Time-to-event
Which time-to-event was indirectly compared	OS+PFS	OS+PFS	OS+PFS
What type of network is being considered?	Multiple comparison (larger network)	Single comparison	Multiple comparison (larger network)
What methodology used for making unanchored indirect comparison	NMA by STC	MAIC	MAIC
How covariates were identified	literature review+ clinical expert	clinical feedback+ Cox regression (univariate and multivariate)	not mentioned
How many variables were included in the model	4 out of 11	4	MAIC 1=6 , MAIC 2=6

Continued on next page

Table A.2 – continued from previous page

TA number	TA 530	TA 628	TA 643
Were all identified prognostic and effect modifier variables included in the model	no	no	no
If no what was the reason	lack of availability	lack of availability	lack of availability
Other than prognostic and effect modifier variables, were other variables also included	no	no	no
Did the model only include main effects	not mentioned	not mentioned	not mentioned
Did the model include second order terms	not mentioned	not mentioned	not mentioned
MAIC effective sample size(%)	not applicable	not mentioned	not mentioned
If NMA was conducted, any attempt made to check if any inconsistencies are found in the connected part of the network?	no	not applicable	not applicable
If NMA was conducted, was heterogeneity among studies was assesses?	no	not applicable	not applicable
If yes, what was the amount of heterogeneity identified?	not mentioned	not applicable	not applicable

Continued on next page

Table A.2 – continued from previous page

TA number	TA 530	TA 628	TA 643
Was at least two studies were available on each contrast for the heterogeneity parameter	no	not applicable	not applicable
Along with the chosen method, were other methods also discussed?	no	yes	no
Was any justification given for the chosen method?	no	no	no
How many events were available for time to event outcome(%)	not available	not available	not available
In cost effectiveness , what approach was used for extrapolation of time to event data	parametric model with unadjusted survival function	parametric model with unadjusted survival function	parametric model with unadjusted survival function
In clinical effectiveness, what adjustment was made for time to event data	adjusted time varying HR from NMA	weighted Cox proportional hazard model	weighted Cox proportional hazard model
Is overlaping between weighed IPD and reconstructed IPD has been checked/ commented on	not applicable	no	no
If no, what procedure was taken to ensure overlapping between weighed IPD and reconstructed IPD	not applicable	no	no

Continued on next page

Table A.2 – continued from previous page

TA number	TA 530	TA 628	TA 643
Was the population for the extrapolation clearly defined?	no	no	no
Treatment effects are estimated for which population?	IPD	IPD	IPD
Had any justification given for transportable treatment effects if they are estimated for IPD population?	no	no	no
If PH assumption was made, was it tested?	not applicable	not mentioned	not mentioned
What procedure has been taken to measure uncertainty	not mentioned	bootstrap	not mentioned
Any attempt made to estimate residual bias?	out sample	no	not mentioned

Table A.3: *NICE STA data extraction table*

TA number	TA 554	TA 567	TA 571
Name of the pivotal study	ENSIGN, ELIANA and B2101J	JULIET trial	ALTA and Study 101
Publication date of the appraisal	21/12/2018	13/03/2019	20/03/2019
Therapeutic Area	oncology	oncology	oncology

Continued on next page

Table A.3 – continued from previous page

TA number	TA 554	TA 567	TA 571
What types of outcome measures were indirectly compared?	Time-to-event	Time-to-event	Time-to-event
Which time-to-event was indirectly compared	EFS+OS	OS+PFS	OS+PFS
What type of network is being considered?	Multiple comparison (larger network)	Single comparison	Multiple comparison (larger network)
What methodology used for making unanchored indirect comparison	MAIC	Unadjusted indirect comparison	MAIC
How covariates were identified	not mentioned	expert opinion	expert opinion
How many variables were included in the model	not mentioned	8	not mentioned
Were all identified prognostic and effect modifier variables included in the model	no	no	no
If no what was the reason	lack of availability	lack of availability+ variation amongst clinician's responses	lack of availability
Other than prognostic and effect modifier variables, were other variables also included	not mentioned	not mentioned	no

Continued on next page

Table A.3 – continued from previous page

TA number	TA 554	TA 567	TA 571
Did the model only include main effects	not mentioned	not applicable	not mentioned
Did the model include second order terms	not mentioned	not applicable	not mentioned
MAIC effective sample size(%)	not mentioned	not applicable	MAIC 1= 67.1 (49.70%) and MAIC 2 =76.5 (56.66%).
If NMA was conducted, any attempt made to check if any inconsistencies are found in the connected part of the network?	not applicable	not applicable	not applicable
If NMA was conducted, was heterogeneity among studies was assesses?	not applicable	not applicable	not applicable
If yes, what was the amount of heterogeneity identified?	not applicable	not applicable	not applicable
Was at least two studies were available on each contrast for the heterogeneity parameter	not applicable	not applicable	not applicable
Along with the chosen method, were other methods also discussed?	no	no	no

Continued on next page

Table A.3 – continued from previous page

TA number	TA 554	TA 567	TA 571
Was any justification given for the chosen method?	not applicable	not applicable	not applicable
How many events were available for time to event outcome(%)	not available	not available	not available
In cost effectiveness , what approach was used for extrapolation of time to event data	Independent parametric model with unadjusted survival function	Independent parametric model with unadjusted survival function	parametric model with unadjusted survival function
In clinical effectiveness, what adjustment was made for time to event data	weighted cox proportional hazard model	not applicable	weighted Cox proportional hazard model
Is overlaping between weighed IPD and reconstructed IPD has been checked/ commented on	not applicable	not applicable	not applicable
If no, what procedure was taken to ensure overlapping between weighed IPD and reconstructed IPD	not applicable	not applicable	not applicable
Was the population for the extrapolation clearly defined?	no	not applicable	no
Treatment effects are estimated for which population?	IPD	not mentioned	IPD

Continued on next page

Table A.3 – continued from previous page

TA number	TA 554	TA 567	TA 571
Had any justification given for transportable treatment effects if they are estimated for IPD population?	not mentioned	no	not mentioned
If PH assumption was made, was it tested?	not mentioned	not applicable	For adjusted comparisons only
What procedure has been taken to measure uncertainty	not mentioned	not applicable	not mentioned
Any attempt made to estimate residual bias?	not mentioned	not applicable	not mentioned

Table A.4: *NICE STA data extraction table*

TA number	TA 522	TA 529	TA 540
Name of the pivotal study	KEYNOTE-052	PROFILE 1001, PROFILE 1014, PROFILE 1007	KEYNOTE-087
Publication date of the appraisal	13/06/2018	04/07/2018	03/09/2018
Therapeutic Area	oncology	oncology	oncology
What types of outcome measures were indirectly compared?	Time-to-event	Time-to-event	Time-to-event

Continued on next page

Table A.4 – continued from previous page

TA number	TA 522	TA 529	TA 540
which time-to-event was indirectly compared	EFS+OS	OS+PFS	PFS
What type of network is being considered?	Multiple comparison (larger network)	Multiple comparison (larger network)	Single comparison
What methodology used for making unanchored indirect comparison	NMA by STC	No indirect comparison	MAIC
How covariates were identified	model predictive performance+ literature review+ expert opinion	not applicable	availability
How many variables were included in the model	5	not applicable	not mentioned
Were all identified prognostic and effect modifier variables included in the model	yes	not applicable	no
If no what was the reason	not applicable	not applicable	lack of availability
Other than prognostic and effect modifier variables, were other variables also included	no	not applicable	not mentioned
Did the model only include main effects	yes	not applicable	not mentioned

Continued on next page

Table A.4 – continued from previous page

TA number	TA 522	TA 529	TA 540
Did the model include second order terms	no	not applicable	not mentioned
MAIC effective sample size(%)	not applicable	not applicable	not mentioned
If NMA was conducted, any attempt made to check if any inconsistencies are found in the connected part of the network?	not mentioned	not applicable	not applicable
If NMA was conducted, was heterogeneity among studies was assesses?	not mentioned	not applicable	not applicable
If yes, what was the amount of heterogeneity identified?	heterogeneity was explored graphically , the digitised survival curves for each trial were overlaid and presented on a single set of axes. The curves were parallel.	not applicable	not applicable
Was at least two studies were available on each contrast for the heterogeneity parameter	yes	not applicable	not applicable
Along with the chosen method, were other methods also discussed?	yes	no	no

Continued on next page

Table A.4 – continued from previous page

TA number	TA 522	TA 529	TA 540
Was any justification given for the chosen method?	yes	not applicable	not applicable
How many events were available for time to event outcome(%)	not available	not available	not available
In cost effectiveness , what approach was used for extrapolation of time to event data	parametric model with unadjusted survival function	not applicable	parametric model with unadjusted survival function
In clinical effectiveness, what adjustment was made for time to event data	adjusted time varying HR from NMA	not applicable	weighted Cox proportional hazard model
Is overlaping between weighed IPD and reconstructed IPD has been checked/ commented on	not applicable	not applicable	not applicable
If no, what procedure was taken to ensure overlapping between weighed IPD and reconstructed IPD	not applicable	not applicable	not applicable
Was the population for the extrapolation clearly defined?	no	not applicable	no
Treatment effects are estimated for which population?	not mentioned	not mentioned	not mentioned

Continued on next page

Table A.4 – continued from previous page

TA number	TA 522	TA 529	TA 540
Had any justification given for transportable treatment effects if they are estimated for IPD population?	not mentioned	not mentioned	not mentioned
If PH assumption was made, was it tested?	not applicable	not applicable	Not mentioned
What procedure has been taken to measure uncertainty	not mentioned	not applicable	not mentioned
Any attempt made to estimate residual bias?	not mentioned	not applicable	not mentioned

Table A.5: *NICE STA data extraction table*

TA number	TA 630	TA 644	TA 592
Name of the pivotal study	LOXO-TRK-14001, NAVIGATE (LOXO-TRK-15002) and SCOUT (LOXO-TRK-15003)	(ALKA, STARTRK-1, STARTRK-2, and STARTRK-NG)	EMPOWER-CSCC 1
Publication date of the appraisal	27/05/2020	12/08/2020	07/08/2019
Therapeutic Area	oncology	oncology	oncology
What types of outcome measures were indirectly compared?	Time-to-event	Time-to-event	Time-to-event

Continued on next page

Table A.5 – continued from previous page

TA number	TA 630	TA 644	TA 592
Which time-to-event was indirectly compared	EFS+OS	OS+PFS	PFS
What type of network is being considered?	Multiple comparison (larger network)	Multiple comparison (larger network)	Single comparison
What methodology used for making unanchored indirect comparison	unadjusted comparison	unadjusted comparison	MAIC+STC
How covariates were identified	not applicable	not applicable	literature review+ clinical expert
How many variables were included in the model	not applicable	not applicable	2 out of 12
Were all identified prognostic and effect modifier variables included in the model	not applicable	not applicable	no
If no what was the reason	not applicable	not applicable	not mentioned
Other than prognostic and effect modifier variables, were other variables also included	not applicable	not applicable	no
Did the model only include main effects	not applicable	not applicable	not mentioned
Did the model include second order terms	not applicable	not applicable	not mentioned

Continued on next page

Table A.5 – continued from previous page

TA number	TA 630	TA 644	TA 592
MAIC effective sample size(%)	not applicable	not applicable	37(34.3%)
If NMA was conducted, any attempt made to check if any inconsistencies are found in the connected part of the network?	not applicable	not applicable	not applicable
If NMA was conducted, was heterogeneity among studies was assesses?	not applicable	not applicable	not applicable
If yes, what was the amount of heterogeneity identified?	not applicable	not applicable	not applicable
Was at least two studies were available on each contrast for the heterogeneity parameter	not applicable	not applicable	not applicable
Along with the chosen method, were other methods also discussed?	no	no	yes
Was any justification given for the chosen method?	no	no	yes
How many events were available for time to event outcome(%)	not available	not available	not available

Continued on next page

Table A.5 – continued from previous page

TA number	TA 630	TA 644	TA 592
In cost effectiveness , what approach was used for extrapolation of time to event data	parametric model with unadjusted survival function	parametric model with unadjusted survival function	one stage
In clinical effectiveness, what adjustment was made for time to event data	adjusted time varying HR from NMA	not applicable	weighted Cox proportional hazard model
Is overlaping between weighed IPD and reconstructed IPD has been checked/ commented on	not applicable	not applicable	not applicable
If no, what procedure was taken to ensure overlapping between weighed IPD and reconstructed IPD	not applicable	not applicable	not applicable
Was the population for the extrapolation clearly defined?	not applicable	not applicable	yes
Treatment effects are estimated for which population?	not mentioned	not mentioned	AgD
Had any justification given for transportable treatment effects if they are estimated for IPD population?	not applicable	not applicable	not mentioned
If PH assumption was made, was it tested?	not mentioned	not mentioned	Not mentioned
What procedure has been taken to measure uncertainty	not mentioned	not mentioned	bootstrap

Continued on next page

Table A.5 – continued from previous page

TA number	TA 630	TA 644	TA 592
Any attempt made to estimate residual bias?	not mentioned	not mentioned	no

Table A.6: *NICE STA data extraction table*

TA number	TA 704	TA 704	TA 722
Name of the pivotal study	DESTINY-Breast01	CheckMate 142	FIGHT-202
Publication date of the appraisal	27/02/2021	28/07/2021	25/08/2021
Therapeutic Area	oncology	oncology	oncology
What types of outcome measures were indirectly compared?	Time-to-event	Time-to-event	Time-to-event
Which time-to-event was indirectly compared	PFS+OS	OS+PFS	PFS
What type of network is being considered?	Multiple comparison (larger network)	Multiple comparison (larger network)	Single comparison
What methodology used for making unanchored indirect comparison	MAIC	MAIC	MAIC
How covariates were identified	literature review+ clinical expert	availability+ clinical expert	no justification

Continued on next page

Table A.6 – continued from previous page

TA number	TA 704	TA 716	TA 722
How many variables were included in the model	8	14	4
Were all identified prognostic and effect modifier variables included in the model	no	no applicable	no
If no what was the reason	lack of availability	lack of availability	not mentioned
Other than prognostic and effect modifier variables, were other variables also included	no	no	no
Did the model only include main effects	not mentioned	not mentioned	not mentioned
Did the model include second order terms	not mentioned	not mentioned	not mentioned
MAIC effective sample size(%)	not available	not available	not available
If NMA was conducted, any attempt made to check if any inconsistencies are found in the connected part of the network?	not applicable	not applicable	not applicable
If NMA was conducted, was heterogeneity among studies was assesses?	not applicable	not applicable	not applicable

Continued on next page

Table A.6 – continued from previous page

TA number	TA 704	TA 716	TA 722
If yes, what was the amount of heterogeneity identified?	not applicable	not applicable	not applicable
Was at least two studies were available on each contrast for the heterogeneity parameter	not applicable	not applicable	not applicable
Along with the chosen method, were other methods also discussed?	no	yes	no
Was any justification given for the chosen method?	not applicable	yes	no
How many events were available for time to event outcome(%)	not available	not available	not available
In cost effectiveness , what approach was used for extrapolation of time to event data	Independent parametric model with unadjusted survival function	Independent parametric model with unadjusted survival function	Independent parametric model with unadjusted survival function
In clinical effectiveness, what adjustment was made for time to event data	weighted Cox proportional hazard model	mean survival from parametric model	weighted Cox proportional hazard model
Is overlaping between weighed IPD and reconstructed IPD has been checked/ commented on	not mentioned	not applicable	not applicable

Continued on next page

Table A.6 – continued from previous page

TA number	TA 704	TA 716	TA 722
If no, what procedure was taken to ensure overlapping between weighed IPD and reconstructed IPD	not mentioned	not mentioned	not applicable
Was the population for the extrapolation clearly defined?	no	yes	no
Treatment effects are estimated for which population?	IPD	IPD	IPD
Had any justification given for transportable treatment effects if they are estimated for IPD population?	not mentioned	no	no
If PH assumption was made, was it tested?	for unadjusted comparison	not applicable	for adjusted comparison
What procedure has been taken to measure uncertainty	bootstrap	not mentioned	bootstrap
Any attempt made to estimate residual bias?	not mentioned	yes	not mentioned

Table A.7: *NICE STA data extraction table*

TA number	TA 742	TA 756
Name of the pivotal study	LIBRETTO-001	142
Publication date of the appraisal	03/11/2021	16/12/2021

Continued on next page

Table A.7 – continued from previous page

TA number	TA 742	TA 756
Therapeutic Area	oncology	oncology
What types of outcome measures were indirectly compared?	Time-to-event	Binary
Which time-to-event was indirectly compared	PFS+OS	SVR+TSS
What type of network is being considered?	single comparison	single comparison
What methodology used for making unanchored indirect comparison	MAIC	MAIC+STC
How covariates were identified	literature review+ clinical expert	clinical expert and univariable + multivariable analysis
How many variables were included in the model	6	3
Were all identified prognostic and effect modifier variables included in the model	no	no
If no what was the reason	lack of availability	lack of availability
Other than prognostic and effect modifier variables, were other variables also included	not mentioned	not mentioned

Continued on next page

Table A.7 – continued from previous page

TA number	TA 742	TA 756
Did the model only include main effects	not mentioned	not mentioned
Did the model include second order terms	not mentioned	not mentioned
MAIC effective sample size(%)	not available	34.4(35.5%)
If NMA was conducted, any attempt made to check if any inconsistencies are found in the connected part of the network?	not applicable	not applicable
If NMA was conducted, was heterogeneity among studies was assesses?	not applicable	not applicable
If yes, what was the amount of heterogeneity identified?	not applicable	not applicable
Was at least two studies were available on each contrast for the heterogeneity parameter	not applicable	not applicable
Along with the chosen method, were other methods also discussed?	no	yes

Continued on next page

Table A.7 – continued from previous page

TA number	TA 742	TA 756
Was any justification given for the chosen method?	no	yes
How many events were available for time to event outcome(%)	not available	not available
In cost effectiveness , what approach was used for extrapolation of time to event data	one stage	parametric model with unadjusted survival function
In clinical effectiveness, what adjustment was made for time to event data	weighted Cox proportional hazard model	weighted risk difference
Is overlaping between weighed IPD and reconstructed IPD has been checked/ commented on	not mentioned	not applicable
If no, what procedure was taken to ensure overlapping between weighed IPD and reconstructed IPD	not applicable	not applicable
Was the population for the extrapolation clearly defined?	no	no
Treatment effects are estimated for which population?	AgD	IPD

Continued on next page

Table A.7 – continued from previous page

TA number	TA 742	TA 756
Had any justification given for transportable treatment effects if they are estimated for IPD population?	not mentioned	not mentioned
If PH assumption was made, was it tested?	for adjusted comparison	no
What procedure has been taken to measure uncertainty	not mentioned	not mentioned
Any attempt made to estimate residual bias?	not mentioned	yes

Appendix B

Review of methods

B.1 Search strategy on MEDLINE

1. Single-arm study$.mp. (1310)
2. Single-arm trial$.mp. (801)
3. Treatment effect$.mp. (41247)
4. Disconnected network$.mp. (35)
5. Historical control$.mp. (7452)
6. No head-to-head.mp. (355)
7. Indirect comparison$.mp. (1955)
8. Network meta-analysis$.mp. (5508)
9. 1 or 2 or 3 or 4 (43342)
10. 5 or 6 or 7 or 8 (14623)
11. 9 and 10 (682)=Search 1

[mp= title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word,

Appendix C

R codes for MAIC-adjusted fixed effect NMA

C.1 MAIC-adjusted fixed effect NMA

```
############################################################
## R codes for data generation to conduct a connected NMA
## and a MAIC adjusted fixed effect NMA
############################################################

library(multinma)
library(dplyr)
library (margins)
library(devtools)
library(MAIC)
library(sandwich)
library(lmtest)
library(boot)
library(parallel)
library(lme4)

################ data generation
rm(list=ls())
set.seed(1128)
### making object for parameter combination

corx<-c(0.20,0.80)                      ##correlation between covariates in a study
b_X1_trt<-c(-log(0.78),-log(0.40))       ## 0.25 and 0.916 (interaction coefficient)
b_X1<-c(-log(0.67),-log(0.33))           ## 0.40 and 1.10  (covariate coefficient)
meanx1<-c(0.45,0.15)                    ## mean of covariates
param.combinations <- expand.grid(corx=corx, b_X1_trt=b_X1_trt,
                                  b_X1=b_X1, meanx1=meanx1)
pc <- param.combinations
```

```
pc<-round(pc, 2)                      ## rounding the values
pc$meanx2<-c(0.48,0.48,0.48,0.48,0.48,0.48,0.48,
             0.48,0.20,0.20,0.20,0.20,0.20,0.20,0.20,0.20) ## adding another column in pc
scenerios<-nrow(pc)                    ### no of scenarios created ## 16scenarios



######## creating a function to generate data for a connected NMA
### Here 10 studies will be genrated which will consist of 40 columns
## Each study is with two arms where
### 150 data will be generated for each arm
### Each study will have two continuous covariates
### in total every study consists of 4 columns in the dataset
### all studies have the same common treatment(trt 1)
### but 6 studies have treatment 2 and four studies have
### treatment 3
#########################################################
d = c(log(1),log(1.5),log(.17))           ## log odds ratio (0, 0.40, -1.77)
ns =10                                    ## no of studies
np = matrix(150, ns, 2)                  ## no  of patient in each study
t = matrix(c(1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,3,3,3,3), ns, 2) ## trt in each arm

gen.data<-function(sdx,corx, meanx1,meanx2, b_X1, b_X1_trt){
  count = 0
  count2=2
  datacel<-matrix(NA, nrow=150, ncol=40)
  for(i in 1:ns){
        n <- 150
        sdX <- 0.4                              # standard deviation of each covariate
        rho <- matrix(corx, nrow=2, ncol=2)    # set correlation matrix
    diag(rho) <- rep(1, 2)
    sd.vec <-rep(sdX, 2)
    cor2cov <- function(R, S) {

      sweep(sweep(R, 1, S, "*"),2,S,"*")
    }
    R <- cor2cov(rho, sd.vec)          # covariance matrix

    if(i == 1){

      mean <- c(X1 =0.60, X2 =0.50)
    }

    if(i == 2){
```

```
  mean <- c(X1 =rnorm(1,meanx1, 0.05), X2 =rnorm(1,meanx2, 0.05))
}

if(i == 3){

  mean <- c(X1 =rnorm(1,meanx1, 0.05), X2 =rnorm(1,meanx2, 0.05))
}

if(i == 4){

  mean <- c(X1 =rnorm(1,meanx1, 0.05), X2 =rnorm(1,meanx2, 0.05))
}

if(i == 5){

  mean <- c(X1 =rnorm(1,meanx1, 0.05), X2 =rnorm(1,meanx2, 0.05))
}

if(i == 6){

  mean <- c(X1 =rnorm(1,meanx1, 0.05), X2 =rnorm(1,meanx2, 0.05))
}
if(i == 7){

  mean <- c(X1 =rnorm(1,meanx1, 0.05), X2 =rnorm(1,meanx2, 0.05))
}
if(i == 8){

  mean <- c(X1 =rnorm(1,meanx1, 0.05), X2 =rnorm(1,meanx2, 0.05))
}

if(i == 9){

  mean <- c(X1 =rnorm(1,meanx1, 0.05), X2 =rnorm(1,meanx2, 0.05))
}

if(i == 10){

  mean <- c(X1 =rnorm(1,meanx1, 0.05), X2 =rnorm(1,meanx2, 0.05))
}

cov<-data.frame(MASS::mvrnorm(n, mu = mean, Sigma = R))
    delta2 =d[t[i,2]] - d[t[i,1]]

mu = 0.85                       ## intercept value in each study
```

```
    prob1=1 / (1 + exp(-( mu + b_X1  * (cov$X2)
                          + b_X1 * (cov$X1)  )))

    prob2=1 / (1 + exp(-( mu + b_X1 * (cov$X2)
                          + b_X1 * (cov$X1) + delta2
                          + b_X1_trt * (cov$X1 ))))


    ytemp1 = rbinom(np[i,1], size=1,prob=prob1) ### outcome for reference arm

    ytemp2 =  rbinom(np[i,2], size=1, prob=prob2) ###  outcome for trt arm
    datacel[ ,1+count]<-ytemp1
    datacel[ ,2+count]<-ytemp2
    datacel[ ,1+count2]<-cov$X1
    datacel[ ,2+count2]<-cov$X2
    count = count + 4
    count2= count2+ 4

  }
  return(datacel)
}

### now generating and saving multiple no of data for scenario 1 using replicate
### generating 3000 datasets using the object pc
### this will store the datasets as list

ipd<-replicate(n=3000,expr=gen.data(corx=pc$corx[1], b_X1_trt=pc$b_X1_trt[1],
        b_X1=pc$b_X1[1],meanx1=pc$meanx1[1], meanx2= pc$meanx2[1]),simplify =FALSE)


#######################################################
##### First estimating true d1 and d2 (parameter values) so
### that the estimates can be used in the calculation of
###  coverage in the simulation
### to get the true d1, d2, running the same data
### generation as before but now the sample size is quite
### large (1 million)
#######################################################
#############################################################

d = c(log(1),log(1.5),log(.17))              ## log odds ratio (0, 0.40, -1.77)
ns =10                                        ## no of studies
np = matrix(1000000, ns, 2)                  ## no  of patient in each study
t = matrix(c(1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,3,3,3,3), ns, 2) ## trt in each arm
```

```
count = 0
count2=2
datacel<-matrix(NA, nrow=1000000, ncol=40)
for(i in 1:ns){

  n <- 1000000

  sdX <- 0.4                              # standard deviation of each covariate

  rho <- matrix(pc[1,1], nrow=2, ncol=2)    # set correlation matrix
  diag(rho) <- rep(1, 2)
  sd.vec <-rep(sdX, 2)
  cor2cov <- function(R, S) {

    sweep(sweep(R, 1, S, "*"),2,S,"*")
  }
  R <- cor2cov(rho, sd.vec)          # covariance matrix

  if(i == 1){

    mean <- c(X1 =0.60, X2 =0.50)
  }

  if(i == 2){

    mean <- c(X1 =rnorm(1,pc[1,4], 0.05), X2 =rnorm(1,pc[1,5], 0.05))
  }

  if(i == 3){

    mean <- c(X1 =rnorm(1,pc[1,4], 0.05), X2 =rnorm(1,pc[1,5], 0.05))
  }

  if(i == 4){

    mean <- c(X1 =rnorm(1,pc[1,4], 0.05), X2 =rnorm(1,pc[1,5], 0.05))
  }

  if(i == 5){

    mean <- c(X1 =rnorm(1,pc[1,4], 0.05), X2 =rnorm(1,pc[1,5], 0.05))
  }

  if(i == 6){
```

```
  mean <- c(X1 =rnorm(1,pc[1,4], 0.05), X2 =rnorm(1,pc[1,5], 0.05))
}
if(i == 7){

  mean <- c(X1 =rnorm(1,pc[1,4], 0.05), X2 =rnorm(1,pc[1,5], 0.05))
}
if(i == 8){

  mean <- c(X1 =rnorm(1,pc[1,4], 0.05), X2 =rnorm(1,pc[1,5], 0.05))
}

if(i == 9){

  mean <- c(X1 =rnorm(1,pc[1,4], 0.05), X2 =rnorm(1,pc[1,5], 0.05))
}

if(i == 10){

  mean <- c(X1 =rnorm(1,pc[1,4], 0.05), X2 =rnorm(1,pc[1,5], 0.05))
}

cov<-data.frame(MASS::mvrnorm(n, mu = mean, Sigma = R))
delta2 = d[t[i,2]] - d[t[i,1]]
mu = 0.85
b_X1 <- pc[1,3]     # conditional effect of variable 1 and 2

b_X1_trt <- pc[1,2] # conditional interaction effect of  effect modifier

prob1=1 / (1 + exp(-( mu + b_X1  * (cov$X2)
                      + b_X1 * (cov$X1)  )))

prob2=1 / (1 + exp(-( mu + b_X1 * (cov$X2)
                      + b_X1 * (cov$X1) + delta2
                      + b_X1_trt * (cov$X1 ))))



ytemp1 = rbinom(np[i,1], size=1,prob=prob1) ### for reference arm

ytemp2 =  rbinom(np[i,2], size=1, prob=prob2) ###  for trt arm
datacel[ ,1+count]<-ytemp1
datacel[ ,2+count]<-ytemp2
datacel[ ,1+count2]<-cov$X1
datacel[ ,2+count2]<-cov$X2
```

```
  count = count + 4
  count2= count2+ 4

}

datacel<-data.frame(datacel)


################################
##  running NMA with the big data
################################

### creating a data frame which will be used in the nma

studyn <- rep(1:10,each=2)

## vector indicates treatment number
trtn <- c(1,2,1,2,1,2,1,2,1,2,1,2,1,3,1,3,1,3,1,3)

## no of events in each arm
r <- c(sum(datacel$X1==1),sum(datacel$X2==1),sum(datacel$X5==1),sum(datacel$X6==1),
       sum(datacel$X9==1),sum(datacel$X10==1),sum(datacel$X13==1),sum(datacel$X14==1),
       sum(datacel$X17==1),sum(datacel$X18==1),sum(datacel$X21==1),sum(datacel$X22==1),
       sum(datacel$X25==1),sum(datacel$X26==1),sum(datacel$X29==1),sum(datacel$X30==1),
       sum(datacel$X33==1),sum(datacel$X34==1),sum(datacel$X37==1),sum(datacel$X38==1))


n <- rep(1000000, 20)             ## no of patients in each arm

datacel <- data.frame(cbind(studyn,trtn,r,n))
colSums(select_if(datacel, is.numeric))

### following code will set up the network (arm based)

true.network.fe<-set_agd_arm(
  datacel,
  study=studyn,
  trt=trtn,
  r = r,
  n = n,
  trt_ref = 1 )


##The model is fitted using the nma() function. nma function will generate
### the value of true d1, d2
```

```
arm_fit_FE_true <- nma(true.network.fe,
                       trt_effects = "fixed",
                       prior_intercept = normal(scale = 100),
                       prior_trt = normal(scale = 10))

#### extracting all the values of d1 and d2
### and storing them in a dataframe
model_true<-as.data.frame(arm_fit_FE_true,pars=c("d"))

### now taking the mean of d1 and d2 values which will be
## used as the estimate of d1 and d2

stan.object.nma.true.mean.d1<-mean(model_true$`d[2]`)
stan.object.nma.true.mean.d2<-mean(model_true$`d[3]`)


##############################################################################
##############################################################################
### now starting the loop which will be implemented on every
###data of a particular scenario
##############################################################################
##############################################################################

out<-matrix(NA,3000,12)             ## matrix to store results
colnames(out) <- c("nma.connected.d1", "nma.connected.d2",
                   "nma.disconnected.d1", "nma.disconnected.d2",
                   "d1.sd.disconnect", "d2.sd.disconnect",
                   "est1", "est2", "nma.connected.sd.d1", "nma.connected.sd.d2",
                   "est3", "est4")
### in this loop for every data first a connected NMA will be conducted and the mean
## for each d1 and d2 will be stored in the matrix out.
## then the data will be converted to single-arm studies
## MAIC will be performed to estimate the treatment contrast
## an NMA will be performed with the estimates from the MAIC
## and mean and S.E for each d1 and d2 will be stored in the matrix out.
##  the coverage, bias, empSE, model SE will be calculated from the matrix out

for(i in 1:length(ipd)){
   datacel<-data.frame(ipd[[i]])

  ################################
  ##  connected NMA codes
  ################################
```

```
  ### creating a data frame which will be used in the nma

studyn <- rep(1:10,each=2)
## vector indicates treatment number
trtn <- c(1,2,1,2,1,2,1,2,1,2,1,2,1,3,1,3,1,3,1,3)
 ## no of events in each arm
r <- c(sum(datacel$X1==1),sum(datacel$X2==1),sum(datacel$X5==1),sum(datacel$X6==1),
       sum(datacel$X9==1),sum(datacel$X10==1),sum(datacel$X13==1),sum(datacel$X14==1),
       sum(datacel$X17==1),sum(datacel$X18==1),sum(datacel$X21==1),sum(datacel$X22==1),
       sum(datacel$X25==1),sum(datacel$X26==1),sum(datacel$X29==1),sum(datacel$X30==1),
       sum(datacel$X33==1),sum(datacel$X34==1),sum(datacel$X37==1),sum(datacel$X38==1))


n <- rep(150, 20)             ## no of patients in each arm

colSums(select_if(datacel, is.numeric))
datacel2 <- data.frame(cbind(studyn,trtn,r,n))

### following code will set up the network (arm based)

connected.network.fe<-set_agd_arm(
  datacel2,
  study=studyn,
  trt=trtn,
  r = r,
  n = n,
  trt_ref = 1 )

##The model is fitted using the nma() function. nma function will generate
### the value of true d1, d2

arm_fit_FE_connected <- nma(connected.network.fe,
                        trt_effects = "fixed",
                        prior_intercept = normal(scale = 100),
                        prior_trt = normal(scale = 10))

#### extracting all the values of d1 and d2
### and storing them in a dataframe
model_connected<-as.data.frame(arm_fit_FE_connected,pars=c("d"))

### now taking the mean of d1 and d2 values which will be
## used as the estimate of d1 and d2

stan.object.nma.connect.mean.d1<-mean(model_connected$`d[2]`)
stan.object.nma.connect.mean.d2<-mean(model_connected$`d[3]`)
```

```
stan.object.nma.connect.sd.d1<-sd(model_connected$‘d[2]‘)
stan.object.nma.connect.sd.d2<-sd(model_connected$‘d[3]‘)


  ##########################################
  ####  renaming datacel to delete arms
  ### and make the studies into single-arm
  ##########################################

  datacel= rename(datacel, std1.ref = "X1", std1.trt = "X2",
                  std1.cov1= "X3", std1.cov2= "X4",

                  std2.ref = "X5", std2.trt = "X6",
                  std2.cov1= "X7", std2.cov2= "X8",

                  std3.ref = "X9", std3.trt = "X10",
                  std3.cov1= "X11", std3.cov2= "X12",

                  std4.ref = "X13", std4.trt = "X14",
                  std4.cov1= "X15", std4.cov2= "X16",

                  std5.ref = "X17", std5.trt = "X18",
                  std5.cov1= "X19", std5.cov2= "X20",

                  std6.ref = "X21", std6.trt = "X22",
                  std6.cov1= "X23", std6.cov2= "X24",

                  std7.ref = "X25", std7.trt = "X26",
                  std7.cov1= "X27", std7.cov2= "X28",

                  std8.ref = "X29", std8.trt = "X30",
                  std8.cov1= "X31", std8.cov2= "X32",


                  std9.ref = "X33", std9.trt = "X34",
                  std9.cov1= "X35", std9.cov2= "X36",

                  std10.ref = "X37", std10.trt = "X38",
                  std10.cov1= "X39", std10.cov2= "X40")

  ### deleting multiple columns and making studies into a single-arm study

  datacel.update<-dplyr::select( datacel, -c('std1.trt',
                 'std2.ref', 'std3.ref',
```

```
              'std4.ref','std5.ref', 'std6.ref',
              'std7.ref', 'std8.ref', 'std9.ref',
            'std10.ref'))


############ making agd dataset

data.agd<-data.frame( mean.std2.trt=mean(datacel.update$std2.trt),
                      mean.std2.cov1=mean(datacel.update$std2.cov1),
                      mean.std2.cov2=mean(datacel.update$std2.cov2),
                      sd.std2.cov1=sd(datacel.update$std2.cov1),
                      sd.std2.cov2=sd(datacel.update$std2.cov2),


                      mean.std3.trt=mean(datacel.update$std3.trt),
                      mean.std3.cov1=mean(datacel.update$std3.cov1),
                      mean.std3.cov2=mean(datacel.update$std3.cov2),
                      sd.std3.cov1=sd(datacel.update$std3.cov1),
                      sd.std3.cov2=sd(datacel.update$std3.cov2),

                      mean.std4.trt=mean(datacel.update$std4.trt),
                      mean.std4.cov1=mean(datacel.update$std4.cov1),
                      mean.std4.cov2=mean(datacel.update$std4.cov2),
                      sd.std4.cov1=sd(datacel.update$std4.cov1),
                      sd.std4.cov2=sd(datacel.update$std4.cov2),

                      mean.std5.trt=mean(datacel.update$std5.trt),
                      mean.std5.cov1=mean(datacel.update$std5.cov1),
                      mean.std5.cov2=mean(datacel.update$std5.cov2),
                      sd.std5.cov1=sd(datacel.update$std5.cov1),
                      sd.std5.cov2=sd(datacel.update$std5.cov2),

                      mean.std6.trt=mean(datacel.update$std6.trt),
                      mean.std6.cov1=mean(datacel.update$std6.cov1),
                      mean.std6.cov2=mean(datacel.update$std6.cov2),
                      sd.std6.cov1=sd(datacel.update$std6.cov1),
                      sd.std6.cov2=sd(datacel.update$std6.cov2),


                      mean.std7.trt=mean(datacel.update$std7.trt),
                      mean.std7.cov1=mean(datacel.update$std7.cov1),
                      mean.std7.cov2=mean(datacel.update$std7.cov2),
                      sd.std7.cov1=sd(datacel.update$std7.cov1),
                      sd.std7.cov2=sd(datacel.update$std7.cov2),
```

```
                      mean.std8.trt=mean(datacel.update$std8.trt),
                      mean.std8.cov1=mean(datacel.update$std8.cov1),
                      mean.std8.cov2=mean(datacel.update$std8.cov2),
                      sd.std8.cov1=sd(datacel.update$std8.cov1),
                      sd.std8.cov2=sd(datacel.update$std8.cov2),


                      mean.std9.trt=mean(datacel.update$std9.trt),
                      mean.std9.cov1=mean(datacel.update$std9.cov1),
                      mean.std9.cov2=mean(datacel.update$std9.cov2),
                      sd.std9.cov1=sd(datacel.update$std9.cov1),
                      sd.std9.cov2=sd(datacel.update$std9.cov2),

                      mean.std10.trt=mean(datacel.update$std10.trt),
                      mean.std10.cov1=mean(datacel.update$std10.cov1),
                      mean.std10.cov2=mean(datacel.update$std10.cov2),
                      sd.std10.cov1=sd(datacel.update$std10.cov1),
                      sd.std10.cov2=sd(datacel.update$std10.cov2)



)


############ making ipd dataset

data.ipd<-dplyr::select(datacel.update, -c(4:30))

## the covariates need to be the same name in ipd and agd study
data.ipd<-rename(data.ipd, cov1 = std1.cov1,cov2 = std1.cov2)


#############################################################
################  MAIC 1 ###################################
################  study 1(ipd) study 2(agd)   ######################

### renaming variables in agd study

data.agd_update<-rename(data.agd, cov1 = mean.std2.cov1, cov2 = mean.std2.cov2)

# List out matching covariates
match_cov <- c("cov1", "cov2")

#### center baseline characteristics
```

```
# matching continuous variables on both mean and standard deviation

data.ipd_update <- data.ipd %>%
  mutate(cov1_centered = cov1- data.agd$mean.std2.cov1,
         cov1_squared_centered = (cov1^2) - (data.agd$mean.std2.cov1^2 +
                                               data.agd$sd.std2.cov1^2),

         cov2_centered = cov2- data.agd$mean.std2.cov2,
         cov2_squared_centered = (cov2^2) - (data.agd$mean.std2.cov2^2 +
                                               data.agd$sd.std2.cov2^2)
  )

cent_match_cov <- c("cov1_centered",
                    "cov1_squared_centered",
                    "cov2_centered",
                    "cov2_squared_centered"
)

#### estimating weights

est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                matching_vars = cent_match_cov)

####################

# Based on the known proportion of respondents, simulate the response data

comparator_n <- 150   # total number of patients in the comparator data

comparator_prop_events <- data.agd$mean.std2.trt # proportion of responders

# Calculate the number with event

n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                              rep(0, comparator_n - n_with_event)))

 comparator_input <-  comparator_binary %>%
  mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1


# Merge comparator and intervention data

combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
```

```
combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

# Estimate weighted OR by fitting a logistic regression model with weights
weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                    family = binomial(link="logit"),
                                    data = combined_data,
                                    weight = wt))

maic1.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
maic1.se<-coeftest(weighted_OR,vcov=vcovHC(weighted_OR,type="HC3"))[2,2]


################  MAIC 2 ###################################
################  study 1(ipd) study 3(agd)   ######################

data.agd_update<-rename(data.agd, cov1 = mean.std3.cov1, cov2 = mean.std3.cov2)

# List out matching covariates
match_cov <- c("cov1","cov2")

#### center baseline characteristics

#matching continuous variables on both mean and standard deviation

data.ipd_update <- data.ipd %>%
  mutate(cov1_centered = cov1- data.agd$mean.std3.cov1,
         cov1_squared_centered = (cov1^2) - (data.agd$mean.std3.cov1^2 +
                                               data.agd$sd.std3.cov1^2),

         cov2_centered = cov2- data.agd$mean.std3.cov2,
         cov2_squared_centered = (cov2^2) - (data.agd$mean.std3.cov2^2 +
                                      data.agd$sd.std3.cov2^2))


cent_match_cov <- c("cov1_centered",
                    "cov1_squared_centered",
                    "cov2_centered",
                    "cov2_squared_centered"
)

#### estimating weights

est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                matching_vars = cent_match_cov)
```

```
# Based on the known proportion of respondents, simulate the response data
comparator_n <- 150 # total number of patients in the comparator data
  comparator_prop_events <- data.agd$mean.std3.trt # proportion of responders

# Calculate number with event

n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                              rep(0, comparator_n - n_with_event)))

comparator_input <-  comparator_binary %>%
  mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

# Join comparator data with the intervention data

combined_data <- bind_rows(est_weights$analysis_data, comparator_input)

combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

# Estimate weighted OR by fitting a logistic regression model with weights
weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                    family = binomial(link="logit"),
                                    data = combined_data,
                                    weight = wt))

maic2.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
maic2.se<-coeftest(weighted_OR,vcov=vcovHC(weighted_OR,type="HC3"))[2,2]

#####################################################################
################  MAIC 3 ###################################
################  study 1(ipd) study 4(agd)   ######################

### renaming variables in agd study

data.agd_update<-rename(data.agd, cov1 = mean.std4.cov1, cov2 = mean.std4.cov2)

# List out matching covariates
match_cov <- c("cov1", "cov2")

#### center baseline characteristics

#matching continuous variables on both mean and standard deviation

data.ipd_update <- data.ipd %>%
```

```
  mutate(cov1_centered = cov1- data.agd$mean.std4.cov1,
         cov1_squared_centered = (cov1^2) - (data.agd$mean.std4.cov1^2 +
                                               data.agd$sd.std4.cov1^2),

         cov2_centered = cov2- data.agd$mean.std4.cov2,
         cov2_squared_centered = (cov2^2) - (data.agd$mean.std4.cov2^2 +
                                               data.agd$sd.std4.cov2^2))

cent_match_cov <- c("cov1_centered",
                    "cov1_squared_centered",
                    "cov2_centered",
                    "cov2_squared_centered"
)

#### estimating weights

est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                matching_vars = cent_match_cov)


# Based on the known proportion of respondents, simulate the response data

comparator_n <- 150 # total number of patients in the comparator data
comparator_prop_events <- data.agd$mean.std4.trt # proportion of responders

# Calculate number with event

n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                              rep(0, comparator_n - n_with_event)))

comparator_input <-  comparator_binary %>%
  mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

# Merge comparator and intervention data

combined_data <- bind_rows(est_weights$analysis_data, comparator_input)

combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

# Estimate weighted OR by fitting a logistic regression model with weights
weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                    family = binomial(link="logit"),
                                    data = combined_data,
                                   weight = wt))
```

```
  maic3.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
  maic3.se<-coeftest(weighted_OR,vcov=vcovHC(weighted_OR,type="HC3"))[2,2]

#####################################################################
  ################  MAIC 4 ###################################
  ################  study 1(ipd) study 5(agd)   ######################

  data.agd_update<-rename(data.agd, cov1 = mean.std5.cov1, cov2 = mean.std5.cov2)

  # List out matching covariates
  match_cov <- c("cov1", "cov2")

  #### center baseline characteristics

  data.ipd_update <- data.ipd %>%
    mutate(cov1_centered = cov1- data.agd$mean.std5.cov1,
           cov1_squared_centered = (cov1^2) - (data.agd$mean.std5.cov1^2 +
                                                 data.agd$sd.std5.cov1^2),

           cov2_centered = cov2- data.agd$mean.std5.cov2,
           cov2_squared_centered = (cov2^2) - (data.agd$mean.std5.cov2^2 +
                                                 data.agd$sd.std5.cov2^2))

  cent_match_cov <- c("cov1_centered",
                      "cov1_squared_centered",
                      "cov2_centered",
                      "cov2_squared_centered"
  )
  #### estimating weights

  est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                  matching_vars = cent_match_cov)

 # Based on the known proportion of respondents, simulate the response data

  comparator_n <- 150 # total number of patients in the comparator data
  comparator_prop_events <- data.agd$mean.std5.trt # proportion of responders

  # Calculate number with event

  n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
  comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                                rep(0, comparator_n - n_with_event)))
```

```
  comparator_input <-  comparator_binary %>%
    mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

  # Merge comparator and intervention data
  combined_data <- bind_rows(est_weights$analysis_data, comparator_input)

  combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

  # Estimate weighted OR by fitting a logistic regression model with weights
  weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                      family = binomial(link="logit"),
                                      data = combined_data,
                                      weight = wt))

  maic4.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
  maic4.se<-coeftest(weighted_OR,vcov=vcovHC(weighted_OR,type="HC3"))[2,2]

#####################################################################
################  MAIC 5 ###################################
################  study 1(ipd) study 6(agd)   ######################

  data.agd_update<-rename(data.agd, cov1 = mean.std6.cov1, cov2 = mean.std6.cov2)

  # List out matching covariates
  match_cov <- c("cov1","cov2")

  #### center baseline characteristics

  #matching continuous variables on both mean and standard deviation

  data.ipd_update <- data.ipd %>%
    mutate(cov1_centered = cov1- data.agd$mean.std6.cov1,
           cov1_squared_centered = (cov1^2) - (data.agd$mean.std6.cov1^2 +
                                                 data.agd$sd.std6.cov1^2),

           cov2_centered = cov2- data.agd$mean.std6.cov2,
           cov2_squared_centered = (cov2^2) - (data.agd$mean.std6.cov2^2 +
                                                 data.agd$sd.std6.cov2^2))

  cent_match_cov <- c("cov1_centered",
                      "cov1_squared_centered",
                      "cov2_centered",
                      "cov2_squared_centered"
  )
```

```
#### estimating weights

est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                matching_vars = cent_match_cov)


# Based on the known proportion of respondents, simulate the response data

comparator_n <- 150 # total number of patients in the comparator data

comparator_prop_events <- data.agd$mean.std6.trt # proportion of responders

# Calculate number with event

n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                              rep(0, comparator_n - n_with_event)))

comparator_input <-  comparator_binary %>%
  mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

# Merge comparator and intervention data

combined_data <- bind_rows(est_weights$analysis_data, comparator_input)

combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")
# Estimate weighted OR by fitting a logistic regression model with weights
weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                    family = binomial(link="logit"),
                                    data = combined_data,
                                    weight = wt))

maic5.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
maic5.se<-coeftest(weighted_OR,vcov=vcovHC(weighted_OR,type="HC3"))[2,2]

#####################################################################
################  MAIC 6 ###################################
################  study 1(ipd) study 8(agd)   ######################

data.agd_update<-rename(data.agd, cov1 = mean.std8.cov1, cov2 = mean.std8.cov2)

# List out matching covariates
match_cov <- c("cov1","cov2")

#### center baseline characteristics
```

```
#matching continuous variables on both mean and standard deviation

data.ipd_update <- data.ipd %>%
  mutate(cov1_centered = cov1- data.agd$mean.std8.cov1,
         cov1_squared_centered = (cov1^2) - (data.agd$mean.std8.cov1^2 +
                                               data.agd$sd.std8.cov1^2),

         cov2_centered = cov2- data.agd$mean.std8.cov2,
         cov2_squared_centered = (cov2^2) - (data.agd$mean.std8.cov2^2 +
                                               data.agd$sd.std8.cov2^2))

cent_match_cov <- c("cov1_centered",
                    "cov1_squared_centered",
                    "cov2_centered",
                    "cov2_squared_centered"
)

#### estimating weights

est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                matching_vars = cent_match_cov)

# Based on the known proportion of respondents, simulate the response data

comparator_n <- 150 # total number of patients in the comparator data

comparator_prop_events <- data.agd$mean.std8.trt # proportion of responders

# Calculate number with event

n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                              rep(0, comparator_n - n_with_event)))

comparator_input <-  comparator_binary %>%
  mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1


# Merge comparator and intervention data
combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

# Estimate weighted OR by fitting a logistic regression model with weights
weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
```

```
                                    family = binomial(link="logit"),
                                    data = combined_data,
                                    weight = wt))
maic6.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
maic6.se<-coeftest(weighted_OR,vcov=vcovHC(weighted_OR,type="HC3"))[2,2]

#####################################################################
################  MAIC 7 ###################################
################  study 1(ipd) study 9(agd)   ######################

data.agd_update<-rename(data.agd, cov1 = mean.std9.cov1, cov2 = mean.std9.cov2)

# List out matching covariates
match_cov <- c("cov1", "cov2")

#### center baseline characteristics

#matching continuous variables on both mean and standard deviation

data.ipd_update <- data.ipd %>%
  mutate(cov1_centered = cov1- data.agd$mean.std9.cov1,
         cov1_squared_centered = (cov1^2) - (data.agd$mean.std9.cov1^2 +
                                               data.agd$sd.std9.cov1^2),

         cov2_centered = cov2- data.agd$mean.std9.cov2,
         cov2_squared_centered = (cov2^2) - (data.agd$mean.std9.cov2^2 +
                                               data.agd$sd.std9.cov2^2))

cent_match_cov <- c("cov1_centered",
                    "cov1_squared_centered",
                    "cov2_centered",
                    "cov2_squared_centered"
)

#### estimating weights

est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                matching_vars = cent_match_cov)

# Based on the known proportion of respondents, simulate the response data
comparator_n <- 150 # total number of patients in the comparator data
comparator_prop_events <- data.agd$mean.std9.trt # proportion of responders

# Calculate number with event
```

```
  n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
  comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                                rep(0, comparator_n - n_with_event)))

  comparator_input <-  comparator_binary %>%
    mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

  # Merge comparator and intervention data
  combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
  combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

  # Estimate weighted OR by fitting a logistic regression model with weights
  weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                      family = binomial(link="logit"),
                                      data = combined_data,
                                      weight = wt))

  maic7.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
  maic7.se<-coeftest(weighted_OR,vcov=vcovHC(weighted_OR,type="HC3"))[2,2]

#####################################################################
################  MAIC 8 ###################################
################  study 1(ipd) study 10(agd)   ######################

  data.agd_update<-rename(data.agd, cov1 = mean.std10.cov1, cov2 = mean.std10.cov2)

  # List out matching covariates
  match_cov <- c("cov1","cov2")

  #### center baseline characteristics
  #matching continuous variables on both mean and standard deviation
  data.ipd_update <- data.ipd %>%
    mutate(cov1_centered = cov1- data.agd$mean.std10.cov1,
           cov1_squared_centered = (cov1^2) - (data.agd$mean.std10.cov1^2 +
                                                 data.agd$sd.std10.cov1^2),
           cov2_centered = cov2- data.agd$mean.std10.cov2,
           cov2_squared_centered = (cov2^2) - (data.agd$mean.std10.cov2^2 +
                                                 data.agd$sd.std10.cov2^2))

  cent_match_cov <- c("cov1_centered",
                      "cov1_squared_centered",
                      "cov2_centered",
                      "cov2_squared_centered"

  )
```

```
#### estimating weights

est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                matching_vars = cent_match_cov)

# Based on the known proportion of respondents, simulate the response data
comparator_n <- 150 # total number of patients in the comparator data
comparator_prop_events <- data.agd$mean.std10.trt # proportion of responders

# Calculate number with event

n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                              rep(0, comparator_n - n_with_event)))


comparator_input <-  comparator_binary %>%
  mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

# Merge comparator and intervention data

combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
  combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

# Estimate weighted OR by fitting a logistic regression model with weights
weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                    family = binomial(link="logit"),
                                    data = combined_data,
                                    weight = wt))

maic8.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
maic8.se<-coeftest(weighted_OR,vcov=vcovHC(weighted_OR,type="HC3"))[2,2]

#####################################################################
################  MAIC 9 ###################################
################  study 1(ipd) study 7(agd)   ######################

data.agd_update<-rename(data.agd, cov1 = mean.std7.cov1, cov2 = mean.std7.cov2)

# List out matching covariates
match_cov <- c("cov1", "cov2")

#### center baseline characteristics
```

```
#matching continuous variables on both mean and standard deviation

data.ipd_update <- data.ipd %>%
  mutate(cov1_centered = cov1- data.agd$mean.std7.cov1,
         cov1_squared_centered = (cov1^2) - (data.agd$mean.std7.cov1^2 +
                                               data.agd$sd.std7.cov1^2),
         cov2_centered = cov2- data.agd$mean.std7.cov2,
         cov2_squared_centered = (cov2^2) - (data.agd$mean.std7.cov2^2 +
                                               data.agd$sd.std7.cov2^2))

cent_match_cov <- c("cov1_centered",
                    "cov1_squared_centered",
                    "cov2_centered",
                    "cov2_squared_centered"
)

#### estimating weights

est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                matching_vars = cent_match_cov)

# Based on the known proportion of respondents, simulate the response data
comparator_n <- 150 # total number of patients in the comparator data
comparator_prop_events <- data.agd$mean.std7.trt # proportion of responders


# Calculate number with event
n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                              rep(0, comparator_n - n_with_event)))

comparator_input <-  comparator_binary %>%
  mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

# Merge comparator and intervention data
combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")
# Estimate weighted OR by fitting a logistic regression model with weights
weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                    family = binomial(link="logit"),
                                    data = combined_data,
                                    weight = wt))

maic9.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
maic9.se<-coeftest(weighted_OR,vcov=vcovHC(weighted_OR,type="HC3"))[2,2]
```

```
 ####conducting an NMA with the estimates from the MAIC

### creating a data frame which will be used in the nma

   studyn <- c(1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8,9,9)
   trtn <- c(1,2,1,2,1,2,1,2,1,2,1,3,1,3,1,3,1,3)

   n <- c(150,150,150,150,150,150,150,150,150,150,150,150,150,150,150,150,150,150)

   diff<-c(NA,maic1.est,
           NA,maic2.est,NA,maic3.est,
           NA,maic4.est,NA,maic5.est,
           NA,maic9.est,NA,maic6.est,
           NA,maic7.est,NA,maic8.est)

   se_diff<-c(NA,maic1.se,
           NA,maic2.se,NA,maic3.se,
           NA,maic4.se,NA,maic5.se,
           NA,maic9.se,NA,maic6.se,
           NA,maic7.se,NA,maic8.se)

   datacel3 <- data.frame(cbind(studyn,trtn,diff,se_diff,n))

   ### following code will set up the network (arm based)

   model.fe.disconect<-set_agd_contrast(
     datacel3,
     study=studyn,
     trt=trtn,
     y=diff,
     se=se_diff,
     sample_size=n)

   ##The model is fitted using the nma() function.

   maic_fit_FE <- nma(model.fe.disconect,
                      trt_effects = "fixed",
                      prior_intercept = normal(scale = 100),
                      prior_trt = normal(scale = 10))

   maic_data_frame<-as.data.frame(maic_fit_FE,pars=c("d"))

   stan.object.nma.discnt.mean.d1<-mean(maic_data_frame$`d[2]`)
   stan.object.nma.discnt.mean.d2<-mean(maic_data_frame$`d[3]`)
```

```
  stan.object.nma.discontd.sd.d1<-sd(maic_data_frame$‘d[2]‘)
  stan.object.nma.discontd.sd.d2<-sd(maic_data_frame$‘d[3]‘)

#### calculation of coverage probability

B1=stan.object.nma.discontd.sd.d1
A1=stan.object.nma.discnt.mean.d1
z.alpha1 <- 1.96
theta.hat.low1=A1-z.alpha1*B1
theta.hat.upp1=A1+z.alpha1*B1
theta1= stan.object.nma.true.mean.d1
est1 <-ifelse(theta1>=theta.hat.low1 & theta1<=theta.hat.upp1,1,0)

B2=stan.object.nma.discontd.sd.d2
A2=stan.object.nma.discnt.mean.d2
z.alpha2 <- 1.96
theta.hat.low2=A2-z.alpha2*B2
theta.hat.upp2=A2+z.alpha2*B2
theta2= stan.object.nma.true.mean.d2
est2 <-ifelse(theta2>=theta.hat.low2 & theta2<=theta.hat.upp2,1,0)

B3=stan.object.nma.connect.sd.d1
A3=stan.object.nma.connect.mean.d1
z.alpha3 <- 1.96
theta.hat.low3=A3-z.alpha3*B3
theta.hat.upp3=A3+z.alpha3*B3
theta3= stan.object.nma.true.mean.d1
est3 <-ifelse(theta3>=theta.hat.low3 & theta3<=theta.hat.upp3,1,0)

B4=stan.object.nma.connect.sd.d2
A4=stan.object.nma.connect.mean.d2
z.alpha4 <- 1.96
theta.hat.low4=A4-z.alpha4*B4
theta.hat.upp4=A4+z.alpha4*B4
theta4= stan.object.nma.true.mean.d2
est4 <-ifelse(theta4>=theta.hat.low4 & theta4<=theta.hat.upp4,1,0)


out[i, 1]<-stan.object.nma.connect.mean.d1
out[i, 2]<-stan.object.nma.connect.mean.d2
out[i, 3]<-stan.object.nma.discnt.mean.d1
out[i, 4]<-stan.object.nma.discnt.mean.d2
out[i, 5]<-stan.object.nma.discontd.sd.d1
out[i, 6]<-stan.object.nma.discontd.sd.d2
```

```
  out[i, 7]<-est1
  out[i, 8]<-est2
  out[i, 9]<-stan.object.nma.connect.sd.d1
  out[i, 10]<-stan.object.nma.connect.sd.d2
    out[i, 11]<- est3
    out[i, 12]<- est4


}

out<-data.frame(out)
```

Appendix D

R codes for MAIC-adjusted random effects NMA

D.1 MAIC-adjusted random effects NMA

```
#####################################
## R codes to conduct a connected NMA
## and a MAIC adjusted random effect NMA

################ data generation
rm(list=ls())
set.seed(1128)
### making object for parameter combination

corx<-c(0.20,0.80)                      ##correlation between covariates in a study
b_X1_trt<-c(-log(0.78),-log(0.40))       ## 0.25 and 0.916 (interaction coefficient)
b_X1<-c(-log(0.67),-log(0.33))           ## 0.40 and 1.10  (covariate coefficient)
meanx1<-c(0.45,0.15)                    ## mean of covariates
param.combinations <- expand.grid(corx=corx, b_X1_trt=b_X1_trt,
                                  b_X1=b_X1, meanx1=meanx1)
pc <- param.combinations
pc<-round(pc, 2)                       ## rounding the values
pc$meanx2<-c(0.48,0.48,0.48,0.48,0.48,0.48,0.48,
             0.48,0.20,0.20,0.20,0.20,0.20,0.20,0.20,0.20) ## adding another column in pc
scenerios<-nrow(pc)                     ### no of scenarios created ## 16scenarios


#################################
d = c(log(1),log(1.5),log(.17))           ## log odds ratio (0, 0.40, -1.77)
tau = 0.3                                 ## hetrogeneity parameter
ns =10                                    ## no of studies
np = matrix(150, ns, 2)                   ## no  of patient in each study
t = matrix(c(1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,3,3,3,3), ns, 2) ## trt in each arm
```

```
gen.data<-function(sdx,corx, meanx1,meanx2, b_X1, b_X1_trt){
  count = 0
  count2=2
  datacel<-matrix(NA, nrow=150, ncol=40)
  for(i in 1:ns){
    n <- 150
    sdX <- 0.4                              # standard deviation of each covariate

    rho <- matrix(corx, nrow=2, ncol=2)    # set correlation matrix
    diag(rho) <- rep(1, 2)
    sd.vec <-rep(sdX, 2)
    cor2cov <- function(R, S) {

      sweep(sweep(R, 1, S, "*"),2,S,"*")
    }
    R <- cor2cov(rho, sd.vec)          # covariance matrix

    if(i == 1){
      mean <- c(X1 =0.60, X2 =0.50)
    }
    if(i == 2){
      mean <- c(X1 =rnorm(1,meanx1, 0.05), X2 =rnorm(1,meanx2, 0.05))
    }

    if(i == 3){
      mean <- c(X1 =rnorm(1,meanx1, 0.05), X2 =rnorm(1,meanx2, 0.05))
    }

    if(i == 4){
      mean <- c(X1 =rnorm(1,meanx1, 0.05), X2 =rnorm(1,meanx2, 0.05))
    }

    if(i == 5){
      mean <- c(X1 =rnorm(1,meanx1, 0.05), X2 =rnorm(1,meanx2, 0.05))
    }

    if(i == 6){
      mean <- c(X1 =rnorm(1,meanx1, 0.05), X2 =rnorm(1,meanx2, 0.05))
    }
    if(i == 7){
      mean <- c(X1 =rnorm(1,meanx1, 0.05), X2 =rnorm(1,meanx2, 0.05))
    }
    if(i == 8){
      mean <- c(X1 =rnorm(1,meanx1, 0.05), X2 =rnorm(1,meanx2, 0.05))
```

```
    }

    if(i == 9){
      mean <- c(X1 =rnorm(1,meanx1, 0.05), X2 =rnorm(1,meanx2, 0.05))
    }

    if(i == 10){
      mean <- c(X1 =rnorm(1,meanx1, 0.05), X2 =rnorm(1,meanx2, 0.05))
    }

    cov<-data.frame(MASS::mvrnorm(n, mu = mean, Sigma = R))

    delta2 =rnorm(1, d[t[i,2]] - d[t[i,1]], tau)  ## treatment effect for trt arm
                                        #in each study

    mu = 0.85                         ## intercept value in each study

    prob1=1 / (1 + exp(-( mu + b_X1  * (cov$X2)
                          + b_X1 * (cov$X1)  )))

    prob2=1 / (1 + exp(-( mu + b_X1 * (cov$X2)
                          + b_X1 * (cov$X1) + delta2
                          + b_X1_trt * (cov$X1 ))))

    ytemp1 = rbinom(np[i,1], size=1,prob=prob1) ### outcome for reference arm

    ytemp2 =  rbinom(np[i,2], size=1, prob=prob2) ###  outcome for trt arm

    datacel[ ,1+count]<-ytemp1
    datacel[ ,2+count]<-ytemp2
    datacel[ ,1+count2]<-cov$X1
    datacel[ ,2+count2]<-cov$X2
    count = count + 4
    count2= count2+ 4

  }
  return(datacel)
}

### now generating and saving multiple no of data for each scenario
ipd<-replicate(n=3000,expr=gen.data(corx=pc$corx[1], b_X1_trt=pc$b_X1_trt[1],
   b_X1=pc$b_X1[1], meanx1=pc$meanx1[1], meanx2= pc$meanx2[1]),simplify =FALSE )
```

```
#####################################################
##### First estimating true d1 and d2 (parameter values) so
### that the estimates can be used in the calculation of
###  coverage in the simulation
### to get the true d1, d2, running the same data
### generation as before but now the sample size is quite
### large (1 million)
#####################################################
#############################################################

d = c(log(1),log(1.5),log(.17))              ## log odds ratio (0, 0.40, -1.77)
tau= 0.3                                   ## heterogeneity parameter
ns =10                                      ## no of studies
np = matrix(1000000, ns, 2)                 ## no  of patient in each study
t = matrix(c(1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,3,3,3,3), ns, 2) ## trt in each arm


count = 0
count2=2
datacel<-matrix(NA, nrow=1000000, ncol=40)
for(i in 1:ns){

  n <- 1000000

  sdX <- 0.4                              # standard deviation of each covariate

  rho <- matrix(pc[1,1], nrow=2, ncol=2)    # set correlation matrix
  diag(rho) <- rep(1, 2)
  sd.vec <-rep(sdX, 2)
  cor2cov <- function(R, S) {

    sweep(sweep(R, 1, S, "*"),2,S,"*")
  }
  R <- cor2cov(rho, sd.vec)          # covariance matrix

  if(i == 1){
    mean <- c(X1 =0.60, X2 =0.50)
  }

  if(i == 2){
    mean <- c(X1 =rnorm(1,pc[1,4], 0.05), X2 =rnorm(1,pc[1,5], 0.05))
  }

  if(i == 3){
    mean <- c(X1 =rnorm(1,pc[1,4], 0.05), X2 =rnorm(1,pc[1,5], 0.05))
```

```
}

if(i == 4){
  mean <- c(X1 =rnorm(1,pc[1,4], 0.05), X2 =rnorm(1,pc[1,5], 0.05))
}

if(i == 5){
  mean <- c(X1 =rnorm(1,pc[1,4], 0.05), X2 =rnorm(1,pc[1,5], 0.05))
}

if(i == 6){
  mean <- c(X1 =rnorm(1,pc[1,4], 0.05), X2 =rnorm(1,pc[1,5], 0.05))
}
if(i == 7){
  mean <- c(X1 =rnorm(1,pc[1,4], 0.05), X2 =rnorm(1,pc[1,5], 0.05))
}
if(i == 8){
  mean <- c(X1 =rnorm(1,pc[1,4], 0.05), X2 =rnorm(1,pc[1,5], 0.05))
}

if(i == 9){
  mean <- c(X1 =rnorm(1,pc[1,4], 0.05), X2 =rnorm(1,pc[1,5], 0.05))
}

if(i == 10){
  mean <- c(X1 =rnorm(1,pc[1,4], 0.05), X2 =rnorm(1,pc[1,5], 0.05))
}

cov<-data.frame(MASS::mvrnorm(n, mu = mean, Sigma = R))

delta2 =rnorm(1, d[t[i,2]] - d[t[i,1]], tau)
mu = 0.85
b_X1 <- pc[1,3] # conditional effect of variable 1 and 2

b_X1_trt <- pc[1,2] # conditional interaction effect of  effect modifier

prob1=1 / (1 + exp(-( mu + b_X1  * (cov$X2)
                      + b_X1 * (cov$X1)  )))

prob2=1 / (1 + exp(-( mu + b_X1 * (cov$X2)
                      + b_X1 * (cov$X1) + delta2
                      + b_X1_trt * (cov$X1 ))))
```

```
  ytemp1 = rbinom(np[i,1], size=1,prob=prob1) ### for reference arm

  ytemp2 =  rbinom(np[i,2], size=1, prob=prob2) ###  for trt arm


  datacel[ ,1+count]<-ytemp1 ## response in every (1,2),(5,6),(9,10)th column
  datacel[ ,2+count]<-ytemp2
  datacel[ ,1+count2]<-cov$X1 ## cov value in every (3,4),(7,8),(11,12)th column
  datacel[ ,2+count2]<-cov$X2
  count = count + 4
  count2= count2+ 4

}

datacel<-data.frame(datacel)

##########################################
##  running NMA with a sample size 1 million
##########################################

### creating a data frame which will be used in the nma

studyn <- rep(1:10,each=2)

trtn <- c(1,2,1,2,1,2,1,2,1,2,1,2,1,3,1,3,1,3,1,3)    ## vector indicates treatment number
## no of events in each arm
r <- c(sum(datacel$X1==1),sum(datacel$X2==1),sum(datacel$X5==1),sum(datacel$X6==1),
       sum(datacel$X9==1),sum(datacel$X10==1),sum(datacel$X13==1),sum(datacel$X14==1),
       sum(datacel$X17==1),sum(datacel$X18==1),sum(datacel$X21==1),sum(datacel$X22==1),
       sum(datacel$X25==1),sum(datacel$X26==1),sum(datacel$X29==1),sum(datacel$X30==1),
       sum(datacel$X33==1),sum(datacel$X34==1),sum(datacel$X37==1),sum(datacel$X38==1))


n <- rep(1000000, 20)             ## no of patients in each arm

datacel <- data.frame(cbind(studyn,trtn,r,n))
colSums(select_if(datacel, is.numeric))

### following code will set up the network (arm based)

true.network.re<-set_agd_arm(
  datacel,
  study=studyn,
  trt=trtn,
  r = r,
```

```
  n = n,
  trt_ref = 1 )


##The model is fitted using the nma() function.

arm_fit_RE_true <- nma(true.network.re,
                       trt_effects = "random",
                       prior_intercept = normal(scale = 100),
                       prior_trt = normal(scale = 10),
                       prior_het = log_normal(-2.56, 0.33),
                      prior_het_type = "var")


#### extracting all the values of d1 and d2
### and storing them in a dataframe
model_true<-as.data.frame(arm_fit_RE_true,pars=c("d","tau"))

### now taking the mean of d1 and d2 values which will be
## used as the estimate of d1 and d2

stan.object.nma.true.mean.d1<-mean(model_true$`d[2]`)
stan.object.nma.true.mean.d2<-mean(model_true$`d[3]`)
stan.object.nma.true.median.tau<-median(model_true$tau)


#############################################################################
### now starting the loop which will be implemented on every data of a particular
## scenario
#############################################################################


out<-matrix(NA,3000,18)          ## matrix to store results
colnames(out) <- c("nma.connected.d1", "nma.connected.d2",
                   "connected.tau", "connected.sd.tau",
                   "nma.disconnected.d1", "nma.disconnected.d2",
                   "disconnect.sd.d1", "disconnect.sd.d2",
                   "est1", "est2", "disconnected.tau",
                   "disconnected.sd.tau", "est3", "est4", "connect.sd.d1",
                   "connect.sd.d2",  "est5", "est6")
### in this loop for every data first a connected NMA will be conducted
## then the data will be converted to single-arm studies
## MAIC will be performed to estimate the treatment contrast
## an NMA will be performed with the estimates from the MAIC
## and mean and S.E for each d1 and d2 will be stored in the matrix out.
```

```
##  the coverage, bias, empSE, model SE will be calculated from the matrix out

for(i in 1:length(ipd)){
   datacel<-data.frame(ipd[[i]])

  ################################
  ##  connected NMA codes
  ################################

  ### creating a data frame which will be used in the nma

studyn <- rep(1:10,each=2)
  ## vector indicates treatment number
trtn <- c(1,2,1,2,1,2,1,2,1,2,1,2,1,3,1,3,1,3,1,3)

 ## no of events in each arm
r <- c(sum(datacel$X1==1),sum(datacel$X2==1),sum(datacel$X5==1),sum(datacel$X6==1),
       sum(datacel$X9==1),sum(datacel$X10==1),sum(datacel$X13==1),sum(datacel$X14==1),
       sum(datacel$X17==1),sum(datacel$X18==1),sum(datacel$X21==1),sum(datacel$X22==1),
       sum(datacel$X25==1),sum(datacel$X26==1),sum(datacel$X29==1),sum(datacel$X30==1),
       sum(datacel$X33==1),sum(datacel$X34==1),sum(datacel$X37==1),sum(datacel$X38==1))


n <- rep(150, 20)             ## no of patients in each arm


datacel2 <- data.frame(cbind(studyn,trtn,r,n))
colSums(select_if(datacel, is.numeric))

### following code will set up the network (arm based)

connected.network.re<-set_agd_arm(
  datacel2,
  study=studyn,
  trt=trtn,
  r = r,
  n = n,
  trt_ref = 1 )

##The model is fitted using the nma() function. nma function will generate
### the value of true d1, d2

arm_fit_RE_connected <- nma(connected.network.re,
                        trt_effects = "random",
                        prior_intercept = normal(scale = 100),
```

```
                      prior_trt = normal(scale = 10),
                      prior_het = log_normal(-2.56, 0.33),
                     prior_het_type = "var")

#### extracting all the values of d1 and d2
### and storing them in a dataframe
model_connected<-as.data.frame(arm_fit_RE_connected,pars=c("d","tau"))

### now taking the mean of d1 and d2 values which will be
## used as the estimate of d1 and d2

stan.object.nma.connect.mean.d1<-mean(model_connected$`d[2]`)
stan.object.nma.connect.mean.d2<-mean(model_connected$`d[3]`)
stan.object.nma.connect.median.tau<-median(model_connected$tau)

stan.object.nma.connect.sd.d1<-sd(model_connected$`d[2]`)
stan.object.nma.connect.sd.d2<-sd(model_connected$`d[3]`)
stan.object.nma.connect.sd.tau<-sd(model_connected$tau)


  ##########################################################
  ####  renaming datacel to
  ###  make the studies into single-arm
  ##########################################


  datacel= rename(datacel, std1.ref = "X1", std1.trt = "X2",
                  std1.cov1= "X3", std1.cov2= "X4",

                  std2.ref = "X5", std2.trt = "X6",
                  std2.cov1= "X7", std2.cov2= "X8",

                  std3.ref = "X9", std3.trt = "X10",
                  std3.cov1= "X11", std3.cov2= "X12",

                  std4.ref = "X13", std4.trt = "X14",
                  std4.cov1= "X15", std4.cov2= "X16",

                  std5.ref = "X17", std5.trt = "X18",
                  std5.cov1= "X19", std5.cov2= "X20",

                  std6.ref = "X21", std6.trt = "X22",
                  std6.cov1= "X23", std6.cov2= "X24",

                  std7.ref = "X25", std7.trt = "X26",
```

```
                std7.cov1= "X27", std7.cov2= "X28",

                std8.ref = "X29", std8.trt = "X30",
                std8.cov1= "X31", std8.cov2= "X32",


                std9.ref = "X33", std9.trt = "X34",
                std9.cov1= "X35", std9.cov2= "X36",

                std10.ref = "X37", std10.trt = "X38",
                std10.cov1= "X39", std10.cov2= "X40")


### deleting multiple columns and making studies into single arm

datacel.update<-dplyr::select( datacel, -c('std1.trt', 'std2.ref', 'std3.ref',
                                           'std4.ref','std5.ref', 'std6.ref',
                                           'std7.ref', 'std8.ref', 'std9.ref',
                                           'std10.ref'))

############ making agd dataset

data.agd<-data.frame( mean.std2.trt=mean(datacel.update$std2.trt),
                      mean.std2.cov1=mean(datacel.update$std2.cov1),
                      mean.std2.cov2=mean(datacel.update$std2.cov2),
                      sd.std2.cov1=sd(datacel.update$std2.cov1),
                      sd.std2.cov2=sd(datacel.update$std2.cov2),

                      mean.std3.trt=mean(datacel.update$std3.trt),
                      mean.std3.cov1=mean(datacel.update$std3.cov1),
                      mean.std3.cov2=mean(datacel.update$std3.cov2),
                      sd.std3.cov1=sd(datacel.update$std3.cov1),
                      sd.std3.cov2=sd(datacel.update$std3.cov2),

                      mean.std4.trt=mean(datacel.update$std4.trt),
                      mean.std4.cov1=mean(datacel.update$std4.cov1),
                      mean.std4.cov2=mean(datacel.update$std4.cov2),
                      sd.std4.cov1=sd(datacel.update$std4.cov1),
                      sd.std4.cov2=sd(datacel.update$std4.cov2),

                      mean.std5.trt=mean(datacel.update$std5.trt),
                      mean.std5.cov1=mean(datacel.update$std5.cov1),
                      mean.std5.cov2=mean(datacel.update$std5.cov2),
                      sd.std5.cov1=sd(datacel.update$std5.cov1),
                      sd.std5.cov2=sd(datacel.update$std5.cov2),
```

```
                      mean.std6.trt=mean(datacel.update$std6.trt),

                      mean.std6.cov1=mean(datacel.update$std6.cov1),
                      mean.std6.cov2=mean(datacel.update$std6.cov2),
                      sd.std6.cov1=sd(datacel.update$std6.cov1),
                      sd.std6.cov2=sd(datacel.update$std6.cov2),

                      mean.std7.trt=mean(datacel.update$std7.trt),
                      mean.std7.cov1=mean(datacel.update$std7.cov1),
                      mean.std7.cov2=mean(datacel.update$std7.cov2),
                      sd.std7.cov1=sd(datacel.update$std7.cov1),
                      sd.std7.cov2=sd(datacel.update$std7.cov2),

                    mean.std8.trt=mean(datacel.update$std8.trt),
                      mean.std8.cov1=mean(datacel.update$std8.cov1),
                      mean.std8.cov2=mean(datacel.update$std8.cov2),
                      sd.std8.cov1=sd(datacel.update$std8.cov1),
                      sd.std8.cov2=sd(datacel.update$std8.cov2),

                    mean.std9.trt=mean(datacel.update$std9.trt),
                      mean.std9.cov1=mean(datacel.update$std9.cov1),
                      mean.std9.cov2=mean(datacel.update$std9.cov2),
                      sd.std9.cov1=sd(datacel.update$std9.cov1),
                      sd.std9.cov2=sd(datacel.update$std9.cov2),

                    mean.std10.trt=mean(datacel.update$std10.trt),
                      mean.std10.cov1=mean(datacel.update$std10.cov1),
                      mean.std10.cov2=mean(datacel.update$std10.cov2),
                      sd.std10.cov1=sd(datacel.update$std10.cov1),
                      sd.std10.cov2=sd(datacel.update$std10.cov2)

)

############ making ipd dataset

data.ipd<-dplyr::select(datacel.update, -c(4:30))

## the covariates need to be the same name in ipd and agd study
data.ipd<-rename(data.ipd, cov1 = std1.cov1,cov2 = std1.cov2)

#############################################################
################  MAIC 1 ###################################
################  study 1(ipd) study 2(agd)   ######################

### renaming variables in agd study
```

```
data.agd_update<-rename(data.agd, cov1 = mean.std2.cov1, cov2 = mean.std2.cov2)

# List out matching covariates
match_cov <- c("cov1", "cov2")
#### center baseline characteristics

#matching continuous variables on both mean and standard deviation

data.ipd_update <- data.ipd %>%
  mutate(cov1_centered = cov1- data.agd$mean.std2.cov1,
         cov1_squared_centered = (cov1^2) - (data.agd$mean.std2.cov1^2 +
                                               data.agd$sd.std2.cov1^2),

         cov2_centered = cov2- data.agd$mean.std2.cov2,
         cov2_squared_centered = (cov2^2) - (data.agd$mean.std2.cov2^2 +
                                               data.agd$sd.std2.cov2^2)
  )

cent_match_cov <- c("cov1_centered",
                    "cov1_squared_centered",
                    "cov2_centered",
                    "cov2_squared_centered"
)

#### estimating weights

est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                matching_vars = cent_match_cov)


####################
# Based on the known proportion of respondents, simulate the response data

comparator_n <- 150 # total number of patients in the comparator data
  comparator_prop_events <- data.agd$mean.std2.trt # proportion of responders

# Calculate number with event

n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                              rep(0, comparator_n - n_with_event)))

comparator_input <-  comparator_binary %>%
  mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1
```

```
 # Merge comparator and intervention data

 combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
 combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")
# Estimate weighted OR by fitting a logistic regression model with weights
 weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                     family = binomial(link="logit"),
                                     data = combined_data,
                                     weight = wt))

 maic1.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
 maic1.se<-coeftest(weighted_OR,vcov=vcovHC(weighted_OR,type="HC3"))[2,2]


 ################  MAIC 2 ###################################
 ################  study 1(ipd) study 3(agd)   ######################

 ### renaming variables in agd study

 data.agd_update<-rename(data.agd, cov1 = mean.std3.cov1, cov2 = mean.std3.cov2)

 # List out matching covariates
 match_cov <- c("cov1", "cov2")

 #### center baseline characteristics

 #matching continuous variables on both mean and standard deviation
 data.ipd_update <- data.ipd %>%
   mutate(cov1_centered = cov1- data.agd$mean.std3.cov1,
          cov1_squared_centered = (cov1^2) - (data.agd$mean.std3.cov1^2 +
                                                data.agd$sd.std3.cov1^2),

          cov2_centered = cov2- data.agd$mean.std3.cov2,
          cov2_squared_centered = (cov2^2) - (data.agd$mean.std3.cov2^2 +
                                                data.agd$sd.std3.cov2^2))


 cent_match_cov <- c("cov1_centered",
                     "cov1_squared_centered",
                     "cov2_centered",
                     "cov2_squared_centered"
 )

 #### estimating weights
```

```
est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                matching_vars = cent_match_cov)

# Based on the known proportion of respondents, simulate the response data
  comparator_n <- 150 # total number of patients in the comparator data
  comparator_prop_events <- data.agd$mean.std3.trt # proportion of responders

# Calculate number with event

n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                              rep(0, comparator_n - n_with_event)))


comparator_input <-  comparator_binary %>%
  mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

# Merge comparator and intervention data

combined_data <- bind_rows(est_weights$analysis_data, comparator_input)

combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

# Estimate weighted OR by fitting a logistic regression model with weights
weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                    family = binomial(link="logit"),
                                    data = combined_data,
                                    weight = wt))

maic2.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
maic2.se<-coeftest(weighted_OR,vcov=vcovHC(weighted_OR,type="HC3"))[2,2]
#####################################################################
################  MAIC 3 ###################################
################  study 1(ipd) study 4(agd)   ######################

### renaming variables in agd study

data.agd_update<-rename(data.agd, cov1 = mean.std4.cov1, cov2 = mean.std4.cov2)

# List out matching covariates
match_cov <- c("cov1","cov2")

#### center baseline characteristics
```

```
#matching continuous variables on both mean and standard deviation

data.ipd_update <- data.ipd %>%
  mutate(cov1_centered = cov1- data.agd$mean.std4.cov1,
         cov1_squared_centered = (cov1^2) - (data.agd$mean.std4.cov1^2 +
                                               data.agd$sd.std4.cov1^2),

         cov2_centered = cov2- data.agd$mean.std4.cov2,
         cov2_squared_centered = (cov2^2) - (data.agd$mean.std4.cov2^2 +
                                               data.agd$sd.std4.cov2^2))


cent_match_cov <- c("cov1_centered",
                    "cov1_squared_centered",
                    "cov2_centered",
                    "cov2_squared_centered"
)

#### estimating weights

est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                matching_vars = cent_match_cov)



# Based on the known proportion of respondents, simulate the response data
  comparator_n <- 150 # total number of patients in the comparator data
  comparator_prop_events <- data.agd$mean.std4.trt # proportion of responders

# Calculate number with event

n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                              rep(0, comparator_n - n_with_event)))


comparator_input <-  comparator_binary %>%
  mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

# Merge comparator and intervention data

combined_data <- bind_rows(est_weights$analysis_data, comparator_input)

combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")
```

```
# Estimate weighted OR by fitting a logistic regression model with weights
weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                    family = binomial(link="logit"),
                                    data = combined_data,
                                    weight = wt))



maic3.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
maic3.se<-coeftest(weighted_OR,vcov=vcovHC(weighted_OR,type="HC3"))[2,2]

#####################################################################
################  MAIC 4 ###################################
################  study 1(ipd) study 5(agd)   ######################

### renaming variables in agd study

data.agd_update<-rename(data.agd, cov1 = mean.std5.cov1, cov2 = mean.std5.cov2)

# List out matching covariates
match_cov <- c("cov1", "cov2")

#### center baseline characteristics

#matching continuous variables on both mean and standard deviation

data.ipd_update <- data.ipd %>%
  mutate(cov1_centered = cov1- data.agd$mean.std5.cov1,
         cov1_squared_centered = (cov1^2) - (data.agd$mean.std5.cov1^2 +
                                               data.agd$sd.std5.cov1^2),

         cov2_centered = cov2- data.agd$mean.std5.cov2,
         cov2_squared_centered = (cov2^2) - (data.agd$mean.std5.cov2^2 +
                                               data.agd$sd.std5.cov2^2))


cent_match_cov <- c("cov1_centered",
                    "cov1_squared_centered",
                    "cov2_centered",
                    "cov2_squared_centered"
)


#### estimating weights
```

```
 est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                 matching_vars = cent_match_cov)

 # Based on the known proportion of respondents, simulate the response data
 comparator_n <- 150 # total number of patients in the comparator data

 comparator_prop_events <- data.agd$mean.std5.trt # proportion of responders

 # Calculate number with event

 n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
 comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                               rep(0, comparator_n - n_with_event)))


 comparator_input <-  comparator_binary %>%
   mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

# Merge comparator and intervention data

 combined_data <- bind_rows(est_weights$analysis_data, comparator_input)

 combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

 # Estimate weighted OR by fitting a logistic regression model with weights
 weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                     family = binomial(link="logit"),
                                     data = combined_data,
                                     weight = wt))


 maic4.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
 maic4.se<-coeftest(weighted_OR,vcov=vcovHC(weighted_OR,type="HC3"))[2,2]

#####################################################################
 ################  MAIC 5 ###################################
 ################  study 1(ipd) study 6(agd)   ######################

 ### renaming variables in agd study
 data.agd_update<-rename(data.agd, cov1 = mean.std6.cov1, cov2 = mean.std6.cov2)

 # List out matching covariates
 match_cov <- c("cov1","cov2")

 #### center baseline characteristics
```

```
#matching continuous variables on both mean and standard deviation

data.ipd_update <- data.ipd %>%
  mutate(cov1_centered = cov1- data.agd$mean.std6.cov1,
         cov1_squared_centered = (cov1^2) - (data.agd$mean.std6.cov1^2 +
                                               data.agd$sd.std6.cov1^2),

         cov2_centered = cov2- data.agd$mean.std6.cov2,
         cov2_squared_centered = (cov2^2) - (data.agd$mean.std6.cov2^2 +
                                               data.agd$sd.std6.cov2^2))


cent_match_cov <- c("cov1_centered",
                    "cov1_squared_centered",
                    "cov2_centered",
                    "cov2_squared_centered"
)

#### estimating weights

est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                matching_vars = cent_match_cov)


# Based on the known proportion of respondents, simulate the response data

comparator_n <- 150 # total number of patients in the comparator data

comparator_prop_events <- data.agd$mean.std6.trt # proportion of responders

# Calculate number with event

n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                              rep(0, comparator_n - n_with_event)))


comparator_input <-  comparator_binary %>%
  mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

# Merge comparator and intervention data

combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
```

```
combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")
# Estimate weighted OR by fitting a logistic regression model with weights
weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                    family = binomial(link="logit"),
                                    data = combined_data,
                                    weight = wt))


maic5.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
maic5.se<-coeftest(weighted_OR,vcov=vcovHC(weighted_OR,type="HC3"))[2,2]

#####################################################################
################  MAIC 6 ###################################
################  study 1(ipd) study 8(agd)   ######################

### renaming variables in agd and

data.agd_update<-rename(data.agd, cov1 = mean.std8.cov1, cov2 = mean.std8.cov2)

# List out matching covariates
match_cov <- c("cov1", "cov2")

#### center baseline characteristics

#matching continuous variables on both mean and standard deviation

data.ipd_update <- data.ipd %>%
  mutate(cov1_centered = cov1- data.agd$mean.std8.cov1,
         cov1_squared_centered = (cov1^2) - (data.agd$mean.std8.cov1^2 +
                                               data.agd$sd.std8.cov1^2),

         cov2_centered = cov2- data.agd$mean.std8.cov2,
         cov2_squared_centered = (cov2^2) - (data.agd$mean.std8.cov2^2 +
                                               data.agd$sd.std8.cov2^2))


cent_match_cov <- c("cov1_centered",
                    "cov1_squared_centered",
                    "cov2_centered",
                    "cov2_squared_centered"
)

#### estimating weights

est_weights <- estimate_weights(intervention_data = data.ipd_update ,
```

```
                                matching_vars = cent_match_cov)


# Based on the known proportion of respondents, simulate the response data

comparator_n <- 150 # total number of patients in the comparator data

comparator_prop_events <- data.agd$mean.std8.trt # proportion of responders

# Calculate number with event

n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                              rep(0, comparator_n - n_with_event)))


comparator_input <-  comparator_binary %>%
  mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

# Merge comparator and intervention data
combined_data <- bind_rows(est_weights$analysis_data, comparator_input)

combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

# Estimate weighted OR by fitting a logistic regression model with weights

weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                    family = binomial(link="logit"),
                                    data = combined_data,
                                    weight = wt))
maic6.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
maic6.se<-coeftest(weighted_OR,vcov=vcovHC(weighted_OR,type="HC3"))[2,2]

#####################################################################
################  MAIC 7 ###################################
################  study 1(ipd) study 9(agd)   ######################

### renaming variables in agd  study

data.agd_update<-rename(data.agd, cov1 = mean.std9.cov1, cov2 = mean.std9.cov2)

# List out matching covariates
match_cov <- c("cov1", "cov2")

#### center baseline characteristics
```

```
#matching continuous variables on both mean and standard deviation

 data.ipd_update <- data.ipd %>%
   mutate(cov1_centered = cov1- data.agd$mean.std9.cov1,
          cov1_squared_centered = (cov1^2) - (data.agd$mean.std9.cov1^2 +
                                                data.agd$sd.std9.cov1^2),

          cov2_centered = cov2- data.agd$mean.std9.cov2,
          cov2_squared_centered = (cov2^2) - (data.agd$mean.std9.cov2^2 +
                                                data.agd$sd.std9.cov2^2))


 cent_match_cov <- c("cov1_centered",
                     "cov1_squared_centered",
                     "cov2_centered",
                     "cov2_squared_centered"
 )

 #### estimating weights

 est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                 matching_vars = cent_match_cov)

 # Based on the known proportion of respondents, simulate the response data

 comparator_n <- 150 # total number of patients in the comparator data
   comparator_prop_events <- data.agd$mean.std9.trt # proportion of responders

 # Calculate number with event

 n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
 comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                               rep(0, comparator_n - n_with_event)))


 comparator_input <-  comparator_binary %>%
   mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

 # Merge comparator and intervention data

 combined_data <- bind_rows(est_weights$analysis_data, comparator_input)

 combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")
```

```
# Estimate weighted OR by fitting a logistic regression model with weights
weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                    family = binomial(link="logit"),
                                    data = combined_data,
                                    weight = wt))

maic7.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
maic7.se<-coeftest(weighted_OR,vcov=vcovHC(weighted_OR,type="HC3"))[2,2]

#####################################################################
################  MAIC 8 ###################################
################  study 1(ipd) study 10(agd)   ######################

### renaming variables in agd  study

data.agd_update<-rename(data.agd, cov1 = mean.std10.cov1, cov2 = mean.std10.cov2)

# List out matching covariates
match_cov <- c("cov1", "cov2")

#### center baseline characteristics

#matching continuous variables on both mean and standard deviation

data.ipd_update <- data.ipd %>%
  mutate(cov1_centered = cov1- data.agd$mean.std10.cov1,
         cov1_squared_centered = (cov1^2) - (data.agd$mean.std10.cov1^2 +
                                               data.agd$sd.std10.cov1^2),

         cov2_centered = cov2- data.agd$mean.std10.cov2,
         cov2_squared_centered = (cov2^2) - (data.agd$mean.std10.cov2^2 +
                                               data.agd$sd.std10.cov2^2))

cent_match_cov <- c("cov1_centered",
                    "cov1_squared_centered",
                    "cov2_centered",
                    "cov2_squared_centered"

)

#### estimating weights

est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                matching_vars = cent_match_cov)
```

```
# Based on the known proportion of respondents, simulate the response data

comparator_n <- 150 # total number of patients in the comparator data
comparator_prop_events <- data.agd$mean.std10.trt # proportion of responders

# Calculate number with event

n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                              rep(0, comparator_n - n_with_event)))


comparator_input <-  comparator_binary %>%
  mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

# Merge comparator and intervention data

combined_data <- bind_rows(est_weights$analysis_data, comparator_input)

combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

# Estimate weighted OR by fitting a logistic regression model with weights
weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                    family = binomial(link="logit"),
                                    data = combined_data,
                                    weight = wt))

maic8.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
maic8.se<-coeftest(weighted_OR,vcov=vcovHC(weighted_OR,type="HC3"))[2,2]

#####################################################################
################  MAIC 9 ###################################
################  study 1(ipd) study 7(agd)   ######################

### renaming variables in agd  study

data.agd_update<-rename(data.agd, cov1 = mean.std7.cov1, cov2 = mean.std7.cov2)

# List out matching covariates
match_cov <- c("cov1", "cov2")

#### center baseline characteristics

#matching continuous variables on both mean and standard deviation
```

```
data.ipd_update <- data.ipd %>%
  mutate(cov1_centered = cov1- data.agd$mean.std7.cov1,
         cov1_squared_centered = (cov1^2) - (data.agd$mean.std7.cov1^2 +
                                               data.agd$sd.std7.cov1^2),

         cov2_centered = cov2- data.agd$mean.std7.cov2,
         cov2_squared_centered = (cov2^2) - (data.agd$mean.std7.cov2^2 +
                                               data.agd$sd.std7.cov2^2))


cent_match_cov <- c("cov1_centered",
                    "cov1_squared_centered",
                    "cov2_centered",
                    "cov2_squared_centered"
)

#### estimating weights

est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                matching_vars = cent_match_cov)

# Based on the known proportion of respondents, simulate the response data
comparator_n <- 150 # total number of patients in the comparator data
comparator_prop_events <- data.agd$mean.std7.trt # proportion of responders

# Calculate number with event

n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                              rep(0, comparator_n - n_with_event)))


comparator_input <-  comparator_binary %>%
  mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

# Merge comparator and intervention data

combined_data <- bind_rows(est_weights$analysis_data, comparator_input)

combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

# Estimate weighted OR by fitting a logistic regression model with weights
weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                    family = binomial(link="logit"),
                                    data = combined_data,
```

```
                                    weight = wt))

maic9.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
maic9.se<-coeftest(weighted_OR,vcov=vcovHC(weighted_OR,type="HC3"))[2,2]

####conducting an NMA with the estimates from the MAIC

     ### creating a data frame which will be used in the nma

  studyn <- c(1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8,9,9)
  trtn <- c(1,2,1,2,1,2,1,2,1,2,1,3,1,3,1,3,1,3)

  n <- rep(150,18)

  diff<-c(NA,maic1.est,
          NA,maic2.est,NA,maic3.est,
          NA,maic4.est,NA,maic5.est,
          NA,maic9.est,NA,maic6.est,
          NA,maic7.est,NA,maic8.est)

  se_diff<-c(NA,maic1.se,
          NA,maic2.se,NA,maic3.se,
          NA,maic4.se,NA,maic5.se,
          NA,maic9.se,NA,maic6.se,
          NA,maic7.se,NA,maic8.se)

  datacel3 <- data.frame(cbind(studyn,trtn,diff,se_diff,n))

  ### following code will set up the network (arm based)

  model.re.disconect<-set_agd_contrast(
    datacel3,
    study=studyn,
    trt=trtn,
    y=diff,
    se=se_diff,
    sample_size=n)

  ##The model is fitted using the nma() function.

  maic_fit_RE <- nma(model.re.disconect,
                     trt_effects = "random",
                     prior_intercept = normal(scale = 100),
                     prior_trt = normal(scale = 10),
                     prior_het = log_normal(-2.56, 0.33),
```

```
                    prior_het_type = "var")

  maic_data_frame<-as.data.frame(maic_fit_RE,pars=c("d","tau"))

  stan.object.nma.discnt.mean.d1<-mean(maic_data_frame$`d[2]`)
  stan.object.nma.discnt.mean.d2<-mean(maic_data_frame$`d[3]`)
  stan.object.nma.discnt.median.tau<-median(maic_data_frame$tau)

  stan.object.nma.discontd.sd.d1<-sd(maic_data_frame$`d[2]`)
  stan.object.nma.discontd.sd.d2<-sd(maic_data_frame$`d[3]`)
  stan.object.nma.discontd.sd.tau<-sd(maic_data_frame$tau)

#### calculation of coverage probability

B1=stan.object.nma.discontd.sd.d1
A1=stan.object.nma.discnt.mean.d1
z.alpha1 <- 1.96
theta.hat.low1=A1-z.alpha1*B1
theta.hat.upp1=A1+z.alpha1*B1
theta1= stan.object.nma.true.mean.d1
est1 <-ifelse(theta1>=theta.hat.low1 & theta1<=theta.hat.upp1,1,0)

B2=stan.object.nma.discontd.sd.d2
A2=stan.object.nma.discnt.mean.d2
z.alpha2 <- 1.96
theta.hat.low2=A2-z.alpha2*B2
theta.hat.upp2=A2+z.alpha2*B2
theta2= stan.object.nma.true.mean.d2
est2 <-ifelse(theta2>=theta.hat.low2 & theta2<=theta.hat.upp2,1,0)

B3=stan.object.nma.connect.sd.d1
A3=stan.object.nma.connect.mean.d1
z.alpha3 <- 1.96
theta.hat.low3=A3-z.alpha3*B3
theta.hat.upp3=A3+z.alpha3*B3
theta3= stan.object.nma.true.mean.d1
est3 <-ifelse(theta3>=theta.hat.low3 & theta3<=theta.hat.upp3,1,0)

B4=stan.object.nma.connect.sd.d2
A4=stan.object.nma.connect.mean.d2
z.alpha4 <- 1.96
theta.hat.low4=A4-z.alpha4*B4
theta.hat.upp4=A4+z.alpha4*B4
theta4= stan.object.nma.true.mean.d2
est4 <-ifelse(theta4>=theta.hat.low4 & theta4<=theta.hat.upp4,1,0)
```

```
  B5=stan.object.nma.connect.sd.tau
  A5=stan.object.nma.connect.median.tau
  z.alpha5 <- 1.96
  theta.hat.low5=A5-z.alpha5*B5
  theta.hat.upp5=A5+z.alpha5*B5
  theta5= stan.object.nma.true.median.tau
  est5 <-ifelse(theta5>=theta.hat.low5 & theta5<=theta.hat.upp5,1,0)

  B6=stan.object.nma.discontd.sd.tau
  A6=stan.object.nma.discnt.median.tau
  z.alpha6 <- 1.96
  theta.hat.low6=A6-z.alpha6*B6
  theta.hat.upp6=A6+z.alpha6*B6
  theta6= stan.object.nma.true.median.tau
  est6 <-ifelse(theta6>=theta.hat.low6 & theta6<=theta.hat.upp6,1,0)


  out[i, 1]<-stan.object.nma.connect.mean.d1
  out[i, 2]<-stan.object.nma.connect.mean.d2
  out[i, 3]<-stan.object.nma.connect.median.tau
  out[i, 4]<-stan.object.nma.connect.sd.tau
  out[i, 5]<-stan.object.nma.discnt.mean.d1
  out[i, 6]<-stan.object.nma.discnt.mean.d2
  out[i, 7]<-stan.object.nma.discontd.sd.d1
  out[i, 8]<-stan.object.nma.discontd.sd.d2
  out[i, 9]<-est1
  out[i, 10]<-est2
  out[i, 11]<-stan.object.nma.discnt.median.tau
  out[i, 12]<-stan.object.nma.discontd.sd.tau
  out[i, 13]<- est3
  out[i, 14]<- est4
  out[i, 15]<-stan.object.nma.connect.sd.d1
  out[i, 16]<-stan.object.nma.connect.sd.d2
  out[i, 17]<-est5
  out[i, 18]<-est6

}

out<-data.frame(out)
```

Appendix E

R codes for double-bootstrapping

E.1 R codes for double-bootstrapping (fixed effect)

```
################################################
# R codes for double bootstrapping (fixed effect)
################################################

library(multinma)
library(dplyr)
library (margins)
library(devtools)
library(MAIC)
library(sandwich)
library(lmtest)
library(boot)
library(foreach)
library(parallel)
library(lme4)


rm(list=ls())
set.seed(1128)

### making R object pc for parameter combination
### which will generate different datasets for
### different parameter combination

corx<-c(0.20,0.80)                      ##correlation between covariates in a study
b_X1_trt<-c(-log(0.78),-log(0.40))       ## 0.25 and 0.916 (interaction coefficient)
b_X1<-c(-log(0.67),-log(0.33))           ## 0.40 and 1.10  (covariate coefficient)
meanx1<-c(0.45,0.15)                    ## mean of covariates
param.combinations <- expand.grid(corx=corx, b_X1_trt=b_X1_trt,
                                  b_X1=b_X1, meanx1=meanx1)
```

```
pc <- param.combinations
pc<-round(pc, 2)                        ## rounding the values
pc$meanx2<-c(0.48,0.48,0.48,0.48,0.48,0.48,0.48,
             0.48,0.20,0.20,0.20,0.20,0.20,0.20,0.20,0.20) ## adding another column in pc
scenerios<-nrow(pc)                      ### no of scenarios created ## 16scenarios

######## Creating a function to generate data
### Generating 10 studies which will consist of 40 columns
## each study is with two arms where
### 150 data will be generated for each arm
### each study will have two continuous covariates
### in total every study consists of 4 columns in the dataset
### all studies have the same common treatment(trt 1)
### but 6 studies have treatment 2 and four studies have
### treatment 3
#########################################################

d = c(log(1),log(1.5),log(.17))           ## log odds ratio (0, 0.40, -1.77)
ns =10                                    ## no of studies
np = matrix(150, ns, 2)                  ## no  of patient in each study
t = matrix(c(1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,3,3,3,3), ns, 2) ## trt in each arm

gen.data<-function(sdx,corx, meanx1,meanx2, b_X1, b_X1_trt){
  count = 0
  count2=2
  datacel<-matrix(NA, nrow=150, ncol=40)
  for(i in 1:ns){
    n <- 150
    sdX <- 0.4                            # standard deviation of each covariate
    rho <- matrix(corx, nrow=2, ncol=2)    # set correlation matrix
    diag(rho) <- rep(1, 2)
    sd.vec <-rep(sdX, 2)
    cor2cov <- function(R, S) {

      sweep(sweep(R, 1, S, "*"),2,S,"*")
    }
    R <- cor2cov(rho, sd.vec)          # covariance matrix

    if(i == 1){                   ## study 1
      mean <- c(X1 =0.60, X2 =0.50)
    }

    if(i >= 2){                     ## study 2 to 10

      mean <- c(X1 =rnorm(1,meanx1, 0.05), X2 =rnorm(1,meanx2, 0.05))
```

```
    }

    cov<-data.frame(MASS::mvrnorm(n, mu = mean, Sigma = R))

    delta2 =d[t[i,2]] - d[t[i,1]]
    mu = 0.85                       ## intercept value in each study

    ### generating outcome variable for control arm
    prob1=1 / (1 + exp(-( mu + b_X1  * (cov$X2)
                          + b_X1 * (cov$X1)  )))

    ### generating outcome variable for treatment arm
    prob2=1 / (1 + exp(-( mu + b_X1 * (cov$X2)
                          + b_X1 * (cov$X1) + delta2
                          + b_X1_trt * (cov$X1 ))))

    ytemp1 = rbinom(np[i,1], size=1,prob=prob1) ### outcome for reference arm
    ytemp2 =  rbinom(np[i,2], size=1, prob=prob2) ###  outcome for trt arm

    datacel[ ,1+count]<-ytemp1
    datacel[ ,2+count]<-ytemp2
    datacel[ ,1+count2]<-cov$X1
    datacel[ ,2+count2]<-cov$X2
    count = count + 4
    count2= count2+ 4

  }
  return(datacel)
}

### now generating and saving multiple no of data for each scenario using replicate
### here for scenario 1

ipd<-replicate(n=1000,expr=gen.data(corx=pc$corx[1], b_X1_trt=pc$b_X1_trt[1],
        b_X1=pc$b_X1[1],meanx1=pc$meanx1[1], meanx2= pc$meanx2[1]),simplify =FALSE )


#####################################################
##### First estimating true d1 and d2 (parameter values) so
### that the estimates can be used in the calculation of
###  coverage in the simulation
### to get the true d1, d2, running the same data
### generation as before but now the sample size is quite
### large (1 million)
#####################################################
```

```
############################################################

d = c(log(1),log(1.5),log(.17))              ## log odds ratio (0, 0.40, -1.77)
ns =10                                        ## no of studies
np = matrix(1000000, ns, 2)                  ## no  of patient in each study
t = matrix(c(1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,3,3,3,3), ns, 2) ## trt in each arm


### Starting loop to generate data

count = 0
count2=2
datacel<-matrix(NA, nrow=1000000, ncol=40)

for(i in 1:ns){
  n <- 1000000
  sdX <- 0.4                                # standard deviation of each covariate
  rho <- matrix(pc[1,1], nrow=2, ncol=2)    # set correlation matrix
  diag(rho) <- rep(1, 2)
  sd.vec <-rep(sdX, 2)
  cor2cov <- function(R, S) {
    sweep(sweep(R, 1, S, "*"),2,S,"*")
  }
  R <- cor2cov(rho, sd.vec)          # covariance matrix
    if(i == 1){
        mean <- c(X1 =0.60, X2 =0.50)
  }
    if(i >= 2){
    mean <- c(X1 =rnorm(1,pc[1,4], 0.05), X2 =rnorm(1,pc[1,5], 0.05))
  }

  cov<-data.frame(MASS::mvrnorm(n, mu = mean, Sigma = R))

  delta2 = d[t[i,2]] - d[t[i,1]]
  mu = 0.85
  b_X1 <- pc[1,3] # conditional effect of variable 1

  b_X1_trt <- pc[1,2] # conditional interaction effect of effect modifier

  prob1=1 / (1 + exp(-( mu + b_X1  * (cov$X2)
                        + b_X1 * (cov$X1)  )))

  prob2=1 / (1 + exp(-( mu + b_X1 * (cov$X2)
                        + b_X1 * (cov$X1) + delta2
                        + b_X1_trt * (cov$X1 ))))
```

```
  ytemp1 = rbinom(np[i,1], size=1,prob=prob1) ### for reference arm

  ytemp2 =  rbinom(np[i,2], size=1, prob=prob2) ###  for trt arm

  datacel[ ,1+count]<-ytemp1
  datacel[ ,2+count]<-ytemp2
  datacel[ ,1+count2]<-cov$X1
  datacel[ ,2+count2]<-cov$X2
  count = count + 4
  count2= count2+ 4

}

datacel<-data.frame(datacel)

### After generating data, a network meta-analysis
### will be conducted to get the true value of d1, d2

### creating a data frame which will be used in the nma
## study no
studyn <- rep(1:10,each=2)
## vector indicates treatment number
trtn <- c(1,2,1,2,1,2,1,2,1,2,1,2,1,3,1,3,1,3,1,3)
 ## no of events in each arm
r <- c(sum(datacel$X1==1),sum(datacel$X2==1),sum(datacel$X5==1),sum(datacel$X6==1),
       sum(datacel$X9==1),sum(datacel$X10==1),sum(datacel$X13==1),sum(datacel$X14==1),
       sum(datacel$X17==1),sum(datacel$X18==1),sum(datacel$X21==1),sum(datacel$X22==1),
       sum(datacel$X25==1),sum(datacel$X26==1),sum(datacel$X29==1),sum(datacel$X30==1),
       sum(datacel$X33==1),sum(datacel$X34==1),sum(datacel$X37==1),sum(datacel$X38==1))

n <- rep(1000000, 20)       ## no of patients in each arm

datacel <- data.frame(cbind(studyn,trtn,r,n))
colSums(select_if(datacel, is.numeric))

### following code will set up the network (arm based)

true.network.fe<-set_agd_arm(
  datacel,
  study=studyn,
  trt=trtn,
  r = r,
  n = n,
  trt_ref = 1 )
```

```
arm_fit_FE_true <- nma(true.network.fe,
                       trt_effects = "fixed",
                       prior_intercept = normal(scale = 100),
                       prior_trt = normal(scale = 100))


#### extracting all the values of d1 and d2
### and storing them in a dataframe
model_true<-as.data.frame(arm_fit_FE_true,pars=c("d"))

### now taking the mean of d1 and d2 values which will be
## used ad the estimate of d1 and d2

stan.object.nma.true.mean.d1<-mean(model_true$`d[2]`)
stan.object.nma.true.mean.d2<-mean(model_true$`d[3]`)

### now starting the loop which will be implemented on every data
##############################################################################
##############################################################################

out<-matrix(NA,1000,10)
colnames(out) <- c("mean.d1", "mean.d2","sd.d1", "sd.d2","est1", "est2",
                   "connect.d1", "connect.d2", "connect.sd.d1", "connect.sd.d2"  )

########################################
########################################
########################################
########################################
########################################

bootstrap <- function(ipd_data) {
  ## datasets are stored as list, converting it to dataframe
  data.main <- data.frame(ipd_data)

 ###############################
  ##  connected NMA codes
  ###############################

  ### creating a data frame which will be used in the nma

studyn <- rep(1:10,each=2)
## vector indicates treatment number
trtn <- c(1,2,1,2,1,2,1,2,1,2,1,2,1,3,1,3,1,3,1,3)
r <- c(sum(data.main $X1==1),sum(data.main$X2==1),sum(data.main$X5==1),sum(data.main$X6==1),
```

```
sum(data.main$X9==1),sum(data.main$X10==1),sum(data.main$X13==1),sum(data.main$X14==1),
sum(data.main$X17==1),sum(data.main$X18==1),sum(data.main$X21==1),sum(data.main$X22==1),
sum(data.main$X25==1),sum(data.main$X26==1),sum(data.main$X29==1),sum(data.main$X30==1),
sum(data.main$X33==1),sum(data.main$X34==1),sum(data.main$X37==1),sum(data.main$X38==1))

n <- rep(150, 20)             ## no of patients in each arm
colSums(select_if(datacel, is.numeric))
datacel2 <- data.frame(cbind(studyn,trtn,r,n))

### following code will set up the network (arm based)

connected.network.fe<-set_agd_arm(
  datacel2,
  study=studyn,
  trt=trtn,
  r = r,
  n = n,
  trt_ref = 1 )

arm_fit_FE_connected <- nma(connected.network.fe,
                        trt_effects = "fixed",
                        prior_intercept = normal(scale = 100),
                        prior_trt = normal(scale = 100))

#### extracting all the values of d1 and d2
### and storing them in a dataframe
model_connected<-as.data.frame(arm_fit_FE_connected,pars=c("d"))

### now taking the mean of d1 and d2 values which will be
## used ad the estimate of d1 and d2

stan.object.nma.connect.mean.d1<-mean(model_connected$`d[2]`)
stan.object.nma.connect.mean.d2<-mean(model_connected$`d[3]`)

stan.object.nma.connect.sd.d1<-sd(model_connected$`d[2]`)
stan.object.nma.connect.sd.d2<-sd(model_connected$`d[3]`)

  datacel= rename(data.main,                        ### renaming the column names

                  std1.ref = "X1", std1.trt = "X2",
                  std1.cov1= "X3", std1.cov2= "X4",

                  std2.ref = "X5", std2.trt = "X6",
                  std2.cov1= "X7", std2.cov2= "X8",
```

```
                  std3.ref = "X9", std3.trt = "X10",
                  std3.cov1= "X11", std3.cov2= "X12",

                  std4.ref = "X13", std4.trt = "X14",
                  std4.cov1= "X15", std4.cov2= "X16",

                  std5.ref = "X17", std5.trt = "X18",
                  std5.cov1= "X19", std5.cov2= "X20",

                  std6.ref = "X21", std6.trt = "X22",
                  std6.cov1= "X23", std6.cov2= "X24",

                  std7.ref = "X25", std7.trt = "X26",
                  std7.cov1= "X27", std7.cov2= "X28",

                  std8.ref = "X29", std8.trt = "X30",
                  std8.cov1= "X31", std8.cov2= "X32",


                  std9.ref = "X33", std9.trt = "X34",
                  std9.cov1= "X35", std9.cov2= "X36",

                  std10.ref = "X37", std10.trt = "X38",
                  std10.cov1= "X39", std10.cov2= "X40")


### deleting multiple columns and making studies into single arm

datacel.update<-dplyr::select( datacel, -c( 'std1.trt','std2.ref', 'std3.ref',
                                            'std4.ref','std5.ref', 'std6.ref',
                                            'std7.ref', 'std8.ref', 'std9.ref',
                                            'std10.ref'))


############ making agd dataset

  data.agd<-data.frame( mean.std2.trt=mean(datacel.update$std2.trt),
                        mean.std2.cov1=mean(datacel.update$std2.cov1),
                        mean.std2.cov2=mean(datacel.update$std2.cov2),
                        sd.std2.cov1=sd(datacel.update$std2.cov1),
                        sd.std2.cov2=sd(datacel.update$std2.cov2),


                        mean.std3.trt=mean(datacel.update$std3.trt),
                        mean.std3.cov1=mean(datacel.update$std3.cov1),
```

```
mean.std3.cov2=mean(datacel.update$std3.cov2),
sd.std3.cov1=sd(datacel.update$std3.cov1),
sd.std3.cov2=sd(datacel.update$std3.cov2),

mean.std4.trt=mean(datacel.update$std4.trt),
mean.std4.cov1=mean(datacel.update$std4.cov1),
mean.std4.cov2=mean(datacel.update$std4.cov2),
sd.std4.cov1=sd(datacel.update$std4.cov1),
sd.std4.cov2=sd(datacel.update$std4.cov2),

mean.std5.trt=mean(datacel.update$std5.trt),
mean.std5.cov1=mean(datacel.update$std5.cov1),
mean.std5.cov2=mean(datacel.update$std5.cov2),
sd.std5.cov1=sd(datacel.update$std5.cov1),
sd.std5.cov2=sd(datacel.update$std5.cov2),

mean.std6.trt=mean(datacel.update$std6.trt),
mean.std6.cov1=mean(datacel.update$std6.cov1),
mean.std6.cov2=mean(datacel.update$std6.cov2),
sd.std6.cov1=sd(datacel.update$std6.cov1),
sd.std6.cov2=sd(datacel.update$std6.cov2),


mean.std7.trt=mean(datacel.update$std7.trt),
mean.std7.cov1=mean(datacel.update$std7.cov1),
mean.std7.cov2=mean(datacel.update$std7.cov2),
sd.std7.cov1=sd(datacel.update$std7.cov1),
sd.std7.cov2=sd(datacel.update$std7.cov2),


mean.std8.trt=mean(datacel.update$std8.trt),
mean.std8.cov1=mean(datacel.update$std8.cov1),
mean.std8.cov2=mean(datacel.update$std8.cov2),
sd.std8.cov1=sd(datacel.update$std8.cov1),
sd.std8.cov2=sd(datacel.update$std8.cov2),


mean.std9.trt=mean(datacel.update$std9.trt),
mean.std9.cov1=mean(datacel.update$std9.cov1),
mean.std9.cov2=mean(datacel.update$std9.cov2),
sd.std9.cov1=sd(datacel.update$std9.cov1),
sd.std9.cov2=sd(datacel.update$std9.cov2),

mean.std10.trt=mean(datacel.update$std10.trt),
mean.std10.cov1=mean(datacel.update$std10.cov1),
```

```
                        mean.std10.cov2=mean(datacel.update$std10.cov2),
                        sd.std10.cov1=sd(datacel.update$std10.cov1),
                        sd.std10.cov2=sd(datacel.update$std10.cov2)



  )

  ############ making ipd dataset by deleting unnecessary columns

  data.ipd<-dplyr::select(datacel.update, -c(4:30))

  #### starting the function for double-bootstrapping

  boot.maic<-function(d,i){

    data.whole.maic <- d[i,]  ##  allows boot to select sample

### Applying 9 unanchored MAICs

#############################################################
 ################  MAIC 1 ###################################
################  study 1(ipd) study 2(agd)   ##################
    boot.std.1.2 <- function(d, i) {

      data.maic1 <- d[i,]

## The covariates need to be the same name in ipd and agd study
      data.maic1<-rename(data.maic1, cov1 = std1.cov1,cov2 = std1.cov2)

### Renaming variables in agd study

      data.agd_update<-rename(data.agd, cov1 = mean.std2.cov1, cov2 = mean.std2.cov2)

# List out matching covariates
      match_cov <- c("cov1", "cov2")

#### center baseline characteristics

#matching continuous variables on both mean and standard deviation

      data.ipd_update <- data.maic1 %>%
        mutate(cov1_centered = cov1- data.agd$mean.std2.cov1,
               cov1_squared_centered = (cov1^2) - (data.agd$mean.std2.cov1^2 +
                                                     data.agd$sd.std2.cov1^2),
```

```
              cov2_centered = cov2- data.agd$mean.std2.cov2,
              cov2_squared_centered = (cov2^2) - (data.agd$mean.std2.cov2^2 +
                                                    data.agd$sd.std2.cov2^2)
       )

     cent_match_cov <- c("cov1_centered",
                         "cov1_squared_centered",
                         "cov2_centered",
                         "cov2_squared_centered"
     )

#### estimating weights

 est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                    matching_vars = cent_match_cov)



# Based on the known proportion of respondents, simulate the response data

comparator_n <- 150 # total number of patients in the comparator data

comparator_prop_events <- data.agd$mean.std2.trt # proportion of responders

# Calculate number with event

     n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
     comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                                   rep(0, comparator_n - n_with_event)))


 comparator_input <-  comparator_binary %>%
       mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

# Merge comparator and intervention data

    combined_data <- bind_rows(est_weights$analysis_data, comparator_input)

     combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

# Estimate weighted OR by fitting a logistic regression model with weights
     weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                         family = binomial(link="logit"),
                                         data = combined_data,
```

```
                                          weight = wt))

maic1.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
 maic1.se<-coeftest(weighted_OR, vcov=vcovHC(weighted_OR,type="HC3"))[2,2]
    return(c(maic1.est,maic1.se))

    }

results.boot.std.1.2 <- boot(data=data.whole.maic, statistic=boot.std.1.2, R=1)

boot.std.1.3 <- function(d, i) {
 data.maic2 <- d[i,]

## The covariates need to be same name in ipd and agd study
    data.maic2<-rename(data.maic2, cov1 = std1.cov1,cov2 = std1.cov2)

 ################  MAIC 2 ###################################
 ################  study 1(ipd) study 3(agd)   ######################

### Renaming variables in agd study

data.agd_update<-rename(data.agd, cov1 = mean.std3.cov1, cov2 = mean.std3.cov2)

# List out matching covariates
    match_cov <- c("cov1", "cov2")

#### center baseline characteristics

#matching continuous variables on both mean and standard deviation

data.ipd_update <- data.maic2 %>%
      mutate(cov1_centered = cov1- data.agd$mean.std3.cov1,
             cov1_squared_centered = (cov1^2) - (data.agd$mean.std3.cov1^2 +
                                                   data.agd$sd.std3.cov1^2),

             cov2_centered = cov2- data.agd$mean.std3.cov2,
             cov2_squared_centered = (cov2^2) - (data.agd$mean.std3.cov2^2 +
                                                   data.agd$sd.std3.cov2^2))

cent_match_cov <- c("cov1_centered",
                        "cov1_squared_centered",
                        "cov2_centered",
                        "cov2_squared_centered"
    )
```

```
#### estimating weights
  est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                  matching_vars = cent_match_cov)


# Based on the known proportion of respondents, simulate the response data
comparator_n <- 150 # total number of patients in the comparator data
comparator_prop_events <- data.agd$mean.std3.trt # proportion of responders

 # Calculate number with event

 n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
    comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                                  rep(0, comparator_n - n_with_event)))

 comparator_input <-  comparator_binary %>%
      mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

# Merge comparator and intervention data

    combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
    combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

# Estimate weighted OR by fitting a logistic regression model with weights
    weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                        family = binomial(link="logit"),
                                        data = combined_data,
                                        weight = wt))

 maic2.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
 maic2.se<-coeftest(weighted_OR, vcov=vcovHC(weighted_OR,type="HC3"))[2,2]
return(c(maic2.est,maic2.se))
  }

 results.boot.std.1.3 <- boot(data=data.whole.maic, statistic=boot.std.1.3, R=1)

boot.std.1.4 <- function(d, i) {
      data.maic3 <- d[i,]

 ## The covariates need to be the same name in ipd and agd study
 data.maic3<-rename(data.maic3, cov1 = std1.cov1,cov2 = std1.cov2)

#####################################################################
 ################  MAIC 3 ###################################
################  study 1(ipd) study 4(agd)   ######################
```

```
### Renaming variables in agd study

data.agd_update<-rename(data.agd, cov1 = mean.std4.cov1, cov2 = mean.std4.cov2)

# List out matching covariates
    match_cov <- c("cov1","cov2")

 #### center baseline characteristics

#matching continuous variables on both mean and standard deviation

    data.ipd_update <- data.maic3 %>%
      mutate(cov1_centered = cov1- data.agd$mean.std4.cov1,
             cov1_squared_centered = (cov1^2) - (data.agd$mean.std4.cov1^2 +
                                                   data.agd$sd.std4.cov1^2),

             cov2_centered = cov2- data.agd$mean.std4.cov2,
             cov2_squared_centered = (cov2^2) - (data.agd$mean.std4.cov2^2 +
                                                   data.agd$sd.std4.cov2^2))

    cent_match_cov <- c("cov1_centered",
                        "cov1_squared_centered",
                        "cov2_centered",
                        "cov2_squared_centered"
    )

 #### estimating weights

est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                    matching_vars = cent_match_cov)


# Based on the known proportion of respondents, simulate the response data
comparator_n <- 150 # total number of patients in the comparator data
 comparator_prop_events <- data.agd$mean.std4.trt # proportion of responders

# Calculate number with event

n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                              rep(0, comparator_n - n_with_event)))


    comparator_input <-  comparator_binary %>%
```

```
      mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

# Merge comparator and intervention data

 combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
 combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

 # Estimate weighted OR by fitting a logistic regression model with weights
    weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                        family = binomial(link="logit"),
                                        data = combined_data,
                                        weight = wt))


 maic3.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
 maic3.se<-coeftest(weighted_OR, vcov=vcovHC(weighted_OR,type="HC3"))[2,2]
    return(c(maic3.est,maic3.se))
    }

    results.boot.std.1.4 <- boot(data=data.whole.maic, statistic=boot.std.1.4, R=1)


boot.std.1.5 <- function(d, i) {
    data.maic4 <- d[i,]

 ## The covariates need to be the same name in ipd and agd study
    data.maic4<-rename(data.maic4, cov1 = std1.cov1,cov2 = std1.cov2)

#####################################################################
 ################  MAIC 4 ###################################
 ################  study 1(ipd) study 5(agd)   ######################

### Renaming variables in agd study

 data.agd_update<-rename(data.agd, cov1 = mean.std5.cov1, cov2 = mean.std5.cov2)

# List out matching covariates
    match_cov <- c("cov1", "cov2")

#### center baseline characteristics

#matching continuous variables on both mean and standard deviation

    data.ipd_update <- data.maic4 %>%
      mutate(cov1_centered = cov1- data.agd$mean.std5.cov1,
```

```
            cov1_squared_centered = (cov1^2) - (data.agd$mean.std5.cov1^2 +
                                                  data.agd$sd.std5.cov1^2),

            cov2_centered = cov2- data.agd$mean.std5.cov2,
            cov2_squared_centered = (cov2^2) - (data.agd$mean.std5.cov2^2 +
                                                  data.agd$sd.std5.cov2^2))

    cent_match_cov <- c("cov1_centered",
                        "cov1_squared_centered",
                        "cov2_centered",
                        "cov2_squared_centered"
    )

#### estimating weights

 est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                    matching_vars = cent_match_cov)

# Based on the known proportion of respondents, simulate the response data
 comparator_n <- 150 # total number of patients in the comparator data
 comparator_prop_events <- data.agd$mean.std5.trt # proportion of responders

 # Calculate number with event

    n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
    comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                                  rep(0, comparator_n - n_with_event)))

 comparator_input <-  comparator_binary %>%
      mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1
# Merge comparator and intervention data
 combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

# Estimate weighted OR by fitting a logistic regression model with weights
    weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                        family = binomial(link="logit"),
                                        data = combined_data,
                                        weight = wt))

 maic4.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
 maic4.se<-coeftest(weighted_OR, vcov=vcovHC(weighted_OR,type="HC3"))[2,2]
return(c(maic4.est,maic4.se))
  }
 results.boot.std.1.5 <- boot(data=data.whole.maic, statistic=boot.std.1.5, R=1)
```

```
 boot.std.1.6 <- function(d, i) {
            data.maic5 <- d[i,]

## The covariates need to be same name in ipd and agd study
      data.maic5<-rename(data.maic5, cov1 = std1.cov1,cov2 = std1.cov2)

#####################################################################
 ################  MAIC 5 ###################################
 ################  study 1(ipd) study 6(agd)   ######################

 ### renaming variables in agd study
    data.agd_update<-rename(data.agd, cov1 = mean.std6.cov1, cov2 = mean.std6.cov2)

# List out matching covariates
    match_cov <- c("cov1","cov2")

 #### center baseline characteristics

#matching continuous variables on both mean and standard deviation

    data.ipd_update <- data.maic5 %>%
      mutate(cov1_centered = cov1- data.agd$mean.std6.cov1,
             cov1_squared_centered = (cov1^2) - (data.agd$mean.std6.cov1^2 +
                                                   data.agd$sd.std6.cov1^2),

             cov2_centered = cov2- data.agd$mean.std6.cov2,
             cov2_squared_centered = (cov2^2) - (data.agd$mean.std6.cov2^2 +
                                                   data.agd$sd.std6.cov2^2))

cent_match_cov <- c("cov1_centered",
                        "cov1_squared_centered",
                        "cov2_centered",
                        "cov2_squared_centered"
    )

#### estimating weights

 est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                    matching_vars = cent_match_cov)

# Based on the known proportion of respondents, simulate the response data
  comparator_n <- 150 # total number of patients in the comparator data
comparator_prop_events <- data.agd$mean.std6.trt # proportion of responders
```

```
 # Calculate number with event

    n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
    comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                                  rep(0, comparator_n - n_with_event)))

    comparator_input <-  comparator_binary %>%
    mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

# Merge comparator and intervention data
 combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")
# Estimate weighted OR by fitting a logistic regression model with weights
    weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                        family = binomial(link="logit"),
                                        data = combined_data,
                                        weight = wt))

    maic5.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
    maic5.se<-coeftest(weighted_OR, vcov=vcovHC(weighted_OR,type="HC3"))[2,2]
return(c(maic5.est,maic5.se))
    }

 results.boot.std.1.6 <- boot(data=data.whole.maic, statistic=boot.std.1.6, R=1)


 boot.std.1.7 <- function(d, i) {

      data.maic6<- d[i,]

 ## The covariates need to be the same name in ipd and agd study
      data.maic6<-rename(data.maic6, cov1 = std1.cov1,cov2 = std1.cov2)

#####################################################################
 ################  MAIC 6 ###################################
 ################  study 1(ipd) study 7(agd)   ######################

 ### Renaming variables in agd  study

      data.agd_update<-rename(data.agd, cov1 = mean.std7.cov1, cov2 = mean.std7.cov2)

      # List out matching covariates
      match_cov <- c("cov1","cov2")

#### center baseline characteristics
```

```
#matching continuous variables on both mean and standard deviation

      data.ipd_update <- data.maic6 %>%
        mutate(cov1_centered = cov1- data.agd$mean.std7.cov1,
               cov1_squared_centered = (cov1^2) - (data.agd$mean.std7.cov1^2 +
                                                     data.agd$sd.std7.cov1^2),

               cov2_centered = cov2- data.agd$mean.std7.cov2,
               cov2_squared_centered = (cov2^2) - (data.agd$mean.std7.cov2^2 +
                                                     data.agd$sd.std7.cov2^2))

      cent_match_cov <- c("cov1_centered",
                          "cov1_squared_centered",
                          "cov2_centered",
                          "cov2_squared_centered"
      )

#### estimating weights

 est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                    matching_vars = cent_match_cov)

# Based on the known proportion of respondents, simulate the response data
 comparator_n <- 150 # total number of patients in the comparator data
 comparator_prop_events <- data.agd$mean.std7.trt # proportion of responders

# Calculate number with event

      n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
      comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                                    rep(0, comparator_n - n_with_event)))

 comparator_input <-  comparator_binary %>%
        mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

# Merge comparator and intervention data

     combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
     combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

# Estimate weighted OR by fitting a logistic regression model with weights
      weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                          family = binomial(link="logit"),
                                          data = combined_data,
```

```
                                          weight = wt))

 maic6.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
maic6.se<-coeftest(weighted_OR, vcov=vcovHC(weighted_OR,type="HC3"))[2,2]
   return(c(maic6.est,maic6.se))
   }

    results.boot.std.1.7 <- boot(data=data.whole.maic, statistic=boot.std.1.7, R=1)

boot.std.1.8 <- function(d, i) {

 data.maic7 <- d[i,]

## The covariates need to be the same name in ipd and agd study
       data.maic7<-rename(data.maic7, cov1 = std1.cov1,cov2 = std1.cov2)

   ########################################################################
   ################  MAIC 7 ###################################
   ################  study 1(ipd) study 8(agd)   ######################

### renaming variables in agd and

data.agd_update<-rename(data.agd, cov1 = mean.std8.cov1, cov2 = mean.std8.cov2)

# List out matching covariates
   match_cov <- c("cov1","cov2")

#### center baseline characteristics

#matching continuous variables on both mean and standard deviation

   data.ipd_update <- data.maic7 %>%
     mutate(cov1_centered = cov1- data.agd$mean.std8.cov1,
            cov1_squared_centered = (cov1^2) - (data.agd$mean.std8.cov1^2 +
                                                  data.agd$sd.std8.cov1^2),

            cov2_centered = cov2- data.agd$mean.std8.cov2,
            cov2_squared_centered = (cov2^2) - (data.agd$mean.std8.cov2^2 +
                                                  data.agd$sd.std8.cov2^2))

       cent_match_cov <- c("cov1_centered",
                           "cov1_squared_centered",
                           "cov2_centered",
                           "cov2_squared_centered"
   )
```

```
 #### estimating weights

 est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                    matching_vars = cent_match_cov)


# Based on the known proportion of respondents, simulate the response data
 comparator_n <- 150 # total number of patients in the comparator data
 comparator_prop_events <- data.agd$mean.std8.trt # proportion of responders

 # Calculate number with event

  n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
    comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                                  rep(0, comparator_n - n_with_event)))

        comparator_input <-  comparator_binary %>%
      mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

# Merge comparator and intervention data

    combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
    combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

# Estimate weighted OR by fitting a logistic regression model with weights
    weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                        family = binomial(link="logit"),
                                        data = combined_data,
                                        weight = wt))
    maic7.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
    maic7.se<-coeftest(weighted_OR, vcov=vcovHC(weighted_OR,type="HC3"))[2,2]
    return(c(maic7.est,maic7.se))
      }

    results.boot.std.1.8 <- boot(data=data.whole.maic, statistic=boot.std.1.8, R=1)

    boot.std.1.9 <- function(d, i) {
     data.maic8<- d[i,]

## The covariates need to be the same name in ipd and agd study
      data.maic8<-rename(data.maic8, cov1 = std1.cov1,cov2 = std1.cov2)

#####################################################################
 ################  MAIC 8 ###################################
```

```
 ################  study 1(ipd) study 9(agd)   ######################

 ### renaming variables in agd  study

    data.agd_update<-rename(data.agd, cov1 = mean.std9.cov1, cov2 = mean.std9.cov2)

 # List out matching covariates
    match_cov <- c("cov1", "cov2")

 #### center baseline characteristics

#matching continuous variables on both mean and standard deviation

    data.ipd_update <- data.maic8%>%
      mutate(cov1_centered = cov1- data.agd$mean.std9.cov1,
             cov1_squared_centered = (cov1^2) - (data.agd$mean.std9.cov1^2 +
                                                   data.agd$sd.std9.cov1^2),

             cov2_centered = cov2- data.agd$mean.std9.cov2,
             cov2_squared_centered = (cov2^2) - (data.agd$mean.std9.cov2^2 +
                                                   data.agd$sd.std9.cov2^2))

    cent_match_cov <- c("cov1_centered",
                        "cov1_squared_centered",
                        "cov2_centered",
                        "cov2_squared_centered"
    )

#### estimating weights

est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                    matching_vars = cent_match_cov)

# Based on the known proportion of respondents, simulate the response data

 comparator_n <- 150 # total number of patients in the comparator data

 comparator_prop_events <- data.agd$mean.std9.trt # proportion of responders

 # Calculate number with event

    n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
    comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                                  rep(0, comparator_n - n_with_event)))
```

```
 comparator_input <-  comparator_binary %>%
      mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

# Merge comparator and intervention data

 combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
  combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

 # Estimate weighted OR by fitting a logistic regression model with weights
    weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                        family = binomial(link="logit"),
                                        data = combined_data,
                                        weight = wt))

 maic8.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
 maic8.se<-coeftest(weighted_OR, vcov=vcovHC(weighted_OR,type="HC3"))[2,2]
return(c(maic8.est,maic8.se))
    }

    results.boot.std.1.9 <- boot(data=data.whole.maic, statistic=boot.std.1.9, R=1)


    boot.std.1.10 <- function(d, i) {
     data.maic9<- d[i,]

## The covariates need to be the same name in ipd and agd study
      data.maic9<-rename(data.maic9, cov1 = std1.cov1,cov2 = std1.cov2)

#####################################################################
 ################  MAIC 9 ###################################
 ################  study 1(ipd) study 10(agd)   ######################

### renaming variables in agd  study

    data.agd_update<-rename(data.agd, cov1 = mean.std10.cov1, cov2 = mean.std10.cov2)

 # List out matching covariates
    match_cov <- c("cov1",  "cov2")

 #### center baseline characteristics
#matching continuous variables on both mean and standard deviation

    data.ipd_update <- data.maic9 %>%
      mutate(cov1_centered = cov1- data.agd$mean.std10.cov1,
```

```
            cov1_squared_centered = (cov1^2) - (data.agd$mean.std10.cov1^2 +
                                                  data.agd$sd.std10.cov1^2),

            cov2_centered = cov2- data.agd$mean.std10.cov2,
            cov2_squared_centered = (cov2^2) - (data.agd$mean.std10.cov2^2 +
                                                  data.agd$sd.std10.cov2^2))

    cent_match_cov <- c("cov1_centered",
                        "cov1_squared_centered",
                        "cov2_centered",
                        "cov2_squared_centered"

    )

#### estimating weights

est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                    matching_vars = cent_match_cov)

# Based on the known proportion of respondents, simulate the response data
    comparator_n <- 150 # total number of patients in the comparator data
    comparator_prop_events <- data.agd$mean.std10.trt # proportion of responders

# Calculate number with event

    n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
    comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                                  rep(0, comparator_n - n_with_event)))


    comparator_input <-  comparator_binary %>%
      mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

# Merge comparator and intervention data

    combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
    combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

# Estimate weighted OR by fitting a logistic regression model with weights
    weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                        family = binomial(link="logit"),
                                        data = combined_data,
                                        weight = wt))

    maic9.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
```

```
    maic9.se<-coeftest(weighted_OR, vcov=vcovHC(weighted_OR,type="HC3"))[2,2]
return(c(maic9.est,maic9.se))
    }

 results.boot.std.1.10 <- boot(data=data.whole.maic, statistic=boot.std.1.10, R=1)

 ####conducting an NMA with the estimates from the MAIC
 ### creating a data frame which will be used in the nma

    studyn <- c(1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8,9,9)
    trtn <- c(1,2,1,2,1,2,1,2,1,2,1,3,1,3,1,3,1,3)

    n <- c(150,150,150,150,150,150,150,150,150,150,150,150,150,150,150,150,150,150)

    diff<-c(NA,results.boot.std.1.2$t[1],
            NA,results.boot.std.1.3$t[1],NA,results.boot.std.1.4$t[1],
            NA,results.boot.std.1.5$t[1],NA,results.boot.std.1.6$t[1],
            NA,results.boot.std.1.7$t[1],NA,results.boot.std.1.8$t[1],
            NA,results.boot.std.1.9$t[1],NA,results.boot.std.1.10$t[1])

    se_diff<-c(NA,results.boot.std.1.2$t[2],
               NA,results.boot.std.1.3$t[2],NA,results.boot.std.1.4$t[2],
               NA,results.boot.std.1.5$t[2],NA,results.boot.std.1.6$t[2],
               NA,results.boot.std.1.7$t[2],NA,results.boot.std.1.8$t[2],
               NA,results.boot.std.1.9$t[2],NA,results.boot.std.1.10$t[2])

    datacel2 <- data.frame(cbind(studyn,trtn,diff,se_diff,n))

    ### following code will set up the network (arm based)

    model.fe<-set_agd_contrast(
      datacel2,
      study=studyn,
      trt=trtn,
      y=diff,
      se=se_diff,
      sample_size=n)

    ##The model is fitted using the nma() function.

    maic_fit_FE <- nma(model.fe,
                       trt_effects = "fixed",
                       prior_intercept = normal(scale = 100),
                       prior_trt = normal(scale = 10))
```

```
  ### extracting the values of d1 and d2 and then calculating the mean

  maic_data_frame<-as.data.frame(maic_fit_FE,pars=c("d"))

  stan.object.nma.discnt.mean.d1<-mean(maic_data_frame$`d[2]`)
  stan.object.nma.discnt.mean.d2<-mean(maic_data_frame$`d[3]`)

  return(c(stan.object.nma.discnt.mean.d1,
           stan.object.nma.discnt.mean.d2))
}


results <- boot(data=data.ipd, statistic=boot.maic, R=300)

mean.d1<-mean(results$t[, 1])  ### mean of first column will give estimate of d1

mean.d2<-mean(results$t[, 2])  ### ### mean of second column will give estimate of d2

sd.d1<-sd(results$t[, 1])    ###sd of first column will give sd of d1

sd.d2<-sd(results$t[, 2])   ###sd of second column will give sd of d2

#### calculation of coverage probability

B1=sd.d1
A1=mean.d1
z.alpha1 <- 1.96
theta.hat.low1=A1-z.alpha1*B1
theta.hat.upp1=A1+z.alpha1*B1
theta1=stan.object.nma.true.mean.d1
est1 <-ifelse(theta1>=theta.hat.low1 & theta1<=theta.hat.upp1,1,0)

B2=sd.d2
A2=mean.d2
z.alpha2 <- 1.96
theta.hat.low2=A2-z.alpha2*B2
theta.hat.upp2=A2+z.alpha2*B2
theta2= stan.object.nma.true.mean.d2
est2 <-ifelse(theta2>=theta.hat.low2 & theta2<=theta.hat.upp2,1,0)

out[i, 1]<-mean.d1
out[i, 2]<-mean.d2
out[i, 3]<-sd.d1
out[i, 4]<-sd.d2
out[i, 5]<-est1
```

```
  out[i, 6]<-est2
out[i, 7]<-stan.object.nma.connect.mean.d1
out[i, 8]<-stan.object.nma.connect.mean.d2
out[i, 9]<-stan.object.nma.connect.sd.d1
out[i, 10]<-stan.object.nma.connect.sd.d2
  out<-data.frame(out)

}

### run in parallel.........

save <- mclapply(ipd, bootstrap)
## saving the output in HPC as a dataframe
saveRDS(save, file="actual.innovtn.scn1.FE.rds")
```

E.2 R codes for double-bootstrapping (random effects)

```
################################################
 # R codes for double bootstrapping (random effects)
################################################
rm(list=ls())
set.seed(1128)

### making R object pc for parameter combination which will
###generate different datasets for
### different parameter combination

corx<-c(0.20,0.80)                      ## correlation between covariates in a study
b_X1_trt<-c(-log(0.78),-log(0.40))       ## 0.25 and 0.916 (interaction coefficient)
b_X1<-c(-log(0.67),-log(0.33))           ## 0.40 and 1.10  (covariate coefficient)

meanx1<-c(0.45,0.15)                    ## mean of covariates
param.combinations <- expand.grid(corx=corx, b_X1_trt=b_X1_trt,
                                  b_X1=b_X1, meanx1=meanx1)
pc <- param.combinations
pc<-round(pc, 2)                       ## rounding the values
pc$meanx2<-c(0.48,0.48,0.48,0.48,0.48,0.48,0.48,
             0.48,0.20,0.20,0.20,0.20,0.20,0.20,0.20,0.20)  ## adding another column in pc
scenerios<-nrow(pc)                     ### no of scenarios created ## 16scenarios



######## creating a function to generate data
```

```
### here generating 10 studies which will consist of 40 columns
## each study is with two arm where
### 150 data will be generated for each arm
### each study will have two continuous covariates
### in total every study is consist of 4 columns in the dataset
### all studies have the same control treatment(trt 1)
###  6 studies have treatment 2 and four studies have
### treatment 3
#########################################################

d = c(log(1),log(1.5),log(.17))           ## log odds ratio (0, 0.40, -1.77)
tau = 0.3                                 ## hetrogeneity parameter
ns =10                                    ## no of studies
np = matrix(150, ns, 2)                   ## no  of patient in each study
t = matrix(c(1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,3,3,3,3), ns, 2) ## trt in each arm

gen.data<-function(sdx,corx, meanx1,meanx2, b_X1, b_X1_trt){
  count = 0
  count2=2
  datacel<-matrix(NA, nrow=150, ncol=40)
  for(i in 1:ns){
    n <- 150
    sdX <- 0.4                             # standard deviation of each covariate

    rho <- matrix(corx, nrow=2, ncol=2)    # set correlation matrix
    diag(rho) <- rep(1, 2)
    sd.vec <-rep(sdX, 2)
    cor2cov <- function(R, S) {

      sweep(sweep(R, 1, S, "*"),2,S,"*")
    }
    R <- cor2cov(rho, sd.vec)          # covariance matrix

    if(i == 1){                        ## study 1

      mean <- c(X1 =0.60, X2 =0.50)
    }

    if(i >= 2){                        ## study 2 to 10

      mean <- c(X1 =rnorm(1,meanx1, 0.05), X2 =rnorm(1,meanx2, 0.05))
    }

    cov<-data.frame(MASS::mvrnorm(n, mu = mean, Sigma = R))
   ## treatment effect for trt arm in each study
```

```
    delta2 =rnorm(1, d[t[i,2]] - d[t[i,1]], tau)

mu = 0.85               ## intercept value in each study
### generating outcome variable for control arm
 prob1=1 / (1 + exp(-( mu + b_X1  * (cov$X2)
                          + b_X1 * (cov$X1)  )))
### generating outcome variable for treatment arm
 prob2=1 / (1 + exp(-( mu + b_X1 * (cov$X2)
                          + b_X1 * (cov$X1) + delta2
                          + b_X1_trt * (cov$X1 ))))

 ytemp1 = rbinom(np[i,1], size=1,prob=prob1)   ### outcome for reference arm

 ytemp2 =  rbinom(np[i,2], size=1, prob=prob2)  ###  outcome for trt arm

    datacel[ ,1+count]<-ytemp1
    datacel[ ,2+count]<-ytemp2
    datacel[ ,1+count2]<-cov$X1
    datacel[ ,2+count2]<-cov$X2
    count = count + 4
    count2= count2+ 4

  }
  return(datacel)
}

### now generating and saving multiple no of data for each scenario using replicate
### scenario 2
ipd<-replicate(n=1000,expr=gen.data(corx=pc$corx[2], b_X1_trt=pc$b_X1_trt[2],

## first 30 datasets
ipd<-ipd[c(1:30)]

#####################################################
##### first estimating true d1 and d2 (parameter values) so
### that the estimates can be used in the calculation of
###  coverage in the simulation
### to get the true d1, d2, running the same data
### generation as before but now the sample size is quite
### large (1 million)
#####################################################
#############################################################

d = c(log(1),log(1.5),log(.17))             ## log odds ratio (0, 0.40, -1.77)
tau= 0.3
```

```
ns =10                                          ## no of studies
np = matrix(1000000, ns, 2)                     ## no  of patient in each study
t = matrix(c(1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,3,3,3,3), ns, 2) ## trt in each arm

count = 0
count2=2
datacel<-matrix(NA, nrow=1000000, ncol=40)
for(i in 1:ns){
  n <- 1000000

  sdX <- 0.4                              # standard deviation of each covariate

  rho <- matrix(pc[2,1], nrow=2, ncol=2)    # set correlation matrix
  diag(rho) <- rep(1, 2)
  sd.vec <-rep(sdX, 2)
  cor2cov <- function(R, S) {

    sweep(sweep(R, 1, S, "*"),2,S,"*")
  }
  R <- cor2cov(rho, sd.vec)          # covariance matrix

  if(i == 1){
    mean <- c(X1 =0.60, X2 =0.50)
  }

  if(i == 2){
    mean <- c(X1 =rnorm(1,pc[2,4], 0.05), X2 =rnorm(1,pc[2,5], 0.05))
  }

  if(i == 3){
    mean <- c(X1 =rnorm(1,pc[2,4], 0.05), X2 =rnorm(1,pc[2,5], 0.05))
  }

  if(i == 4){
    mean <- c(X1 =rnorm(1,pc[2,4], 0.05), X2 =rnorm(1,pc[2,5], 0.05))
  }

  if(i == 5){
    mean <- c(X1 =rnorm(1,pc[2,4], 0.05), X2 =rnorm(1,pc[2,5], 0.05))
  }

  if(i == 6){
    mean <- c(X1 =rnorm(1,pc[2,4], 0.05), X2 =rnorm(1,pc[2,5], 0.05))
  }
  if(i == 7){
```

```
    mean <- c(X1 =rnorm(1,pc[2,4], 0.05), X2 =rnorm(1,pc[2,5], 0.05))
  }
  if(i == 8){
    mean <- c(X1 =rnorm(1,pc[2,4], 0.05), X2 =rnorm(1,pc[2,5], 0.05))
  }

  if(i == 9){
    mean <- c(X1 =rnorm(1,pc[2,4], 0.05), X2 =rnorm(1,pc[2,5], 0.05))
  }

  if(i == 10){

    mean <- c(X1 =rnorm(1,pc[2,4], 0.05), X2 =rnorm(1,pc[2,5], 0.05))
  }

  cov<-data.frame(MASS::mvrnorm(n, mu = mean, Sigma = R))

  delta2 =rnorm(1, d[t[i,2]] - d[t[i,1]], tau)
  mu = 0.85
  b_X1 <- pc[2,3]       # conditional effect of variable 1 and 2

  b_X1_trt <- pc[2,2]  # conditional interaction effect of  effect modifier

  prob1=1 / (1 + exp(-( mu + b_X1  * (cov$X2)
                        + b_X1 * (cov$X1)  )))

  prob2=1 / (1 + exp(-( mu + b_X1 * (cov$X2)
                        + b_X1 * (cov$X1) + delta2
                        + b_X1_trt * (cov$X1 ))))

  ytemp1 = rbinom(np[i,1], size=1,prob=prob1) ### for reference arm

  ytemp2 =  rbinom(np[i,2], size=1, prob=prob2) ###  for trt arm

  datacel[ ,1+count]<-ytemp1 ## response in every (1,2),(5,6),(9,10)th column
  datacel[ ,2+count]<-ytemp2
  datacel[ ,1+count2]<-cov$X1 ## cov value in every (3,4),(7,8),(11,12)th column
  datacel[ ,2+count2]<-cov$X2
  count = count + 4
  count2= count2+ 4

}

datacel<-data.frame(datacel)
```

```
##############################
## running NMA with the big data
##############################

### after generating data, now will conduct a network meta analysis
### with the data to extract the value of true
### d1, d2 which will be used in calculating the confidence interval
### and coverage

### creating a data frame which will be used in the nma

studyn <- rep(1:10,each=2)
## vector indicates treatment number
trtn <- c(1,2,1,2,1,2,1,2,1,2,1,3,1,3,1,3,1,3)
## no of events in each arm
r <- c(sum(datacel$X1==1),sum(datacel$X2==1),sum(datacel$X5==1),sum(datacel$X6==1),
       sum(datacel$X9==1),sum(datacel$X10==1),sum(datacel$X13==1),sum(datacel$X14==1),
       sum(datacel$X17==1),sum(datacel$X18==1),sum(datacel$X21==1),sum(datacel$X22==1),
       sum(datacel$X25==1),sum(datacel$X26==1),sum(datacel$X29==1),sum(datacel$X30==1),
       sum(datacel$X33==1),sum(datacel$X34==1),sum(datacel$X37==1),sum(datacel$X38==1))
n <- rep(1000000, 20)             ## no of patients in each arm
colSums(select_if(datacel, is.numeric))
datacel <- data.frame(cbind(studyn,trtn,r,n))

### following code will set up the network (arm based)

true.network.re<-set_agd_arm(
  datacel,
  study=studyn,
  trt=trtn,
  r = r,
  n = n,
  trt_ref = 1 )


##The model is fitted using the nma() function. nma function will generate
### the value of true d1, d2

arm_fit_RE_true <- nma(true.network.re,
                       trt_effects = "random",
                       prior_intercept = normal(scale = 100),
                       prior_trt = normal(scale = 10),
                       prior_het = log_normal(-2.56, 0.33),
                      prior_het_type = "var")
                      )
```

```
#### extracting all the values of d1 and d2
### and storing them in a dataframe
model_true<-as.data.frame(arm_fit_RE_true,pars=c("d","tau"))

### now taking the mean of d1 and d2 values which will be
## used as the estimate of d1 and d2

stan.object.nma.true.mean.d1<-mean(model_true$`d[2]`)
stan.object.nma.true.mean.d2<-mean(model_true$`d[3]`)

stan.object.nma.true.median.tau<-median(model_true$tau)

### now starting the loop which will be implemented on every data
##############################################################################
##############################################################################


## before running the loop, making a matrix to store results.
## this matrix will store result for the previously
## generated  datsets

out<-matrix(NA,30,10)
colnames(out) <- c("mean.d1", "mean.d2","sd.d1", "sd.d2","est1", "est2",
                   "connect.d1", "connect.d2", "connect.sd.d1", "connect.sd.d2"  )


########################################
########################################
##########################################

bootstrap <- function(ipd_data) {
 ## datasets are stored as list, converting it to dataframe
  data.main <- data.frame(ipd_data)

##############################
##  connected NMA codes
##############################

### creating a data frame which will be used in the nma

studyn <- rep(1:10,each=2)
trtn <- c(1,2,1,2,1,2,1,2,1,2,1,2,1,3,1,3,1,3,1,3)    ## vector indicates treatment number

## no of events in each arm
```

```
r <- c(sum(data.main $X1==1),sum(data.main$X2==1),sum(data.main$X5==1),
     sum(data.main$X6==1), sum(data.main$X9==1),sum(data.main$X10==1),
     sum(data.main$X13==1),  sum(data.main$X14==1), sum(data.main$X17==1),
     sum(data.main$X18==1), sum(data.main$X21==1), sum(data.main$X22==1),
     sum(data.main$X25==1), sum(data.main$X26==1),sum(data.main$X29==1),
     sum(data.main$X30==1),  sum(data.main$X33==1),sum(data.main$X34==1),
     sum(data.main$X37==1),  sum(data.main$X38==1))

n <- rep(150, 20)             ## no of patients in each arm

datacel2 <- data.frame(cbind(studyn,trtn,r,n))

### following code will set up the network (arm based)

connected.network.re<-set_agd_arm(
  datacel2,
  study=studyn,
  trt=trtn,
  r = r,
  n = n,
  trt_ref = 1 )

arm_fit_RE_connected <- nma(connected.network.re,
                        trt_effects = "random",
                        prior_intercept = normal(scale = 100),
                        prior_trt = normal(scale = 10),
                        prior_het = log_normal(-2.56, 0.33),
                       prior_het_type = "var")

#### extracting all the values of d1 and d2
### and storing them in a dataframe
model_connected<-as.data.frame(arm_fit_RE_connected,pars=c("d"))

### now taking the mean of d1 and d2 values which will be
## used as the estimate of d1 and d2

stan.object.nma.connect.mean.d1<-mean(model_connected$`d[2]`)
stan.object.nma.connect.mean.d2<-mean(model_connected$`d[3]`)

stan.object.nma.connect.sd.d1<-sd(model_connected$`d[2]`)
stan.object.nma.connect.sd.d2<-sd(model_connected$`d[3]`)

  datacel= rename(data.main,                        ### renaming the column names

                  std1.ref = "X1", std1.trt = "X2",
```

```
                std1.cov1= "X3", std1.cov2= "X4",

                std2.ref = "X5", std2.trt = "X6",
                std2.cov1= "X7", std2.cov2= "X8",

                std3.ref = "X9", std3.trt = "X10",
                std3.cov1= "X11", std3.cov2= "X12",

                std4.ref = "X13", std4.trt = "X14",
                std4.cov1= "X15", std4.cov2= "X16",

                std5.ref = "X17", std5.trt = "X18",
                std5.cov1= "X19", std5.cov2= "X20",

                std6.ref = "X21", std6.trt = "X22",
                std6.cov1= "X23", std6.cov2= "X24",

                std7.ref = "X25", std7.trt = "X26",
                std7.cov1= "X27", std7.cov2= "X28",

                std8.ref = "X29", std8.trt = "X30",
                std8.cov1= "X31", std8.cov2= "X32",


                std9.ref = "X33", std9.trt = "X34",
                std9.cov1= "X35", std9.cov2= "X36",

                std10.ref = "X37", std10.trt = "X38",
                std10.cov1= "X39", std10.cov2= "X40")


### deleting multiple columns and making studies into single arm

datacel.update<-dplyr::select( datacel, -c( 'std1.trt','std2.ref', 'std3.ref',
                                            'std4.ref','std5.ref', 'std6.ref',
                                            'std7.ref', 'std8.ref', 'std9.ref',
                                            'std10.ref'))

## now only keeping study 1  IPD and will make other study into
## agd data

############ making agd dataset

  data.agd<-data.frame( mean.std2.trt=mean(datacel.update$std2.trt),
                        mean.std2.cov1=mean(datacel.update$std2.cov1),
```

```
mean.std2.cov2=mean(datacel.update$std2.cov2),
sd.std2.cov1=sd(datacel.update$std2.cov1),
sd.std2.cov2=sd(datacel.update$std2.cov2),


mean.std3.trt=mean(datacel.update$std3.trt),
mean.std3.cov1=mean(datacel.update$std3.cov1),
mean.std3.cov2=mean(datacel.update$std3.cov2),
sd.std3.cov1=sd(datacel.update$std3.cov1),
sd.std3.cov2=sd(datacel.update$std3.cov2),

mean.std4.trt=mean(datacel.update$std4.trt),
mean.std4.cov1=mean(datacel.update$std4.cov1),
mean.std4.cov2=mean(datacel.update$std4.cov2),
sd.std4.cov1=sd(datacel.update$std4.cov1),
sd.std4.cov2=sd(datacel.update$std4.cov2),

mean.std5.trt=mean(datacel.update$std5.trt),
mean.std5.cov1=mean(datacel.update$std5.cov1),
mean.std5.cov2=mean(datacel.update$std5.cov2),
sd.std5.cov1=sd(datacel.update$std5.cov1),
sd.std5.cov2=sd(datacel.update$std5.cov2),

mean.std6.trt=mean(datacel.update$std6.trt),
mean.std6.cov1=mean(datacel.update$std6.cov1),
mean.std6.cov2=mean(datacel.update$std6.cov2),
sd.std6.cov1=sd(datacel.update$std6.cov1),
sd.std6.cov2=sd(datacel.update$std6.cov2),


mean.std7.trt=mean(datacel.update$std7.trt),
mean.std7.cov1=mean(datacel.update$std7.cov1),
mean.std7.cov2=mean(datacel.update$std7.cov2),
sd.std7.cov1=sd(datacel.update$std7.cov1),
sd.std7.cov2=sd(datacel.update$std7.cov2),


mean.std8.trt=mean(datacel.update$std8.trt),
mean.std8.cov1=mean(datacel.update$std8.cov1),
mean.std8.cov2=mean(datacel.update$std8.cov2),
sd.std8.cov1=sd(datacel.update$std8.cov1),
sd.std8.cov2=sd(datacel.update$std8.cov2),


mean.std9.trt=mean(datacel.update$std9.trt),
```

```
                        mean.std9.cov1=mean(datacel.update$std9.cov1),
                        mean.std9.cov2=mean(datacel.update$std9.cov2),
                        sd.std9.cov1=sd(datacel.update$std9.cov1),
                        sd.std9.cov2=sd(datacel.update$std9.cov2),

                        mean.std10.trt=mean(datacel.update$std10.trt),
                        mean.std10.cov1=mean(datacel.update$std10.cov1),
                        mean.std10.cov2=mean(datacel.update$std10.cov2),
                        sd.std10.cov1=sd(datacel.update$std10.cov1),
                        sd.std10.cov2=sd(datacel.update$std10.cov2)


 )

  ############ making ipd dataset by deleting unnecessary columns

    data.ipd<-dplyr::select(datacel.update, -c(4:30))

  ### upto now the ipd dataset and aggregate dataset is complete
  ### now will start a function for bootstraping of the data.ipd

  #### starting the function for bootstrapping

  boot.maic<-function(d,i){

    data.whole.maic <- d[i,]     ##  allows boot to select sample

### Now 9 unanchored MAICs will be applied

#############################################################

################  MAIC 1 ###################################
################  study 1(ipd) study 2(agd)   ######################
    boot.std.1.2 <- function(d, i) {

      data.maic1 <- d[i,]

## The covariates need to be same name in ipd and agd study
      data.maic1<-rename(data.maic1, cov1 = std1.cov1,cov2 = std1.cov2)

### Renaming variables in agd study
data.agd_update<-rename(data.agd, cov1 = mean.std2.cov1, cov2 = mean.std2.cov2)

# List out matching covariates
      match_cov <- c("cov1",   "cov2")
```

```
#### center baseline characteristics

#matching continuous variables on both mean and standard deviation

           data.ipd_update <- data.maic1 %>%
        mutate(cov1_centered = cov1- data.agd$mean.std2.cov1,
               cov1_squared_centered = (cov1^2) - (data.agd$mean.std2.cov1^2 +
                                                     data.agd$sd.std2.cov1^2),

               cov2_centered = cov2- data.agd$mean.std2.cov2,
               cov2_squared_centered = (cov2^2) - (data.agd$mean.std2.cov2^2 +
                                                     data.agd$sd.std2.cov2^2)
        )

      cent_match_cov <- c("cov1_centered",
                          "cov1_squared_centered",
                          "cov2_centered",
                          "cov2_squared_centered"
      )

#### estimating weights
      est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                      matching_vars = cent_match_cov)

# Based on the known proportion of respondents, simulate the response data

comparator_n <- 150 # total number of patients in the comparator data

comparator_prop_events <- data.agd$mean.std2.trt # proportion of responders

# Calculate number with event
      # Use round() to ensure we end up with a whole number of people

 n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                              rep(0, comparator_n - n_with_event)))

comparator_input <-  comparator_binary %>%
        mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

# Merge comparator and intervention data
combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
      combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")
```

```
# Estimate weighted OR by fitting a logistic regression model with weights
      weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                          family = binomial(link="logit"),
                                          data = combined_data,
                                          weight = wt))

maic1.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
      maic1.se<-coeftest(weighted_OR, vcov=vcovHC(weighted_OR,type="HC3"))[2,2]

      return(c(maic1.est,maic1.se))

    }

    results.boot.std.1.2 <- boot(data=data.whole.maic, statistic=boot.std.1.2, R=1)

################  MAIC 2 ###################################
################  study 1(ipd) study 3(agd)   ######################

 boot.std.1.3 <- function(d, i) {

data.maic2 <- d[i,]
## The covariates needs to be same name in ipd and agd study
    data.maic2<-rename(data.maic2, cov1 = std1.cov1,cov2 = std1.cov2)

### Renaming variables in agd study

data.agd_update<-rename(data.agd, cov1 = mean.std3.cov1, cov2 = mean.std3.cov2)

# List out matching covariates
    match_cov <- c("cov1",  "cov2")

#matching continuous variables on both mean and standard deviation

 data.ipd_update <- data.maic2 %>%
      mutate(cov1_centered = cov1- data.agd$mean.std3.cov1,
             cov1_squared_centered = (cov1^2) - (data.agd$mean.std3.cov1^2 +
                                                   data.agd$sd.std3.cov1^2),

             cov2_centered = cov2- data.agd$mean.std3.cov2,
             cov2_squared_centered = (cov2^2) - (data.agd$mean.std3.cov2^2 +
                                                   data.agd$sd.std3.cov2^2))

    cent_match_cov <- c("cov1_centered",
                        "cov1_squared_centered",
                        "cov2_centered",
```

```
                        "cov2_squared_centered"
    )

#### estimating weights

est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                    matching_vars = cent_match_cov)

# Based on the known proportion of respondents, simulate the response data
comparator_n <- 150 # total number of patients in the comparator data

comparator_prop_events <- data.agd$mean.std3.trt # proportion of responders

# Calculate number with event
# Use round() to ensure we end up with a whole number of people

n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
    comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                                  rep(0, comparator_n - n_with_event)))

comparator_input <-  comparator_binary %>%
      mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

# Merge comparator and intervention data
    combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
    combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

# Estimate weighted OR by fitting a logistic regression model with weights
    weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                        family = binomial(link="logit"),
                                        data = combined_data,
                                        weight = wt))

maic2.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
    maic2.se<-coeftest(weighted_OR, vcov=vcovHC(weighted_OR,type="HC3"))[2,2]
    return(c(maic2.est,maic2.se))
  }

results.boot.std.1.3 <- boot(data=data.whole.maic, statistic=boot.std.1.3, R=1)

#####################################################################
 ################  MAIC 3 ###################################
 ################  study 1(ipd) study 4(agd)   ######################

boot.std.1.4 <- function(d, i) {
```

```
 data.maic3 <- d[i,]

## The covariates need to be same name in ipd and agd study
      data.maic3<-rename(data.maic3, cov1 = std1.cov1,cov2 = std1.cov2)

### Renaming variables in agd study
    data.agd_update<-rename(data.agd, cov1 = mean.std4.cov1, cov2 = mean.std4.cov2)

# List out matching covariates
    match_cov <- c("cov1",  "cov2")

#### center baseline characteristics
#matching continuous variables on both mean and standard deviation

data.ipd_update <- data.maic3 %>%
      mutate(cov1_centered = cov1- data.agd$mean.std4.cov1,
             cov1_squared_centered = (cov1^2) - (data.agd$mean.std4.cov1^2 +
                                                   data.agd$sd.std4.cov1^2),

             cov2_centered = cov2- data.agd$mean.std4.cov2,
             cov2_squared_centered = (cov2^2) - (data.agd$mean.std4.cov2^2 +
                                                   data.agd$sd.std4.cov2^2))

cent_match_cov <- c("cov1_centered",
                        "cov1_squared_centered",
                        "cov2_centered",
                        "cov2_squared_centered"
    )

#### estimating weights

est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                    matching_vars = cent_match_cov)

# Based on the known proportion of respondents, simulate the response data

comparator_n <- 150 # total number of patients in the comparator data
    comparator_prop_events <- data.agd$mean.std4.trt # proportion of responders

# Calculate number with event
 # Use round() to ensure we end up with a whole number of people
    n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
    comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                                  rep(0, comparator_n - n_with_event)))
```

```
 comparator_input <-  comparator_binary %>%
      mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

 combined_data <- bind_rows(est_weights$analysis_data, comparator_input)

combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")


# Estimate weighted OR by fitting a logistic regression model with weights
    weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                        family = binomial(link="logit"),
                                        data = combined_data,
                                        weight = wt))

maic3.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
    maic3.se<-coeftest(weighted_OR, vcov=vcovHC(weighted_OR,type="HC3"))[2,2]
    return(c(maic3.est,maic3.se))
    }

results.boot.std.1.4 <- boot(data=data.whole.maic, statistic=boot.std.1.4, R=1)

#####################################################################
################  MAIC 4 ###################################
################  study 1(ipd) study 5(agd)   ######################

  boot.std.1.5 <- function(d, i) {
    data.maic4 <- d[i,]
 ## The covariates need to be same name in ipd and agd study
    data.maic4<-rename(data.maic4, cov1 = std1.cov1,cov2 = std1.cov2)

### Renaming variables in agd study
    data.agd_update<-rename(data.agd, cov1 = mean.std5.cov1, cov2 = mean.std5.cov2)

# List out matching covariates
    match_cov <- c("cov1",     "cov2")

#### center baseline characteristics
#matching continuous variables on both mean and standard deviation

    data.ipd_update <- data.maic4 %>%
      mutate(cov1_centered = cov1- data.agd$mean.std5.cov1,
             cov1_squared_centered = (cov1^2) - (data.agd$mean.std5.cov1^2 +
                                                   data.agd$sd.std5.cov1^2),
```

```
            cov2_centered = cov2- data.agd$mean.std5.cov2,
            cov2_squared_centered = (cov2^2) - (data.agd$mean.std5.cov2^2 +
                                                  data.agd$sd.std5.cov2^2))

 cent_match_cov <- c("cov1_centered",
                       "cov1_squared_centered",
                       "cov2_centered",
                       "cov2_squared_centered"
    )

#### estimating weights

 est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                   matching_vars = cent_match_cov)

# Based on the known proportion of respondents, simulate the response data
 comparator_n <- 150 # total number of patients in the comparator data
 comparator_prop_events <- data.agd$mean.std5.trt # proportion of responders

# Calculate number with event
    # Use round() to ensure we end up with a whole number of people

 n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
    comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                                  rep(0, comparator_n - n_with_event)))

 comparator_input <-  comparator_binary %>%
      mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

 combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
    combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

 # Estimate weighted OR by fitting a logistic regression model with weights
    weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                        family = binomial(link="logit"),
                                        data = combined_data,
                                        weight = wt))
maic4.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
    maic4.se<-coeftest(weighted_OR, vcov=vcovHC(weighted_OR,type="HC3"))[2,2]
return(c(maic4.est,maic4.se))
  }
results.boot.std.1.5 <- boot(data=data.whole.maic, statistic=boot.std.1.5, R=1)


#####################################################################
```

```
################  MAIC 5 ###################################
 ################  study 1(ipd) study 6(agd)   ######################

 boot.std.1.6 <- function(d, i) {

 data.maic5 <- d[i,]
## The covariates need to be same name in ipd and agd study
      data.maic5<-rename(data.maic5, cov1 = std1.cov1,cov2 = std1.cov2)

### Renaming variables in agd study
    data.agd_update<-rename(data.agd, cov1 = mean.std6.cov1, cov2 = mean.std6.cov2)

# List out matching covariates
    match_cov <- c("cov1",    "cov2")

 #### center baseline characteristics
#matching continuous variables on both mean and standard deviation

    data.ipd_update <- data.maic5 %>%
      mutate(cov1_centered = cov1- data.agd$mean.std6.cov1,
             cov1_squared_centered = (cov1^2) - (data.agd$mean.std6.cov1^2 +
                                                   data.agd$sd.std6.cov1^2),

             cov2_centered = cov2- data.agd$mean.std6.cov2,
             cov2_squared_centered = (cov2^2) - (data.agd$mean.std6.cov2^2 +
                                                   data.agd$sd.std6.cov2^2))

    cent_match_cov <- c("cov1_centered",
                        "cov1_squared_centered",
                        "cov2_centered",
                        "cov2_squared_centered"
    )

 #### estimating weights

est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                    matching_vars = cent_match_cov)

# Based on the known proportion of respondents, simulate the response data

comparator_n <- 150 # total number of patients in the comparator data
    comparator_prop_events <- data.agd$mean.std6.trt # proportion of responders

# Calculate number with event
# Use round() to ensure we end up with a whole number of people
```

```
    n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
    comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                                  rep(0, comparator_n - n_with_event)))

comparator_input <-  comparator_binary %>%
      mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
    combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

# Estimate weighted OR by fitting a logistic regression model with weights

weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                    family = binomial(link="logit"),
                                    data = combined_data,
                                    weight = wt))
maic5.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
 maic5.se<-coeftest(weighted_OR, vcov=vcovHC(weighted_OR,type="HC3"))[2,2]
return(c(maic5.est,maic5.se))
    }

 results.boot.std.1.6 <- boot(data=data.whole.maic, statistic=boot.std.1.6, R=1)


#####################################################################
################  MAIC 6 ###################################
################  study 1(ipd) study 7(agd)   ######################
 boot.std.1.7 <- function(d, i) {

data.maic6<- d[i,]

## The covariates need to be same name in ipd and agd study
      data.maic6<-rename(data.maic6, cov1 = std1.cov1,cov2 = std1.cov2)

### Renaming variables in agd  study
     data.agd_update<-rename(data.agd, cov1 = mean.std7.cov1, cov2 = mean.std7.cov2)

 # List out matching covariates
match_cov <- c("cov1",    "cov2")

 #### center baseline characteristics
#matching continuous variables on both mean and standard deviation

      data.ipd_update <- data.maic6 %>%
```

```
        mutate(cov1_centered = cov1- data.agd$mean.std7.cov1,
               cov1_squared_centered = (cov1^2) - (data.agd$mean.std7.cov1^2 +
                                                     data.agd$sd.std7.cov1^2),

               cov2_centered = cov2- data.agd$mean.std7.cov2,
               cov2_squared_centered = (cov2^2) - (data.agd$mean.std7.cov2^2 +
                                                     data.agd$sd.std7.cov2^2))

      cent_match_cov <- c("cov1_centered",
                          "cov1_squared_centered",
                          "cov2_centered",
                          "cov2_squared_centered"
      )

#### estimating weights

est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                      matching_vars = cent_match_cov)

# Based on the known proportion of respondents, simulate the response data

 comparator_n <- 150 # total number of patients in the comparator data
 comparator_prop_events <- data.agd$mean.std7.trt # proportion of responders

 # Calculate number with event
 # Use round() to ensure we end up with a whole number of people

 n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
      comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                                    rep(0, comparator_n - n_with_event)))

 comparator_input <-  comparator_binary %>%
        mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

 combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
      combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

# Estimate weighted OR by fitting a logistic regression model with weights
      weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                          family = binomial(link="logit"),
                                          data = combined_data,
                                          weight = wt))

 maic6.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
  maic6.se<-coeftest(weighted_OR, vcov=vcovHC(weighted_OR,type="HC3"))[2,2]
```

```
return(c(maic6.est,maic6.se))
    }


 results.boot.std.1.7 <- boot(data=data.whole.maic, statistic=boot.std.1.7, R=1)


#####################################################################
 ################  MAIC 7 ###################################
################  study 1(ipd) study 8(agd)   ######################

 boot.std.1.8 <- function(d, i) {

 data.maic7 <- d[i,]

## The covariates need to be same name in ipd and agd study
        data.maic7<-rename(data.maic7, cov1 = std1.cov1,cov2 = std1.cov2)
     ### renaming variables in agd and
    ##
data.agd_update<-rename(data.agd, cov1 = mean.std8.cov1, cov2 = mean.std8.cov2)

 # List out matching covariates
 match_cov <- c("cov1",   "cov2")

#### center baseline characteristics
#matching continuous variables on both mean and standard deviation

data.ipd_update <- data.maic7 %>%
      mutate(cov1_centered = cov1- data.agd$mean.std8.cov1,
             cov1_squared_centered = (cov1^2) - (data.agd$mean.std8.cov1^2 +
                                                   data.agd$sd.std8.cov1^2),

             cov2_centered = cov2- data.agd$mean.std8.cov2,
             cov2_squared_centered = (cov2^2) - (data.agd$mean.std8.cov2^2 +
                                                   data.agd$sd.std8.cov2^2))

    cent_match_cov <- c("cov1_centered",
                        "cov1_squared_centered",
                        "cov2_centered",
                        "cov2_squared_centered"
    )

 #### estimating weights

est_weights <- estimate_weights(intervention_data = data.ipd_update ,
```

```
                                    matching_vars = cent_match_cov)

# Based on the known proportion of respondents, simulate the response data

comparator_n <- 150 # total number of patients in the comparator data
comparator_prop_events <- data.agd$mean.std8.trt # proportion of responders

 # Calculate number with event
    # Use round() to ensure we end up with a whole number of people

n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
    comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                                  rep(0, comparator_n - n_with_event)))

comparator_input <-  comparator_binary %>%
      mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

 combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
    combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

# Estimate weighted OR by fitting a logistic regression model with weights
    weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                        family = binomial(link="logit"),
                                        data = combined_data,
                                        weight = wt))
maic7.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
 maic7.se<-coeftest(weighted_OR, vcov=vcovHC(weighted_OR,type="HC3"))[2,2]

return(c(maic7.est,maic7.se))

      }

results.boot.std.1.8 <- boot(data=data.whole.maic, statistic=boot.std.1.8, R=1)


#####################################################################
 ################  MAIC 8 ###################################
################  study 1(ipd) study 9(agd)   ######################

 boot.std.1.9 <- function(d, i) {
 data.maic8<- d[i,]
 ## The covariates need to be same name in ipd and agd study
      data.maic8<-rename(data.maic8, cov1 = std1.cov1,cov2 = std1.cov2)

 ### Renaming variables in agd  study
```

```
    data.agd_update<-rename(data.agd, cov1 = mean.std9.cov1, cov2 = mean.std9.cov2)
 # List out matching covariates
    match_cov <- c("cov1",  "cov2")

#### center baseline characteristics
#matching continuous variables on both mean and standard deviation

data.ipd_update <- data.maic8%>%
      mutate(cov1_centered = cov1- data.agd$mean.std9.cov1,
             cov1_squared_centered = (cov1^2) - (data.agd$mean.std9.cov1^2 +
                                                   data.agd$sd.std9.cov1^2),

             cov2_centered = cov2- data.agd$mean.std9.cov2,
             cov2_squared_centered = (cov2^2) - (data.agd$mean.std9.cov2^2 +
                                                   data.agd$sd.std9.cov2^2))

    cent_match_cov <- c("cov1_centered",
                        "cov1_squared_centered",
                        "cov2_centered",
                        "cov2_squared_centered"
    )

#### estimating weights

 est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                 matching_vars = cent_match_cov)

# Based on the known proportion of respondents, simulate the response data

 comparator_n <- 150 # total number of patients in the comparator data
 comparator_prop_events <- data.agd$mean.std9.trt # proportion of responders

 # Calculate number with event
 # Use round() to ensure we end up with a whole number of people

n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                              rep(0, comparator_n - n_with_event)))

comparator_input <-  comparator_binary %>%
      mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
    combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")
```

```
# Estimate weighted OR by fitting a logistic regression model with weights
    weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                        family = binomial(link="logit"),
                                        data = combined_data,
                                        weight = wt))

 maic8.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
 maic8.se<-coeftest(weighted_OR, vcov=vcovHC(weighted_OR,type="HC3"))[2,2]
return(c(maic8.est,maic8.se))
    }

 results.boot.std.1.9 <- boot(data=data.whole.maic, statistic=boot.std.1.9, R=1)


#####################################################################
 ################  MAIC 9 ###################################
 ################  study 1(ipd) study 10(agd)   ######################

 boot.std.1.10 <- function(d, i) {
     data.maic9<- d[i,]
  ## The covariates need to be same name in ipd and agd study
     data.maic9<-rename(data.maic9, cov1 = std1.cov1,cov2 = std1.cov2)

 ### Renaming variables in agd  study

 data.agd_update<-rename(data.agd, cov1 = mean.std10.cov1, cov2 = mean.std10.cov2)

 # List out matching covariates
    match_cov <- c("cov1",    "cov2")

 #### center baseline characteristics
 #matching continuous variables on both mean and standard deviation

    data.ipd_update <- data.maic9 %>%
      mutate(cov1_centered = cov1- data.agd$mean.std10.cov1,
             cov1_squared_centered = (cov1^2) - (data.agd$mean.std10.cov1^2 +
                                                   data.agd$sd.std10.cov1^2),

             cov2_centered = cov2- data.agd$mean.std10.cov2,
             cov2_squared_centered = (cov2^2) - (data.agd$mean.std10.cov2^2 +
                                                   data.agd$sd.std10.cov2^2))

    cent_match_cov <- c("cov1_centered",
                        "cov1_squared_centered",
                        "cov2_centered",
```

```
                        "cov2_squared_centered"

    )

 #### estimating weights

 est_weights <- estimate_weights(intervention_data = data.ipd_update ,
                                    matching_vars = cent_match_cov)

# Based on the known proportion of respondents, simulate the response data

 comparator_n <- 150 # total number of patients in the comparator data
 comparator_prop_events <- data.agd$mean.std10.trt # proportion of responders

 # Calculate number with event
 # Use round() to ensure we end up with a whole number of people

    n_with_event <- round(comparator_n*comparator_prop_events, digits = 0)
    comparator_binary <- data.frame("std1.ref"= c(rep(1, n_with_event),
                                                  rep(0, comparator_n - n_with_event)))
    comparator_input <-  comparator_binary %>%
      mutate(wt=1, wt_rs=1, ARM="Comparator") # All patients have weight= 1

 # Merge comparator and intervention data
    combined_data <- bind_rows(est_weights$analysis_data, comparator_input)
    combined_data$ARM <- relevel(as.factor(combined_data$ARM), ref="Intervention")

 # Estimate weighted OR by fitting a logistic regression model with weights
    weighted_OR <- suppressWarnings(glm(formula =std1.ref~ARM,
                                        family = binomial(link="logit"),
                                        data = combined_data,
                                        weight = wt))

 maic9.est<-summary(weighted_OR)$coefficients["ARMComparator", "Estimate"]
maic9.se<-coeftest(weighted_OR, vcov=vcovHC(weighted_OR,type="HC3"))[2,2]
return(c(maic9.est,maic9.se))
    }

results.boot.std.1.10 <- boot(data=data.whole.maic, statistic=boot.std.1.10, R=1)

 ####conducting an NMA with the estimates from the MAIC

### creating a data frame which will be used in the nma

    studyn <- c(1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8,9,9)
```

```
    trtn <- c(1,2,1,2,1,2,1,2,1,2,1,3,1,3,1,3,1,3)

    n <- rep(150,18)

    diff<-c(NA,results.boot.std.1.2$t[1],
            NA,results.boot.std.1.3$t[1],NA,results.boot.std.1.4$t[1],
            NA,results.boot.std.1.5$t[1],NA,results.boot.std.1.6$t[1],
            NA,results.boot.std.1.7$t[1],NA,results.boot.std.1.8$t[1],
            NA,results.boot.std.1.9$t[1],NA,results.boot.std.1.10$t[1])

    se_diff<-c(NA,results.boot.std.1.2$t[2],
               NA,results.boot.std.1.3$t[2],NA,results.boot.std.1.4$t[2],
               NA,results.boot.std.1.5$t[2],NA,results.boot.std.1.6$t[2],
               NA,results.boot.std.1.7$t[2],NA,results.boot.std.1.8$t[2],
               NA,results.boot.std.1.9$t[2],NA,results.boot.std.1.10$t[2])

datacel2 <- data.frame(cbind(studyn,trtn,diff,se_diff,n))

### following code will set up the network (arm based)

    model.re<-set_agd_contrast(
      datacel2,
      study=studyn,
      trt=trtn,
      y=diff,
      se=se_diff,
      sample_size=n)

    ##The model is fitted using the nma() function.

    maic_fit_RE <- nma(model.re,
                       trt_effects = "random",
                       prior_intercept = normal(scale = 100),
                       prior_trt = normal(scale = 10),
                       prior_het = log_normal(-2.56, 0.33),
                      prior_het_type = "var")


### extracting the values of d1 and d2 and then calculating the mean

maic_data_frame<-as.data.frame(maic_fit_RE,pars=c("d"))

stan.object.nma.discnt.mean.d1<-mean(maic_data_frame$`d[2]`)
stan.object.nma.discnt.mean.d2<-mean(maic_data_frame$`d[3]`)
```

```
return(c(stan.object.nma.discnt.mean.d1,
             stan.object.nma.discnt.mean.d2))
  }

  results <- boot(data=data.ipd, statistic=boot.maic, R=1000)

  mean.d1<-mean(results$t[, 1])  ### mean of first column will give estimate of d1

  mean.d2<-mean(results$t[, 2])  ### ### mean of second column will give estimate of d2

  sd.d1<-sd(results$t[, 1])    ###sd of first column will give sd of d1

  sd.d2<-sd(results$t[, 2])   ###sd of second column will give sd of d2

  #### calculation of coverage probability

  B1=sd.d1
  A1=mean.d1
  z.alpha1 <- 1.96
  theta.hat.low1=A1-z.alpha1*B1
  theta.hat.upp1=A1+z.alpha1*B1
  theta1=stan.object.nma.true.mean.d1
  est1 <-ifelse(theta1>=theta.hat.low1 & theta1<=theta.hat.upp1,1,0)

  B2=sd.d2
  A2=mean.d2
  z.alpha2 <- 1.96
  theta.hat.low2=A2-z.alpha2*B2
  theta.hat.upp2=A2+z.alpha2*B2
  theta2= stan.object.nma.true.mean.d2
  est2 <-ifelse(theta2>=theta.hat.low2 & theta2<=theta.hat.upp2,1,0)

  out[i, 1]<-mean.d1
  out[i, 2]<-mean.d2
  out[i, 3]<-sd.d1
  out[i, 4]<-sd.d2
  out[i, 5]<-est1
  out[i, 6]<-est2
out[i, 7]<-stan.object.nma.connect.mean.d1
out[i, 8]<-stan.object.nma.connect.mean.d2
out[i, 9]<-stan.object.nma.connect.sd.d1
out[i, 10]<-stan.object.nma.connect.sd.d2

  out<-data.frame(out)
```

```
}

### run in parallel..........

library(parallel)
library(lme4)

save <- mclapply(ipd, bootstrap)
saveRDS(save , file="inovatn.0.3.RE.scn2.1.rds")
```

Appendix F

Graphs and R codes for asthma data

F.1 Graphs of weight distribution for multiple MAICs

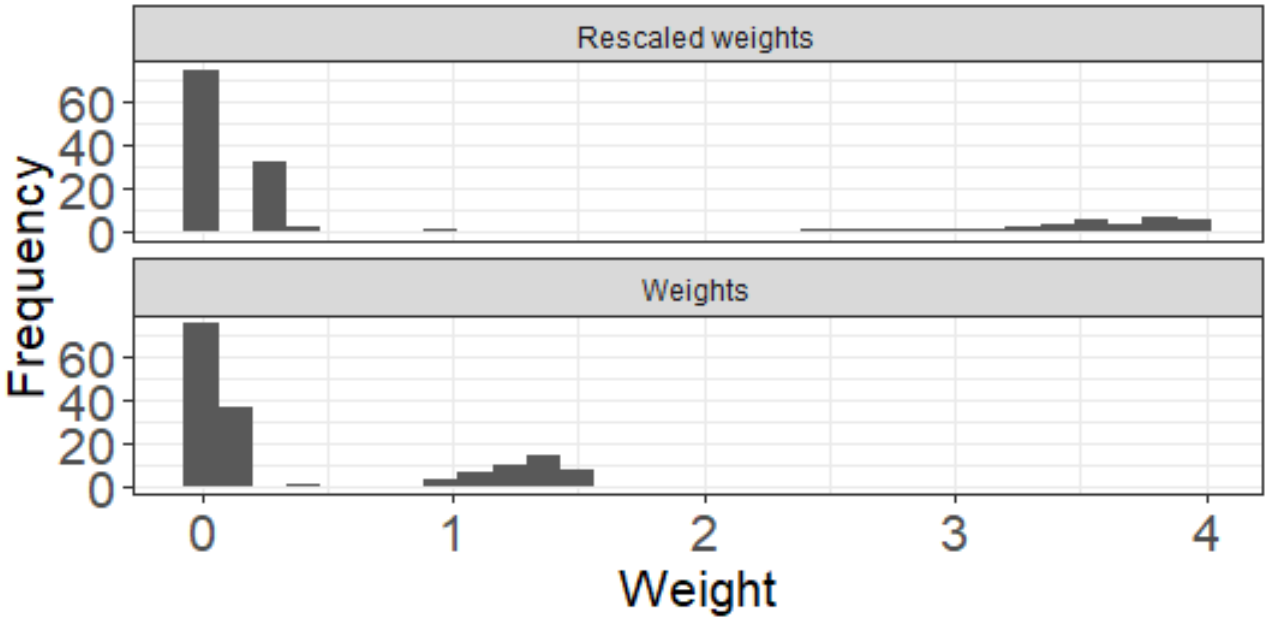


Figure F.1: *Weight distribution for MAIC 1*

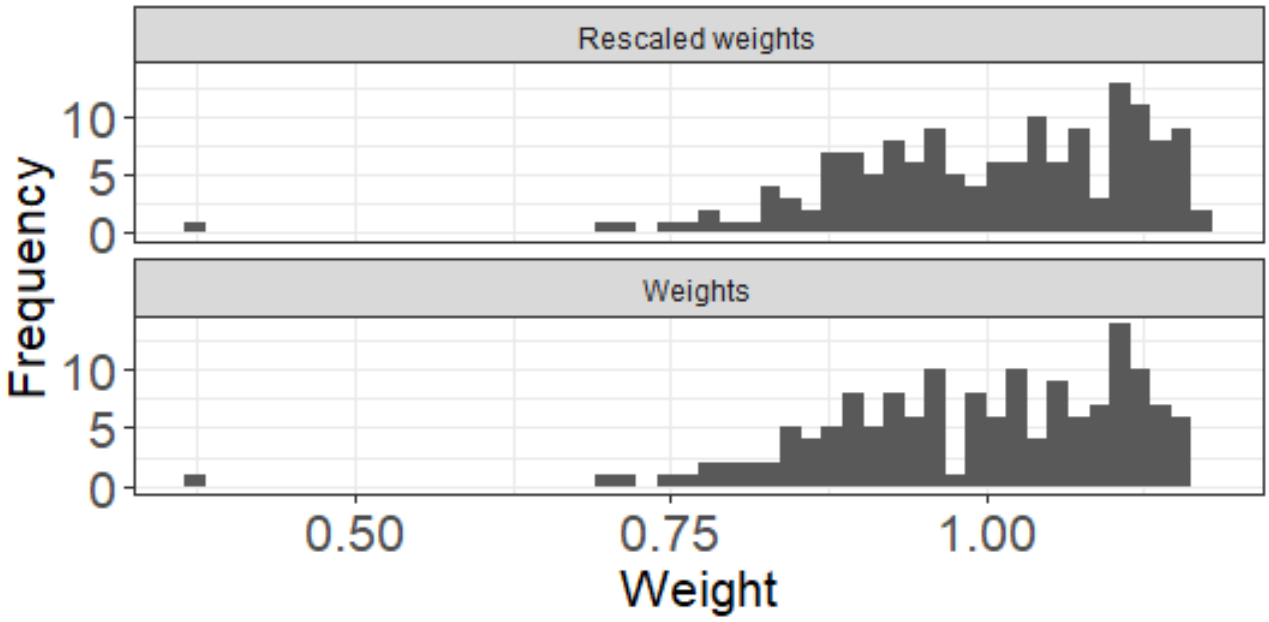


Figure F.2: *Weight distribution for MAIC 2*

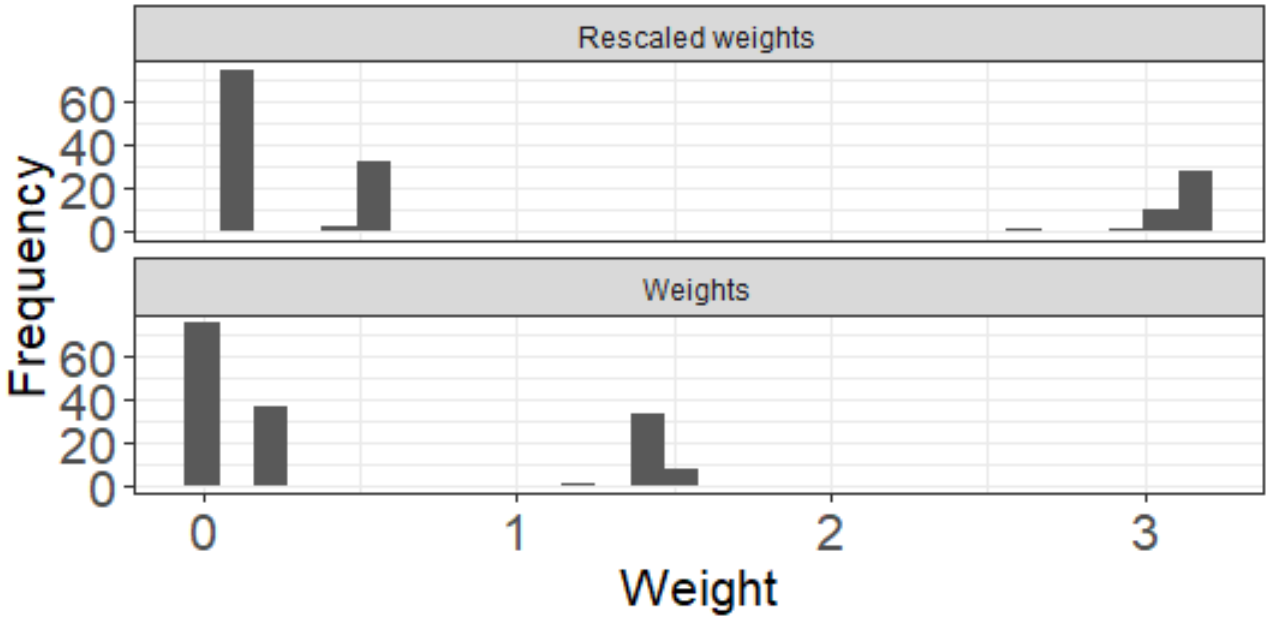


Figure F.3: *Weight distribution for MAIC 3*

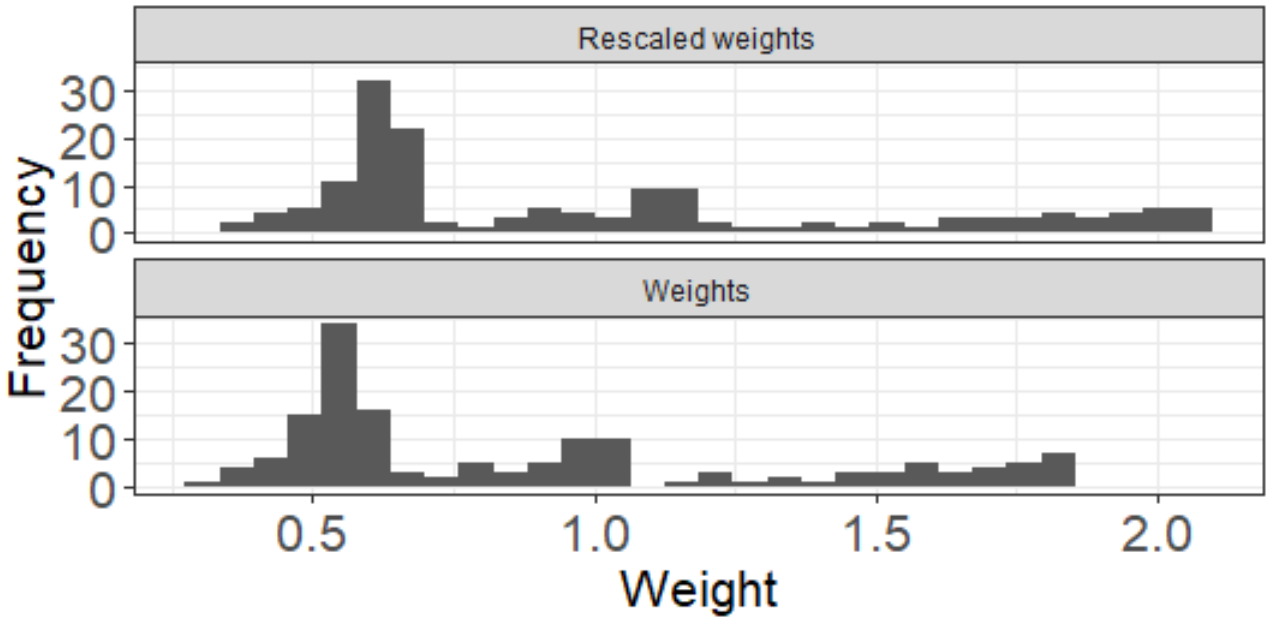


Figure F.4: *Weight distribution for MAIC 4*

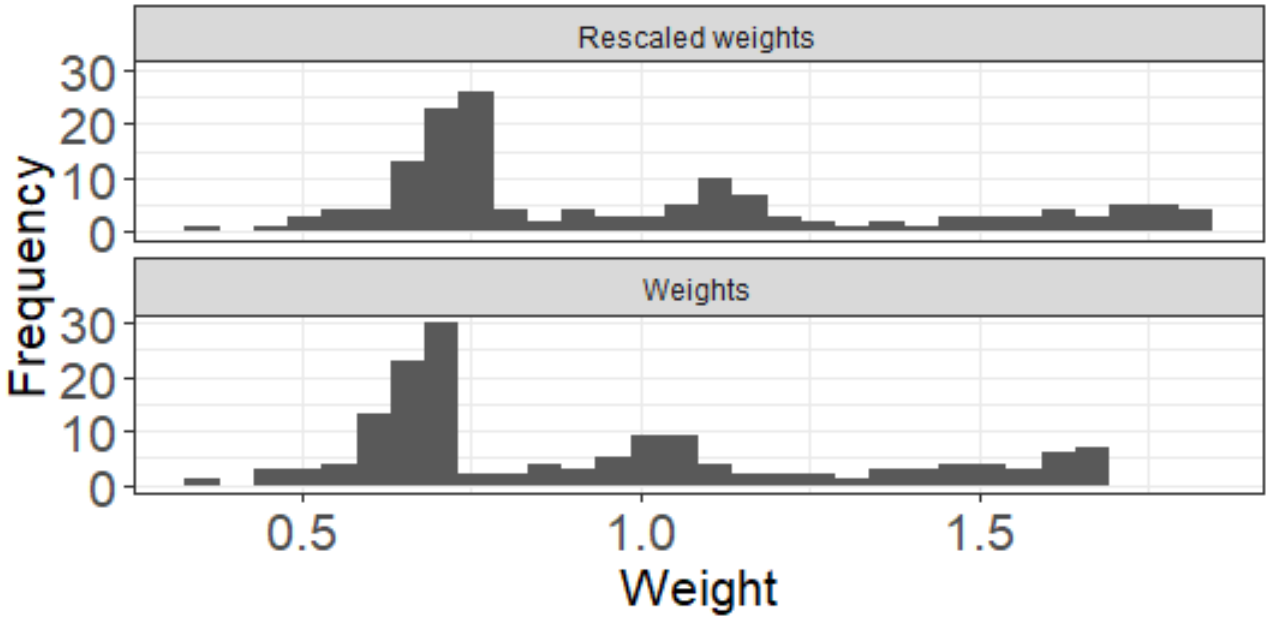


Figure F.5: *Weight distribution for MAIC 5*

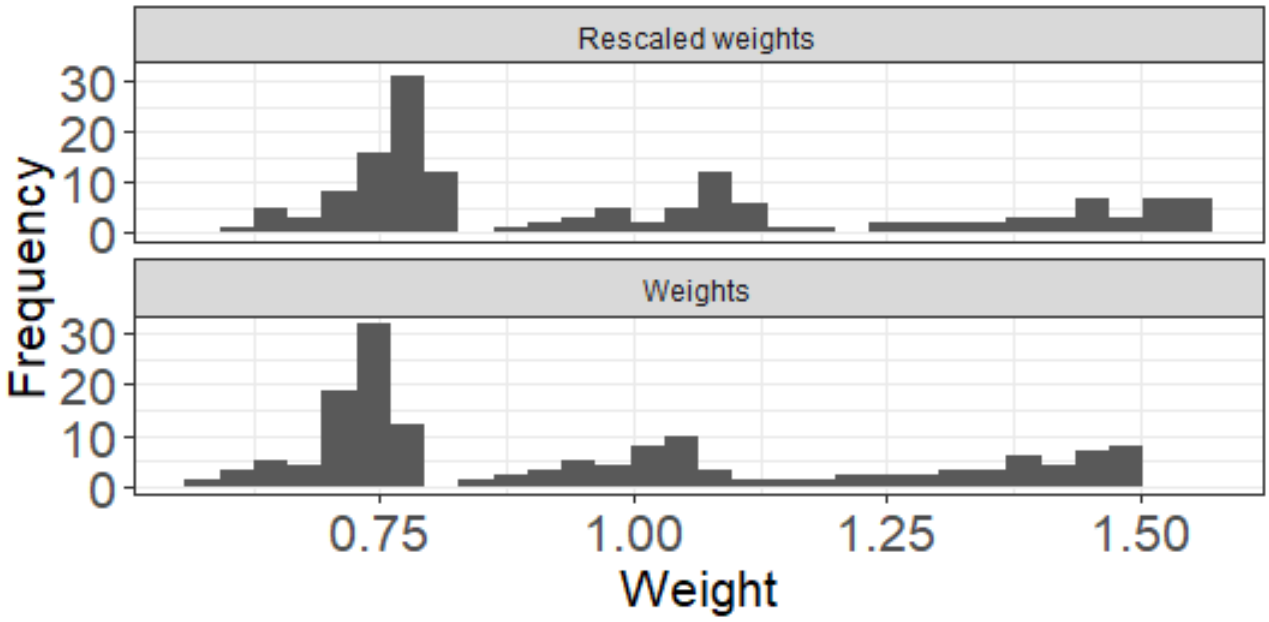


Figure F.6: *Weight distribution for MAIC 6*

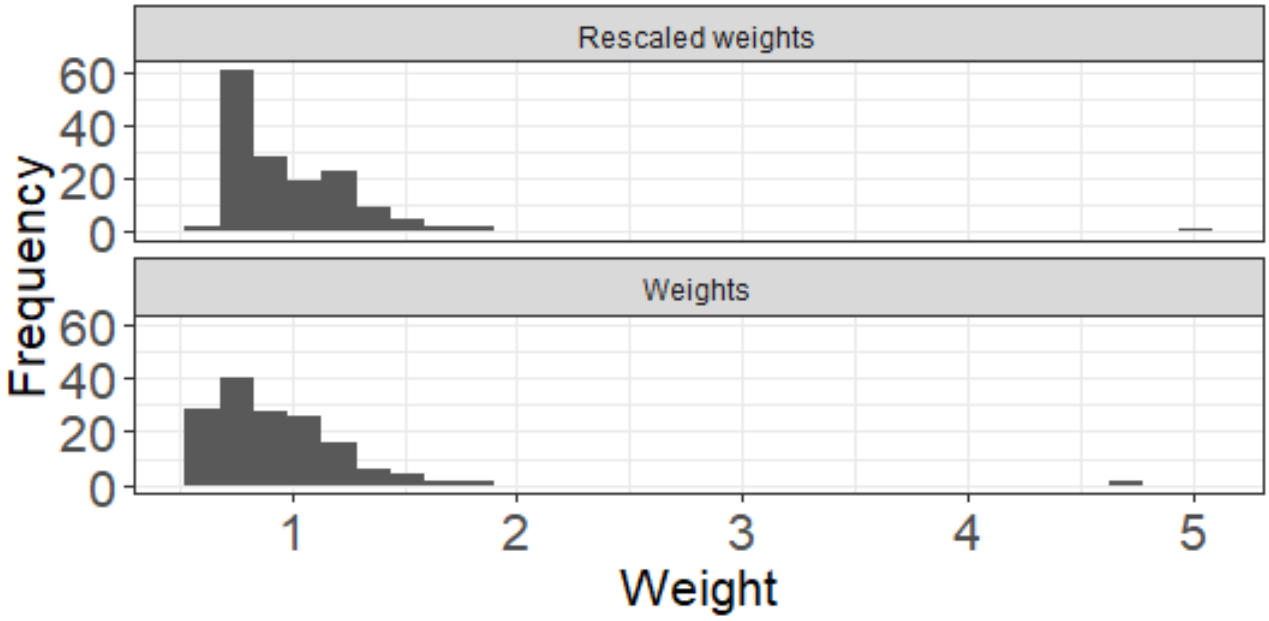


Figure F.7: *Weight distribution for MAIC 7*

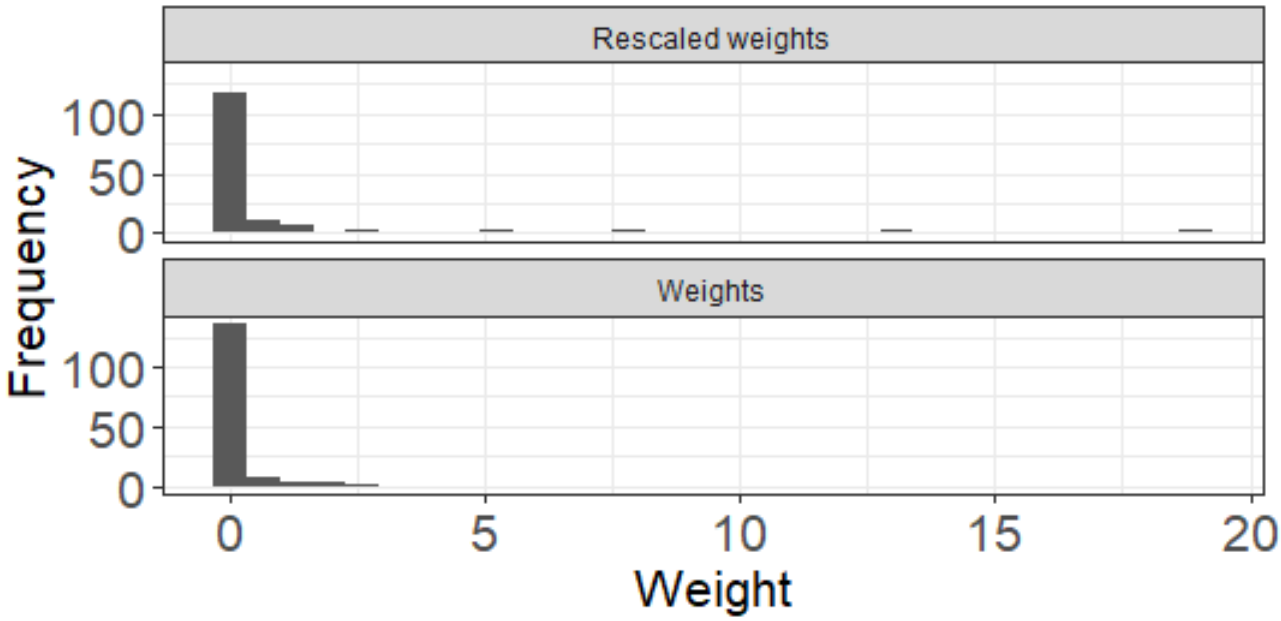


Figure F.8: *Weight distribution for MAIC 8*

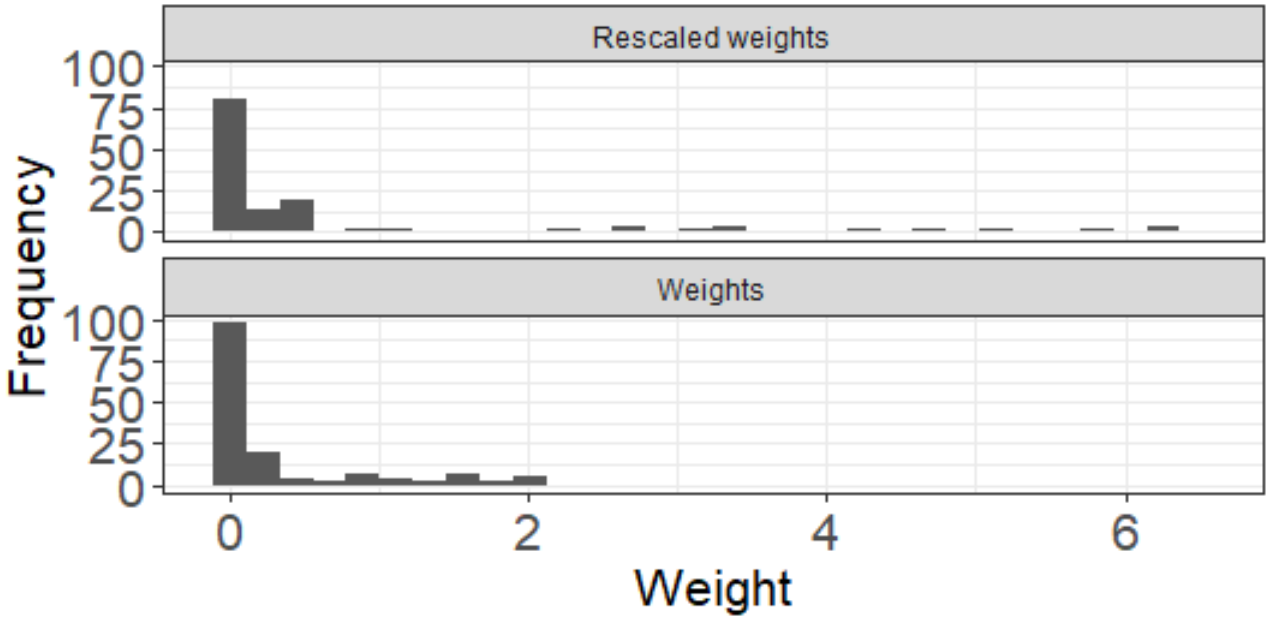


Figure F.9: *Weight distribution for MAIC 9*

F.2 MAIC-adjusted NMA for asthma data

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
library(haven)
library(dplyr)
library(devtools)
library(MAIC)
library(tidyverse)
library(multinma)
library(gtsummary)
```

```
data<-read_sas("V:\\112997\\Dataset\\AR_UNKNOWN_-99\\gsk_112997_exacanal.sas7bdat")
```

```
##subsetting with only two groups
data2<-data %>% filter (TRTGRP=="Placebo"| TRTGRP=="Mepolizumab 250mg" )
```

```
 ##reading eisonifil data
bleos<-read_sas("V:\\112997\\Dataset\\AR_UNKNOWN_-99\\gsk_112997_bleos.sas7bdat")
 ## subsetting for two grp
bleos<-bleos %>% filter (TRTGRP=="Placebo"| TRTGRP=="Mepolizumab 250mg" )
 ##taking only baseline eisonofil
bleos<-bleos %>% filter (VISIT=="Baseline")
## discarding percentage value
bleos<-bleos %>% filter (LBTEST=="Eosinophils")
```

```
exacag<-read_sas("V:\\112997\\Dataset\\AR_UNKNOWN_-99\\gsk_112997_exacag.sas7bdat")
```

```
## subsetting for two group
exacag<-exacag %>% filter (TRTGRP=="Placebo"| TRTGRP=="Mepolizumab 250mg" )
exacag2<-exacag %>% group_by(SUBJID)  %>% mutate(max_value=max(STATUS))
exacag2<-exacag2 %>% distinct (SUBJID, .keep_all = T)
```

```
################## merging the two datasets

join_with_bleos<-left_join(data2,bleos, by="SUBJID")

join_with_bleos<-left_join(join_with_bleos,exacag2, by="SUBJID")
```

```
### Converting eisonofil column into another unit
```

```
join_with_bleos<-join_with_bleos %>% mutate(eisonofil=LBSTRESN*1000)

## Taking only mepolizumab arm
join_with_bleos<-join_with_bleos %>% filter(TRTGRP=="Mepolizumab 250mg")

#######################################

####################### MAIC 1 with matching only mean

agd<-read.csv("V:\\112997\\Dataset\\test.csv")
agd.update<-rename(agd, cov1=PRENEXCD.std1, cov2=eisonofil.std1)
join_with_bleos <- rename( join_with_bleos, cov1=PRENEXCD, cov2=eisonofil)

match_cov<-c("cov1", "cov2")

data.maic<- join_with_bleos %>%
mutate(cov1_cen=cov1-agd$PRENEXCD.std1,cov2_cen=cov2-agd$eisonofil.std1,)

cent_match_cov<-c("cov1_cen","cov2_cen")
est_weights<-estimate_weights(intervention_data = data.maic,
                              matching_vars =cent_match_cov )


## chacking if opotimization working

baseline_summary<-list('intervention'=NA,
                       'intervention_weighted'=NA,
                       'comparator'=NA)

baseline_summary$intervention_weighted<-est_weights$analysis_data%>%
    transmute(cov1,cov2,wt) %>% summarise_at(match_cov,list(~weighted.mean(.,wt)))

baseline_summary$intervention<-est_weights$analysis_data%>%
  transmute(cov1,cov2,wt) %>% summarise_at(match_cov,list(~mean(.)))

baseline_summary$comparator <- transmute(agd.update, cov1, cov2)

trt<-names(baseline_summary)
baseline_summary<-bind_rows(baseline_summary)
%>% transmute_all(sprintf, fmt="%.2f")
%>% transmute(ARM=as.character(trt),cov1,cov2)

baseline_summary
analysis_data<-data.frame(est_weights[[2]])
```

```
## attaching the weight column to ipd dataset
join_with_bleos<-cbind(join_with_bleos,analysis_data$wt)
join_with_bleos<-rename(join_with_bleos, wt=‘analysis_data$wt‘)

## adding a new column for exposure time
join_with_bleos<-join_with_bleos %>%
add_column (exposure.time=join_with_bleos$TOTDAYS/365.25)

## making a new variable called weighted exposure time
join_with_bleos<-join_with_bleos %>% mutate(wetd.exposr.time=wt*exposure.time)

print(sum(join_with_bleos$wetd.exposr.time))

## making a new variable called weighted event
join_with_bleos<-join_with_bleos %>% mutate(wetd.event=wt*max_value)

print(sum(join_with_bleos$wetd.event))

sum(join_with_bleos$wt)


######################## MAIC 2 with matching only mean

data.maic<- join_with_bleos %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std2,


                                       cov2_cen=cov2-agd$eisonofil.std2,
)

agd.update<-rename(agd, cov1=PRENEXCD.std2, cov2=eisonofil.std2)

cent_match_cov<-c("cov1_cen","cov2_cen")
est_weights<-estimate_weights(intervention_data = data.maic,
                              matching_vars =cent_match_cov )


## chacking if potimization working

baseline_summary<-list('intervention'=NA,
                       'intervention_weighted'=NA,
                       'comparator'=NA)


baseline_summary$intervention_weighted<-est_weights$analysis_data%>%
  transmute(cov1,cov2,wt) %>% summarise_at(match_cov,list(~weighted.mean(.,wt)))
```

```
baseline_summary$intervention<-est_weights$analysis_data%>%
  transmute(cov1,cov2,wt) %>% summarise_at(match_cov,list(~mean(.)))

baseline_summary$comparator <- transmute(agd.update, cov1, cov2)

trt<-names(baseline_summary)
baseline_summary<-bind_rows(baseline_summary)
%>% transmute_all(sprintf, fmt="%.2f")
%>% transmute(ARM=as.character(trt),cov1,cov2)

baseline_summary

analysis_data<-data.frame(est_weights[[2]])

## attaching the weight column to ipd dataset
join_with_bleos<-cbind(join_with_bleos,analysis_data$wt)

join_with_bleos<-rename(join_with_bleos, wt=`analysis_data$wt`)

## adding a new column for exposure time
join_with_bleos<-join_with_bleos
%>% add_column (exposure.time=join_with_bleos$TOTDAYS/365.25)

## making a new variable called weighted exposure time
join_with_bleos<-join_with_bleos %>% mutate(wetd.exposr.time=wt*exposure.time)


print(sum(join_with_bleos$wetd.exposr.time))

## making a new variable called weighted event
join_with_bleos<-join_with_bleos %>% mutate(wetd.event=wt*max_value)

print(sum(join_with_bleos$wetd.event))

sum(join_with_bleos$wt)



######################## MAIC 3 with matching only mean


data.maic<- join_with_bleos %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std3,
```

```
                                        cov2_cen=cov2-agd$eisonofil.std3,
)

agd.update<-rename(agd, cov1=PRENEXCD.std3, cov2=eisonofil.std3)

cent_match_cov<-c("cov1_cen","cov2_cen")
est_weights<-estimate_weights(intervention_data = data.maic,
                              matching_vars =cent_match_cov )


## chacking if potimization working

baseline_summary<-list('intervention'=NA,
                       'intervention_weighted'=NA,
                       'comparator'=NA)


baseline_summary$intervention_weighted<-est_weights$analysis_data%>%
  transmute(cov1,cov2,wt) %>% summarise_at(match_cov,list(~weighted.mean(.,wt)))


baseline_summary$intervention<-est_weights$analysis_data%>%
  transmute(cov1,cov2,wt) %>% summarise_at(match_cov,list(~mean(.)))

baseline_summary$comparator <- transmute(agd.update, cov1, cov2)


trt<-names(baseline_summary)
baseline_summary<-bind_rows(baseline_summary)
%>% transmute_all(sprintf, fmt="%.2f")
%>% transmute(ARM=as.character(trt),cov1,cov2)

baseline_summary

analysis_data<-data.frame(est_weights[[2]])

## attaching the weight column to ipd dataset
join_with_bleos<-cbind(join_with_bleos,analysis_data$wt)

join_with_bleos<-rename(join_with_bleos, wt=`analysis_data$wt`)

## adding a new column for exposure time
join_with_bleos<-join_with_bleos
%>% add_column (exposure.time=join_with_bleos$TOTDAYS/365.25)
```

```
## making a new variable called weighted exposure time
join_with_bleos<-join_with_bleos %>% mutate(wetd.exposr.time=wt*exposure.time)

print(sum(join_with_bleos$wetd.exposr.time))

## making a new variable called weighted event
join_with_bleos<-join_with_bleos %>% mutate(wetd.event=wt*max_value)

print(sum(join_with_bleos$wetd.event))

sum(join_with_bleos$wt)


######################## MAIC 4 with matching only mean


data.maic<- join_with_bleos %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std4,


                                       cov2_cen=cov2-agd$eisonofil.std4,
)

agd.update<-rename(agd, cov1=PRENEXCD.std4, cov2=eisonofil.std4)

cent_match_cov<-c("cov1_cen","cov2_cen")
est_weights<-estimate_weights(intervention_data = data.maic,
                              matching_vars =cent_match_cov )


## chacking if potimization working

baseline_summary<-list('intervention'=NA,
                       'intervention_weighted'=NA,
                       'comparator'=NA)


baseline_summary$intervention_weighted<-est_weights$analysis_data%>%
  transmute(cov1,cov2,wt) %>% summarise_at(match_cov,list(~weighted.mean(.,wt)))


baseline_summary$intervention<-est_weights$analysis_data%>%
  transmute(cov1,cov2,wt) %>% summarise_at(match_cov,list(~mean(.)))

baseline_summary$comparator <- transmute(agd.update, cov1, cov2)
```

```
trt<-names(baseline_summary)
baseline_summary<-bind_rows(baseline_summary)
%>% transmute_all(sprintf, fmt="%.2f")
%>% transmute(ARM=as.character(trt),cov1,cov2)

baseline_summary

analysis_data<-data.frame(est_weights[[2]])

## attaching the weight column to ipd dataset
join_with_bleos<-cbind(join_with_bleos,analysis_data$wt)

join_with_bleos<-rename(join_with_bleos, wt=`analysis_data$wt`)

## adding a new column for exposure time
join_with_bleos<-join_with_bleos %>% mutate (exposure.time=TOTDAYS/365.25)

## making a new variable called weighted exposure time
join_with_bleos<-join_with_bleos %>% mutate(wetd.exposr.time=wt*exposure.time)

print(sum(join_with_bleos$wetd.exposr.time))     ##weighted person year

## making a new variable called weighted event
join_with_bleos<-join_with_bleos %>% mutate(wetd.event=wt*max_value)

print(sum(join_with_bleos$wetd.event))

sum(join_with_bleos$wt)


######################## MAIC 5 with matching only mean


data.maic<- join_with_bleos %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std5,


                                       cov2_cen=cov2-agd$eisonofil.std5,
)

agd.update<-rename(agd, cov1=PRENEXCD.std5, cov2=eisonofil.std5)

cent_match_cov<-c("cov1_cen","cov2_cen")
est_weights<-estimate_weights(intervention_data = data.maic,
```

```
                              matching_vars =cent_match_cov )


## chacking if potimization working

baseline_summary<-list('intervention'=NA,
                       'intervention_weighted'=NA,
                       'comparator'=NA)


baseline_summary$intervention_weighted<-est_weights$analysis_data%>%
  transmute(cov1,cov2,wt) %>% summarise_at(match_cov,list(~weighted.mean(.,wt)))


baseline_summary$intervention<-est_weights$analysis_data%>%
  transmute(cov1,cov2,wt) %>% summarise_at(match_cov,list(~mean(.)))

baseline_summary$comparator <- transmute(agd.update, cov1, cov2)


trt<-names(baseline_summary)
baseline_summary<-bind_rows(baseline_summary)
%>% transmute_all(sprintf, fmt="%.2f")
%>% transmute(ARM=as.character(trt),cov1,cov2)

baseline_summary


analysis_data<-data.frame(est_weights[[2]])

## attaching the weight column to ipd dataset
join_with_bleos<-cbind(join_with_bleos,analysis_data$wt)

join_with_bleos<-rename(join_with_bleos, wt=`analysis_data$wt`)

## adding a new column for exposure time
join_with_bleos<-join_with_bleos
%>% add_column (exposure.time=join_with_bleos$TOTDAYS/365.25)

## making a new variable called weighted exposure time
join_with_bleos<-join_with_bleos %>% mutate(wetd.exposr.time=wt*exposure.time)

print(sum(join_with_bleos$wetd.exposr.time))

## making a new variable called weighted event
```

```
join_with_bleos<-join_with_bleos %>% mutate(wetd.event=wt*max_value)

print(sum(join_with_bleos$wetd.event))

sum(join_with_bleos$wt)

######################## MAIC 6 with matching only mean

data.maic<- join_with_bleos %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std6,


                                       cov2_cen=cov2-agd$eisonofil.std6,
)

agd.update<-rename(agd, cov1=PRENEXCD.std6, cov2=eisonofil.std6)

cent_match_cov<-c("cov1_cen","cov2_cen")
est_weights<-estimate_weights(intervention_data = data.maic,
                              matching_vars =cent_match_cov )


## chacking if potimization working

baseline_summary<-list('intervention'=NA,
                       'intervention_weighted'=NA,
                       'comparator'=NA)


baseline_summary$intervention_weighted<-est_weights$analysis_data%>%
  transmute(cov1,cov2,wt) %>% summarise_at(match_cov,list(~weighted.mean(.,wt)))


baseline_summary$intervention<-est_weights$analysis_data%>%
  transmute(cov1,cov2,wt) %>% summarise_at(match_cov,list(~mean(.)))

baseline_summary$comparator <- transmute(agd.update, cov1, cov2)
trt<-names(baseline_summary)
baseline_summary<-bind_rows(baseline_summary)
%>% transmute_all(sprintf, fmt="%.2f")
%>% transmute(ARM=as.character(trt),cov1,cov2)

baseline_summary
analysis_data<-data.frame(est_weights[[2]])

## attaching the weight column to ipd dataset
```

```
join_with_bleos<-cbind(join_with_bleos,analysis_data$wt)

join_with_bleos<-rename(join_with_bleos, wt=‘analysis_data$wt‘)

## adding a new column for exposure time
join_with_bleos<-join_with_bleos %>% add_column (exposure.time=join_with_bleos$TOTDAYS/365.2

## making a new variable called weighted exposure time
join_with_bleos<-join_with_bleos %>% mutate(wetd.exposr.time=wt*exposure.time)

print(sum(join_with_bleos$wetd.exposr.time))

## making a new variable called weighted event
join_with_bleos<-join_with_bleos %>% mutate(wetd.event=wt*max_value)

print(sum(join_with_bleos$wetd.event))
sum(join_with_bleos$wt)

######################## MAIC 7 with matching only mean

data.maic<- join_with_bleos %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std7,


                                       cov2_cen=cov2-agd$eisonofil.std7,
)

agd.update<-rename(agd, cov1=PRENEXCD.std7, cov2=eisonofil.std7)

cent_match_cov<-c("cov1_cen","cov2_cen")
est_weights<-estimate_weights(intervention_data = data.maic,
                              matching_vars =cent_match_cov )


## chacking if potimization working

baseline_summary<-list('intervention'=NA,
                       'intervention_weighted'=NA,
                       'comparator'=NA)


baseline_summary$intervention_weighted<-est_weights$analysis_data%>%
  transmute(cov1,cov2,wt) %>% summarise_at(match_cov,list(~weighted.mean(.,wt)))


baseline_summary$intervention<-est_weights$analysis_data%>%
```

```
  transmute(cov1,cov2,wt) %>% summarise_at(match_cov,list(~mean(.)))

baseline_summary$comparator <- transmute(agd.update, cov1, cov2)

trt<-names(baseline_summary)
baseline_summary<-bind_rows(baseline_summary)
%>% transmute_all(sprintf, fmt="%.2f")
%>% transmute(ARM=as.character(trt),cov1,cov2)

baseline_summary
analysis_data<-data.frame(est_weights[[2]])

## attaching the weight column to ipd dataset
join_with_bleos<-cbind(join_with_bleos,analysis_data$wt)

join_with_bleos<-rename(join_with_bleos, wt=`analysis_data$wt`)

## adding a new column  exposure time
join_with_bleos<-join_with_bleos
%>% add_column (exposure.time=join_with_bleos$TOTDAYS/365.25)


## making a new variable called weighted exposure time
join_with_bleos<-join_with_bleos %>% mutate(wetd.exposr.time=wt*exposure.time)

print(sum(join_with_bleos$wetd.exposr.time))     ##weighted person year

## making a new variable called weighted event
join_with_bleos<-join_with_bleos %>% mutate(wetd.event=wt*max_value)

print(sum(join_with_bleos$wetd.event))
sum(join_with_bleos$wt)

######################## MAIC 8 with matching only mean


data.maic<- join_with_bleos %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std8,


                                       cov2_cen=cov2-agd$eisonofil.std8,
)

agd.update<-rename(agd, cov1=PRENEXCD.std8, cov2=eisonofil.std8)

cent_match_cov<-c("cov1_cen","cov2_cen")
```

```
est_weights<-estimate_weights(intervention_data = data.maic,
                              matching_vars =cent_match_cov )


## chacking if opotimization working

baseline_summary<-list('intervention'=NA,
                       'intervention_weighted'=NA,
                       'comparator'=NA)


baseline_summary$intervention_weighted<-est_weights$analysis_data%>%
  transmute(cov1,cov2,wt) %>% summarise_at(match_cov,list(~weighted.mean(.,wt)))


baseline_summary$intervention<-est_weights$analysis_data%>%
  transmute(cov1,cov2,wt) %>% summarise_at(match_cov,list(~mean(.)))

baseline_summary$comparator <- transmute(agd.update, cov1, cov2)


trt<-names(baseline_summary)
baseline_summary<-bind_rows(baseline_summary)
%>% transmute_all(sprintf, fmt="%.2f")
%>% transmute(ARM=as.character(trt),cov1,cov2)

baseline_summary
analysis_data<-data.frame(est_weights[[2]])

## attaching the weight column to ipd dataset
join_with_bleos<-cbind(join_with_bleos,analysis_data$wt)

join_with_bleos<-rename(join_with_bleos, wt=`analysis_data$wt`)

## adding a new column exposure time
join_with_bleos<-join_with_bleos
%>% add_column (exposure.time=join_with_bleos$TOTDAYS/365.25)

## making a new variable called weighted exposure time
join_with_bleos<-join_with_bleos %>% mutate(wetd.exposr.time=wt*exposure.time)

print(sum(join_with_bleos$wetd.exposr.time))     ##weighted person year

## making a new variable called weighted event
join_with_bleos<-join_with_bleos %>% mutate(wetd.event=wt*max_value)
```

```
print(sum(join_with_bleos$wetd.event))
sum(join_with_bleos$wt)

############################## NMA Fixed effect

asthma<-read.csv("V:\\112997\\Dataset\\2nd version\\agd_data_disconnected.csv")

asthma_net<-set_agd_arm(asthma ,
                        study=study.no ,
                        trt= trt.no,
                        r=r,
                        E=E,
                        sample_size=n,
                        trt_ref = "2")

asthma_net
plot(asthma_net)

asthma_net_FE<-nma(asthma_net,
                   trt_effects = "fixed",
                   prior_intercept = normal(scale=100),
                   prior_trt = normal(scale=100))

asthma_net_FE
rank<-posterior_ranks(asthma_net_FE)
plot(rank)

rank2<-posterior_rank_probs(asthma_net_FE)
plot(rank2)

############################## NMA random effect

asthma_net_RE<-nma(asthma_net,
                   trt_effects = "random",
                   prior_intercept = normal(scale=100),
                   prior_trt = normal(scale=100),
                   prior_het = half_normal(scale=0.15))

asthma_net_RE
rank3<-posterior_ranks(asthma_net_RE)
plot(rank3)
rank4<-posterior_rank_probs(asthma_net_RE)
rank4<-plot(rank4)
```

F.3 Double-bootstrap with asthma data (fixed effect model)

```
library(haven)
library(boot)
library(dplyr)
library(devtools)
library(MAIC)
library(tidyverse)
library(multinma)


data<-read_sas("V:\\112997\\Dataset\\AR_UNKNOWN_-99\\gsk_112997_exacanal.sas7bdat")
##subsetting with only two groups
data2<-data %>% filter (TRTGRP=="Placebo"| TRTGRP=="Mepolizumab 250mg" )

data2<-select(data2,-c("STUDYID" ,"INVID", "CENTREID", "USUBJID" , "RACECD", "RACE",
"TRTCD", "ATRTCD", "ATRTGRP", "LTOTAL" , "CS_TIME", "HED_NUM",  "HED_RATE", "HED_TIME" ,
"ALL_NUM","ALL_RATE", "ALL_TIME", "HOS_NUM", "HOS_RATE" ,"HOS_TIME" ,"STRATCD" ,"REGIONCD",

##reading eisonifil data
bleos<-read_sas("V:\\112997\\Dataset\\AR_UNKNOWN_-99\\gsk_112997_bleos.sas7bdat")
## subsetting for two grp
bleos<-bleos %>% filter (TRTGRP=="Placebo"| TRTGRP=="Mepolizumab 250mg" )
 ##taking only baseline eisonofil
bleos<-bleos %>% filter (VISIT=="Baseline")            ## discarding percentage value
bleos<-bleos %>% filter (LBTEST=="Eosinophils")

## deleting extra columns

bleos<-select(bleos,-c("CENTREID", "USUBJID" ,"AGE",  "SEX", "RACECD", "RACE",
"TRTCD","TRTGRP", "ATRTCD",   "ATRTGRP", "VISITNUM", "AVISNUM", "AVISIT", "PTMNUM",
"PTM", "LBACTDY","LBDT","LBACTTM", "LBTESTCD","LBTEST", "LBSTUNIT", "LBSTNRLO", "LBSTNRHI",
"LBORRES", "LBORRESN", "LBORUNIT", "LBORNRLO", "LBORNRHI", "LBNRCD", "LBNRIND", "LBSTCCLO",
"LBSTCCHI", "LBCCCD", "LBCCIND", "LBSTDBL", "STDCHGBL", "ILBRES", "ILBBL", "ILBCHBL",
"_ILBRES", "_ILBBL", "_ILBCHBL", "LBIDCD","LBAGE" ,"LBID" , "LBCAT" , "LBACSNUM", "LBFAST",
"LBTSTCOM", "LBUDIFCD", "EVALFLAG","STUDYID",  "INVID" ))

### reading data to create binary variable

exacag<-read_sas("V:\\112997\\Dataset\\AR_UNKNOWN_-99\\gsk_112997_exacag.sas7bdat")
## subsetting for two grp
exacag<-exacag %>% filter (TRTGRP=="Placebo"| TRTGRP=="Mepolizumab 250mg" )
```

```
exacag2<-exacag %>% group_by(SUBJID)  %>% mutate(max_value=max(STATUS))

exacag2<-exacag2 %>% distinct (SUBJID, .keep_all = T)

exacag2<-select(exacag2,-c("STUDYID", "INVID","CENTREID", "USUBJID","AGE",
"SEX", "RACECD", "RACE", "TRTCD", "TRTGRP" ,   "ATRTCD", "ATRTGRP", "ESTART", "ESTOP",
"STATUS", "STRATCD","STRATUM", "REGIONCD", "REGION" ,"PRENEXCD", "PRENEX", "BPRBDVAL"
))

################## merging the two datasets

join_with_bleos<-left_join(data2,bleos, by="SUBJID")
join_with_bleos<-left_join(join_with_bleos,exacag2, by="SUBJID")

### converting to eisonofil column into another unit

join_with_bleos<-join_with_bleos %>% mutate(eisonofil=LBSTRESN*1000)

## taking only mepolizumab arm
join_with_bleos<-join_with_bleos %>% filter(TRTGRP=="Mepolizumab 250mg")


######################## MAIC 1 with matching only mean

agd<-read.csv("V:\\112997\\Dataset\\2nd version\\agd_data_maic.csv")

boot.maic<-function(d,i){

  data.whole.maic<-d[i,]

  ####maic 1

  boot.std.1.2<-function(d,i){
    data.maic1<- d[i,]

    agd.update<-rename(agd, cov1=PRENEXCD.std1, cov2=eisonofil.std1)

    data.maic1 <- rename( data.maic1, cov1=PRENEXCD, cov2=eisonofil)

    match_cov<-c("cov1", "cov2")

  data.maic1<- data.maic1 %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std1,
  cov2_cen=cov2-agd$eisonofil.std1)
```

```
   cent_match_cov<-c("cov1_cen","cov2_cen")
   est_weights<-estimate_weights(intervention_data = data.maic1,
                                 matching_vars =cent_match_cov )

   analysis_data<-data.frame(est_weights[[2]])

   ## attaching the weight column to ipd dataset
   data.maic1<-cbind(data.maic1,analysis_data$wt)
   data.maic1<-rename(data.maic1, wt=`analysis_data$wt`)

   ## adding a new column by giving exposure time 1 to all the ipd patient
   data.maic1<-data.maic1 %>% add_column (exposure.time=data.maic1$TOTDAYS/365.25)

   ## making a new variable called weighted exposure time
   data.maic1<-data.maic1 %>% mutate(wetd.exposr.time=wt*exposure.time)

   print(sum(data.maic1$wetd.exposr.time))     ##weighted person year

   ## making a new variable called weighted event
   data.maic1<-data.maic1 %>% mutate(wetd.event=wt*max_value)

   print(sum(data.maic1$wetd.event))
    print(sum(data.maic1$wt))

 return(c(sum(data.maic1$wetd.exposr.time),sum(data.maic1$wetd.event),sum(data.maic1$wt)))

 }

 results.boot.std.1.2<-boot(data=data.whole.maic, statistic=boot.std.1.2,R=1)


#####  MAIC 2

  boot.std.1.3<-function(d,i){
   data.maic2<- d[i,]

   agd.update<-rename(agd, cov1=PRENEXCD.std2, cov2=eisonofil.std2)

   data.maic2 <- rename(data.maic2, cov1=PRENEXCD, cov2=eisonofil)

   match_cov<-c("cov1", "cov2")

   data.maic2<- data.maic2 %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std2,
   cov2_cen=cov2-agd$eisonofil.std2)
```

```
  cent_match_cov<-c("cov1_cen","cov2_cen")
  est_weights<-estimate_weights(intervention_data = data.maic2,
                                matching_vars =cent_match_cov )

  analysis_data<-data.frame(est_weights[[2]])

  ## attaching the weight column to ipd dataset
  data.maic2<-cbind(data.maic2,analysis_data$wt)
  data.maic2<-rename(data.maic2, wt=`analysis_data$wt`)

  ## adding a new column by giving exposure time 1 to all the ipd patient
  data.maic2<-data.maic2 %>% add_column (exposure.time=data.maic2$TOTDAYS/365.25)

  ## making a new variable called weighted exposure time
  data.maic2<-data.maic2 %>% mutate(wetd.exposr.time=wt*exposure.time)

  print(sum(data.maic2$wetd.exposr.time))     ##weighted person year

  ## making a new variable called weighted event
  data.maic2<-data.maic2%>% mutate(wetd.event=wt*max_value)

  print(sum(data.maic2$wetd.event))

  print(sum(data.maic2$wt))

  return(c(sum(data.maic2$wetd.exposr.time),sum(data.maic2$wetd.event),sum(data.maic2$wt))

}

results.boot.std.1.3<-boot(data=data.whole.maic, statistic=boot.std.1.3,R=1)


#####  MAIC 3

boot.std.1.4<-function(d,i){
  data.maic3<- d[i,]

  agd.update<-rename(agd, cov1=PRENEXCD.std3, cov2=eisonofil.std3)

  data.maic3 <- rename( data.maic3, cov1=PRENEXCD, cov2=eisonofil)
   match_cov<-c("cov1", "cov2")

  data.maic3<- data.maic3 %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std3,
  cov2_cen=cov2-agd$eisonofil.std3)
```

```
  cent_match_cov<-c("cov1_cen","cov2_cen")
  est_weights<-estimate_weights(intervention_data = data.maic3,
                                matching_vars =cent_match_cov )

  analysis_data<-data.frame(est_weights[[2]])

  ## attaching the weight column to ipd dataset
  data.maic3<-cbind(data.maic3,analysis_data$wt)
  data.maic3<-rename(data.maic3, wt=`analysis_data$wt`)

  ## adding a new column by giving exposure time 1 to all the ipd patient
  data.maic3<-data.maic3 %>% add_column (exposure.time=data.maic3$TOTDAYS/365.25)

  ## making a new variable called weighted exposure time
  data.maic3<-data.maic3 %>% mutate(wetd.exposr.time=wt*exposure.time)

  print(sum(data.maic3$wetd.exposr.time))     ##weighted person year

  ## making a new variable called weighted event
  data.maic3<-data.maic3 %>% mutate(wetd.event=wt*max_value)
   print(sum(data.maic3$wetd.event))
   print(sum(data.maic3$wt))

  return(c(sum(data.maic3$wetd.exposr.time),sum(data.maic3$wetd.event),sum(data.maic3$wt))

}

results.boot.std.1.4<-boot(data=data.whole.maic, statistic=boot.std.1.4,R=1)


#####  MAIC 4

boot.std.1.5<-function(d,i){
  data.maic4<- d[i,]

  agd.update<-rename(agd, cov1=PRENEXCD.std4, cov2=eisonofil.std4)

  data.maic4 <- rename( data.maic4, cov1=PRENEXCD, cov2=eisonofil)

  match_cov<-c("cov1", "cov2")

  data.maic4<- data.maic4 %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std4,
  cov2_cen=cov2-agd$eisonofil.std4)

  cent_match_cov<-c("cov1_cen","cov2_cen")
```

```
  est_weights<-estimate_weights(intervention_data = data.maic4,
                                matching_vars =cent_match_cov )

  analysis_data<-data.frame(est_weights[[2]])

  ## attaching the weight column to ipd dataset
  data.maic4<-cbind(data.maic4,analysis_data$wt)
  data.maic4<-rename(data.maic4, wt=`analysis_data$wt`)

  ## adding a new column by giving exposure time 1 to all the ipd patient
  data.maic4<-data.maic4 %>% add_column (exposure.time=data.maic4$TOTDAYS/365.25)

  ## making a new variable called weighted exposure time
  data.maic4<-data.maic4 %>% mutate(wetd.exposr.time=wt*exposure.time)
  print(sum(data.maic4$wetd.exposr.time))     ##weighted person year

  ## making a new variable called weighted event
  data.maic4<-data.maic4 %>% mutate(wetd.event=wt*max_value)
  print(sum(data.maic4$wetd.event))
  print(sum(data.maic4$wt))

  return(c(sum(data.maic4$wetd.exposr.time),sum(data.maic4$wetd.event),sum(data.maic4$wt))

}

results.boot.std.1.5<-boot(data=data.whole.maic, statistic=boot.std.1.5,R=1)


#####  MAIC 5

boot.std.1.6<-function(d,i){
  data.maic5<- d[i,]

  agd.update<-rename(agd, cov1=PRENEXCD.std5, cov2=eisonofil.std5)
  data.maic5 <- rename(data.maic5, cov1=PRENEXCD, cov2=eisonofil)
  match_cov<-c("cov1", "cov2")

  data.maic5<- data.maic5 %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std5,
  cov2_cen=cov2-agd$eisonofil.std5)

  cent_match_cov<-c("cov1_cen","cov2_cen")
  est_weights<-estimate_weights(intervention_data = data.maic5,
                                matching_vars =cent_match_cov )

  analysis_data<-data.frame(est_weights[[2]])
```

```
  ## attaching the weight column to ipd dataset
  data.maic5<-cbind(data.maic5,analysis_data$wt)
  data.maic5<-rename(data.maic5, wt=‘analysis_data$wt‘)

  ## adding a new column by giving exposure time 1 to all the ipd patient
  data.maic5<-data.maic5 %>% add_column (exposure.time=data.maic5$TOTDAYS/365.25)

  ## making a new variable called weighted exposure time
  data.maic5<-data.maic5%>% mutate(wetd.exposr.time=wt*exposure.time)

  print(sum(data.maic5$wetd.exposr.time))     ##weighted person year

  ## making a new variable called weighted event
  data.maic5<-data.maic5 %>% mutate(wetd.event=wt*max_value)
  print(sum(data.maic5$wetd.event))
  print(sum(data.maic5$wt))

  return(c(sum(data.maic5$wetd.exposr.time),sum(data.maic5$wetd.event),sum(data.maic5$wt))

}
results.boot.std.1.6<-boot(data=data.whole.maic, statistic=boot.std.1.6,R=1)

#####  MAIC 6

boot.std.1.7<-function(d,i){
  data.maic6<- d[i,]

  agd.update<-rename(agd, cov1=PRENEXCD.std6, cov2=eisonofil.std6)
  data.maic6 <- rename( data.maic6, cov1=PRENEXCD, cov2=eisonofil)
  match_cov<-c("cov1", "cov2")

  data.maic6<- data.maic6 %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std6,
  cov2_cen=cov2-agd$eisonofil.std6)

  cent_match_cov<-c("cov1_cen","cov2_cen")
  est_weights<-estimate_weights(intervention_data = data.maic6,
                                matching_vars =cent_match_cov )

  analysis_data<-data.frame(est_weights[[2]])

  ## attaching the weight column to ipd dataset
  data.maic6<-cbind(data.maic6,analysis_data$wt)
  data.maic6<-rename(data.maic6, wt=‘analysis_data$wt‘)
```

```
  ## adding a new column by giving exposure time 1 to all the ipd patient
  data.maic6<-data.maic6 %>% add_column (exposure.time=data.maic6$TOTDAYS/365.25)

  ## making a new variable called weighted exposure time
  data.maic6<-data.maic6 %>% mutate(wetd.exposr.time=wt*exposure.time)
  print(sum(data.maic6$wetd.exposr.time))     ##weighted person year

  ## making a new variable called weighted event
  data.maic6<-data.maic6 %>% mutate(wetd.event=wt*max_value)
  print(sum(data.maic6$wetd.event))
  print(sum(data.maic6$wt))

  return(c(sum(data.maic6$wetd.exposr.time),sum(data.maic6$wetd.event),sum(data.maic6$wt))

}

results.boot.std.1.7<-boot(data=data.whole.maic, statistic=boot.std.1.7,R=1)


#####  MAIC 7

boot.std.1.8<-function(d,i){
  data.maic7<- d[i,]

  agd.update<-rename(agd, cov1=PRENEXCD.std7, cov2=eisonofil.std7)
  data.maic7 <- rename( data.maic7, cov1=PRENEXCD, cov2=eisonofil)
  match_cov<-c("cov1", "cov2")

  data.maic7<- data.maic7 %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std7,
  cov2_cen=cov2-agd$eisonofil.std7)

  cent_match_cov<-c("cov1_cen","cov2_cen")
  est_weights<-estimate_weights(intervention_data = data.maic7,
                                matching_vars =cent_match_cov )


  analysis_data<-data.frame(est_weights[[2]])

  ## attaching the weight column to ipd dataset
  data.maic7<-cbind(data.maic7,analysis_data$wt)
  data.maic7<-rename(data.maic7, wt=`analysis_data$wt`)

  ## adding a new column by giving exposure time 1 to all the ipd patient
  data.maic7<-data.maic7 %>% add_column (exposure.time=data.maic7$TOTDAYS/365.25)
```

```
  ## making a new variable called weighted exposure time
  data.maic7<-data.maic7 %>% mutate(wetd.exposr.time=wt*exposure.time)

  print(sum(data.maic7$wetd.exposr.time))     ##weighted person year

  ## making a new variable called weighted event
  data.maic7<-data.maic7 %>% mutate(wetd.event=wt*max_value)
  print(sum(data.maic7$wetd.event))
  print(sum(data.maic7$wt))

  return(c(sum(data.maic7$wetd.exposr.time),sum(data.maic7$wetd.event),sum(data.maic7$wt))

}

results.boot.std.1.8<-boot(data=data.whole.maic, statistic=boot.std.1.8,R=1)


#####  MAIC 8

boot.std.1.9<-function(d,i){
  data.maic8<- d[i,]

  agd.update<-rename(agd, cov1=PRENEXCD.std8, cov2=eisonofil.std8)
  data.maic8 <- rename( data.maic8, cov1=PRENEXCD, cov2=eisonofil)
  match_cov<-c("cov1", "cov2")

  data.maic8<-  data.maic8 %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std8,
  cov2_cen=cov2-agd$eisonofil.std8)

  cent_match_cov<-c("cov1_cen","cov2_cen")
  est_weights<-estimate_weights(intervention_data = data.maic8,
                                matching_vars =cent_match_cov )

  analysis_data<-data.frame(est_weights[[2]])

  ## attaching the weight column to ipd dataset
  data.maic8<-cbind(data.maic8,analysis_data$wt)
  data.maic8<-rename(data.maic8, wt=‘analysis_data$wt‘)

  ## adding a new column by giving exposure time 1 to all the ipd patient
  data.maic8<-data.maic8 %>% add_column (exposure.time=data.maic8$TOTDAYS/365.25)

  ## making a new variable called weighted exposure time
  data.maic8<-data.maic8 %>% mutate(wetd.exposr.time=wt*exposure.time)
  print(sum(data.maic8$wetd.exposr.time))     ##weighted person year
```

```
  ## making a new variable called weighted event
  jdata.maic8<-data.maic8%>% mutate(wetd.event=wt*max_value)
  print(sum(data.maic8$wetd.event))
  print(sum(data.maic8$wt))

  return(c(sum(data.maic8$wetd.exposr.time),sum(data.maic8$wetd.event),sum(data.maic8$wt))

}

results.boot.std.1.9<-boot(data=data.whole.maic, statistic=boot.std.1.9,R=1)


#####  MAIC 9

boot.std.1.10<-function(d,i){
  data.maic9<- d[i,]

  agd.update<-rename(agd, cov1=PRENEXCD.std9, cov2=eisonofil.std9)
  data.maic9 <- rename( data.maic9, cov1=PRENEXCD, cov2=eisonofil)
  match_cov<-c("cov1", "cov2")

  data.maic9<- data.maic9 %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std9,
  cov2_cen=cov2-agd$eisonofil.std9)

  cent_match_cov<-c("cov1_cen","cov2_cen")
  est_weights<-estimate_weights(intervention_data = data.maic9,
                                matching_vars =cent_match_cov )

  analysis_data<-data.frame(est_weights[[2]])

  ## attaching the weight column to ipd dataset
  data.maic9<-cbind(data.maic9,analysis_data$wt)
  data.maic9<-rename(data.maic9, wt=`analysis_data$wt`)

  ## adding a new column by giving exposure time 1 to all the ipd patient
  data.maic9<-data.maic9 %>% add_column (exposure.time=data.maic9$TOTDAYS/365.25)

  ## making a new variable called weighted exposure time
  data.maic9<-data.maic9%>% mutate(wetd.exposr.time=wt*exposure.time)
  print(sum(data.maic9$wetd.exposr.time))     ##weighted person year

  ## making a new variable called weighted event
  data.maic9<-data.maic9 %>% mutate(wetd.event=wt*max_value)
  print(sum(data.maic9$wetd.event))
```

```
  print(sum(data.maic9$wt))

  return(c(sum(data.maic9$wetd.exposr.time),sum(data.maic9$wetd.event),sum(data.maic9$wt))

}

results.boot.std.1.10<-boot(data=data.whole.maic, statistic=boot.std.1.10,R=1)

study.no<-c(1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8,9,9)
trt.no<-c(1,2,2,3,2,3,2,3,2,4,2,4,2,4,2,5,2,5)
E<-c(25.3836,results.boot.std.1.10$t[1],
     results.boot.std.1.2$t[1],400,
     results.boot.std.1.3$t[1],454.75 ,
      results.boot.std.1.4$t[1],39.1937  ,
     results.boot.std.1.5$t[1], 11.8118  ,
     results.boot.std.1.6$t[1],631 ,
      results.boot.std.1.7$t[1],47.408222  ,
     results.boot.std.1.8$t[1],393.073996 ,
     results.boot.std.1.9$t[1],72.263018
     )

r<-c(54,results.boot.std.1.10$t[2],

  results.boot.std.1.2$t[2],292,
     results.boot.std.1.3$t[2],273,
     results.boot.std.1.4$t[2],33,
     results.boot.std.1.5$t[2], 5,
     results.boot.std.1.6$t[2],290,
      results.boot.std.1.7$t[2],31,
     results.boot.std.1.8$t[2], 259,
     results.boot.std.1.9$t[2],15
     )
r<-as.integer(r)

n<-c(66,results.boot.std.1.10$t[3],
  results.boot.std.1.2$t[3], 400,
     results.boot.std.1.3$t[3],425,
     results.boot.std.1.4$t[3],73,
     results.boot.std.1.5$t[3], 154,
     results.boot.std.1.6$t[3],631,
      results.boot.std.1.7$t[3],103,
     results.boot.std.1.8$t[3], 427,
     results.boot.std.1.9$t[3],157
     )
```

```
  n<-as.integer(n)
  datacel<-data.frame(cbind(study.no,trt.no, E,r,n ))
  asthma_net<-set_agd_arm(datacel ,
                          study=study.no ,
                          trt= trt.no,
                          r=r,
                          E=E,
                          sample_size=n,
                          trt_ref = "2")

  asthma_net_FE<-nma(asthma_net,
                     trt_effects = "fixed",
                     prior_intercept = normal(scale=100),
                     prior_trt = normal(scale=100))

  as<-as.data.frame(asthma_net_FE, pars=c("d"))

  d1<-mean(as$`d[1]`)
  d2<-mean(as$`d[3]`)
  d3<-mean(as$`d[4]`)
  d4<-mean(as$`d[5]`)
  ci<-as.data.frame(summary(asthma_net_FE,pars,include,stat="pointinterval"))
  lowerci.d1<-mean(ci[10,4])
  upperci.d1<-mean(ci[10,8])

  lowerci.d2<-mean(ci[11,4])
  upperci.d2<-mean(ci[11,8])

  lowerci.d3<-mean(ci[12,4])
  upperci.d3<-mean(ci[12,8])

  lowerci.d4<-mean(ci[13,4])
  upperci.d4<-mean(ci[13,8])

  dic_FE<-dic(asthma_net_FE)

  return(c(d1,d2,d3,d4,lowerci.d1,upperci.d1,
           lowerci.d2,upperci.d2,
           lowerci.d3,upperci.d3,
           lowerci.d4,upperci.d4,
           dic_FE$dic,dic_FE$pd,dic_FE$resdev))

}

results<-boot(data=join_with_bleos, statistic=boot.maic, R=2000)
```

```
(mean.d1<-mean(results$t[,1]))
(mean.d2<-mean(results$t[,2]))
(mean.d3<-mean(results$t[,3]))
(mean.d4<-mean(results$t[,4]))
(sd.d1<-sd(results$t[,1]))
(sd.d2<-sd(results$t[,2]))
(sd.d3<-sd(results$t[,3]))
(sd.d4<-sd(results$t[,4]))


(lowerci.d1<-mean(results$t[,5]))
(upperci.d1<-mean(results$t[,6]))
(lowerci.d2<-mean(results$t[,7]))
(upperci.d2<-mean(results$t[,8]))
(lowerci.d3<-mean(results$t[,9]))
(upperci.d3<-mean(results$t[,10]))
(lowerci.d4<-mean(results$t[,11]))
(upperci.d4<-mean(results$t[,12]))
(dic<-mean(results$t[,13]))
(pd<-mean(results$t[,14]))
(resdev<-mean(results$t[,15]))
```

F.4 Double-bootstrap with asthma data (random effects model)

```
data<-read_sas("V:\\112997\\Dataset\\AR_UNKNOWN_-99\\gsk_112997_exacanal.sas7bdat")
##subsetting with only two groups
data2<-data %>% filter (TRTGRP=="Placebo"| TRTGRP=="Mepolizumab 250mg" )

data2<-select(data2,-c("STUDYID" ,"INVID", "CENTREID", "USUBJID" , "RACECD", "RACE",
"TRTCD", "ATRTCD", "ATRTGRP", "LTOTAL" , "CS_TIME", "HED_NUM",  "HED_RATE", "HED_TIME" ,
"ALL_NUM","ALL_RATE", "ALL_TIME", "HOS_NUM", "HOS_RATE" ,"HOS_TIME" ,"STRATCD" ,"REGIONCD",

##reading eisonifil data
bleos<-read_sas("V:\\112997\\Dataset\\AR_UNKNOWN_-99\\gsk_112997_bleos.sas7bdat")
## subsetting for two grp
bleos<-bleos %>% filter (TRTGRP=="Placebo"| TRTGRP=="Mepolizumab 250mg" )
 ##taking only baseline eisonofil
bleos<-bleos %>% filter (VISIT=="Baseline")            ## discarding percentage value
bleos<-bleos %>% filter (LBTEST=="Eosinophils")

## deleting extra columns
```

```
bleos<-select(bleos,-c("CENTREID", "USUBJID" ,"AGE",  "SEX", "RACECD", "RACE",
"TRTCD","TRTGRP", "ATRTCD",   "ATRTGRP", "VISITNUM", "AVISNUM", "AVISIT", "PTMNUM",
"PTM", "LBACTDY","LBDT","LBACTTM", "LBTESTCD","LBTEST", "LBSTUNIT", "LBSTNRLO", "LBSTNRHI",
"LBORRES", "LBORRESN", "LBORUNIT", "LBORNRLO", "LBORNRHI", "LBNRCD", "LBNRIND", "LBSTCCLO",
"LBSTCCHI", "LBCCCD", "LBCCIND", "LBSTDBL", "STDCHGBL", "ILBRES", "ILBBL", "ILBCHBL",
"_ILBRES", "_ILBBL", "_ILBCHBL", "LBIDCD","LBAGE" ,"LBID" , "LBCAT" , "LBACSNUM", "LBFAST",
"LBTSTCOM", "LBUDIFCD", "EVALFLAG","STUDYID",  "INVID" ))

### reading data to create binary variable

exacag<-read_sas("V:\\112997\\Dataset\\AR_UNKNOWN_-99\\gsk_112997_exacag.sas7bdat")
## subsetting for two grp
exacag<-exacag %>% filter (TRTGRP=="Placebo"| TRTGRP=="Mepolizumab 250mg" )

exacag2<-exacag %>% group_by(SUBJID)  %>% mutate(max_value=max(STATUS))

exacag2<-exacag2 %>% distinct (SUBJID, .keep_all = T)

exacag2<-select(exacag2,-c("STUDYID", "INVID","CENTREID", "USUBJID","AGE",
"SEX", "RACECD", "RACE", "TRTCD", "TRTGRP" ,   "ATRTCD", "ATRTGRP", "ESTART", "ESTOP",
"STATUS", "STRATCD","STRATUM", "REGIONCD", "REGION" ,"PRENEXCD", "PRENEX", "BPRBDVAL"
))

################## merging the two datasets

join_with_bleos<-left_join(data2,bleos, by="SUBJID")
join_with_bleos<-left_join(join_with_bleos,exacag2, by="SUBJID")

### converting eisonofil column into another unit

join_with_bleos<-join_with_bleos %>% mutate(eisonofil=LBSTRESN*1000)

join_with_bleos<-join_with_bleos %>% filter(TRTGRP=="Mepolizumab 250mg")


######################## MAIC 1 with matching only mean

agd<-read.csv("V:\\112997\\Dataset\\2nd version\\agd_data_maic.csv")

boot.maic<-function(d,i){

  data.whole.maic<-d[i,]

  ####MAIC 1
```

```
boot.std.1.2<-function(d,i){
  data.maic1<- d[i,]

  agd.update<-rename(agd, cov1=PRENEXCD.std1, cov2=eisonofil.std1)

  data.maic1 <- rename( data.maic1, cov1=PRENEXCD, cov2=eisonofil)

  match_cov<-c("cov1", "cov2")

data.maic1<- data.maic1 %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std1,
cov2_cen=cov2-agd$eisonofil.std1)

  cent_match_cov<-c("cov1_cen","cov2_cen")
  est_weights<-estimate_weights(intervention_data = data.maic1,
                                matching_vars =cent_match_cov )

  analysis_data<-data.frame(est_weights[[2]])

  ## attaching the weight column to ipd dataset
  data.maic1<-cbind(data.maic1,analysis_data$wt)
  data.maic1<-rename(data.maic1, wt=`analysis_data$wt`)

  ## adding a new column by giving exposure time 1 to all the ipd patient
  data.maic1<-data.maic1 %>% add_column (exposure.time=data.maic1$TOTDAYS/365.25)

  ## making a new variable called weighted exposure time
  data.maic1<-data.maic1 %>% mutate(wetd.exposr.time=wt*exposure.time)

  print(sum(data.maic1$wetd.exposr.time))     ##weighted person year

  ## making a new variable called weighted event
  data.maic1<-data.maic1 %>% mutate(wetd.event=wt*max_value)

  print(sum(data.maic1$wetd.event))
   print(sum(data.maic1$wt))

return(c(sum(data.maic1$wetd.exposr.time),sum(data.maic1$wetd.event),sum(data.maic1$wt)))

}

results.boot.std.1.2<-boot(data=data.whole.maic, statistic=boot.std.1.2,R=1)


#####  MAIC 2
```

```
 boot.std.1.3<-function(d,i){
  data.maic2<- d[i,]

  agd.update<-rename(agd, cov1=PRENEXCD.std2, cov2=eisonofil.std2)

  data.maic2 <- rename(data.maic2, cov1=PRENEXCD, cov2=eisonofil)

  match_cov<-c("cov1", "cov2")

  data.maic2<- data.maic2 %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std2,
  cov2_cen=cov2-agd$eisonofil.std2)

  cent_match_cov<-c("cov1_cen","cov2_cen")
  est_weights<-estimate_weights(intervention_data = data.maic2,
                                matching_vars =cent_match_cov )

  analysis_data<-data.frame(est_weights[[2]])

  ## attaching the weight column to ipd dataset
  data.maic2<-cbind(data.maic2,analysis_data$wt)
  data.maic2<-rename(data.maic2, wt=`analysis_data$wt`)

  ## adding a new column by giving exposure time 1 to all the ipd patient
  data.maic2<-data.maic2 %>% add_column (exposure.time=data.maic2$TOTDAYS/365.25)

  ## making a new variable called weighted exposure time
  data.maic2<-data.maic2 %>% mutate(wetd.exposr.time=wt*exposure.time)

  print(sum(data.maic2$wetd.exposr.time))     ##weighted person year

  ## making a new variable called weighted event
  data.maic2<-data.maic2%>% mutate(wetd.event=wt*max_value)

  print(sum(data.maic2$wetd.event))

  print(sum(data.maic2$wt))

  return(c(sum(data.maic2$wetd.exposr.time),sum(data.maic2$wetd.event),sum(data.maic2$wt))

}

results.boot.std.1.3<-boot(data=data.whole.maic, statistic=boot.std.1.3,R=1)
```

```
#####  MAIC 3

boot.std.1.4<-function(d,i){
  data.maic3<- d[i,]

  agd.update<-rename(agd, cov1=PRENEXCD.std3, cov2=eisonofil.std3)

  data.maic3 <- rename( data.maic3, cov1=PRENEXCD, cov2=eisonofil)
   match_cov<-c("cov1", "cov2")

  data.maic3<- data.maic3 %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std3,
  cov2_cen=cov2-agd$eisonofil.std3)

  cent_match_cov<-c("cov1_cen","cov2_cen")
  est_weights<-estimate_weights(intervention_data = data.maic3,
                                matching_vars =cent_match_cov )

  analysis_data<-data.frame(est_weights[[2]])

  ## attaching the weight column to ipd dataset
  data.maic3<-cbind(data.maic3,analysis_data$wt)
  data.maic3<-rename(data.maic3, wt=`analysis_data$wt`)

  ## adding a new column by giving exposure time 1 to all the ipd patient
  data.maic3<-data.maic3 %>% add_column (exposure.time=data.maic3$TOTDAYS/365.25)

  ## making a new variable called weighted exposure time
  data.maic3<-data.maic3 %>% mutate(wetd.exposr.time=wt*exposure.time)

  print(sum(data.maic3$wetd.exposr.time))     ##weighted person year

  ## making a new variable called weighted event
  data.maic3<-data.maic3 %>% mutate(wetd.event=wt*max_value)
   print(sum(data.maic3$wetd.event))
   print(sum(data.maic3$wt))

  return(c(sum(data.maic3$wetd.exposr.time),sum(data.maic3$wetd.event),sum(data.maic3$wt))

}

results.boot.std.1.4<-boot(data=data.whole.maic, statistic=boot.std.1.4,R=1)


#####  MAIC 4
```

```
boot.std.1.5<-function(d,i){
  data.maic4<- d[i,]

  agd.update<-rename(agd, cov1=PRENEXCD.std4, cov2=eisonofil.std4)

  data.maic4 <- rename( data.maic4, cov1=PRENEXCD, cov2=eisonofil)

  match_cov<-c("cov1", "cov2")

  data.maic4<- data.maic4 %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std4,
  cov2_cen=cov2-agd$eisonofil.std4)

  cent_match_cov<-c("cov1_cen","cov2_cen")
  est_weights<-estimate_weights(intervention_data = data.maic4,
                                matching_vars =cent_match_cov )

  analysis_data<-data.frame(est_weights[[2]])

  ## attaching the weight column to ipd dataset
  data.maic4<-cbind(data.maic4,analysis_data$wt)
  data.maic4<-rename(data.maic4, wt=`analysis_data$wt`)

  ## adding a new column by giving exposure time 1 to all the ipd patient
  data.maic4<-data.maic4 %>% add_column (exposure.time=data.maic4$TOTDAYS/365.25)

  ## making a new variable called weighted exposure time
  data.maic4<-data.maic4 %>% mutate(wetd.exposr.time=wt*exposure.time)
  print(sum(data.maic4$wetd.exposr.time))     ##weighted person year

  ## making a new variable called weighted event
  data.maic4<-data.maic4 %>% mutate(wetd.event=wt*max_value)
  print(sum(data.maic4$wetd.event))
  print(sum(data.maic4$wt))

  return(c(sum(data.maic4$wetd.exposr.time),sum(data.maic4$wetd.event),sum(data.maic4$wt))

}

results.boot.std.1.5<-boot(data=data.whole.maic, statistic=boot.std.1.5,R=1)


#####  MAIC 5

boot.std.1.6<-function(d,i){
  data.maic5<- d[i,]
```

```
  agd.update<-rename(agd, cov1=PRENEXCD.std5, cov2=eisonofil.std5)
  data.maic5 <- rename(data.maic5, cov1=PRENEXCD, cov2=eisonofil)
  match_cov<-c("cov1", "cov2")

  data.maic5<- data.maic5 %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std5,
  cov2_cen=cov2-agd$eisonofil.std5)

  cent_match_cov<-c("cov1_cen","cov2_cen")
  est_weights<-estimate_weights(intervention_data = data.maic5,
                                matching_vars =cent_match_cov )

  analysis_data<-data.frame(est_weights[[2]])

  ## attaching the weight column to ipd dataset
  data.maic5<-cbind(data.maic5,analysis_data$wt)
  data.maic5<-rename(data.maic5, wt=`analysis_data$wt`)

  ## adding a new column by giving exposure time 1 to all the ipd patient
  data.maic5<-data.maic5 %>% add_column (exposure.time=data.maic5$TOTDAYS/365.25)

  ## making a new variable called weighted exposure time
  data.maic5<-data.maic5%>% mutate(wetd.exposr.time=wt*exposure.time)

  print(sum(data.maic5$wetd.exposr.time))     ##weighted person year

  ## making a new variable called weighted event
  data.maic5<-data.maic5 %>% mutate(wetd.event=wt*max_value)
  print(sum(data.maic5$wetd.event))
  print(sum(data.maic5$wt))

  return(c(sum(data.maic5$wetd.exposr.time),sum(data.maic5$wetd.event),sum(data.maic5$wt))

}
results.boot.std.1.6<-boot(data=data.whole.maic, statistic=boot.std.1.6,R=1)

#####  MAIC 6

boot.std.1.7<-function(d,i){
  data.maic6<- d[i,]

  agd.update<-rename(agd, cov1=PRENEXCD.std6, cov2=eisonofil.std6)
  data.maic6 <- rename( data.maic6, cov1=PRENEXCD, cov2=eisonofil)
  match_cov<-c("cov1", "cov2")
```

```
  data.maic6<- data.maic6 %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std6,
  cov2_cen=cov2-agd$eisonofil.std6)

  cent_match_cov<-c("cov1_cen","cov2_cen")
  est_weights<-estimate_weights(intervention_data = data.maic6,
                                matching_vars =cent_match_cov )

  analysis_data<-data.frame(est_weights[[2]])

  ## attaching the weight column to ipd dataset
  data.maic6<-cbind(data.maic6,analysis_data$wt)
  data.maic6<-rename(data.maic6, wt=`analysis_data$wt`)

  ## adding a new column by giving exposure time 1 to all the ipd patient
  data.maic6<-data.maic6 %>% add_column (exposure.time=data.maic6$TOTDAYS/365.25)

  ## making a new variable called weighted exposure time
  data.maic6<-data.maic6 %>% mutate(wetd.exposr.time=wt*exposure.time)
  print(sum(data.maic6$wetd.exposr.time))     ##weighted person year

  ## making a new variable called weighted event
  data.maic6<-data.maic6 %>% mutate(wetd.event=wt*max_value)
  print(sum(data.maic6$wetd.event))
  print(sum(data.maic6$wt))

  return(c(sum(data.maic6$wetd.exposr.time),sum(data.maic6$wetd.event),sum(data.maic6$wt))

}

results.boot.std.1.7<-boot(data=data.whole.maic, statistic=boot.std.1.7,R=1)



#####  MAIC 7

boot.std.1.8<-function(d,i){
  data.maic7<- d[i,]

  agd.update<-rename(agd, cov1=PRENEXCD.std7, cov2=eisonofil.std7)
  data.maic7 <- rename( data.maic7, cov1=PRENEXCD, cov2=eisonofil)
  match_cov<-c("cov1", "cov2")

  data.maic7<- data.maic7 %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std7,
  cov2_cen=cov2-agd$eisonofil.std7)
```

```
  cent_match_cov<-c("cov1_cen","cov2_cen")
  est_weights<-estimate_weights(intervention_data = data.maic7,
                                matching_vars =cent_match_cov )


  analysis_data<-data.frame(est_weights[[2]])

  ## attaching the weight column to ipd dataset
  data.maic7<-cbind(data.maic7,analysis_data$wt)
  data.maic7<-rename(data.maic7, wt=`analysis_data$wt`)

  ## adding a new column by giving exposure time 1 to all the ipd patient
  data.maic7<-data.maic7 %>% add_column (exposure.time=data.maic7$TOTDAYS/365.25)

  ## making a new variable called weighted exposure time
  data.maic7<-data.maic7 %>% mutate(wetd.exposr.time=wt*exposure.time)

  print(sum(data.maic7$wetd.exposr.time))     ##weighted person year

  ## making a new variable called weighted event
  data.maic7<-data.maic7 %>% mutate(wetd.event=wt*max_value)
  print(sum(data.maic7$wetd.event))
  print(sum(data.maic7$wt))

  return(c(sum(data.maic7$wetd.exposr.time),sum(data.maic7$wetd.event),sum(data.maic7$wt))

}

results.boot.std.1.8<-boot(data=data.whole.maic, statistic=boot.std.1.8,R=1)


#####  MAIC 8

boot.std.1.9<-function(d,i){
  data.maic8<- d[i,]

  agd.update<-rename(agd, cov1=PRENEXCD.std8, cov2=eisonofil.std8)
  data.maic8 <- rename( data.maic8, cov1=PRENEXCD, cov2=eisonofil)
  match_cov<-c("cov1", "cov2")

  data.maic8<-  data.maic8 %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std8,
  cov2_cen=cov2-agd$eisonofil.std8)

  cent_match_cov<-c("cov1_cen","cov2_cen")
  est_weights<-estimate_weights(intervention_data = data.maic8,
```

```
                                matching_vars =cent_match_cov )

  analysis_data<-data.frame(est_weights[[2]])

  ## attaching the weight column to ipd dataset
  data.maic8<-cbind(data.maic8,analysis_data$wt)
  data.maic8<-rename(data.maic8, wt=‘analysis_data$wt‘)

  ## adding a new column by giving exposure time 1 to all the ipd patient
  data.maic8<-data.maic8 %>% add_column (exposure.time=data.maic8$TOTDAYS/365.25)

  ## making a new variable called weighted exposure time
  data.maic8<-data.maic8 %>% mutate(wetd.exposr.time=wt*exposure.time)
  print(sum(data.maic8$wetd.exposr.time))     ##weighted person year

  ## making a new variable called weighted event
  jdata.maic8<-data.maic8%>% mutate(wetd.event=wt*max_value)
  print(sum(data.maic8$wetd.event))
  print(sum(data.maic8$wt))

  return(c(sum(data.maic8$wetd.exposr.time),sum(data.maic8$wetd.event),sum(data.maic8$wt))

}

results.boot.std.1.9<-boot(data=data.whole.maic, statistic=boot.std.1.9,R=1)


#####  MAIC 9

boot.std.1.10<-function(d,i){
  data.maic9<- d[i,]

  agd.update<-rename(agd, cov1=PRENEXCD.std9, cov2=eisonofil.std9)
  data.maic9 <- rename( data.maic9, cov1=PRENEXCD, cov2=eisonofil)
  match_cov<-c("cov1", "cov2")

  data.maic9<- data.maic9 %>% mutate(cov1_cen=cov1-agd$PRENEXCD.std9,
  cov2_cen=cov2-agd$eisonofil.std9)

  cent_match_cov<-c("cov1_cen","cov2_cen")
  est_weights<-estimate_weights(intervention_data = data.maic9,
                                matching_vars =cent_match_cov )

  analysis_data<-data.frame(est_weights[[2]])
```

```
  ## attaching the weight column to ipd dataset
  data.maic9<-cbind(data.maic9,analysis_data$wt)
  data.maic9<-rename(data.maic9, wt=`analysis_data$wt`)

  ## adding a new column by giving exposure time 1 to all the ipd patient
  data.maic9<-data.maic9 %>% add_column (exposure.time=data.maic9$TOTDAYS/365.25)

  ## making a new variable called weighted exposure time
  data.maic9<-data.maic9%>% mutate(wetd.exposr.time=wt*exposure.time)
  print(sum(data.maic9$wetd.exposr.time))     ##weighted person year

  ## making a new variable called weighted event
  data.maic9<-data.maic9 %>% mutate(wetd.event=wt*max_value)
  print(sum(data.maic9$wetd.event))
  print(sum(data.maic9$wt))

  return(c(sum(data.maic9$wetd.exposr.time),sum(data.maic9$wetd.event),sum(data.maic9$wt))

}

results.boot.std.1.10<-boot(data=data.whole.maic, statistic=boot.std.1.10,R=1)

study.no<-c(1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8,9,9)
trt.no<-c(1,2,2,3,2,3,2,3,2,4,2,4,2,4,2,5,2,5)
E<-c(25.3836,results.boot.std.1.10$t[1],
     results.boot.std.1.2$t[1],400,
     results.boot.std.1.3$t[1],454.75 ,
      results.boot.std.1.4$t[1],39.1937  ,
     results.boot.std.1.5$t[1], 11.8118  ,
     results.boot.std.1.6$t[1],631 ,
      results.boot.std.1.7$t[1],47.408222  ,
     results.boot.std.1.8$t[1],393.073996 ,
     results.boot.std.1.9$t[1],72.263018
     )

r<-c(54,results.boot.std.1.10$t[2],

  results.boot.std.1.2$t[2],292,
     results.boot.std.1.3$t[2],273,
     results.boot.std.1.4$t[2],33,
     results.boot.std.1.5$t[2], 5,
     results.boot.std.1.6$t[2],290,
      results.boot.std.1.7$t[2],31,
     results.boot.std.1.8$t[2], 259,
     results.boot.std.1.9$t[2],15
```

```
     )
r<-as.integer(r)

n<-c(66,results.boot.std.1.10$t[3],
  results.boot.std.1.2$t[3], 400,
     results.boot.std.1.3$t[3],425,
     results.boot.std.1.4$t[3],73,
     results.boot.std.1.5$t[3], 154,
     results.boot.std.1.6$t[3],631,
      results.boot.std.1.7$t[3],103,
     results.boot.std.1.8$t[3], 427,
     results.boot.std.1.9$t[3],157
     )

n<-as.integer(n)
datacel<-data.frame(cbind(study.no,trt.no, E,r,n ))
asthma_net<-set_agd_arm(datacel ,
                        study=study.no ,
                        trt= trt.no,
                        r=r,
                        E=E,
                        sample_size=n,
                        trt_ref = "2")

asthma_net_RE<-nma(asthma_net,
                   trt_effects = "random",
                   prior_intercept = normal(scale=100),
                   prior_trt = normal(scale=100),
                   prior_het = half_normal(scale=0.15),iter=3000,thin=5)

as<-as.data.frame(asthma_net_RE, pars=c("d","tau"))

d1<-mean(as$`d[1]`)
d2<-mean(as$`d[3]`)
d3<-mean(as$`d[4]`)
d4<-mean(as$`d[5]`)
tau<-mean(as$tau)

ci<-as.data.frame(summary(asthma_net_RE,pars,include,stat="pointinterval"))
lowerci.d1<-mean(ci[10,4])
upperci.d1<-mean(ci[10,8])

lowerci.d2<-mean(ci[11,4])
upperci.d2<-mean(ci[11,8])
```

```
  lowerci.d3<-mean(ci[12,4])
  upperci.d3<-mean(ci[12,8])

  lowerci.d4<-mean(ci[13,4])
  upperci.d4<-mean(ci[13,8])

  dic_RE<-dic(asthma_net_RE)

  return(c(d1,d2,d3,d4,lowerci.d1,upperci.d1,
           lowerci.d2,upperci.d2,
           lowerci.d3,upperci.d3,
           lowerci.d4,upperci.d4,
           dic_RE$dic,dic_RE$pd,dic_RE$resdev,tau))

}

results<-boot(data=join_with_bleos, statistic=boot.maic, R=2000)

(mean.d1<-mean(results$t[,1]))
(mean.d2<-mean(results$t[,2]))
(mean.d3<-mean(results$t[,3]))
(mean.d4<-mean(results$t[,4]))
(sd.d1<-sd(results$t[,1]))
(sd.d2<-sd(results$t[,2]))
(sd.d3<-sd(results$t[,3]))
(sd.d4<-sd(results$t[,4]))

(lowerci.d1<-mean(results$t[,5]))
(upperci.d1<-mean(results$t[,6]))
(lowerci.d2<-mean(results$t[,7]))
(upperci.d2<-mean(results$t[,8]))
(lowerci.d3<-mean(results$t[,9]))
(upperci.d3<-mean(results$t[,10]))
(lowerci.d4<-mean(results$t[,11]))
(upperci.d4<-mean(results$t[,12]))
(dic<-mean(results$t[,13]))
(pd<-mean(results$t[,14]))
(resdev<-mean(results$t[,15]))
(tau<-mean(results$t[,16]))
```

Glossary

bias When the outcome of a study systematically differs from the 'true' outcome. 5

blinding In a study, the information on which intervention have been assigned to which patient is concealed from patients, caregivers, researchers and outcome assessors. 1

clinical effectiveness A measure to estimate the overall health advantage of an intervention considering both benefit and adverse effects. Not the same as efficacy . 13

collapsibility After a modification that doesn't alter a value, the value becomes collapsible or invariant post-modification. Conversely, when a modification does alter a value, the value is considered non-collapsible. . 43

conditional treatment effect Moving a subject from untreated to treated produces a conditional effect, at the subject level. The estimate of a conditional or adjusted effect comes from the regression coefficient of a treatment assignment indicator variable in a multivariable regression model. . 43

confounders Variables that are related to the intervention or outcome in such a way that it misinterprets the effect of the intervention on the outcome . 1

consistency consistency is the ability to compare effects in a consistent manner, whether direct or indirect. . 4

control In a study, a control is a comparator treatment that is used to estimate the effect of an intervention . 8

effect-modifier covariates that alter the effect of treatment on outcomes, so that the treatment is more or less effective in different subgroups formed by levels of the effect-modifier. 5, 28

efficacy When an intervention is examined under controlled research conditions to discover its effectiveness . 1

endpoint In a research study, a measurable event or outcome that makes up one of the study's objectives. 8

extrapolation Forecasting the value of a parameter beyond the range of observed values . 16

fixed effect In the fixed-effect model, we seek to calculate the mean of a single population, with the standard error reflecting the precision of this calculation. . 50

fractional polynomial network meta-analysis (FP NMA) The proportional hazard assumption has been increasingly challenged when conducting network meta-analyses (NMAs) of survival data. This issue has been addressed by the fractional polynomial NMA model which has been used in HTA submissions. It's a flexible parametric modeling approach for analyzing time-to-event data. In FP, the log hazard of an event can be fitted as a function of time using parametric models . 20

heterogeneity A term frequently used in meta-analyses and systematic reviews. It describes to what extent the treatment effect estimates from different studies differ (e.g. some studies may indicate beneficial treatment effects and others suggest negative effects). Result discrepancies may be the result of differences in study quality, populations, interventions, or outcomes measured in the included studies. 17

indirect comparison A investigation of competing interventions when they have not been compared in a head-to-head randomised study. 1

MAIC As opposed to conventional meta-analytic methods, Matching-Adjusted Indirect Comparison (MAIC) provides a robust comparison by re-weighting Individual Patient Data (IPD) from one study to baseline summary statistics from another, which provides a greater adjustment for observed study differences. . 12

marginal treatment effect The average effect of moving an entire population from untreated to treated, at the population level, is known as a marginal effect. . 43

meta-analysis A statistical analysis that includes several studies and analyses the results from the studies that are dealing with the same question and narrating the same outcomes. It is capable of producing a more precise estimate of the effect on a given outcome. 3

NMA Using a network of studies, network meta-analyses combine direct and indirect evidence to evaluate three or more interventions simultaneously. . 3

outcomes This is the measure of the possible results of a preventive or therapeutic intervention. It can be described either as an intermediary or as a terminal endpoints. 5

prognostic variable Covariates that affect (or is prognostic of) outcome . 5, 28

propensity score In the context of an individual's covariate values, this is the conditional probability of their participation in a trial. They are usually estimated through logistic regression. 32

proportional hazard (PH) In proportional hazards, variables are multiplicatively related to hazards . 20

quality-adjusted life year (QALY) In healthcare, Quality-Adjusted Life Years (QALYs) represent the value and benefit of health outcomes. Policy makers use it to guide their policy decisions and health economists use it as a health economic outcome measure . 15

random effects In the random-effects model, we are determining the mean across various populations, and the standard error denotes the precision of this calculation. . 50

randomisation When participants in a study are assigned to two or more alternative groups by implementing a random procedure, e.g. computer-generated random numbers. Randomisation tries to ensure a balance distribution of participants with distinct attributes between groups in order to reduces bias and confounding . 1

relative treatment effect What a treatment accomplishes relative to another treatment or control, for example, relative risk (RR). 10

sandwich estimator A variance estimator that is derived empirically from the data. It does not rely upon strong assumptions about the data (or in the case of MAIC, the weights). "Sandwich" refers to how the estimator is constructed, with the empirical approximation "sandwiched" between other matrices. 22

similarity In a network meta-analysis, any characteristics that may modify the treatment effect should be similar across all studies in the network. . 4

STC An analysis of STC involves estimating a linear regression model of population characteristics and outcomes in a trial that has individual patient data and then using that model to estimate outcomes in other studies. . 12

systematic review With a predefined protocol, when evidence on a clearly formulated question is summarised. In a systematic review, systematic methods are used to identify, select and appraise relevant studies as well as to extract and report their findings. Statistical meta-analysis may or may not be used . 8

transitivity The validity of logical inference is covered by transitivity, whereas the methodological feasibility of comparison is covered by similarity. For instance, if A is more effective than B, and B is more effective than C in treating the same illness, then A likely will be more effective than C, even if they were never directly compared. The transitivity requirement must be met in every case in an NMA. . 4

Acronyms

ACD appraisal consultation document. 15, 16

AFT accelerated failure time. 26

AgD aggregate data. ii, 2, 11, 15, 19, 24, 25, 28, 31, 32, 33, 34, 35, 37, 38, 40, 44, 46, 47, 53, 54, 55, 57, 106, 150, 155, 156, 177, 178, 184, 185, 196, 199, 203, 204, 206, 208, 209, 214, 215

ALM Aggregate level matching. 45, 46, 47, 203, 204, 205

CADTH Canadian Agency for Drugs and Technologies in Health. 8, 9, 11

CI confidence interval. 152, 153, 154, 163, 173, 179, 211

CrI credible interval. 186, 191, 195, 197, 209, 210

CS company submission. 15

DGM Data generating machanism. 54, 59, 61, 68, 71, 77, 88, 97, 99, 103, 105, 107, 110, 116, 119, 122, 126, 128, 134, 137, 144, 146, 148, 149, 152, 157, 160, 166, 170, 173, 177, 179, 207, 211

DREAM Dose Ranging Efficacy And Safety with Mepolizumab. 179, 180, 181, 182, 183, 184, 185, 186, 193, 196, 197, 198, 212

EAG external assessment group. 15, 19, 25, 26

EB Entropy Balancing. 32, 33, 57, 106, 156

EM Effect modifier. xi, xii, xiii, xiv, xv, xvi, xvii, 60, 67, 70, 76, 79, 86, 87, 90, 96, 109, 115, 118, 124, 127, 133, 136, 142, 143, 159, 165, 168, 175

EMA European Medical Agency. 8, 9

ESS effective sample size. 16, 19, 24, 25, 26, 33, 55, 189, 201

FAD final appraisal document. 15, 16

FDA Food and Drug Administration. 8, 9

FP fractional polynomial. 20, 21, 22, 24, 201

HPC high performance computer. 206, 211

HR hazard ratio. 8, 11, 22, 26

HTA Health technology assessment. 2, 9, 11, 13, 14, 15, 23, 24, 28, 35, 43, 46, 47, 48, 53, 151, 199, 200, 204, 206, 213

IPD individual patient data. ii, 2, 8, 10, 11, 12, 15, 16, 17, 18, 19, 20, 23, 24, 25, 26, 28, 30, 31, 32, 33, 34, 35, 37, 38, 44, 45, 46, 47, 48, 49, 53, 54, 55, 57, 103, 106, 149, 150, 151, 154, 155, 156, 177, 178, 179, 181, 184, 185, 186, 193, 196, 197, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 212, 213, 215, 216

IQWiG German Institute for Quality and Efficiency in Health Care. 8, 9, 11

ITC indirect treatment comparison. 2, 11, 12, 14, 15, 44, 48, 49

K-M Kaplan-Meier. 17, 20, 22, 23, 26

MAIC matching adjusted indirect comparison. ii, iii, xvi, xvii, xx, 12, 13, 16, 18, 19, 20, 21, 22, 23, 24, 25, 26, 32, 33, 34, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 77, 84, 88, 94, 97, 99, 102, 103, 104, 105, 106, 107, 125, 131, 134, 140, 144, 146, 148, 149, 150, 151, 152, 154, 155, 156, 157, 160, 163, 166, 170, 173, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 189, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216

MCMC Markov Chain Monte Carlo. 39, 58, 150, 157, 166, 211

ML-NMR multi level network meta-regression. 56, 178, 196, 198, 208, 215

MoM Method of Moment. 32, 33, 57, 106, 156, 189

MTC mixed treatment comparison. 3, 4, 6

NHS National Health Service. 2, 15

NICE National Institute for Health and Care Excellence. ii, 2, 8, 9, 11, 12, 13, 15, 23, 26, 34, 46, 48, 49, 151, 181, 199, 200, 201, 202, 204, 205, 211

NMA network meta-analysis. ii, iii, xvi, xvii, xx, 3, 4, 6, 7, 13, 17, 20, 21, 22, 24, 25, 35, 36, 37, 39, 40, 41, 42, 43, 44, 45, 46, 48, 49, 50, 51, 53, 54, 55, 56, 57, 58, 61, 65, 71, 74, 77, 80, 84, 88, 94, 97, 99, 102, 103, 104, 105, 106, 107, 110, 113, 116, 119, 122, 125, 128, 131, 134, 137, 140, 144, 146, 148, 149, 150, 151, 152, 154, 155, 156, 157, 160, 163, 166, 170, 173, 177, 178, 179, 180, 181, 182, 183, 185, 186, 187, 188, 189, 191, 192, 193, 194, 195, 196, 197, 198, 200, 201, 202, 203, 204, 205, 206, 207, 209, 210, 211, 212, 213, 214, 215, 216

ORR objective response rate. 8

ORs odds ratios. 2, 10, 11

OS overall survival. 8, 9, 10, 18, 26

PFS progression free survival. 9, 10, 18, 26

PH proportional hazard. 17, 26

PICO participants, intervention, comparator, and outcome framework. 44

PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses. 18, 30

PV Prognostic variable. xi, xii, xiii, xiv, xv, xvi, xvii, 61, 68, 71, 77, 80, 87, 88, 91, 96, 97, 110, 116, 119, 125, 128, 134, 137, 143, 160, 166, 169, 176

RCT Randomised Control Trial. 1, 3, 6, 8, 9, 11, 14, 19, 22, 32, 36, 37, 39, 40, 41, 42, 43, 45, 51, 53, 55, 58, 68, 102, 116, 154, 155, 179, 180, 181, 196, 199, 200, 203, 204, 206, 212, 213

RCTs randomised control trials. ii, 1, 50, 51, 53, 54, 77, 88, 125, 134, 151, 206

SD standard deviation. 52, 53, 55, 106, 152, 153, 155, 186, 209

SE standard error. 33, 57, 58, 61, 71, 80, 91, 106, 107, 110, 119, 128, 137, 149, 152, 153, 154, 156, 157, 160, 170, 186, 197, 206, 207, 208, 209

SLR systematic literature review. 28, 182

SOC standard of care. 183, 186, 188, 191, 192, 194, 195, 196, 197, 198

STA single technology appraisal. ii, xi, 12, 13, 15, 16, 18, 23, 46, 48, 49, 151, 181, 199, 200, 201, 202, 204, 205

STC simulated treatment comparison. ii, 12, 16, 18, 20, 21, 22, 23, 24, 25, 26, 34, 35, 44, 45, 47, 48, 49, 151, 196, 199, 200, 201, 202, 203, 204, 205, 206, 208, 214, 215

SVR spleen volume reduction. 18

TA technical appraisal. 9, 10, 11, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 49, 50, 55, 200, 201, 202

TSD technical support document. 12, 207

TSS total symptom score. 18