Methods Research Report

Findings of Bayesian Mixed Treatment Comparison Meta-Analyses: Comparison and Exploration Using Real-World Trial Data and Simulation

Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • **www.ahrq.gov**

Methods Research Report

Findings of Bayesian Mixed Treatment Comparison Meta-Analyses: Comparison and Exploration Using Real-World Trial Data and Simulation

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-2007-10056-I

Prepared by:
RTI-UNC Evidence-based Practice Center
Research Triangle Park, NC

Investigators:
Daniel E. Jonas, M.D., M.P.H.
Tania M. Wilkins, M.S.
Shrikant Bangdiwala, Ph.D.
Carla M. Bann, Ph.D.
Laura C. Morgan, M.A.
Kylie J. Thaler, M.D., M.P.H.
Halle R. Amick, M.S.P.H.
Gerald Gartlehner, M.D., M.P.H.

AHRQ Publication No. 13-EHC039-EF
February 2013

This report is based on research conducted by the RTI-UNC Evidence-based Practice Center, (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-2007-10056-I). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact EffectiveHealthCare@ahrq.hhs.gov.

None of the investigators have any affiliations or financial involvement that conflicts with material presented in this report.

Suggested Citation: Jonas DE, Wilkins TM, Bangdiwala S, Bann CM, Morgan LC, Thaler KJ, Amick HR, Gartlehner G. Findings of Bayesian Mixed Treatment Comparison Meta-Analyses: Comparison and Exploration Using Real-World Trial Data and Simulation. (Prepared by RTI-UNC Evidence-based Practice Center under Contract No. 290-2007-10056-I.) AHRQ Publication No. 13-EHC039-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-Based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To improve the scientific rigor of these evidence reports, AHRQ supports empiric research by the EPCs to help understand or improve complex methodologic issues in systematic reviews. These methods research projects are intended to contribute to the research base in and be used to improve the science of systematic reviews. They are not intended to be guidance to the EPC program, although may be considered by EPCs along with other scientific research when determining EPC program methods guidance.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality. The reports undergo peer review prior to their release as a final report.

We welcome comments on this Methods Research Project. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by email to epc@ahrq.hhs.gov.

Carolyn M. Clancy, M.D.
Director
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director
Evidence-based Practice Program
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Parivash Nourjah, Ph.D.
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Acknowledgments

The authors gratefully acknowledge the following individuals for their contributions to this project: We thank RTI International colleague Loraine Monroe for technical word processing.

Peer Reviewers

Brad Carlin, Ph.D.
University of Minnesota
Minneapolis, MN

Andrea Cipriani, Ph.D.
University of Verona
Verona, Italy

Christopher Schmid, Ph.D.
Tufts Evidence-based Practice Center
Boston, MA

Fujian Song, Ph.D.
University of East Anglia
Norwich, UK

Findings of Bayesian Mixed Treatment Comparison Meta-Analyses: Comparison and Exploration Using Real-World Trial Data and Simulation

Structured Abstract

Objectives. Specific objectives were to examine the following: (1a) how results of Bayesian mixed treatment comparison (MTC) methods compare with several commonly considered frequentist indirect methods; (1b) how Bayesian MTC methods perform for different evidence network patterns; (2) how meta-regression can be used with Bayesian MTC meta-analysis to explore heterogeneity; and (3) how findings of Bayesian MTC meta-analyses compare for different numbers of studies and different network pattern assumptions. For objectives 1 and 2, we aimed to conduct case studies using data from two recent comparative effectiveness reviews (CERs). For objective 3, we aimed to use simulated data.

Methods. For objectives 1 and 2, we used data from CERs that examined second-generation antidepressants (SGAs) and biologic disease-modifying antirheumatic drugs (DMARDs) for rheumatoid arthritis (RA). For objective 1, we compared results of Bayesian MTC methods with those of three frequentist indirect methods: meta-regression, the Bucher method, and logistic regression for dichotomous and continuous outcomes. For objective 2, we conducted two types of meta-regression. One explored subgroup effects with a binary covariate to assess whether efficacy of SGAs differs between older adults (≥55 years) and adults of any age. The other explored a continuous covariate to assess whether treatment efficacy varies by disease duration of RA. For objective 3, we used simulated data to examine the Bayesian MTC method's ability to produce valid results for two data scenarios when varying numbers of studies were available for each comparison for various network patterns.

Results. Bayesian MTC methods permitted the calculation of results for more comparisons of interest than frequentist meta-regression or the Bucher method (when applied as they would typically be used). When comparisons were calculated, the findings generally agreed but differed for a small proportion (less than 10%) of comparisons. Regarding precision, logistic regression produced the most precise estimates, followed by the Bayesian MTC method.
Our meta-regressions found a trend toward lesser efficacy for SGAs in older adults and a trend toward greater efficacy of biologic DMARDs for those with greater mean disease duration.
Our simulations supported the validity of Bayesian MTC methods for star and ladder network patterns but raised some concerns about one closed loop (and possibly loop) network patterns. Simulations generally found similar probabilities for which drug was the best treatment for scenarios when only 1 study was available for each comparison and those when more studies (2, 3, 5, or 10) were available; precision increased as the number of available studies increased.

Conclusions. Bayesian MTC methods offer several advantages over frequentist indirect methods, including the ability to produce results for all comparisons of interest in a single analysis. Results of Bayesian MTC methods and those of frequentist indirect methods may differ for a small proportion of comparisons, which could lead to differences in conclusions when using different methods. Our findings raise some concerns about the validity of the results of Bayesian MTC methods for certain network patterns. Further research is needed to explore

additional real-world datasets and simulated data to determine if our findings are reproducible or generalizable and to better understand the validity of Bayesian MTC methods for various scenarios.

Contents

Tables

Figures

Appendixes

Introduction

Background

Comparative effectiveness reviews (CERs) often aim to compare the benefits and harms of multiple available approaches for treating a health condition[1] with the ultimate goal of informing clinical practice and other decisionmaking. To this end, analysts conducting CERs aim to find studies conducting direct head-to-head comparisons. However, direct head-to-head evidence on competing interventions is often scant. As a result, several methods to conduct indirect comparisons have been proposed.[2-7] These include meta-regression, logistic regression, the Bucher method,[2-4] and, more recently, Bayesian mixed treatment comparison (MTC) meta-analysis.[5-7]

MTC meta-analysis is a relatively new methodology.[8] Various other terms have been used to describe the approach, including multiple treatment comparisons[9-11] and network meta-analysis.[12, 13] Terminology has evolved to where most experts in that field now refer to the broad area of comparison of different treatments as network meta-analysis and restrict the use of MTC to describe methods that explicitly look at combining direct and indirect evidence.[14] One of the most compelling reasons to use MTC meta-analysis is that it allows for the combination of both direct head-to-head and indirect evidence (e.g., placebo-controlled trials) in one modeling framework. The use of all potentially relevant available evidence is an appealing feature for analysts, because other methods rely solely on one type of evidence. In addition, unlike other indirect analysis methods, MTC meta-analysis allows all relevant comparisons to be made through a single analysis, providing the information to calculate an effect size for each comparison of interest and to rank treatments based on the probability of being the best treatment.

The history of MTC meta-analysis dates back to 1996 when Higgins and Whitehead first described likelihood-based methods for indirect comparisons, focusing on Bayesian methods and providing an illustration of the methodology using data from 26 clinical trials that investigated the prevention of cirrhosis using beta-blockers and sclerotherapy.[15] Lu and Ades subsequently published additional information on the theoretical underpinnings of the MTC method,[8] as well as methods for assessing evidence inconsistency[16] (which is not addressed in this report). But more research is needed to better understand how these methods operate in real-world scenarios. As investigators conducting systematic reviews use these methods and continue to develop new techniques based on these methods, it is important for us to have a better understanding of how these methods compare with other indirect methods and how MTC meta-analyses perform in various situations.

Relatively little information on the validity of MTC meta-analysis exists (in comparison with other indirect methods or for various types of evidence networks), and further research is needed.[12, 17] Some analysts have validated frequentist approaches for indirect comparisons using artificial/simulated data.[3]

One underlying question is whether the extra complexity of Bayesian MTC meta-analyses is worth the investment. MTC meta-analysis requires greater statistical expertise and, until recently, could only be run using programs (e.g., WinBUGS) that are unfamiliar to many analysts. (Recently, Higgins et al. published an example of likelihood-based MTC using Stata).[18] In contrast, several of the frequentist methods for indirect comparisons can be conducted by analysts with less statistical expertise and using programs that are more familiar. It is important to gain a better understanding of the consequences of choosing various analytic approaches.

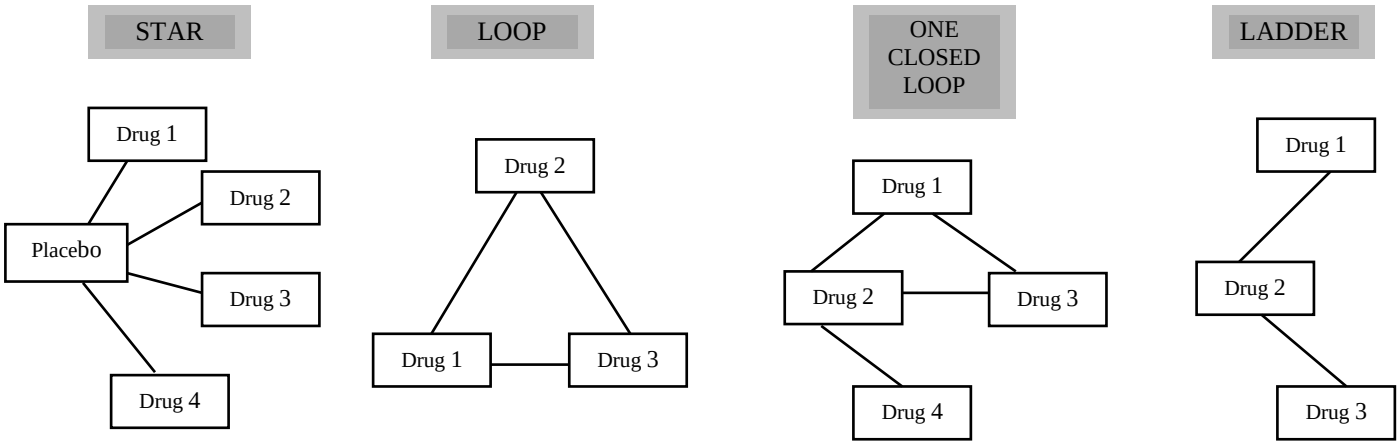
Some questions include the following: Do results differ for various analytic approaches? Do certain methods yield more precise estimates? Do the findings and validity of various analytic approaches vary depending on the evidence network pattern?

We used MTC meta-analyses in two recent CERs for the Agency for Healthcare Research and Quality (AHRQ)—one on second-generation antidepressants (SGAs)[19] and one on treatments for rheumatoid arthritis (RA).[20] In this report, we use the real-world literature from these two reports to address our objectives and Key Questions (KQs).

Evidence Networks

An evidence network refers to the linkage of treatment comparisons that exists in the literature for a given population. It can take on many shapes or patterns.[21, 22] The basic premise underlying MTC methods is that the network must be a connected one. A network can include any of the four patterns in Figure 1. A network pattern can resemble a star, with one common comparator at the center and other treatments connected through this comparator. This is a common scenario for pharmacotherapies, because randomized controlled trials (RCTs) often include only drug comparisons with placebo (and no direct head-to-head comparisons). Another pattern is a loop design, where all drugs or treatments are connected to one another through one other treatment. A third common pattern is a variation on this loop design, but this network also includes one or more drugs or treatments outside of the loop and is referred to as one closed loop. Another network pattern resembles a ladder, with no treatment compared with any other treatment more than once.

Figure 1. Evidence network patterns



Second-Generation Antidepressants

Our report on SGAs compared the benefits and harms of 13 SGAs approved for use in the United States[19] (Table 1).

Table 1. Second-generation antidepressants approved for use in the United States

Generic Name	U.S. Trade Name[a]	Dosage Forms	Therapeutic Classification
Bupropion[b]	Wellbutrin®; Wellbutrin SR®; Wellbutrin XL®	75, 100 mg tabs; 100, 150, 200 mg SR tabs 150, 300 mg XL tabs	Other
Citalopram[b]	Celexa®	10, 20, 40 mg tabs; 2 mg/ml solution	SSRI
Desvenlafaxine	Pristiq®	50, 100 mg tabs	SNRI
Duloxetine	Cymbalta®	20, 30, 60 mg caps	SSNRI
Escitalopram	Lexapro®	5, 10, 20 mg tabs 1 mg/ml solution	SSRI
Fluoxetine[b]	Prozac®; Prozac Weekly®	10, 20, 40 mg caps; 4 mg/ml solution 90 mg caps	SSRI
Fluvoxamine[b]	Luvox®	25, 50, 100 mg tabs	SSRI
Mirtazapine[c]	Remer on® Remer on Sol tab®	15, 30, 45 mg tabs; 15, 30, 45 mg orally Disintegrating tabs	SNRIc
Nefazodone[b]	Serzone®[d]	50, 100, 150, 200, 250 mg tabs	Other
Paroxetine[b]	Paxil®; Paxil CR®	10, 20, 30, 40 mg tabs; 2 mg/ml solution; 12.5, 25, 37.5 mg CR tabs	SSRI
Sertraline[b]	Zoloft®	25, 50, 100 mg tabs; 20 mg/ml solution	SSRI
Trazodone[b]	Desyrel®	50, 100, 150, 300 mg tabs	Other
Venlafaxine[b]	Effexor®; Effexor XR®	25, 37.5, 50, 75, 100 mg tabs; 37.5, 75, 150 mg XR caps	SNRI

caps = capsules; mg = milligram; ml = milliliter; SNRI = serotonin and norepinephrine reuptake inhibitor; SSRI = selective serotonin reuptake inhibitor; tabs = tablets

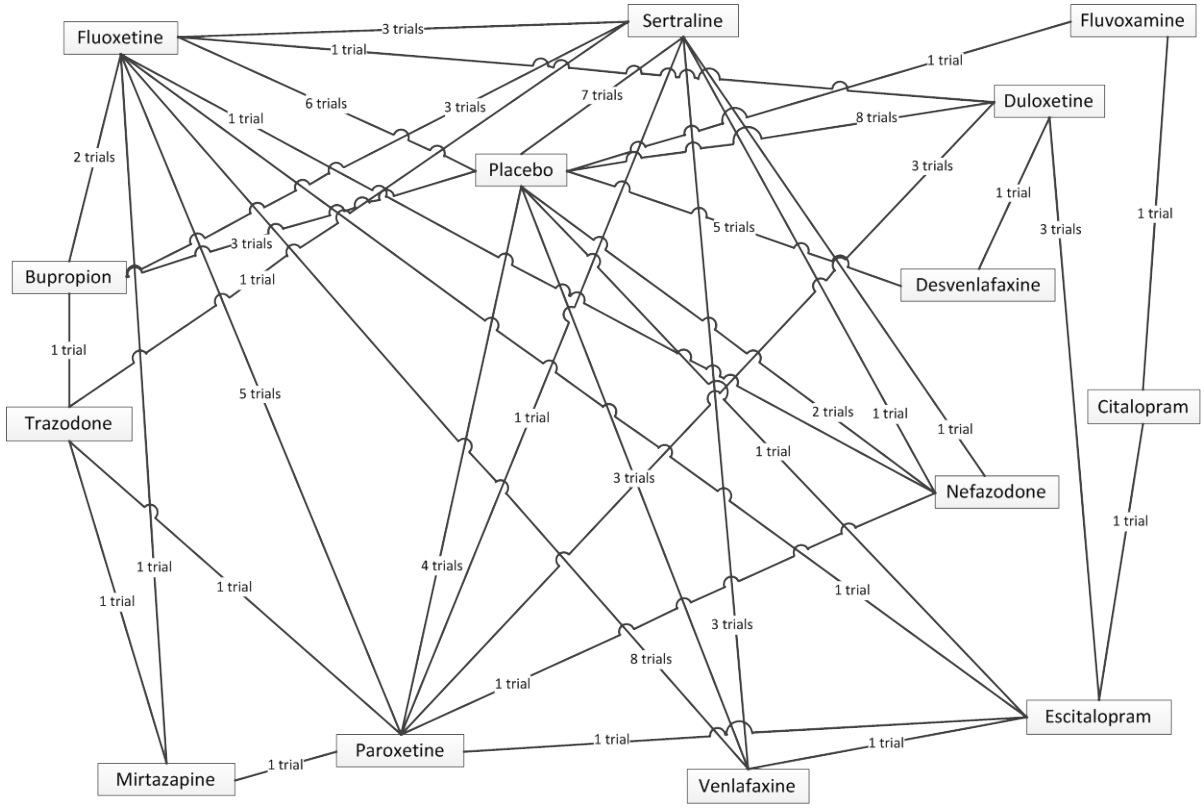
[a]CR, SR, XL, and XR are registered trademarks referring to controlled, sustained, or extended-release dosage forms, respectively.

[b]Generic available for some dosage forms.

[c]Mirtazapine's mechanism of action is not clearly an SNRI, but it was grouped in this class owing to similarities.

[d]Only generic nefazodone is available in the United States.In our recent CER, we conducted MTC meta-analysis to derive estimates of the comparative efficacy among all SGAs for the treatment of major depressive disorder. Our primary efficacy outcome was the rate of response on the Hamilton Depression rating scale (HAM-D), defined as a 50 percent or greater improvement of scores from baseline. Figure 2 shows the evidence network that contributed data to the analysis.

Figure 2. Evidence network for mixed treatment comparison meta-analysis of second-generation antidepressants for achieving response for adults with depression



Our MTC meta-analysis found that SGAs had similar efficacy. There were some differences (based on the 95% credible intervals) for some pairwise comparisons that are likely not clinically relevant.

Biologic Disease-Modifying Antirheumatic Drugs for Rheumatoid Arthritis

Our report on treatments for RA compared the benefits and harms of corticosteroids and oral and biologic disease-modifying antirheumatic drugs (DMARDs) for adults with RA.[20] Nine biologic DMARDs were included in the CER (Table 2).

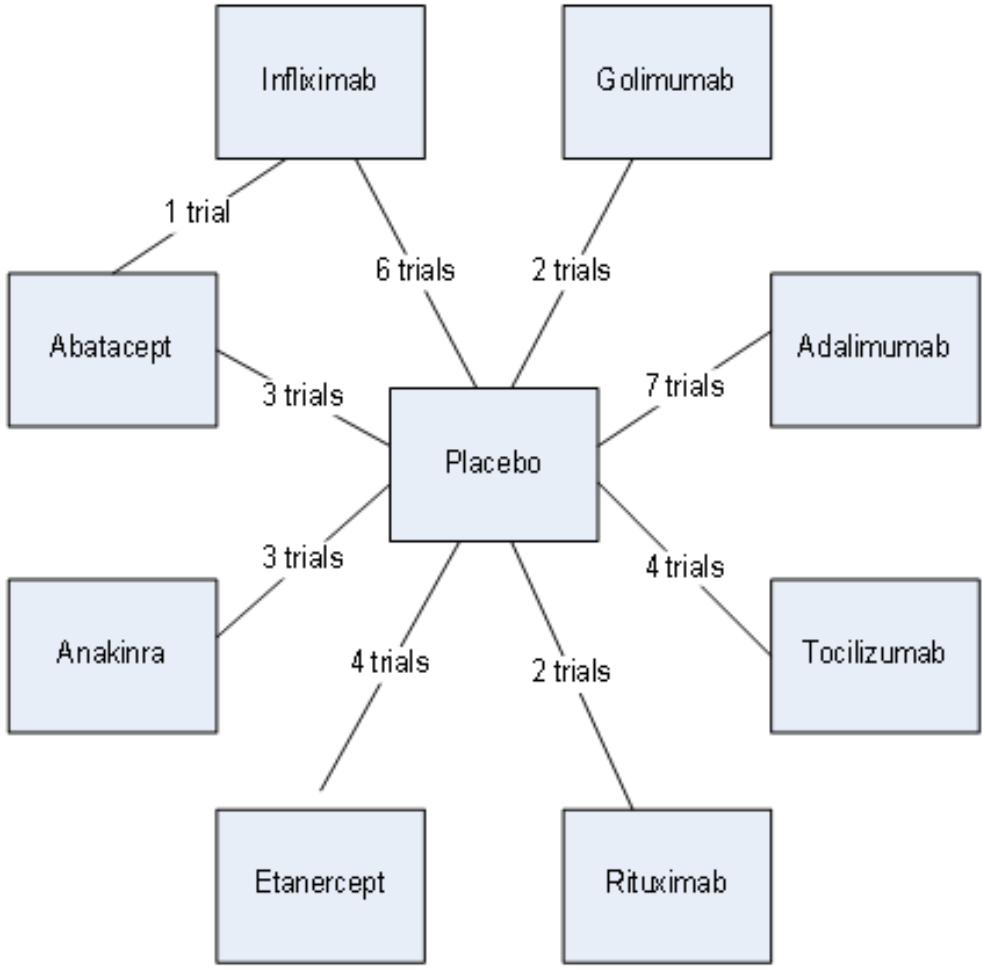
Table 2. Biologic DMARD treatments for rheumatoid arthritis

Generic Name	Manufacturer U.S. Trade Name(s)*	Injectable Supply	Usual Adult Dose
Abatacept	Bristol Myers Squibb Orencia®	250 mg powder in single-use vial, 125 mg/ml solution in a prefilled syringe	IV—Dosed according to body weight (< 60 kg = 500 mg; 60-100 kg = 750 mg; > 100 kg = 1,000 mg); dose repeated at 2 weeks and 4 weeks after initial dose, and every 4 weeks thereafter SQ—After a single IV infusion as a loading dose (as per body weight categories above), 125 mg should be given within a day, followed by 125 mg once a week
Adalimumab	Abbott Humira®	40 mg/0.8 ml prefilled pen or syringe, 20 mg/0.4 ml prefilled syringe	SQ—40 mg every other week; may increase to 40 mg every week in patients not taking concomitant MTX
Anakinra	Amgen Kineret®	100 mg/0.67 ml syringe	SQ—100 mg/day; dose should be decreased to 100 mg every other day in renal insufficiency or end-stage renal disease
Certolizumab Pegol	UCB Cimzia®	200 mg powder for reconstitution, 200 mg/ml solution in a pre-filled syringe	SQ—Initial dose of 400 mg, repeat dose 2 and 4 weeks after initial dose, followed by 200 mg every other week; for maintenance dosing, consider 400 mg every 4 weeks
Etanercept	Amgen Pfizer Immunex Enbrel®	50 mg/ml autoinjector or prefilled syringe, or as two 25 mg/0.5 mL single-use prefilled syringes or free-hand vials	SQ— 50 mg once weekly with or without MTX
Golimumab	Centocor Ortho Biotech Janssen Biotech, Inc. Simponi®	50 mg/0.5 ml prefilled syringe or autoinjector	SQ—50 mg once a month, in combination with MTX
Infliximab	Centocor Ortho Biotech Remicade®	100 mg lyophilized in a 20 ml vial	IV—3 mg/kg in combination with MTX at 0, 2, and 6 weeks followed by maintenance every 8 weeks thereafter; may increase to maximum of 10 mg/kg or treat as often as every 4 weeks
Rituximab	Biogen Idec/Genentech Rituxan®	100 mg/10 ml and 500 mg/50 ml vial	IV—In combination with MTX, two 1,000 mg IV infusions separated by 2 weeks (one course) every 24 weeks or based on clinical evaluation, but not sooner than every 16 weeks
Tocilizumab	Genentech/Roche Actemra®, RoActemra®	80 mg/4 ml, 200 mg/10 ml, 400 mg/20 ml vial	IV—4 mg/kg followed by an increase to 8 mg/kg based on clinical response; given every 4 weeks with or without MTX

IV = intravenous; kg = kilogram; mg = milligram; ml = milliliter; MTX = methotrexate; SQ = subcutaneous

*Listed trade names are limited to commonly prescribed U.S. products when multiple trade names are available.To compare the effectiveness of biologic DMARDs with each other, we conducted MTC meta-analyses of trials enrolling methotrexate-resistant patients with active RA. The primary efficacy outcome of our MTC meta-analysis was the American College of Rheumatology 50 percent response (ACR 50). We also conducted analyses using ACR 20 and ACR 70. Figure 3 shows the evidence network that contributed data to the analysis. The majority of evidence was from placebo-controlled trials; only one trial included a direct head-to-head comparison. Due to heterogeneity in study design of the included studies for certolizumab, it was excluded from the MTC meta-analysis. More information can be found in the full report.[20]

Figure 3. Evidence network for mixed treatment comparison meta-analysis of biologic DMARDs for achieving ACR 50 for adults with rheumatoid arthritis



Note: The total number of trials does not appear to equal 30 (the total number of studies included in the analysis) because some trials have multiple arms that were included.

Our MTC meta-analyses found higher odds of achieving ACR 50 response for etanercept compared with most other biologic DMARDs (abatacept, adalimumab, anakinra, infliximab, rituximab, tocilizumab) for methotrexate-resistant patients with active RA. The ACR 50 odds ratio range for etanercept compared with most other biologic DMARDs was 2.4 to 5.2. The differences were potentially important (based on 95% credible intervals) for etanercept compared with abatacept, adalimumab, anakinra, infliximab, rituximab, or tocilizumab, but were not when compared with golimumab. Anakinra had the lowest mean response, and point estimates favored other biologic DMARDs over anakinra, but the differences were only potentially important (based on 95% credible intervals) when compared with adalimumab and etanercept for ACR 50.

Scope and Key Questions

The main objectives of this report are to contribute to the body of literature on MTC meta-analysis by examining (1a) how results of Bayesian MTC methods compare with several frequentist indirect methods for various types of outcome measures, (1b) how Bayesian MTC

methods perform for different types of evidence network patterns, (2) how study-level covariates can be incorporated with Bayesian MTC meta-analysis to explore heterogeneity through meta-regression, and (3) how findings of Bayesian MTC meta-analysis compare for different numbers of studies and different network pattern assumptions. For objectives 1 and 2, we aimed to conduct case studies using data from two recent CERs. For objective 3, we aimed to use simulated data. We address the KQs listed below.

KQ 1. How do the results of Bayesian MTC meta-analysis methods compare with those of several frequentist indirect methods? Related questions of interest included the following: For each of the common evidence network patterns, how do the Bayesian MTC methods compare with frequentist indirect methods? How do Bayesian MTC methods perform (e.g., precision, convergence) for different types of evidence network patterns?

KQ 2. How can meta-regression be used with Bayesian MTC meta-analysis to explore sources of heterogeneity?

KQ 3. How do findings of Bayesian MTC meta-analysis compare for different numbers of studies and network pattern assumptions?

For KQ 1, our choice of frequentist analytic methods to compare with the Bayesian MTC approach was based on our judgment regarding the methods most commonly considered by analysts conducting CERs. In addition, we selected frequentist methods with some evidence to support their validity. These included frequentist meta-regression, the Bucher method, and frequentist logistic regression. Of note, the frequentist methods used are not the analogue of the Bayesian methods implemented. We did not compare findings with the frequentist network meta-analysis method (i.e., the Lumley method).[23] Our experience indicates that it is much more rarely used than the other methods, and comparisons between Bayesian MTC methods and frequentist network meta-analysis were not our intention.

We applied Bayesian MTC methods to a variety of different evidence networks using data from recent systematic reviews (of SGAs and biologic DMARDs) and using simulated data. For the first two KQs, we used the real-world bodies of literature described above (for SGAs and biologic DMARDs). For KQ 3, we use simulated data sets.

For KQ 2, we focused on clinically important issues to guide the meta-regressions. For the SGAs literature, some have questioned whether the medications are equally or less effective in older adults.[24-27] To address this question, we conducted meta-regression by assessing whether efficacy differed in trials that enrolled older adults compared with trials that enrolled adults of any age. For biologic DMARDs, some have questioned whether treatment efficacy varies by disease duration of RA.[28-30] To address this question, we conducted meta-regression using mean disease duration of subjects enrolled in each study as a continuous covariate.

Methods

In this chapter, we first describe the data from two real-world bodies of trial literature that we used for our analyses for Key Questions (KQs) 1 and 2. We then describe the methods used for each KQ. The focus of this report is on mixed treatment comparison (MTC) meta-analysis implemented in a Bayesian framework; for comparison, we performed adjusted indirect comparisons using four different approaches with a frequentist framework (each described below).

The National Institute for Health and Clinical Excellence (NICE) Decision Support Unit has released several Technical Support Documents (TSDs) detailing use of MTC in a Bayesian framework, including several illustrative examples with annotated WinBUGS code. We implemented the methods illustrated in TSD 2[6] and TSD 3[5] under different scenarios for the MTC meta-analyses described below. For both of our recent comparative effectiveness reviews (CERs) including MTC meta-analyses, we used WinBUGS Version 1.4.3, a Bayesian software package that uses Markov chain Monte Carlo (MCMC) techniques. WinBUGS code used for analyses is available in Appendix A.

Data Included in This Report

For this report, as described in the introduction, we used datasets from MTC meta-analyses from two recent CERs. An underlying assumption for the validity of MTC meta-analyses is transitivity (sometimes referred to as *similarity*).[31] Assessing whether the included studies had sufficiently comparable compositions was undertaken in the original CERs rather than for this report. Briefly, to assess similarity and to determine whether to combine studies in MTC meta-analyses for the CERs, we evaluated the populations, interventions, comparators, outcomes, timing, and settings of the trials under consideration.

Dataset 1. Second-Generation Antidepressants

For this report, we extracted one binary and one continuous outcome from the evidence base of our CER on second-generation antidepressants (SGAs). Our binary outcome was treatment response as measured by at least a 50 percent improvement from baseline on the Hamilton Rating Scale for Depression (HAM-D). We recalculated response rates for each study using the number of all randomized patients as the denominator to reflect a true ITT analysis. With this approach we attempted to correct variations in results of modified ITT analyses encountered in individual studies. A total of 64 studies with adequate reporting of treatment response were included in subsequent analyses. Eight additional studies were identified in the older adult populations (age 55 or older) and used in the meta-regression for KQ 2. Characteristics of the included studies and data used in the meta-analyses are listed in Appendix Table B-1. The dataset includes many multi-arm trials (Appendix Table B-1). Additional descriptions of study populations and other eligibility criteria can be found in the full report.[19]

Our continuous outcome was mean change from baseline to endpoint on the HAM-D. For studies not reporting a variance for mean change from baseline, we calculated one using the baseline and endpoint variances and assumed a correlation of 0.5. We chose a value of 0.5 for the correlation coefficient as a reasonable assumption for the similarity of baseline and endpoint values across patients, because it assumes neither a weak nor strong correlation. In lieu of more complete reporting, it has been suggested that a value of 0.5 is a reasonable assumption.[32, 33] We included a total of 40 studies in the analyses for this outcome; data and characteristics of the

included studies are listed in Appendix Table B-2. The dataset includes several multi-arm trials (Appendix Table B-2).

Dataset 2. Biologic Disease-Modifying Antirheumatic Drugs for Treatment of Rheumatoid Arthritis

We extracted one binary and one continuous outcome from the studies included in our CER on biologic DMARDs and other treatments for RA. The binary outcome considered was treatment response as measured by achievement of ACR 50 after 12 weeks of treatment. Again, we recalculated response rates for each study using the number of all randomized patients as the denominator to reflect a true ITT analysis. With this approach we attempted to correct variations in results of modified ITT analyses encountered in individual studies. A total of 31 studies covering eight biologic DMARDs with adequate reporting of treatment response were included. Characteristics and data of the included studies are shown in Appendix Table B-3 below. The dataset includes one multi-arm trial (Appendix Table B-3).

We also extracted a continuous outcome, mean change from baseline in Health Assessment Questionnaire Disability Index (HAQ-DI); however, because few eligible studies reported adequate data, we were not able to perform any meaningful MTC meta-analysis on the continuous outcome.

KQ 1. Comparison of Bayesian MTC Meta-Analysis With Frequentist Indirect Methods

Our objectives for this KQ were to examine how results of Bayesian MTC methods compare with commonly used frequentist indirect methods and how Bayesian MTC methods perform for different types of evidence network patterns. We chose four analytic methods—Bayesian MTC meta-analysis, frequentist meta-regression, the Bucher method, and frequentist logistic regression. These methods were chosen because they are among the most common approaches considered when conducting CERs. For both of our datasets (SGAs and biologic DMARDs), we compared the findings from these four methods—first, for the full networks; second, for subcomponents of the full networks representing specific evidence network patterns (star, loop, one closed loop, and ladder). To compare the four analytic methods, we used several measures: (1) the proportion of drug-drug comparisons (out of the total possible number of comparisons) for which each method was unable to calculate a result, either because of model convergence issues or the lack of a common comparator; (2) the percent agreement, with findings considered to agree if both methods produced a non-statistically significant (for frequentist methods) or unimportant (for Bayesian methods, based on 95% credible intervals) result for the comparison or if both analyses found a statistically significant or important result favoring the same treatment; (3) the precision of findings—assessed by comparing the width of credible intervals and confidence intervals; and (4) kappa statistics. The kappa statistic is a measure of inter-rater agreement[34] that attempts to take into account agreement beyond chance. It can range between -1 and 1, but usually ranges from zero to 1, as kappa is negative when the observed agreement is less than chance. When the observed agreement exceeds chance agreement, kappa is positive, with its magnitude reflecting the strength of agreement. Landis and Koch propose the following as guidelines for strength of agreement for the kappa coefficient: <0.20 = poor, 0.21 to 0.40 = fair, 0.41 to 0.60 = moderate, 0.61 to 0.80 = good, and 0.81 to 1.00 = very good.[35] For each comparison (e.g., Bayesian MTC vs. frequentist meta-regression), we calculated kappa statistics

using SAS version 9.2. We did not calculate kappa statistics for comparisons in which either method had less than two levels (i.e., when all results for that method found no statistically significant or important difference for each drug-drug comparison).

Finally, we describe measures of model fit for the Bayesian MTC analyses for the full networks and for subcomponents. For all dichotomous data, we used odds ratios, and for all continuous data, we used weighted mean differences as outcome measures.

Bayesian Mixed Treatment Comparison Meta-Analysis

For all Bayesian MTC meta-analyses included in this report, we used the methods developed in TSDs 2 and 3, which detail use of the generalized linear modeling (GLM) framework for Bayesian MTC and use of meta-regression to explore sources of heterogeneity, respectively. We used random effects models in this report because in meta-analyses of randomized controlled trials, we assume the study effect is sampled from a distribution of effect sizes.[36] Because studies will not include exactly the same mix of participants or carry out the interventions in an identical way, there may be different underlying effect sizes for different studies.

GLM theory[37] allows for likelihood-based statistical inference. It is also a flexible approach because it can be constructed to model data arising from a large range of distributions within the exponential family, allowing us to model various binary and continuous outcomes that we would encounter in conducting CERs. GLM theory can be implemented in both frequentist and Bayesian frameworks. Spiegelhalter et al.[38] make the case for a Bayesian framework, stating that such an approach is more flexible, efficient, and useful. MTC could in theory be implemented in a frequentist framework, but the literature reflects a preference for a Bayesian framework, in part because of the availability of WinBUGS code that has been developed for this purpose.[6, 39] Reasons to use a Bayesian framework (rather than a frequentist approach) include inferential superiority and modeling flexibility. Whether Bayesian approaches have inferential superiority is a contentious subject and may have little to do with the reason the Bayesian MTC methods are used more widely. Modeling flexibility, however, is likely a major reason the Bayesian approach is more popular. With WinBUGS, one has simply to specify the likelihood and the prior distributions; the estimation is handled by MCMC methods that are now fairly robust. With maximum likelihood estimation using a frequentist approach, some of the likelihoods can be very challenging to maximize and perhaps amenable to approximate solutions only.

In a Bayesian framework, specification of prior distributions accompanies the specification of the likelihoods, in addition to the data. The complete model specification is detailed below. For this report, binary outcomes are treatment response measured by the number of people achieving a prespecified level in each arm. We used noninformative (flat) prior distributions throughout the report, which allow the data to drive the posterior distributions, in absence of more informative priors. This choice follows the recommendations laid out in TSD 2[6] and in absence of rationale for specifying informative priors with these data. For all analyses, we modeled study effect and treatment effect parameters by noninformative (flat) prior distributions that were Normal (0, 10000). For the heterogeneity of the random-effects model, we used a uniform prior distribution with sufficiently large variance. Unless otherwise specified, we discarded the first 20,000 simulations to allow for model convergence, and we used an additional 100,000 simulations in estimating the posterior probabilities. Satisfactory convergence was verified by trace plots, monitoring the Monte Carlo error, and with Gelman-Rubin diagnostics.[40] If multi-arm studies were available, we included the appropriate adjustments to the likelihood to

account for correlations between the treatment differences. We did not use statistical methods to assess inconsistency.

For the initial values to start the Markov chain simulations, we chose values that were relatively widely dispersed. We reviewed previous publications and TSDs 2 and 3 to help inform our choice of initial values. To help determine whether our choice of initial values influenced findings, we conducted some sensitivity analyses for our MTC analyses of the full networks by changing the initial values (e.g., for our ACR 50 MTC analysis, we [1] changed the initial values from chains of 0 and −3 to chains of 1 and −3 and [2] added chains with a variety of values ranging from −5 to 7). The sensitivity analyses did not produce appreciably different results.

Bayesian Mixed Treatment Comparison: Model Specification

Most of the data used in this report consisted of outcomes that measured the number of events, in all cases treatment response, occurring out of the total number of patients. These types of data were assumed to arise from a binomial likelihood. The likelihood was specified as

$$r_{i,k} \sim Binomial(p_{i,k}, n_{i,k})$$

where $r_{i,k}$ is the number achieving treatment response out of the total number of patients in each arm and $p_{i,k}$ is the probability of treatment response, and where i denotes trial number and k denotes each arm of the trial. For this likelihood, the logit link function was used to map the probability of treatment response onto plus and minus infinity. The logit link was specified as follows:

$$\text{logit}(p_{i,k}) = \mu_i + \delta_{i,k} I_{(k \neq 1)}$$

The trial-specific log-odds ratios from a random effect model come from a common distribution, $\delta_{i,k} \sim \text{N}(d_k, \sigma^2)$, where d represents the relative treatment effect and σ^2 is the common variance term.

In all cases, parameters were given vague or noninformative prior distributions. The choice of the uniform prior between 0 and 5 reflects a between-study variability that allows for a wide range of treatment effects.

Prior specification:
$\mu_i \sim \text{Normal}\ (0, 10000)$
$d_k \sim \text{Normal}\ (0, 10000)$
$\sigma \sim \text{Uniform}\ (0, 5)$

For continuous outcomes, the data were assumed to arise from a normal likelihood, and under generalized linear modeling theory, the model specified above becomes a normal likelihood with the identity link function. The observed data $y_{i,k}$ arise from a normal distribution with mean $\theta_{i,k}$ and variance $se_{i,k}^2$.

$$y_{i,k} = N(\theta_{i,k}, se_{i,k}^2)$$

For these data, the prior specifications were identical to the ones used with the dichotomous data.

For KQ 2, in order to introduce a dichotomous study-level covariate to the logistic regression model above, the model was simply extended to include a covariate x_i. The model specification and resulting WinBUGS code were taken from the Decision Support Unit's TSD 3. Again, the subscript i denotes trials and the subscript k refers to the arms within a trial. The full model structure is specified below.

$$r_{i,k} \sim Binomial(p_{i,k}, n_{i,k})$$
$$\text{logit}(p_{i,k}) = \mu_i + \delta_{i,k} I_{(k \neq 1)} + (\beta_k - \beta_1) x_i$$
$$\beta_1 = 0$$

Prior specification:
In addition to the above, now a vague normal prior on the common covariate effect

$$\beta_k \sim \text{Normal}\ (0, 10000)$$

Again, the extension to introducing a continuous covariate results in the same model above, but to improve the efficiency of estimation, the continuous covariate was centered around its mean, as shown below. All other pieces of the model specification remained the same. Changes to the logistic model are shown below.

$$\text{logit}(p_{i,k}) = \mu_i + \delta_{i,k} I_{(k \neq 1)} + (\beta_k - \beta_1)(x_i - mean(x_i))$$

Model Statistics

When conducting the Bayesian MTC meta-analyses, we output several statistics to compare relative efficacy and assess model fit, as described below.

Outcome Measures

We assessed relative efficacy between treatments with odds ratios and 95 percent credible intervals for dichotomous data, and with mean differences and 95 percent credible intervals for continuous data. We also calculated the probability that each treatment was the best, by ranking the drugs on a relative scale. Because we used random effects models, we also output the estimate of the between-studies standard deviation in each scenario.

Model Fit

We assessed model fit with the Deviance Information Criterion (DIC) and the posterior mean of the total residual deviance.[41] Deviance measures the fit of the model to the data using the likelihood function. A good model fit is indicated by a total residual deviance approximately equal to the number of data points available. The DIC is a statistic that measures Bayesian model fit and penalizes the deviance by the model complexity. When comparing two DIC values, a difference of 5 or more is regarded as a meaningful difference.[42]

Frequentist Indirect Comparisons

We compared the results from Bayesian MTC meta-analyses with three frequentist indirect methods for binary data and two frequentist indirect methods for continuous data. We used only two for continuous data because logistic regression cannot be used with continuous data. We

chose these methods because they are among the most common indirect analyses currently used in CERs and because they have been validated with data simulation.[3]

Bucher Method

For one indirect comparisons approach, we used methods proposed by Bucher and colleagues.[2] To derive indirect comparisons of two treatments, the method compares the magnitude of treatment effects of two interventions relative to a common comparator. For binary data, we calculated the pooled odds ratios for each drug of interest relative to a common comparator using random effects meta-analyses (as proposed by DerSimonian and Laird). For continuous data we used weighted mean differences.

Specifically, as described by Bucher and colleagues,[2] to compare treatment A versus B, we first estimated the pooled odds ratios for treatment A versus C and treatment B vs. C with:

$$OR_{AC} = \frac{\frac{p_A}{1-p_A}}{\frac{p_C}{1-p_C}}$$

$$OR_{BC} = \frac{\frac{p_B}{1-p_B}}{\frac{p_C}{1-p_C}}$$

where p is the probability of response to treatment (i.e., p_A is the probability of response given treatment A). The odds ratios for the indirect comparison of treatments A versus B may then be estimated by taking the ratio of the two odds ratios:

$$OR_{AB} = \frac{OR_{AC}}{OR_{BC}}$$

The indirect comparison may be calculated as follows:

$$\ln(OR_{AB}) = \ln(OR_{AC}) - \ln(OR_{BC})$$

Given the odds ratios (OR_{AC} and OR_{BC}) are estimated based on different studies and therefore are independent, the variance for this effect is the pooled variance:

$$Variance(\ln(OR_{AB})) = Variance(\ln(OR_{AC})) + Variance(\ln(OR_{BC}))$$

For continuous outcomes we used the same analytic approach without conversion to a logarithmic scale. The effect measure of choice for continuous outcomes was the weighted mean difference of the treatments relative to a common comparator. We conducted all analyses with Comprehensive MetaAnalysis version 2.2.050 (Biostat, Englewood NJ).

Frequentist Meta-Regression

For the frequentist meta-regression approach, we followed a two-step process to indirectly estimate the comparative efficacy of two drugs. We first conducted random effects meta-

analyses as proposed by DerSimonian and Laird,[4] combining two drugs of interest to estimate an overall pooled effect.

Second, we estimated the comparative treatment effects between two drugs with random effects meta-regression. We implemented this method using the Stata "metareg" command (Harbord & Higgins, 2008).[43] Stata version 11 was used for all frequentist meta-regression analyses. The unit of analysis for meta-regression was a study. The predictor in the regression model was a binary variable indicating the presence or absence of the drug of interest. Random effects meta-regression extends random effects meta-analysis by replacing the mean with a linear predictor. The response to treatment given study i is:

$$y_i = x_i\beta + u_i + \varepsilon_i$$

where i = study,
y_i = response to treatment in study i,
x_i = study-level covariate,
β = fixed effect parameter,
u_i = treatment effect in study i, and
ε_i = error.

For dichotomous outcomes, y_i is the log odds ratio and for continuous outcomes, y_i is the weighted mean difference.

The treatment effects may vary across studies and assumed to be normally distributed where τ^2 is the between study variance:

$$u_i \sim N(0, \tau^2)$$

Errors are also assumed to be normally distributed:

$$\varepsilon_i \sim N(0, \sigma^2)$$

Logistic Regression Modeling

The logistic regression modeling approach is similar to the meta-regression approach; however, an individual rather than a study is the unit of analysis. For the indirect comparisons using logistic regression, we incorporated the approach outlined by Glenny et al. (2005).[3] We constructed datasets with the number of participants responding to the treatment and the number of participants not responding to the treatment matching those reported in each study. In other words, if a study reported that 100 of 300 patients responded to treatment A and 150 of 300 patients responded to treatment B, we constructed a dataset with a total of 600 patients (100 responders and 200 nonresponders receiving treatment A and 150 responders and 150 nonresponders receiving treatment B). We then estimated logistic regression models using PROC GLIMMIX in SAS version 9.2 with a dichotomous outcome of response versus no response to treatment and including treatment as a fixed effect and study as a random effect.

PROC GLIMMIX is used for computing generalized linear mixed effect models and can be applied to various types of outcome variables. The general form of the model is:

$$E(Y|\gamma) = g^{-1}(X\beta + Z\gamma)$$

where Y = outcome variable,
X = matrix of fixed effects,
β = vector of fixed effect parameters,
γ = matrix of random effect parameters,
Z = matrix of random effect, and
g = link function.

In this model, E(Y| γ) represents the expected value for the outcome based on the model. In our analyses, we used a mixed effects model where the outcome (Y) was response to treatment (yes vs. no) with treatment as a fixed effect (X) and study as a random effect (Z) as shown below:

$$E(Response\ to\ treatment|\gamma) = g^{-1}(\beta_0 + \beta_1 Treatment_1 \ldots + \beta_j Treatment_j + \gamma_1 Study_1 \ldots . + \gamma_i Study_i)$$

where i is the number of studies and j is the number of treatments.

Because the outcome is dichotomous, we selected a logit link function and when applied to the equation, the final model is:

$$E(Response\ to\ treatment|\gamma) = \frac{1}{1 + e^{-(\beta_0 + \beta_0 + \beta_1 Treatment_1 \ldots + \beta_j Treatment_j + \gamma_1 Study_1 \ldots . + \gamma_i Study_i)}}$$

A binomial variance function was specified along with an unstructured covariance matrix that permits the estimation of all elements of the variance/covariance matrix without constraints. The indirect comparisons were estimated within these models by applying contrast statements to compute odds ratios comparing response between two treatments of interest.

Handling of Multi-Arm Studies in Frequentist Analyses

None of the frequentist indirect methods used in this report accounted for correlations between treatment responses within multi-arm studies. This is an advantage of the Bayesian MTC methods when multi-arm studies are included, as the correction to the likelihood is easily implemented in the available code. However, in most cases this was not an issue because multi-arm studies were generally not included when conducting the indirect frequentist analyses.

Choice of Network Patterns

As part of KQs 1 and 2, we aimed to investigate how Bayesian MTC methods perform for different network patterns. An evidence network's geometry may be shaped by clinical context, regulatory pressure, or other factors. The evidence network may change over time, as new drugs are added to the market or different comparators are chosen.[21] We chose four simple network patterns as the basis for comparisons within this report: star, loop, one closed loop, and ladder (see Introduction). These four patterns reflect different scenarios that exist in the context of real-world data. Within each of the real-world datasets used, we selected a subset of the studies for each of the four patterns. Although many different subsets of data were available for each pattern that we could have selected, we selected examples that maximized the amount of data available for each sub-network. We selected one of each type of network pattern for each of the datasets.

Because of limited time and resources, we could not evaluate every possible network pattern within each dataset. Appendix C illustrates the network patterns evaluated. Other choices could have been made, such as one reflecting the entry of drugs into the market or chronologically by publication date, but for the purposes of this report, we believe that an approach maximizing the amount of data available is the most useful.

KQ 2. Meta-Regression

Our objective for this KQ was to introduce study-level covariates in the models used with Bayesian MTC meta-analysis to explore heterogeneity, attempting to answer two clinically important questions—one for each of our datasets. For the SGAs dataset, we conducted meta-regression by assessing whether efficacy differed in trials that enrolled older adults compared with trials that enrolled adults of any age. For the biologic DMARDs dataset, we conducted meta-regression using mean disease duration of subjects enrolled in each study as a continuous covariate to determine whether treatment efficacy varies by disease duration of RA. The two analyses involved exploring two different types of covariates in the meta-regression: one relates to exploring subgroup effects with a binary covariate (for SGAs) and the other to exploring interaction effects with a continuous covariate (disease duration of RA).

There is often a need to explore heterogeneity in treatment effects in terms of another variable, and many software packages facilitate meta-regression for pair-wise meta-analysis. But, until recently, there has not been a readily available process for Bayesian MTC in WinBUGS. Incorporation of covariates into network models was first discussed by Cooper et al.,[44] describing three different types of models for covariates. The NICE Decision Support Unit published TSD 3[5] to illustrate the use of different types of meta-regression within Bayesian MTC (using the work of Cooper et al. as a template). Bayesian MTC allows us to incorporate all the available evidence into one analysis, which, combined with the ability to explore potential effect modifiers or confounders, can be a powerful tool when conducting CERs.

We used the methods described in TSD 3, which build on the generalized linear modeling framework established in TSD 2. TSD 3 describes three general approaches to meta-regression models in a multiple treatment context: separate and unrelated interaction terms for each treatment, exchangeable and related interaction terms, and one single interaction effect for all treatments. We used the latter. WinBUGs code and the datasets used are provided in Appendix A. For exploring subgroup effects through a meta-regression, a trial-level binary indicator was created to indicate study population. This indicator variable was added to the random effects logistic regression model used in KQ 1. In this way, the interaction term describes the effect of the population on the outcome. In the second example, a continuous covariate was added to the same random effects logistic regression model, assuming a common covariate effect for each treatment. The continuous covariate was centered at its mean value to improve model convergence. The magnitude and direction of the interaction term was examined in each example, along with its effect on the odds of treatment response in each case.

Meta-Regression With a Subgroup Indicator Covariate: Efficacy of SGAs and Older Adults

From our report on SGAs, eight trials provided data on efficacy in older adult patients. We hypothesized that differences in efficacy may exist between studies enrolling older adults (≥55 years) and those enrolling adults of any age. We explored this by introducing a covariate to the

model that indicated the subgroup population. The analysis included 72 trials; 64 trials addressing efficacy in the adult population and eight trials conducted in the older adult population. The outcome was treatment response as measured by 50 percent or greater improvement from baseline on the HAM-D. We intended to test the interaction effect and compare the odds ratios for the treatments within each population. One advantage of this type of meta-regression is that it allows for the exploration of differences within one model, rather than performing multiple analyses separately and comparing results qualitatively.

Meta-Regression With a Continuous Covariate: Efficacy of Biologic DMARDs and Disease Duration of Rheumatoid Arthritis

Our recent CER on biologic DMARDs included a total of 31 studies in the MTC meta-analysis (see descriptions above and data from the included studies provided in Appendix Table B-3). The outcome was response as measured by ACR 50 after a minimum treatment period of 12 weeks. The trials included RA patients who failed methotrexate, a first-line therapy, but were naïve to treatment with a biologic DMARD. One trial-level factor hypothesized to have an effect on treatment response is mean disease duration of RA. Longer disease durations are hypothesized to be associated with higher levels of treatment response. We aimed to test the effects of this covariate (mean disease duration) on our findings (with etanercept yielding greater treatment response than most other biologic DMARDs: see Introduction).

KQ 3. Methods Section for Simulation Study

Because of inherent limitations when using real-world data, we proposed a simulation study to investigate the role of the number of studies available for each comparison on the MTC meta-analysis model's ability to detect true relative efficacy under different network patterns. The motivation for a simulation study arose from uncertainty about how the number of studies available for each comparison affects the ability of the model to produce valid results and whether this varied under different network patterns. In performing Bayesian MTC meta-analyses for CERs, the question of whether there are enough studies to make valid conclusions about comparative efficacy often arises. Many times, only one study is identified that links two treatments (or a treatment and placebo) together in a network. We hypothesized that findings may vary depending on the underlying data structure (i.e., network pattern).

In order to assess the ability of the Bayesian MTC method to produce valid results under the four network patterns, we first created simulated datasets comprising two scenarios for comparative efficacy. Simulated datasets will lend flexibility in choosing the number of studies available for each comparison and also allow us to determine the true comparative efficacy a priori. The basic process of the simulation is illustrated in Figure 4. We restricted the simulations to four treatments in each case and set the sample size *within* each study to be 100 patients per treatment arm. Our outcome was a dichotomous measure of treatment response (1 = response, 0 = no response).

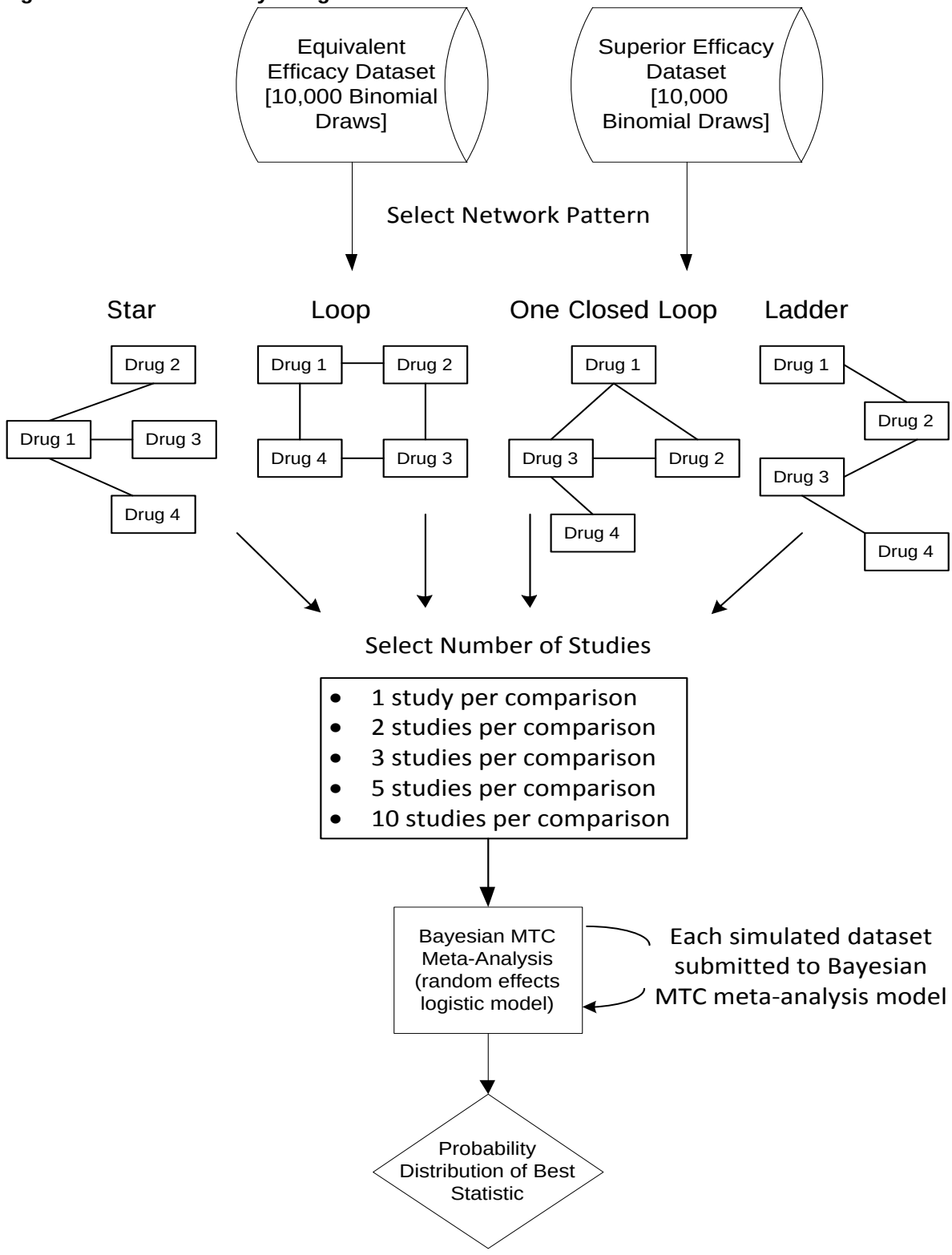
Two master simulated datasets were created in SAS to test the method in one case where analyses should not find important differences and in another case where analyses should find important differences. The first dataset is a scenario of equivalent efficacy of three drugs compared with placebo. A dataset of 10,000 "studies" and four treatments was created by sampling from the binomial distribution so that placebo had a mean response of 0.10, and drugs

2 through 4 had a mean response of 0.50. In this way, drugs 2 through 4 have equivalent efficacy, but all have greater efficacy than placebo.

The second of the two datasets is a scenario of superior efficacy of one drug compared with three other drugs. Again, a dataset of 10,000 "studies" and four treatments was created by sampling from the binomial distribution so that drugs 1 through 3 had a mean response of 0.20, and drug 4 had a mean response of 0.80. In this case, drug 4 is superior to the other drugs in the network, which all have equal efficacy compared with each other. We did not include any multi-arm trials in either of the simulated datasets.

Next, to address variations in the number of studies available for each comparison, we selected five cases: 1, 2, 3, 5, and 10 studies per comparison. That is, the link connecting any two drugs is based on 1, 2, 3, etc., studies. We fixed this number so that the overall number of studies per link is the same for each drug comparison. We sampled successive studies from the two master datasets to produce 1,000 sample datasets for each of the four network patterns (star, loop, one closed loop, and ladder) and the five cases, resulting in a total of 40,000 simulated datasets. The datasets were then submitted to the Bayesian MTC meta-analysis model used in KQ 1. In each model run, there was a burn-in of 5,000 iterations, followed by 15,000 iterations from which we monitored and output the probability that each drug was the best (i.e., most efficacious). Satisfactory convergence was verified by trace plots and by monitoring the Monte Carlo error. Convergence was generally achieved quickly (as might be expected with each individual dataset being fairly small, with just four drugs). The output of 1,000 "best" statistics for each pattern by case (determined by number of studies for each comparison) formed a distribution to allow us to assess the model estimates versus the predetermined treatment efficacy. To estimate bias, we computed the difference between the percentage of times a drug was considered most effective based on the models and the expected percentage based on the assumptions used to generate the data. The SAS code and WinBUGs code used to generate the master datasets and Bayesian models are provided in Appendix A.

Figure 4. Simulation study design



Results

Key Question (KQ) 1. Bayesian Mixed Treatment Comparison (MTC) Methods Compared With Frequentist Indirect Methods, and Performance for Different Types of Evidence Network Patterns

We organized this section in two main parts: (1) comparison of results of Bayesian mixed treatment comparison (MTC) methods with those of various frequentist methods for the full evidence networks and (2) comparison of and performance for different types of evidence network patterns. For the second part, we divided the results into two subsections—one comparing results of Bayesian MTC methods with those of various frequentist methods for subcomponents (i.e., evidence network patterns) of the full networks and one presenting results of Bayesian MTC methods for the full networks and those of Bayesian MTC methods for subcomponents of the full networks. Within each part, we first address results using data from our second-generation antidepressants (SGAs) report and then results using data from our report on treatments for rheumatoid arthritis (RA).

Comparison of Bayesian MTC Results With Those of Various Frequentist Analyses for the Full Networks

In this section, we provide results of comparisons between Bayesian MTC methods and three frequentist methods for the full networks illustrated in Appendix C. Tables in Appendix D provide detailed results for each analysis. Tables in Appendix E provide a comparison of precision of findings from the various analyses (determined by width of the 95% credible interval or confidence interval), with the darkest shading indicating the most precise result and the lightest indicating the least precise for each drug-drug comparison. The tables in this section provide a summary and some comparison of the data in Appendix D and Appendix E.

Second-Generation Antidepressants: Response (Binary Outcome)

Appendix Table D-1 provides results of our analyses for the full SGA network for each drug-drug comparison for Bayesian MTC meta-analyses, frequentist meta-regression, the Bucher method, and frequentist logistic regression (Appendix D). For 15 out of 78 drug-drug comparisons, neither frequentist meta-regression nor the Bucher method produced a result, either because no studies included a common comparator or because an insufficient number of studies included a common comparator for the program to run the analysis.

Table 3 summarizes the number of comparisons for which each frequentist method was unable to produce a result (the Bayesian MTC method produced results for all 78 comparisons), the percent agreement and kappa between each frequentist method and the Bayesian MTC method, and comparative precision. The three frequentist methods were unable to produce a result for 0 percent (0/78), 32 percent (25/78), and 45 percent (35/75) of the drug-drug comparisons (for the logistic regression, meta-regression, and Bucher methods, respectively). Logistic regression always produced the most precise result and Bayesian MTC was the next most precise for all but one drug-drug comparison (Appendix Table E-1).

Table 3. Comparison of findings for SGA response between Bayesian MTC meta-analysis findings and those of frequentist methods: agreement and kappas

Outcome	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
No result produced[a]	0 of 78 comparisons	25 of 78 comparisons	35 of 78 comparisons	0 of 78 comparisons
% agreement with Bayesian MTC[b]	NA	94.3 (50/53)	93.0 (40/43)	100
Kappa[c]	NA	NA	NA	1.00
Precision	2nd most precise	Least precise	3rd most precise	Most precise

MTC = mixed treatment comparison; NA = not applicable; SGA = second-generation antidepressant

[a]No result produced indicated the number of drug-drug comparisons for which the method was unable to produce a result, either because no studies had a common comparator or an insufficient number of studies had a common comparator.

[b]Percent agreement calculated only for drug-drug comparisons that both methods were able to produce a result. For example, meta-regression did not produce a result for 25 of 78 drug-drug comparisons. Therefore, the percent agreement was calculated using the 53 comparisons for which both methods produced a result. Results were considered to agree if both methods produced a non-statistically significant (for frequentist methods) or unimportant (for Bayesian MTC) result for the comparison or if both analyses found a statistically significant or important result favoring the same treatment.

[c]Kappa was calculated only for comparisons when both methods were able to produce a result, as with percent agreement. Not applicable (NA) indicates that the statistical program was unable to calculate a kappa because of insufficient data.

For the results of the three drug-drug comparisons that were not in agreement between the Bayesian MTC and the meta-regression and Bucher methods (duloxetine vs. escitalopram, escitalopram vs. fluoxetine, and fluoxetine vs. venlafaxine), the Bayesian MTC meta-analysis found a potentially important difference (based on 95% credible intervals) between treatments, whereas the other two methods did not. In addition, the results of the Bayesian MTC had greater precision than the other two methods for all three comparisons. Point estimates were very similar for the three methods for one of the comparisons (fluoxetine vs. venlafaxine, odds rations [ORs] ranged from 0.75 to 0.77) but not for the other two comparisons (duloxetine vs. escitalopram: ORs 0.74 vs. 1.23 vs. 1.19, respectively; escitalopram vs. fluoxetine: ORs 1.44 vs. 0.94 vs. 0.94).

Second-Generation Antidepressants: Mean Change in HAM-D (Continuous Outcome)

Appendix Table D-2 provides results of our analyses for the full SGA network for each drug-drug comparison for Bayesian MTC meta-analyses, frequentist meta-regression, and the Bucher method (Appendix D). There are no results from frequentist logistic regression for this outcome because it is a continuous outcome.

For 33 out of 78 drug-drug comparisons, neither of the frequentist methods produced a result either because no studies included a common comparator or because an insufficient number of studies included a common comparator for the program to run the analysis. Table 4 summarizes the number of comparisons for which each frequentist method was unable to produce a result (the Bayesian MTC method produced results for all 78 comparisons) as well as the percent agreement between each frequentist method and the Bayesian MTC method. Bayesian MTC meta-analysis produced the most precise result for 67 percent of the comparisons; the Bucher method was the most precise for 33 percent of the comparisons (Appendix Table E-2).

Table 4. Comparison of findings for mean change in HAM-D between Bayesian MTC meta-analysis findings and those of frequentist methods: agreement and kappas

Outcome	Bayesian MTC	Meta-Regression	Bucher Method
No result produced[a]	0 of 78 comparisons	46 of 78 comparisons	33 of 78 comparisons
% agreement with Bayesian MTC[b]	NA	100 (32/32)	84.4 (38/45)
Kappa[c]	NA	NA	NA
Precision	Most precise	Least precise	2nd most precise

HAM-D = Hamilton Depression rating scale; MTC = mixed treatment comparison; NA = not applicable

[a]No result produced indicated the number of drug-drug comparisons for which the method was unable to produce a result, either because no studies had a common comparator or an insufficient number of studies had a common comparator.

[b]Percent agreement calculated only for drug-drug comparisons that both methods were able to produce a result. Results were considered to agree if both methods produced a non-statistically significant (for frequentist methods) or unimportant (for Bayesian MTC) result for the comparison or if both analyses found a statistically significant or important result favoring the same treatment.

[c]Not applicable (NA) indicates that the statistical program was unable to calculate a kappa because of insufficient data.

For the results of the seven drug-drug comparisons that were not in agreement between the Bayesian MTC and the Bucher methods, the Bucher method found a statistically significant difference between treatments, whereas the Bayesian MTC method did not find an important difference (based on the 95% credible interval). The seven comparisons were duloxetine versus mirtazapine, duloxetine versus venlafaxine, escitalopram versus trazodone, escitalopram versus venlafaxine, mirtazapine versus venlafaxine, paroxetine versus venlafaxine, and sertraline versus venlafaxine. The Bayesian MTC method usually produced a more precise result (based on comparison of credible intervals with confidence intervals), but the point estimates differed in magnitude and sometimes in the direction (three of the seven) of effect (i.e., point estimates trended in opposite directions, favoring different treatments).

Biologic Disease-Modifying Antirheumatic Drugs for Rheumatoid Arthritis: ACR 50 (Binary Outcome)

Appendix Table D-3 provides results of our analyses for the full biologic DMARDs network for each drug-drug comparison for Bayesian MTC meta-analyses, frequentist meta-regression, the Bucher method, and frequentist logistic regression (Appendix D). For all 28 drug-drug comparisons, all four methods were able to produce results. Table 5 summarizes the percent agreement and kappa between each frequentist method and the Bayesian MTC method. Logistic regression produced the most precise result for 89 percent of the comparisons; the Bayesian MTC meta-analysis produced the third most precise result for 86 percent of the comparisons (Appendix Table E-3).

Table 5. Comparison of findings for ACR 50 between Bayesian MTC meta-analysis findings and those of frequentist methods: agreement and kappas

Outcome	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
No result produced	0 of 28 comparisons	0 of 28 comparisons	0 of 28 comparisons	0 of 28 comparisons
% agreement with Bayesian MTC[a]	NA	96.4 (27/28)	92.9 (26/28)	82.1 (23/28)
Kappa	NA	0.90	0.83	0.62
Precision	3rd most precise	Least precise	2nd most precise	Most precise

ACR 50 = American College of Rheumatology 50 percent response; MT C= mixed treatment comparison; NA = not applicable

[a]Results were considered to agree if both methods produced a non-statistically significant (for frequentist methods) or unimportant (for Bayesian MTC) result for the comparison or if both analyses found a statistically significant or important result favoring the same treatment.

For the results of the one drug-drug comparison (adalimumab vs. anakinra) that was not in agreement between the Bayesian MTC meta-analysis and meta-regression, the Bayesian MTC method found a potentially important result favoring adalimumab over anakinra (OR, 1.88, 95% CrI 1.01 to 3.98). Meta-regression found a similar point estimate but with a wider confidence interval not reaching statistical significance (OR, 1.95, 95%, CI, 0.82 to 4.61).

For the results of both drug-drug comparisons (anakinra vs. golimumab and etanercept vs. golimumab) that were not in agreement between the Bayesian MTC and the Bucher method, the Bucher method found a statistically significant difference between treatments, whereas the Bayesian MTC did not find an important difference. Point estimates for both were fairly similar and in the same direction, but the confidence intervals were more narrow (reaching statistical significance) for the Bucher method.

For the results of the five drug-drug comparisons that were not in agreement between the Bayesian MTC and logistic regression, logistic regression found a statistically significant difference between treatments, whereas the Bayesian MTC did not find an important difference. The five comparisons were abatacept versus anakinra, anakinra versus golimumab, anakinra versus infliximab, anakinra versus rituximab, and anakinra versus tocilizumab. Point estimates were fairly similar and were in the same direction for all five for the Bayesian MTC meta-analysis and logistic regression, but the confidence intervals were more narrow (reaching statistical significance) for logistic regression.

Performance for Different Types of Evidence Network Patterns

Comparison of Bayesian MTC Results With Those of Various Frequentist Analyses for Subcomponents of the Full Networks

In this section, we provide results of comparisons between Bayesian MTC methods and three frequentist methods for the subcomponents of the full networks illustrated in Appendix C. These subcomponents represent at least one of each of the following network patterns: placebo star, loop, one closed loop, and ladder. Tables in Appendix D provide detailed results for each analysis by network pattern. Tables in Appendix E provide a comparison of precision of findings from the various analyses (determined by width of the 95% credible interval or confidence interval), with the darkest shading indicating the most precise result and the lightest indicating the least precise for each drug-drug comparison. The tables in this section provide a summary and some comparison of the data in Appendix D and Appendix E.

Second-Generation Antidepressants: Response (Binary Outcome)

Appendix Tables D-4 through D-7 provide results of our analyses for each of the network patterns (sub-components of the full network) selected, including a placebo star, loop, one closed loop, and ladder for each drug-drug comparison for Bayesian MTC meta-analyses, frequentist meta-regression, the Bucher method, and frequentist logistic regression (Appendix D). Table 6 summarizes the number of comparisons for which each frequentist method was unable to produce a result as well as the percent agreement and kappa between each frequentist method and the Bayesian MTC method by network pattern. On average, logistic regression produced the most precise results for the star, one closed loop, and ladder, whereas the Bayesian MTC meta-analysis produced the most precise results for the loop (Appendix Tables E-4 through E-7).

Table 6. Comparison of findings for SGA response between Bayesian MTC meta-analysis findings and those of frequentist methods for various network patterns

Network Pattern	Outcome	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
Star	No result produced[a]	0 of 45 comparisons	3 of 45 comparisons	2 of 45 comparisons	0 of 45 comparisons
	% agreement with Bayesian MTC[b]	NA	100	100	100
	Kappa[c]	NA	NA	NA	NA
	Precision	3rd most precise	Least precise	2nd most precise	Most precise
Loop	No result produced[a]	0 of 3	0 of 3	0 of 3	0 of 3
	% agreement with Bayesian MTC[b]	NA	66.7 (2/3)	66.7 (2/3)	66.7 (2/3)
	Kappa[c]	NA	NA	NA	NA
	Precision	Most precise	Least precise	3rd most precise	2nd most precise
One closed loop	No result produced[a]	0 of 32	25 of 32	25 of 32	0 of 32
	% agreement with Bayesian MTC[b]	NA	85.7	85.7	93.8
	Kappa[c]	NA	NA	NA	0.6364
	Precision	2nd most precise	Least precise	3rd most precise	Most precise
Ladder	No result produced[a]	0 of 55	39 of 55	36 of 55	0 of 55
	% agreement with Bayesian MTC[b]	NA	100	100	96.4
	Kappa[c]	NA	NA	NA	0.6474
	Precision	3rd most precise	Least precise	2nd most precise	Most precise

MTC = mixed treatment comparison; NA = not applicable; SGA = second-generation antidepressant

[a]No result produced indicated the number of drug-drug comparisons for which the method was unable to produce a result, either because no studies had a common comparator or an insufficient number of studies had a common comparator.

[b]Percent agreement calculated only for drug-drug comparisons that both methods were able to produce a result. Results were considered to agree if both methods produced a non-statistically significant (for frequentist methods) or unimportant (for Bayesian MTC) result for the comparison or if both analyses found a statistically significant or important result favoring the same treatment.

[c]Kappa was calculated only for comparisons when both methods were able to produce a result, as with percent agreement. Not applicable (NA) indicates that the statistical program was unable to calculate a kappa because of insufficient data (because there were fewer than two levels).

Second-Generation Antidepressants: Mean Change in HAM-D (Continuous Outcome)

Appendix Tables D-8 through D-11 provide results of our analyses for each of the network patterns (sub-components of the full network) selected, including a placebo star, loop, one closed loop, and ladder for each drug-drug comparison for Bayesian MTC meta-analyses, frequentist meta-regression, and the Bucher method (Appendix D). There are no results from frequentist logistic regression for this outcome because it is a continuous outcome.

Table 7 summarizes the number of comparisons for which each frequentist method was unable to produce a result (the Bayesian MTC method produced results for all comparisons) as well as the percent agreement and kappa between each frequentist method and the Bayesian MTC method, by network pattern. For the one closed loop and ladder patterns, the frequentist methods did not produce a result for the majority of drug-drug comparisons. On average, logistic regression produced the most precise results for the star and ladder, the Bayesian MTC meta-analysis produced the most precise results for the loop, and meta-regression produced the least precise results for all network patterns (Appendix Tables E-8 through E-11).

Table 7. Comparison of findings for mean change in HAM-D between Bayesian MTC meta-analysis findings and those of frequentist methods for various network patterns

Network Pattern	Outcome	Bayesian MTC	Meta-Regression	Bucher Method
Star	No result produced[a]	0 of 21 comparisons	1 of 21 comparisons	0 of 21 comparisons
	% agreement with Bayesian MTC[b]	NA	100 (20/20)	81 (17/21)
	Kappa[c]	NA	NA	NA
	Precision	2nd most precise	Least precise	Most precise
Loop	No result produced[a]	0 of 3 comparisons	2 of 3 comparisons	0 of 3 comparisons
	% agreement with Bayesian MTC[b]	NA	100 (1/1)	100 (3/3)
	Kappa[c]	NA	NA	NA
	Precision	Most precise	Least precise	2nd most precise
One closed loop	No result produced[a]	0 of 21 comparisons	17 of 21 comparisons	14 of 21 comparisons
	% agreement with Bayesian MTC[b]	NA	100 (4/4)	85.7 (6/7)
	Kappa[c]	NA	NA	NA
	Precision	Most precise[d]	Least precise	Most precise[d]
Ladder	No result produced[a]	0 of 28 comparisons	25 of 28 comparisons	23 of 28 comparisons
	% agreement with Bayesian MTC[b]	NA	100 (3/3)	100 (5/5)
	Kappa[c]	NA	NA	NA
	Precision	2nd most precise	Least precise	Most precise

HAM-D = Hamilton Depression rating scale; MTC = mixed treatment comparison; NA = not applicable

[a]No result produced indicated the number of drug-drug comparisons for which the method was unable to produce a result, either because no studies had a common comparator or an insufficient number of studies had a common comparator.

[b]Percent agreement calculated only for drug-drug comparisons that both methods were able to produce a result. Results were considered to agree if both methods produced a non-statistically significant (for frequentist methods) or unimportant (for Bayesian MTC) result for the comparison or if both analyses found a statistically significant or important result favoring the same treatment.

[c]Kappa was calculated only for comparisons when both methods were able to produce a result. Not applicable (NA) indicates that the statistical program was unable to calculate a kappa because of insufficient data (because there were fewer than two levels).

[d]Neither the Bayesian MTC method nor the Bucher method was clearly more precise than the other. Each of them reported the most precise results for about half of the drug-drug comparisons.

Biologic Disease-Modifying Antirheumatic Drugs for Rheumatoid Arthritis: ACR 50 (Binary Outcome)

Appendix Tables D-12 through D-16 provide results of our analyses for each of the network patterns (subcomponents of the full network) selected, including a placebo star, loop, one closed loop, and ladder for each drug-drug comparison for Bayesian MTC meta-analyses, frequentist meta-regression, the Bucher method, and frequentist logistic regression (Appendix D). Table 8 summarizes the number of comparisons for which each method was unable to produce a result, as well as the percent agreement and kappa between each frequentist method and the Bayesian MTC method by network pattern. On average, logistic regression produced the most precise results for all of the network patterns; the Bayesian MTC meta-analysis produced the third most precise or the least precise results, depending on the network pattern (Appendix Tables E-12 through E-16).

Table 8. Comparison of findings for ACR 50 response between Bayesian MTC meta-analysis findings and those of frequentist methods for various network patterns

Network Pattern	Outcome	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
Star	No result produced[a]	0 of 28 comparisons	0 of 28 comparisons	0 of 28 comparisons	0 of 28 comparisons
	% agreement with Bayesian MTC[b]	NA	100 (28/28)	89.3 (25/28)	78.6 (22/28)
	Kappa[c]	NA	1.000	0.7308	0.5333
	Precision	3rd most precise	Least precise	2nd most precise	Most precise
Loop	No result produced[a]	0 comparisons	0 of 1 comparisons	0 of 1 comparisons	0 of 1 comparisons
	% agreement with Bayesian MTC[b]	NA	100 (1/1)	100 (1/1)	100 (1/1)
	Kappa[c]	NA	NA	NA	NA
	Precision	3rd most precise	Least precise	2nd most precise	Most precise
One closed loop with adalimumab	No result produced[a]	0 of 3 comparisons	0 of 3 comparisons	0 of 3 comparisons	0 of 3 comparisons
	% agreement with Bayesian MTC[b]	NA	100 (3/3)	100 (3/3)	100 (3/3)
	Kappa[c]	NA	NA	NA	NA
	Precision	Least precise	3rd most precise	2nd most precise	Most precise
One closed loop with etanercept	No result produced[a]	0 of 3 comparisons	0 of 3 comparisons	0 of 3 comparisons	0 of 3 comparisons
	% agreement with Bayesian MTC[b]	NA	100 (3/3)	100 (3/3)	100 (3/3)
	Kappa[c]	NA	1.000	1.000	1.000
	Precision	3rd most precise	Least precise	2nd most precise	Most precise
Ladder	No result produced[a]	0 of 3 comparisons	1 of 3 comparisons	1 of 3 comparisons	0 of 3 comparisons
	% agreement with Bayesian MTC[b]	NA	100 (2/2)	100 (2/2)	66.7 (2/32)
	Kappa[c]	NA	1.000	NA	NA
	Precision	Least precise	3rd most precise	2nd most precise	Most precise

ACR 50 = American College of Rheumatology 50 percent response; MTC= mixed treatment comparison; NA = not applicable

[a]No result produced indicated the number of drug-drug comparisons for which the method was unable to produce a result, either because no studies had a common comparator or an insufficient number of studies had a common comparator.

[b]Percent agreement calculated only for drug-drug comparisons that both methods were able to produce a result. Results were considered to agree if both methods produced a non-statistically significant (for frequentist methods) or unimportant (for Bayesian MTC) result for the comparison or if both analyses found a statistically significant or important result favoring the same treatment.

[c]Kappa was calculated only for comparisons when both methods were able to produce a result. Not applicable (NA) indicates that the statistical program was unable to calculate a kappa because of insufficient data (because there were fewer than two levels).

Comparison of Bayesian MTC Results for the Full Networks With Bayesian MTC Results for Subcomponents

In this section, we present findings of the Bayesian MTC meta-analyses for each dataset for the full network and for subcomponents of the network. Because the data differ for each scenario, the measures are not intended to be directly comparable for the full network and for the various subcomponents of the network but are presented here to show model fit under each scenario. For example, the deviance information criterion should not be used to compare the fit

of models using different datasets. The findings for the full network represent those for the complete literature on each set of medications (at the time of our CERs on each topic).[19, 20] Those for the subcomponents examined hypothetical scenarios that would occur if the bodies of evidence were more limited.

Second-Generation Antidepressants: Binary Outcome

We ran MTC meta-analyses for the binary outcome of response to SGAs under five different network scenarios, the full network and four subcomponent network patterns: star, loop, one closed loop, and ladder (as described in the Methods chapter). Network figures in Appendix C detail the geometry of these subcomponents.

Measures of model fit are reported in Table 9. Estimates of between-study heterogeneity (tau squared) ranged from 0.12 to 0.44. When examining the total residual deviance for each scenario, all of them reasonably approximate the number of data points available, suggesting good model fit. The ladder and star pattern models most closely approximated the number of data points available.

Table 9. Measures of model fit for second-generation antidepressants for response

Statistic	Full Network	Star	Loop	One Closed Loop	Ladder
Total residual deviance (total number of data points available)	135.3 (140)	35.96 (36)	25.64 (30)	37.67 (42)	39.36 (40)
Deviance information criterion	259.45	246.67	188.42	271.67	270.31
Between-study heterogeneity	0.2571	0.4427	0.1172	0.1196	0.3019

One feature of the Bayesian framework for MTC is the ability to directly calculate the probability that each drug is the best treatment. Table 10 shows the results of this statistic for the five scenarios. Because not every treatment was available in each network pattern, some comparisons were not applicable for every network pattern.

Table 10. Probability of best treatment for second-generation antidepressants for response

Statistic	Full Network	Star	Loop	One Closed Loop	Ladder
Bupropion	0.005	NA	NA	0.057	0.034
Citalopram	0.366	NA	NA	0.218	0.268
Desvenlafaxine	0.023	0.019	NA	NA	NA
Duloxetine	0.000	0.005	NA	NA	0.002
Escitalopram	0.090	NA	NA	0.056	0.110
Fluoxetine	0.000	0.185	0.019	0.003	0.004
Fluvoxamine	0.280	0.255	NA	0.202	0.262
Mirtazapine	0.158	NA	NA	0.077	0.061
Nefazodone	0.038	0.077	NA	NA	NA
Paroxetine	0.003	0.210	0.182	0.070	0.021
Sertraline	0.003	0.007	0.798	0.317	0.069
Trazodone	0.008	NA	NA	0.001	0.002
Venlafaxine	0.028	0.243	NA	NA	0.168

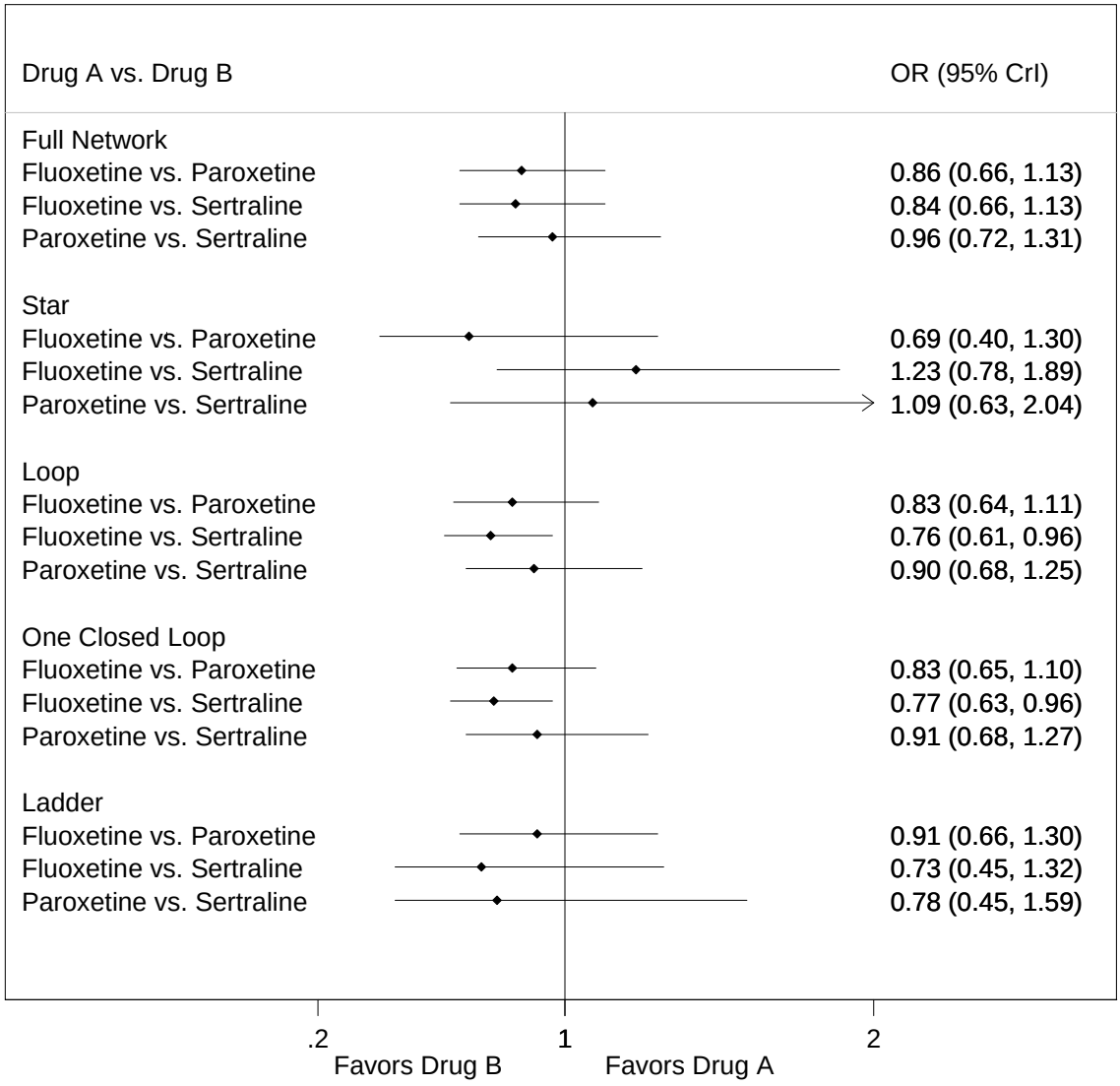
NA = not applicable

For the full network, citalopram had the greatest probability of being the best treatment for achieving response (36.6%), followed by fluvoxamine. However, none of the treatments were particularly dominant to where we would have high confidence in that treatment truly having greater efficacy—as we might suggest if one reached a probability of 90 percent or more, for example. For the star network pattern, fluvoxamine had the greatest probability, followed by venlafaxine. But, again, none of the treatments were particularly dominant. The medication with

the greatest probability of being the best treatment varied for each of the five scenarios. Only the one closed loop had a fairly dominant medication: sertraline had a probability of almost 80 percent for being the best treatment.

Given that not all of the comparisons were available in each scenario, we focused on the three drugs that were represented in each pattern: fluoxetine, paroxetine, and sertraline. Figure 5 shows the odds ratios and 95% credible intervals for the relative response of the antidepressants for each scenario. The full set of all pairwise comparisons is reported in Appendix D (Appendix Tables D-1, D-4, D-5, D-6, and D-7). Generally, the results show that the three drugs are not significantly different in odds of response. The two models with the least amount of connected data (star and ladder) also had the greatest heterogeneity and thus wider credible intervals around the mean. The response profiles for the loop and one closed loop mirrored the full network (with very similar point estimates) but found a significant difference between fluoxetine and sertraline, in part because of the lower between-studies heterogeneity of these reduced datasets.

Figure 5. Results of Bayesian MTC meta-analysis for five scenarios: odds ratio (95% credible interval) comparing fluoxetine, paroxetine, and sertraline for achieving response



CrI = credible interval; OR = odds ratio; vs. = versus

Second-Generation Antidepressants: Continuous Outcome

For the second illustration with the antidepressant data, response is now represented on a continuous scale by using mean change from baseline in HAM-D in each treatment arm. We ran MTC meta-analyses using a normal likelihood and identify link function for each of the five different network scenarios. Network figures in Appendix C detail the geometry of these scenarios.

Out of the 78 pairwise comparisons available in the full network, 21 were available in the star pattern, 3 in the loop, 21 in the one closed loop, and 28 in the ladder. The full results are reported in Appendix D (Tables D-2, D-8, D-9, D-10, and D-11). Although the point estimates vary across the patterns, likely because of the availability of direct head-to-head studies, the confidence intervals cross the line of no difference in all but one of the comparisons. As in the

case of the binary outcome (response), there is little evidence from these data to support any important differences in the efficacy of the 13 antidepressants.

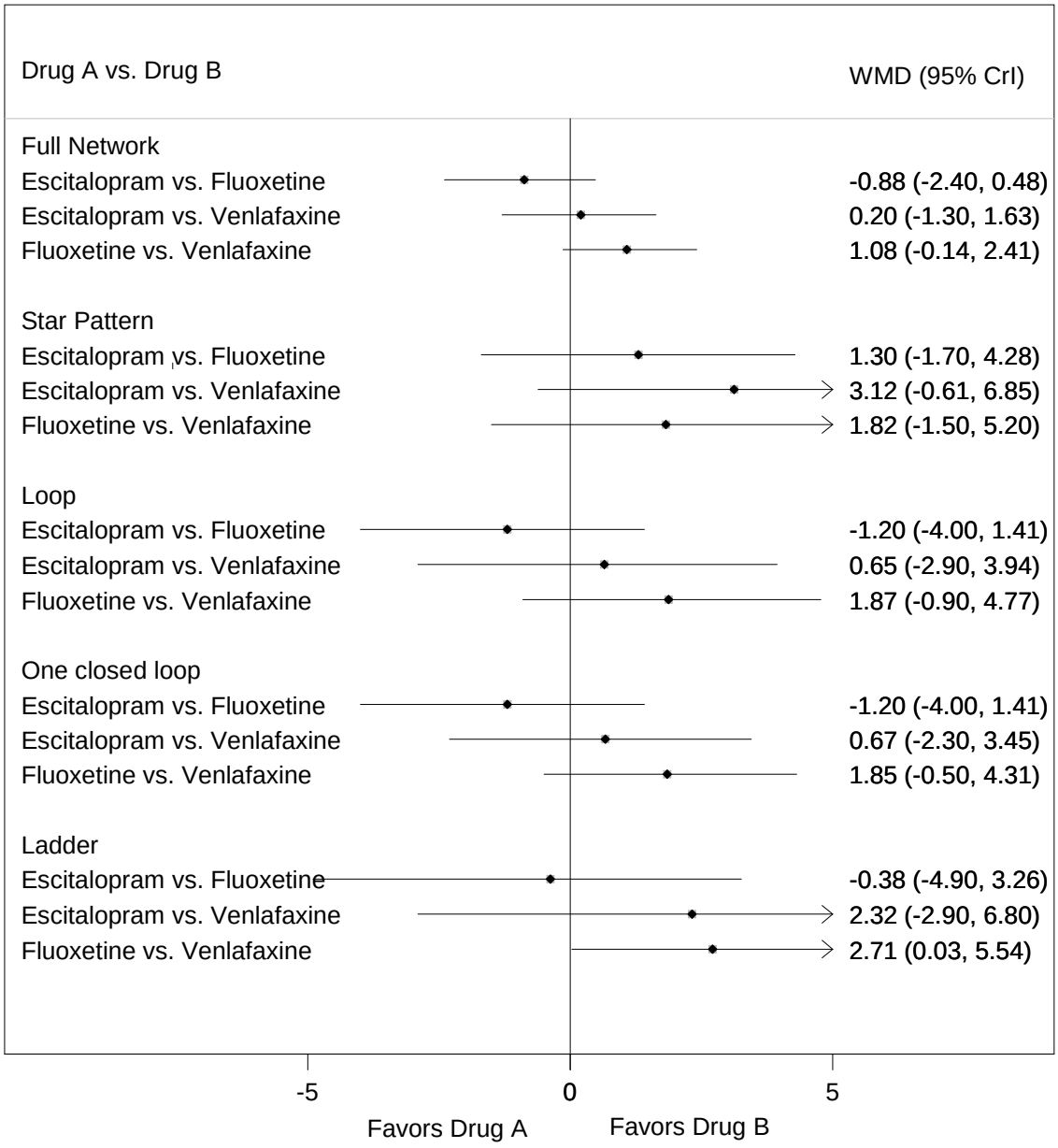
Measures of model fit are reported in Table 11. Estimates of between-study heterogeneity (tau squared) ranged from 0.86 to 1.97. For the total residual deviance for each scenario, the star, loop, and ladder most closely approximated the number of data points available, but fairly close approximations were found for all scenarios, suggesting good model fit.

Table 11. Measures of model fit for second-generation antidepressants for mean change from baseline in HAM-D

Statistic	Full Network	Star	Loop	One Closed Loop	Ladder
Total residual deviance (total number of data points available)	86.93 (80)	41.27 (40)	18.81 (18)	39.64 (38)	26.71 (26)
Deviance information criterion	254.08	120.76	59.866	122.637	75.126
Between-study heterogeneity	0.8564	0.8637	1.974	1.683	1.192

As in the example with a binary outcome, we looked closer at three pairwise comparisons available under each scenario (Figure 6). Generally, the results show that the three drugs are not significantly different in mean change from baseline in HAM-D.

Figure 6. Results of Bayesian MTC meta-analysis for five scenarios: weighted mean difference (95% credible interval) comparing escitalopram, fluoxetine, and venlafaxine for mean change from baseline in HAM-D



CrI = credible interval; vs. = versus; WMD = weighted means difference

Biologic DMARDs for Rheumatoid Arthritis: Binary Outcome

For the RA dataset, the majority of the trials were placebo-controlled; only one trial included a head-to-head comparison, between adalimumab and infliximab. Therefore, the choice of subcomponent network patterns was more limited than for the SGA dataset. We ran MTC meta-analyses on the binary outcome of ACR 50 response for five different network scenarios; the full network; and four sub-network patterns: star, loop, one closed loop, and ladder (as described in the Methods chapter). Network figures in Appendix C detail the geometry of these scenarios.

Measures of model fit are reported in Table 12. Estimates of between-study heterogeneity (tau squared) were very similar across scenarios (range 0.27 to 0.33). Total residual deviance data show that the full network and the loop most closely approximated the number of data points available, suggesting good model fit. The other scenarios had more sizeable differences between total residual deviance and the number of data points.

Table 12. Measures of model fit for biologic DMARDs for rheumatoid arthritis for achieving ACR 50

Statistic	Full Network	Star	Loop	One Closed Loop	Ladder
Total residual deviance (total number of data points available)	65.54 (62)	62.29 (40)	17.01 (18)	22.87 (38)	16.03 (26)
Deviance information criterion	394.56	374.04	110.95	149.89	101.45
Between-study heterogeneity	0.3006	0.3193	0.3315	0.2668	0.3899

ACR 50 = American College of Rheumatology 50 percent response; DMARDs = disease-modifying antirheumatic drugs

Table 13 shows the probability that each drug is the best treatment for achieving ACR 50 response for each of the five scenarios. Because not every treatment was available in each network pattern, the probabilities for some treatments were not reported (i.e., not applicable) for some network patterns. For all five scenarios, etanercept had the greatest probability of being the best treatment for achieving ACR 50 response, with probabilities of 94.9 percent or higher.

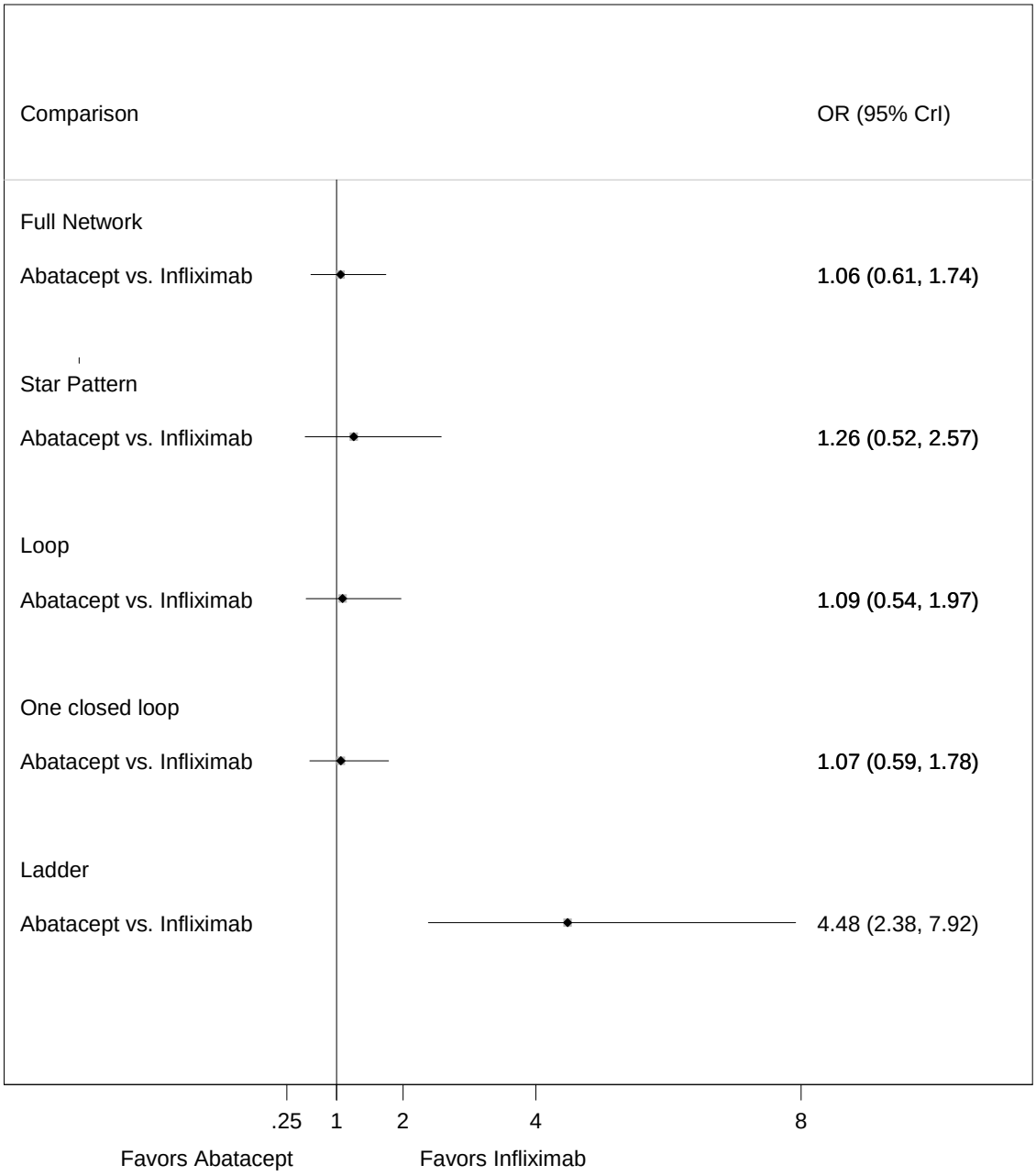
Table 13. Probability of best treatment for biologic DMARDs for rheumatoid arthritis for achieving ACR 50

Statistic	Full Network	Star	Loop	One Closed Loop (ETA)	Ladder
Abatacept	0.000	0.001	0.4526	0.001	0.000
Adalimumab	0.003	0.003	NA	NA	NA
Anakinra	0.000	0.000	NA	NA	NA
Etanercept	0.953	0.949	NA	0.999	0.993
Golimumab	0.040	0.041	NA	NA	NA
Infliximab	0.000	0.001	0.5473	0.001	0.007
Rituximab	0.003	0.004	NA	NA	NA
Tocilizumab	0.000	0.000	NA	NA	NA

DMARDs = disease-modifying antirheumatic drugs; ETA = etanercept; NA = not applicable

Given that not all of the comparisons were available in each scenario, we focused on the two drugs that are represented in each pattern. Figure 7 shows the odds ratios and 95 percent credible intervals for the relative response for each scenario. The full set of all pairwise comparisons is reported in Appendix D (Appendix Tables D-3, D-12, D-13, D-14, D-15, and D-16). Generally, the results show that the two drugs in Figure 7 are not significantly different in odds of ACR 50 response for four of the scenarios; the ladder found greater response for infliximab than for abatacept.

Figure 7. Results of Bayesian MTC meta-analysis for five scenarios: odds ratio (95% credible interval) comparing abatacept and infliximab for treatment response (ACR 50)



CrI = credible interval; OR = odds ratio; vs. = versus

KQ 2. Meta-Regression With Bayesian MTC Meta-Analysis

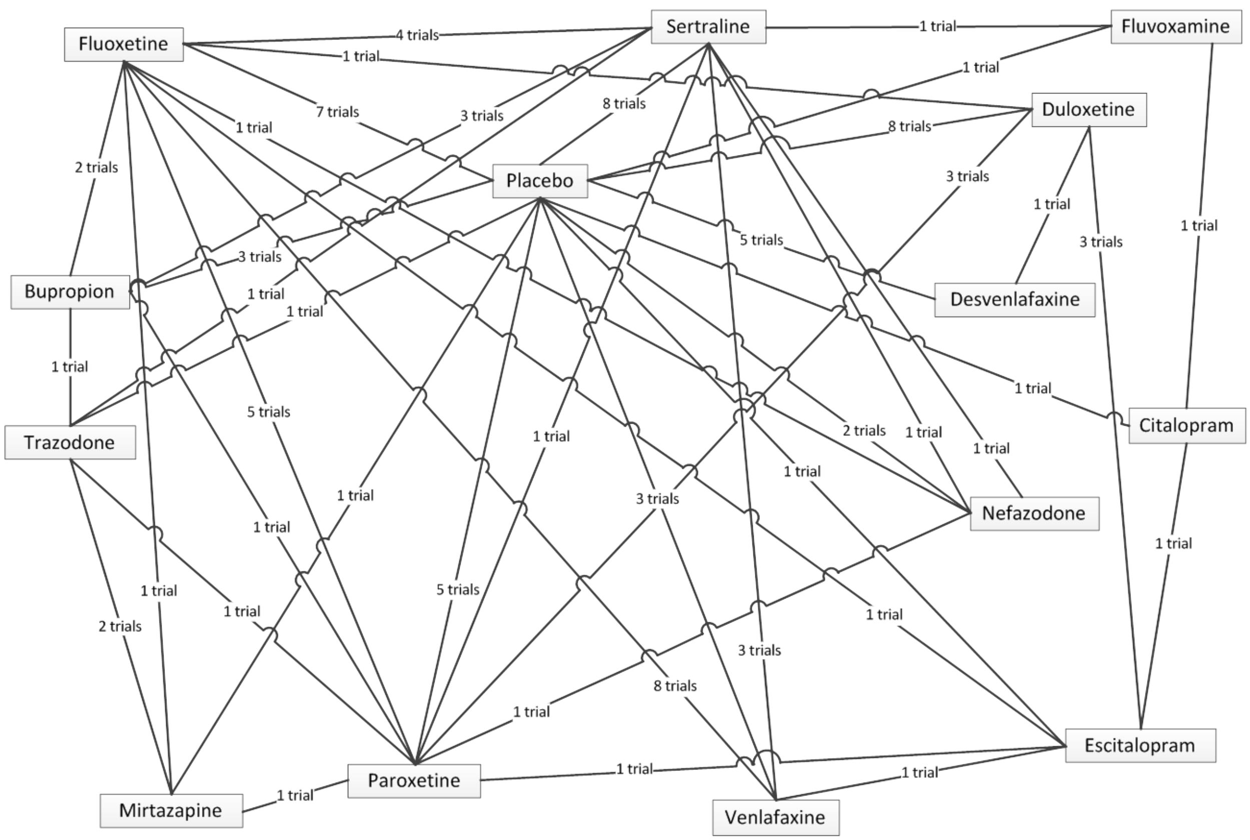
This KQ aims to illustrate the use of meta-regression as a technique for exploring sources of heterogeneity of treatment effects in CERs. We examine two types of meta-regression in this KQ: one relates to exploring subgroup effects with a binary covariate and the other to exploring interaction effects with a continuous covariate. Our purpose here is to use meta-regression with

Bayesian MTC meta-analysis within the context of two real-world scenarios from recent CERs and to examine two clinically important questions.

Meta-Regression With a Subgroup Indicator Covariate

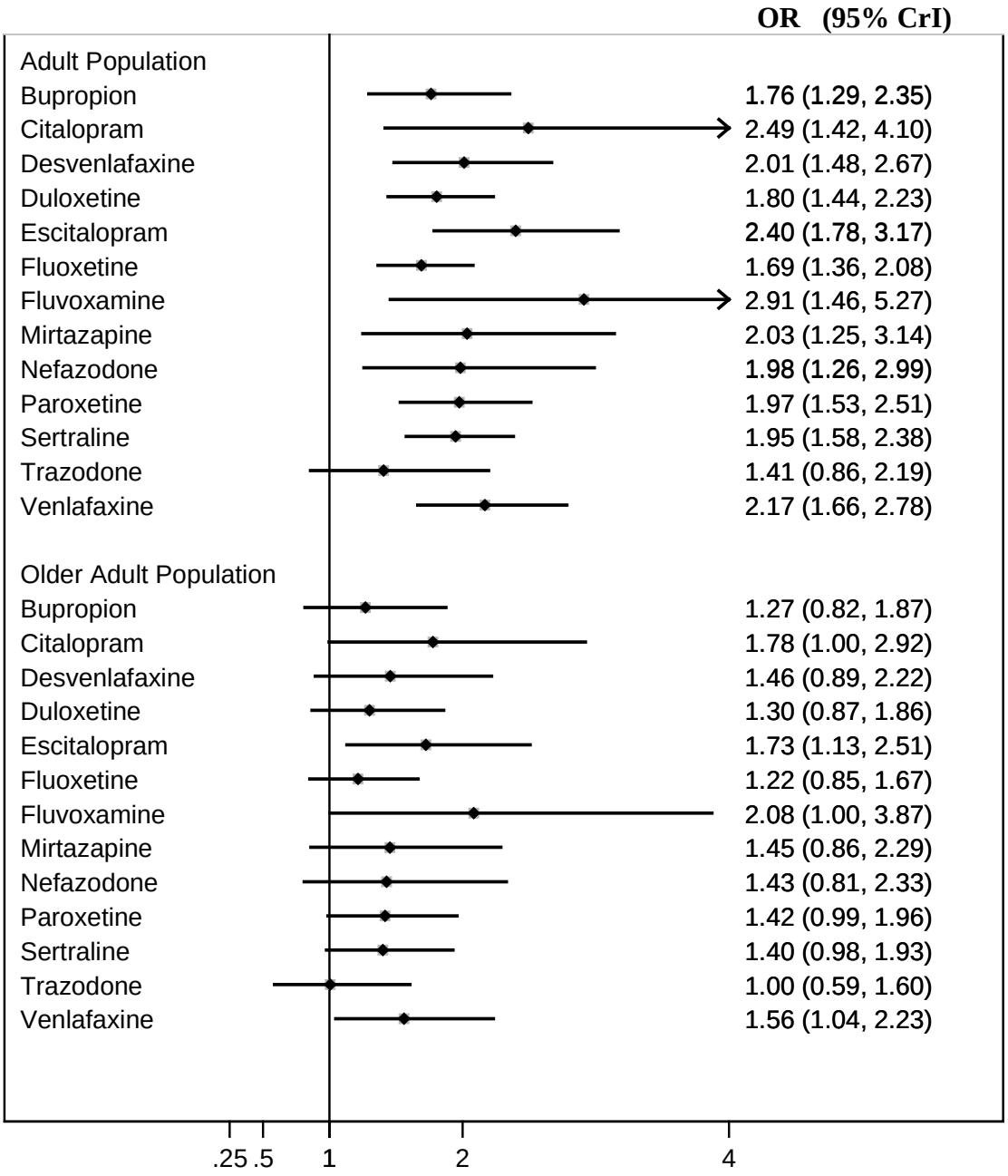
We hypothesized that differences in efficacy of SGAs may exist between older adults (≥55 years) and adults of any age. We used 72 trials of SGAs; 64 trials addressed efficacy in the general adult population and eight trials conducted exclusively in the older adult population. The outcome was treatment response as measured by 50 percent or greater improvement from baseline on the HAM-D. Figure 8 shows the evidence network for the 72 trials.

Figure 8. Evidence network for subgroup meta-regression to assess whether efficacy of second-generation antidepressants differs for older adults



For this analysis, we used a random effects logistic regression model, including the appropriate adjustment for correlations within multi-arm trials. Convergence was checked via trace plots, and posterior means were calculated after 100,000 iterations following a burn-in of 20,000 iterations. We used a common (single) interaction term. Figure 9 shows the odds ratios of treatment response for each of the SGAs, relative to placebo, within each age subgroup.

Figure 9. Odds ratios (95% credible interval) of treatment response for second-generation antidepressants, by patient population



CrI = credible interval; OR = odds ratio; vs. = versus

Within the general adult population, efficacy was supported for all antidepressants with the exception of trazodone. When estimating these effects for the older adult population within the same meta-regression model, the efficacy for each drug appears to be diminished with only escitalopram and venlafaxine maintaining statistical superiority compared with placebo. This trend is supported by the interaction effect estimate of -0.34 (95% CrI, -0.696 to 0.006). This represents, on average, a -0.34 reduction in the log odds of response (not for the odds ratio scale). While approaching marginal statistical significance, the interaction estimate indicates a trend toward lower response rates in the older adult population. We compared the meta-

regression model of 72 studies with the main adult analysis including 64 studies. The estimates for between-study heterogeneity were similar, although slightly reduced in the model containing the additional older adult studies (Table 14). Both model estimates of total residual deviance were appropriate when compared with the number of data points available in each analysis, suggesting a good model fit in each case.

Table 14. Measures of model fit for main analysis and meta-regression

Statistic	Main Adult Analysis (n=64)	Meta-Regression Including Older Adults (n=72)
Total residual deviance (total number of datapoints available)	135.3 (140)	153.8 (158)
DIC	928.73	933.38
Between-study heterogeneity	0.2571	0.2506

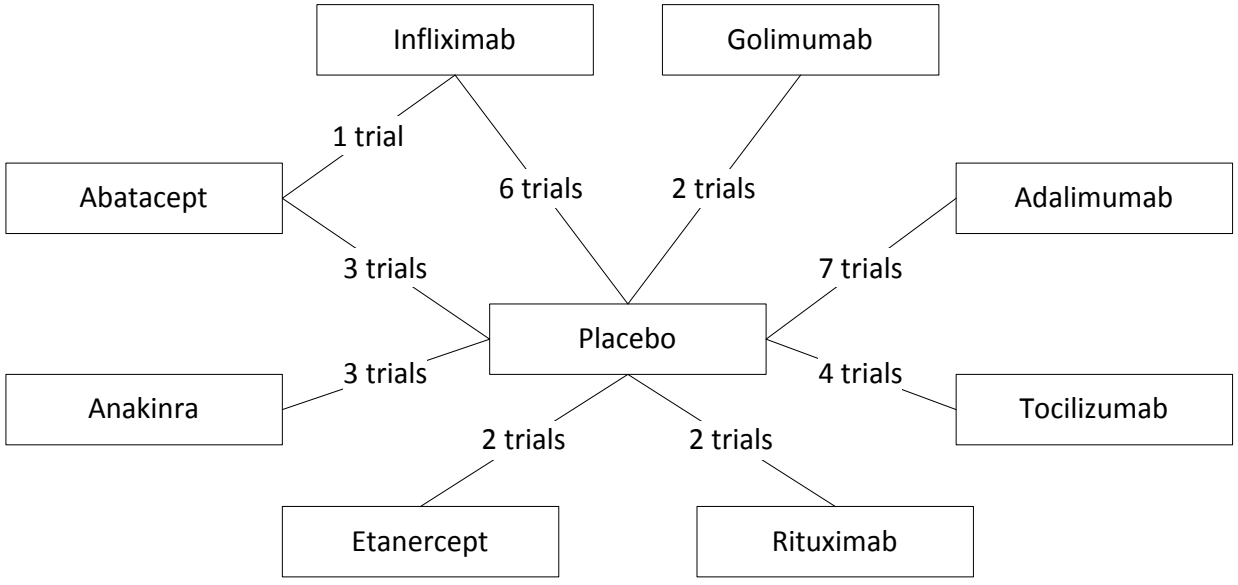
DIC = Deviance Information Criterion

Meta-Regression With a Continuous Covariate

We hypothesized that differences in efficacy of treatments for RA may exist for patients with varying durations of disease. Longer disease durations were hypothesized to be associated with higher levels of treatment response. We aimed to test the effects of this covariate (mean disease duration) on the findings from our recent CER—with our Bayesian MTC meta-analysis finding that etanercept resulted in greater treatment response than most other biologic DMARDs.

Twenty-eight of the 31 trials included in our Bayesian MTC meta-analysis reported data on mean disease duration (in years). Both trials that did not report data on mean disease duration compared etanercept with placebo. If disease duration was presented as only an arm-level mean, the average of the arms was taken for the trial-level mean. All eight biologic DMARDs were represented in the 28 trials (Figure 10). The mean trial-level disease duration ranged from 3.9 to 13 years.

Figure 10. Evidence network for continuous covariate meta-regression to assess whether efficacy of biologic DMARDs differs for patients with longer disease duration



A random effects logistic regression model was used, including the appropriate adjustment for correlations within multi-arm trials. The model was fitted assuming a common interaction effect for all treatments. Convergence was checked via trace plots, and posterior means were taken after 100,000 iterations following a burn-in of 20,000 iterations.

Table 15 shows the probability of best treatment with and without the meta-regression covariate. The interaction effect was found to be statistically important and supported the hypothesis of increased treatment response with increased disease duration. When including the effect of mean disease duration, the probability that etanercept was the best treatment dropped from 0.961 to 0.678, indicating that treatments with higher reported response rates, including etanercept, may have appeared to be more efficacious due to the inclusion of studies enrolling patient populations with longer disease durations.

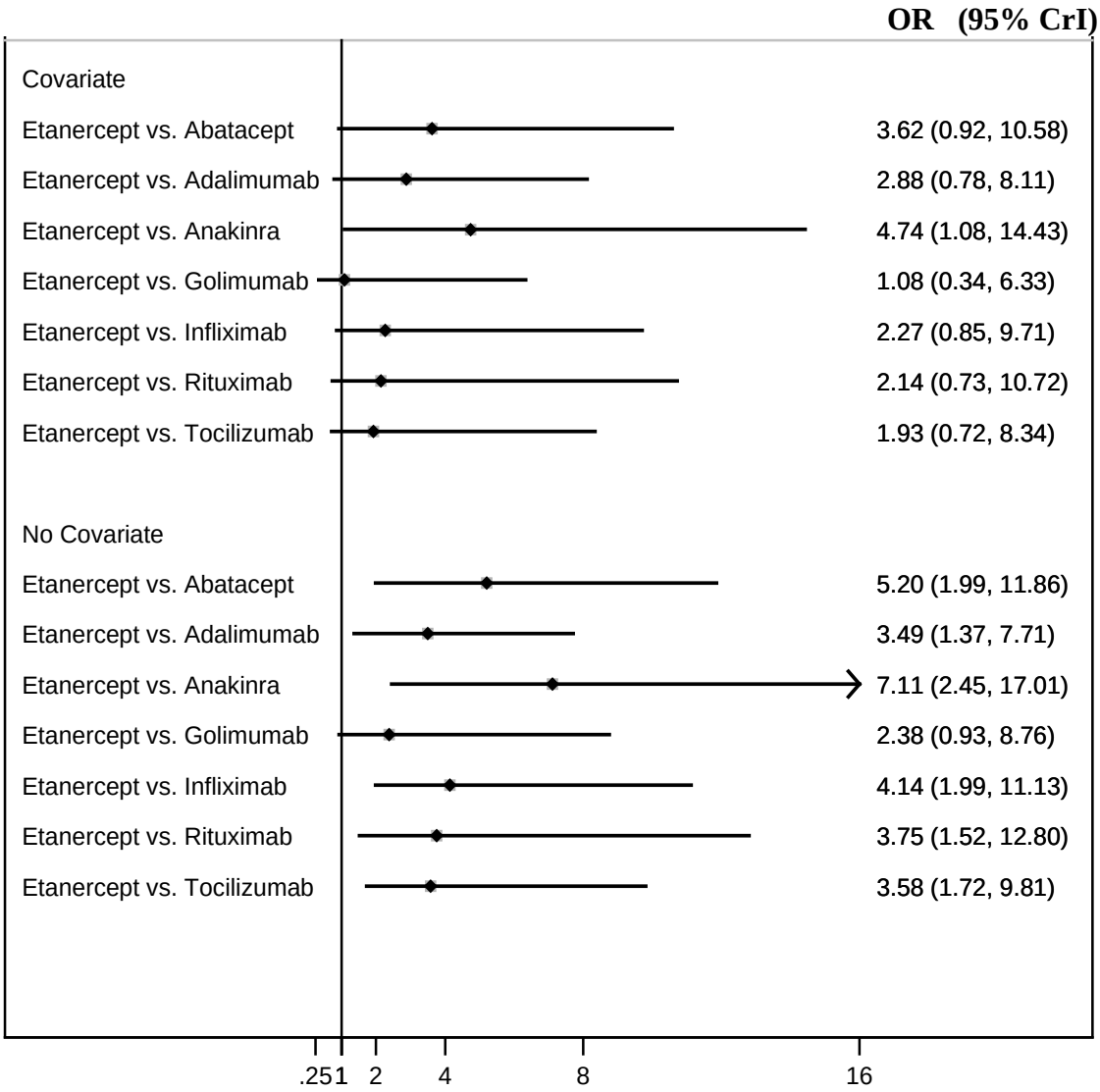
Table 15. Interaction effect and probability of best treatment

Probability of Best Treatment	No Covariate	With Disease Duration Covariate
Interaction Effect	NA	0.093 (95% CrI: 0.005 to 0.182)
Abatacept	0.001	0.001
Adalimumab	0.003	0.007
Anakinra	0.000	0.000
Etanercept	0.961	0.678
Golimumab	0.033	0.286
Infliximab	0.000	0.002
Rituximab	0.003	0.015
Tocilizumab	0.001	0.010

CrI = credible interval; NA = not applicable

Figure 11 presents odds ratios for etanercept response compared with other biologic DMARDs from the model with and without the covariate. When examining the relative effects of etanercept compared with the other biologic DMARDs, we see that when controlling for years of disease severity, the relative efficacy of etanercept is reduced. While trending in the direction of greater efficacy compared with all of the other biologic DMARDs, statistical significance was only reached for the comparison with anakinra. However, even with the muted effect of etanercept in the model controlling for disease duration level, when looking at the probability of best treatment it appears the only other biologic with competing efficacy is golimumab. When predicting the estimated odds ratios for different disease durations, we found that odds ratios for etanercept compared with placebo ranged from 7.35 with one year of RA to 14.99 with 10 years of the disease.

Figure 11. Odds ratios (95% credible interval) of treatment response for Etanercept compared with other biologic DMARDs, with and without disease duration covariate



CrI = credible interval; OR = odds ratio; vs. = versus

KQ 3. Stability of the Bayesian MTC Analyses for Different Numbers of Studies and Network Pattern Assumptions

The motivation for this question arose from uncertainty about how the number of studies per comparison and different network scenarios interact with the ability of the model to produce valid results. We aimed to explore this question through an empirical study with simulated data. We anticipated the simulation study would provide insights about how many studies are needed in order for the Bayesian MTC meta-analysis to produce valid results. We first describe the results from our equivalent efficacy dataset and then describe those from our superior efficacy dataset.

Equivalent Efficacy Scenario

Table 16 shows the mean, standard deviation, and standard error of the probability that each drug was the best in the MTC meta-analysis runs for the 1,000 sample datasets for each network pattern and each available number of studies for the equivalent efficacy scenario. In the equivalent efficacy scenario, the pre-determined mean response (standard deviation) of drug one (placebo) was 0.10 (0.03); the pre-determined mean response (standard deviation) for each of drugs two, three, and four was 0.50 (0.05).

Table 16. Simulation results: Probability of best treatment under equivalent efficacy scenario

Network Pattern	Number of Studies for Each Drug-Drug Comparison	Drug 1 (placebo) Mean (SD, SE)	Drug 2 Mean (SD, SE)	Drug 3 Mean (SD, SE)	Drug 4 Mean (SD, SE)
Star	1	0.01 (0.004, 0.004)	0.34 (0.045, 0.045)	0.33 (0.045, 0.045)	0.32 (0.048, 0.048)
	2	2×10^{-3} (2×10^{-4}, 1×10^{-4})	0.34 (0.115, 0.081)	0.34 (0.112, 0.079)	0.33 (0.116, 0.082)
	3	1×10^{-6} (1×10^{-5}, 6×10^{-6})	0.33 (0.131, 0.076)	0.33 (0.133, 0.077)	0.34 (0.136, 0.079)
	5	0.00 (0.000, 0.000)	0.33 (0.144, 0.064)	0.34 (0.141, 0.063)	0.33 (0.139, 0.062)
	10	0.00 (0.000, 0.000)	0.33 (0.153, 0.048)	0.34 (0.150, 0.047)	0.33 (0.150, 0.047)
Loop	1	0.03 (0.009, 0.009)	0.24 (0.054, 0.054)	0.34 (0.049, 0.049)	0.38 (0.040, 0.040)
	2	5×10^{-4} (5×10^{-4}, 4×10^{-4})	0.27 (0.152, 0.107)	0.33 (0.140, 0.099)	0.40 (0.116, 0.082)
	3	3×10^{-6} (2×10^{-5}, 1×10^{-5})	0.27 (0.174, 0.100)	0.33 (0.168, 0.097)	0.40 (0.132, 0.076)
	5	0.00 (0.000, 0.000)	0.27 (0.190, 0.085)	0.33 (0.180, 0.080)	0.40 (0.139, 0.062)
	10	0.00 (0.000, 0.000)	0.27 (0.201, 0.064)	0.33 (0.190, 0.060)	0.40 (0.150, 0.047)
One Closed Loop	1	0.01 (0.004, 0.004)	0.25 (0.077, 0.077)	0.25 (0.077, 0.077)	0.49 (0.004, 0.004)
	2	1×10^{-4} (2×10^{-4}, 1×10^{-4})	0.26 (0.126, 0.089)	0.25 (0.126, 0.089)	0.49 (0.005, 0.004)
	3	7×10^{-7} (1×10^{-5}, 6×10^{-6})	0.25 (0.137, 0.079)	0.26 (0.137, 0.079)	0.49 (0.005, 0.003)
	5	0.00 (0.000, 0.000)	0.25 (0.146, 0.065)	0.26 (0.146, 0.065)	0.49 (0.006, 0.003)
	10	0.00 (0.000, 0.000)	0.25 (0.148, 0.047)	0.26 (0.148, 0.047)	0.49 (0.006, 0.002)
Ladder	1	0.11 (0.019, 0.019)	0.30 (0.054, 0.054)	0.24 (0.062, 0.062)	0.35 (0.053, 0.053)
	2	0.003 (0.002, 0.001)	0.37 (0.169, 0.120)	0.28 (0.186, 0.132)	0.36 (0.168, 0.119)
	3	4×10^{-5} (1×10^{-4}, 6×10^{-5})	0.37 (0.195, 0.113)	0.28 (0.218, 0.126)	0.35 (0.195, 0.113)
	5	0.00 (0.000, 0.000)	0.36 (0.209, 0.093)	0.29 (0.234, 0.105)	0.35 (0.203, 0.091)
	10	0.00 (0.000, 0.000)	0.36 (0.225, 0.071)	0.30 (0.246, 0.078)	0.34 (0.215, 0.068)

SD = standard deviation; SE = standard error

In the equivalent efficacy scenario, we would expect to find that there are no differences between drugs two, three, and four, while drug one (placebo) should clearly be found the least efficacious. Table 17 shows the bias of the simulation results, which is the difference between observed and expected probabilities. For this data scenario, the bias ranged from 0.00 to 0.16, the latter being mainly in the one closed loop pattern. In the following bullets, we summarize our findings:

- For the star and ladder network patterns the correct conclusion was generally supported for each scenario (with varying numbers of studies for each drug-drug comparison). Even in cases where one drug had a slightly higher probability than expected, the difference was not sufficiently large to lead to the wrong conclusion (that one drug was superior to other drugs).
- The loop and one closed loop patterns had higher predicted means for the fourth drug in the network, even though drug four is not more efficacious than drugs two or three. The one closed loop pattern found drug four to be the best drug in almost half of the iterations.

- Findings of analyses when one study was available for each comparison were generally very similar to findings when more studies (two, three, five, or ten) were available. Differences between results with one study available and those with ten studies available were often 0.01 or less and were never greater than 0.06 (for drugs two and three in the ladder pattern).

Table 17. Bias of simulation results: Difference between observed and expected probability of best treatment under equivalent efficacy scenario

Expected Value		**0.00**	**0.33**	**0.33**	**0.33**
Network pattern	Number of Studies for Each Drug-Drug Comparison	Drug 1 (placebo) Bias	Drug 2 Bias	Drug 3 Bias	Drug 4 Bias
Star	1	0.01	0.01	0.00	0.01
	2	0.00	0.01	0.01	0.00
	3	0.00	0.00	0.00	0.01
	5	0.00	0.00	0.01	0.00
	10	0.00	0.00	0.01	0.00
Loop	1	0.03	0.09	0.01	0.05
	2	0.00	0.06	0.00	0.07
	3	0.00	0.06	0.00	0.07
	5	0.00	0.06	0.00	0.07
	10	0.00	0.06	0.00	0.07
One Closed Loop	1	0.01	0.08	0.08	0.16
	2	0.00	0.07	0.08	0.16
	3	0.00	0.08	0.07	0.16
	5	0.00	0.08	0.07	0.16
	10	0.00	0.08	0.07	0.16
Ladder	1	0.11	0.03	0.09	0.02
	2	0.00	0.04	0.05	0.02
	3	0.00	0.04	0.05	0.02
	5	0.00	0.03	0.04	0.02
	10	0.00	0.03	0.03	0.01

Superior Efficacy Scenario

Table 18 shows the mean, standard deviation, and standard error of the probability that each drug was the best in the MTC meta-analysis runs for the 1,000 sample datasets for each network pattern and each available number of studies for the superior efficacy scenario. In the superior efficacy scenario, the pre-determined mean response (standard deviation) for each of drugs one, two, and three was 0.20 (0.04); the predetermined mean response (standard deviation) of drug four was 0.80 (0.04).

Table 18. Simulation results: Probability of best treatment under superior efficacy scenario

Network Pattern	Number of Studies for Each Drug-Drug Comparison	Drug 1 Mean (SD, SE)	Drug 2 Mean (SD, SE)	Drug 3 Mean (SD, SE)	Drug 4 Mean (SD, SE)
Star	1	0.03 (0.009, 0.009)	0.13 (0.016, 0.016)	0.13 (0.017, 0.017)	0.71 (0.023, 0.023)
	2	8×10^{-4} (0.008, 0.006)	0.01 (0.006, 0.004)	0.010 (0.006, 0.004)	0.98 (0.012, 0.008)
	3	2×10^{-5} (5×10^{-4}, 3×10^{-4})	2×10^{-4} (5×10^{-4}, 3×10^{-4})	0.0002 (0.0005, 0.0002)	0.9995 (0.001, 6×10^{-4})
	5	0.00 (0.000, 0.000)	4×10^{-7} (7×10^{-6}, 3×10^{-6})	9×10^{-7} (9×10^{-6}, 4×10^{-6})	0.9999 (1×10^{-5}, 5×10^{-6})
	10	0.00 (0.000, 0.000)	0.00 (0.000, 0.000)	0.00 (0.000, 0.000)	1.00 (0.00, 0.00)
Loop	1	0.05 (0.011, 0.011)	0.08 (0.016, 0.016)	0.16 (0.020, 0.020)	0.71 (0.025, 0.025)
	2	0.001 (0.001, 7×10^{-4})	0.004 (0.004, 0.003)	0.01 (0.010, 0.007)	0.98 (0.014, 0.10)
	3	2×10^{-5} (7×10^{-5}, 4×10^{-5})	0.0001 (0.0003, 0.0002)	0.0005 (0.001, 6×10^{-4})	0.999 (0.001, 6×10^{-4})
	5	0.00 (0.000, 0.000)	0.00 (0.000, 0.000)	1.4×10^{-6} (1×10^{-5}, 5×10^{-6})	0.99999 (1×10^{-5}, 5×10^{-6})
	10	0.00 (0.000, 0.000)	0.00 (0.000, 0.000)	0.00 (0.000, 0.000)	1.00 (0.00, 0.00)
One Closed Loop	1	0.17 (0.076, 0.076)	0.17 (0.079, 0.079)	0.17 (0.077, 0.077)	0.49 (0.003, 0.003)
	2	0.17 (0.118, 0.083)	0.17 (0.119, 0.084)	0.17 (0.120, 0.085)	0.50 (0.005, 0.004)
	3	0.16 (0.129, 0.074)	0.17 (0.131, 0.076)	0.17 (0.129, 0.074)	0.50 (0.005, 0.003)
	5	0.16 (0.132, 0.059)	0.17 (0.133, 0.059)	0.16 (0.131, 0.059)	0.50 (0.005, 0.002)
	10	0.17 (0.143, 0.045)	0.17 (0.137, 0.043)	0.16 (0.136, 0.043)	0.50 (0.007, 0.002)
Ladder	1	0.17 (0.023, 0.023)	0.09 (0.018, 0.018)	0.05 (0.011, 0.011)	0.69 (0.030, 0.030)
	2	0.01 (0.009, 0.006)	0.004 (0.004, 0.003)	0.001 (0.001, 7×10^{-4})	0.98 (0.013, 0.009)
	3	6×10^{-4} (0.001, 6×10^{-4})	1×10^{-4} (0.0003, 0.0002)	2×10^{-5} (7×10^{-5}, 4×10^{-5})	0.999 (0.001, 6×10^{-4})
	5	1×10^{-6} (1×10^{-5}, 5×10^{-6})	1×10^{-7} (3×10^{-6}, 1×10^{-6})	0.00 (0.000, 0.000)	1.0 (0.00, 0.00)
	10	0.00 (0.000, 0.000)	0.00 (0.000, 0.000)	0.00 (0.000, 0.000)	1.0 (0.00, 0.00)

SD = standard deviation; SE = standard error

In the superior data scenario, we would expect to find no differences between drugs one, two, and three, while drug four should clearly be found to be the best treatment. Table 19 shows the bias of the simulation results for this data scenario. The differences between observed and expected probabilities ranged from 0.00 to 0.51, again with the bias in the one closed loop pattern being the most pronounced. In the following bullets, we summarize our findings:

- For each scenario (with varying numbers of studies and network patterns), the correct conclusion was generally supported, with one notable exception. In the one closed loop pattern, the model failed to definitively find drug four to be the most efficacious.
- Generally, there were no significant differences between the estimates generated with 2, 3, 5, or 10 studies available for each comparison.
- For the scenarios with one study available per comparison, although probabilities of best treatment differed numerically compared with those scenarios with more studies

available, the findings were still indicative of superior efficacy of drug four. For drug four, differences between estimates (of the probability of being the best treatment) from analyses with one study available per comparison and those from analyses with two studies available per comparison ranged up to 0.29.

The estimates from tables 16 and 18 are also presented graphically, as histograms in Appendix F.

Table 19. Bias of simulation results: Difference between observed and expected probability of best treatment under superior efficacy scenario

Expected Value		0.00	0.00	0.00	1.00
Network Pattern	**Number of Studies for Each Drug-Drug Comparison**	**Drug 1 (placebo) Bias**	**Drug 2 Bias**	**Drug 3 Bias**	**Drug 4 Bias**
Star	1	0.03	0.13	0.13	0.29
	2	0.00	0.01	0.01	0.02
	3	0.00	0.00	0.00	0.00
	5	0.00	0.00	0.00	0.00
	10	0.00	0.00	0.00	0.00
Loop	1	0.05	0.08	0.16	0.29
	2	0.00	0.00	0.01	0.02
	3	0.00	0.00	0.00	0.00
	5	0.00	0.00	0.00	0.00
	10	0.00	0.00	0.00	0.00
One Closed Loop	1	0.17	0.17	0.17	0.51
	2	0.17	0.17	0.17	0.50
	3	0.16	0.17	0.17	0.50
	5	0.16	0.17	0.16	0.50
	10	0.17	0.17	0.16	0.50
Ladder	1	0.17	0.09	0.05	0.31
	2	0.01	0.00	0.00	0.02
	3	0.00	0.00	0.00	0.00
	5	0.00	0.00	0.00	0.00
	10	0.00	0.00	0.00	0.00

Discussion

Main Findings by Key Question (KQ)

In this report, we addressed three KQs using real-world bodies of trial literature (for KQs 1 and 2) from recent comparative effectiveness reviews (CERs) of second-generation antidepressants (SGAs) and treatments for rheumatoid arthritis (RA) and using simulated data (for KQ 3). Below, we summarize the main findings by KQ. We then address the implications, limitations, research gaps, and conclusions.

KQ 1. Bayesian Mixed Treatment Comparison (MTC) Methods Compared With Several Frequentist Indirect Methods, and Performance for Different Types of Evidence Network Patterns

We compared the results of the Bayesian MTC approach with results of several frequentist analytic methods—specifically, frequentist meta-regression, the Bucher method (adjusted indirect comparisons), and frequentist logistic regression. Of note, the frequentist methods used are not the analogues of the Bayesian methods implemented; i.e., we did not compare findings with a frequentist approach to MTC. Our choice of methods to compare, and our manner of conducting the analyses, was based on our judgment regarding the methods most commonly used by analysts conducting CERs. Our results for these comparisons, such as those related to precision, are not necessarily generalizable to other datasets (e.g., comparative effectiveness of other medication classes).

We found one of the main differences between Bayesian MTC meta-analysis and the typically-applied approach for frequentist meta-regression and the Bucher method to be that the Bayesian MTC approach was able to calculate a result for all drug-drug comparisons of interest whereas the other methods were unable to produce a result for many comparisons of interest. This was not surprising; it is because Bayesian MTC meta-analysis is able to produce a result for all comparisons in a connected network, whereas we only calculated results for these other frequentist indirect methods when there was a common comparator. Our results showed that some frequentist methods were unable to compute results for substantial proportions of the drug-drug comparisons of interest for some of the datasets. Despite that, our results may actually underestimate the proportion that one might determine for inability to calculate a result if running these analyses for a real-world CER because some guidance for certain methods would suggest not even attempting some of the analyses without a certain minimum number of studies. For example, we conducted meta-regression for all comparisons with a direct common comparator for which the statistical program was able to calculate a result; however, some research indicates that a certain minimal number of studies is needed for the meta-regression to produce a reliable result.[33, 45] In other words, although technically we could produce results for meta-regression and the Bucher method for all of the comparisons for which we reported results, many of those, perhaps, should not be calculated if conducting a CER for the purpose of informing decisionmaking.

We were able to calculate results for all methods (Bayesian MTC and the three frequentist methods) for the RA dataset, but not for the SGA dataset. This is not surprising given the geometry of the two networks. The RA network is largely a star pattern with all treatments

connected via placebo and with only one head-to-head study. The data is thus set up to allow any of the methods to determine results for each drug-drug comparison of interest (because of the common placebo comparator for each drug-drug comparison of interest). The SGA network, however, is a much more complex network. In this complex network, there were often too few studies with a common comparator to allow some frequentist methods to calculate results for many comparisons of interest.

Regarding various network patterns, the frequentist meta-regression and the Bucher method (in the manner we applied them) were least likely to be able to compute results for comparisons of interest for ladder patterns and most likely to be able to compute results for star patterns. The geometry of ladder patterns, by definition, does not include common comparators for most comparisons of interest, as each treatment will be linked to two other treatments at most.

In the cases of comparisons for which we were able to produce results, the majority of these results were in general agreement (i.e., they all found no statistically significant difference [for frequentist analyses] or no important difference [for Bayesian analyses] or they found a significant or important difference favoring the same drug). However, for each of the full networks, and for each outcome of interest, there were instances when some of the frequentist methods produced results that did not agree with those of the Bayesian MTC analysis. In each case, we do not know for certain which of the results represents the truth; instead, we just know that the findings do not agree. One might presume that the Bayesian MTC analysis is more likely to approximate the truth because it is able to incorporate more information from the full dataset; however, this assumption has not been proven. Validating such a conclusion would require a large number of real-world examples or data simulations using numerous situations and assumptions.

When considering precision, we speculated whether the Bayesian MTC method might generally produce more precise results because of its ability to incorporate all data (from both direct head-to-head trials and placebo-controlled trials). Our results did not find this to be the case. Across all the different network patterns (including the full networks), the logistic regression method produced the most precise result for a greater number of drug-drug comparisons than any of the other methods, followed by the Bayesian MTC method, and then the Bucher method. The meta-regression method fairly consistently produced the least precise result. The differences in precision between the various analyses were sometimes very small (on the order of hundredths), but were larger in some cases (e.g., on the order of 2 to 5 points different for an odds ratio), and very large in rare cases (e.g., a difference of 20 to 50).

For some specific networks (5 of the 16 networks or subcomponent networks that we analyzed), the Bayesian MTC results were the most precise. From our findings, we are unable to determine if any particular network geometry makes the Bayesian MTC method more or less likely to produce the most precise results. However, we should note that greater precision is only a good thing if the treatment effect estimated by the method is reflective of the truth (which is not something that we can confirm for these KQ 1 analyses as they are based on actual data rather than simulations).

KQ 2. Meta-Regression With Bayesian MTC Meta-Analysis

For the first meta-regression, we hypothesized that differences in efficacy of SGAs may exist for older adults (≥55 years) than for younger adults. We explored this hypothesis through a subgroup meta-regression. We found a trend toward lack of efficacy or lesser efficacy for all of

the SGAs in older adults. However, this finding could be due to case mix, as studies of older adults may enroll subjects with less severe depression at baseline.

For the second meta-regression, we hypothesized that treatment response may differ for patients with longer disease duration. We speculated that longer disease duration was related to more severe RA (and potentially more room for improvement). We explored this hypothesis through a continuous covariate meta-regression with mean disease duration as the continuous covariate. We focused on the response to etanercept to determine whether controlling for this covariate might alter findings because our MTC meta-analysis showed etanercept yielding a greater treatment response than most other biologic DMARDs. We found a trend toward greater efficacy for those with greater mean disease duration. In addition, we found that the superior relative efficacy of etanercept compared with other biologic DMARDs was reduced; furthermore, it was no longer significantly greater (based on the 95% credible intervals) than most of the other biologic DMARDs (with the exception of anakinra). However, reporting bias might underlie this finding because the only 2 trials (out of 31) that were not included in the analysis (because they did not report mean disease duration) compared etanercept and placebo.

KQ 3. Stability of the Bayesian MTC Analyses for Different Numbers of Studies and Network Pattern Assumptions

Our simulations for KQ 3 revealed several important findings. We note that the differences hypothesized between drugs and placebo (for the equivalent efficacy scenario) and between drug 4 and drugs 1 through 3 (for the superior efficacy scenario) were relatively large, and our findings do not necessarily apply to other scenarios (e.g., smaller differences between treatments). First, the simulations validated the ability of the Bayesian MTC meta-analyses to produce results that would yield conclusions that reflect the truth for two scenarios—one with three medications of equivalent efficacy and a placebo arm and one with one medication of superior efficacy. The analyses were generally able to produce findings reflecting the underlying truth with only one study per comparison.

Second, for the equivalent efficacy scenario, results changed very little based on the number of studies available for each comparison. For the superior efficacy scenario, however, we found larger differences between scenarios with one available study for each comparison and scenarios with two studies for each comparison, and similar results with little variation between two and 10 studies per comparison.

Third, networks with a one closed loop pattern did not perform as expected and often produced results that might yield inaccurate conclusions. These networks produced results for the equivalent efficacy scenario that might lead investigators to conclude one of the drugs (drug four) to have the greatest likelihood of efficacy, regardless of the number of studies available for each comparison (i.e., the accuracy of results did not improve even with 10 studies per comparison). In contrast, for the superior efficacy scenario, the one closed loop found the correct drug to have the greatest likelihood of efficacy, but with much lower probability compared with the other network patterns (approximately 50% vs. well over 90%), regardless of the number of studies per comparison. We are uncertain as to why this occurred and it was an unexpected finding. Perhaps it could be related to the position of the superior drug within the network (it was the “dangling” treatment in our simulations, and thus the least connected with the rest of the network; see the following section). A drug could have three potentially different positions, even in our relatively simple one closed loop network. The superior drug could be the “dangling” treatment (as in our simulation), it could be connected to everything else (like drug 3 in the

figure in the following section), or it could be connected to two of the other three drugs. Regardless, future research should attempt to explore this further and to reproduce the findings using different simulated data.

Implications for Comparative Effectiveness Reviews

Investigators have a number of analytic strategies from which to choose when conducting indirect comparisons (Bayesian MTC meta-analysis, frequentist meta-regression, the Bucher method, logistic regression, and others). Our findings indicate that the choice of method can have an important impact on the results. Although the various methods we compared using our two real-world datasets found results that generally agreed for the majority of drug-drug comparisons, we also found some instances where results differed for the same comparison. Even though we do not know with certainty which method is correct from these real-world comparisons, our findings identified several advantages of the Bayesian MTC approach (Table 20).

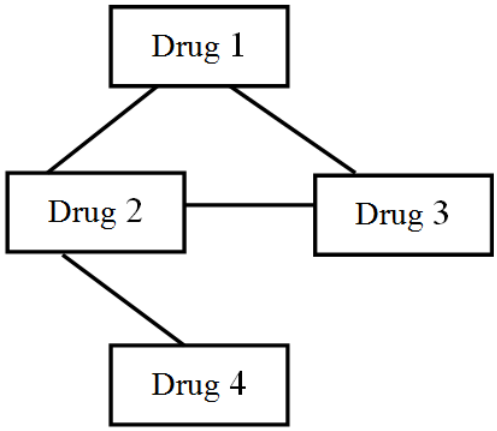
Table 20. Advantages and disadvantages of the Bayesian MTC approach

Advantages	Disadvantages
Able to incorporate both direct evidence (from head-to-head trials) and indirect evidence (e.g., placebo-controlled trials) into a single analysis	Less accessible to many investigators because the analyses are usually run using software that is unfamiliar or less familiar to many of them (usually run using WinBUGS)
Able to produce results for all comparisons of interest within a connected network (even for ladder network patterns or complex networks that limit the ability of other methods to get any results for some comparisons of interest)	Requires greater statistical expertise than some other methods
Able to directly calculate the probability that each drug is the best treatment	Further research is needed to evaluate performance in specific scenarios[a]
Able to adjust for correlations within multi-arm trials	Might not produce accurate results for one closed loop networks
Able to incorporate meta-regression to assess heterogeneity (e.g., for subgroups or to control for covariates), all within one model	Possibly sensitive to the prior probabilities chosen (therefore recommended to generally use flat priors)[46]
Appears to produce valid, accurate results for star and ladder network patterns	

[a]Such as those with various numbers of studies available per comparison, various sample sizes of studies, and complex network patterns.

Our simulations (KQ 3) supported the validity of the Bayesian MTC method for star and ladder network patterns. However, they raised some concerns about the validity of the Bayesian MTC method for one closed loop networks and possibly for loop patterns (albeit lesser concerns), as results did not converge on the pre-determined truth. These findings demonstrate the need for additional exploration of more hypothetical scenarios, and analysts should be cognizant of the fact that Bayesian MTC methods may not produce accurate results for one closed loop networks such as the one used in our simulation (Figure 12). We are uncertain as to why the results did not converge on the truth; we wonder whether a "dangling" treatment (where 1 treatment is peripheral to the rest of the network, like drug 4 in the figure) might reduce accuracy.

Figure 12. One closed loop network pattern used in simulation



Surprisingly, our simulations did not find much difference between scenarios when one study was available for each comparison and those when more studies (2, 3, 5, or 10) were available. We initially expected that results from analyses with one study available for each comparison would be much less likely to approximate the underlying truth, but found that this was not the case. We previously wondered whether Bayesian MTC meta-analyses should not be attempted when there is only one study per comparison, hypothesizing that results would be much less accurate. However, this hypothesis was not supported by our simulations. It is possible that our choice of relatively large differences between drugs and placebo (for the equivalent efficacy scenario) and between drug 4 and drugs 1 through 3 (for the superior efficacy scenario) influenced this finding. In other words, if differences were smaller, analyses with one study available for each comparison could be less likely to approximate the underlying truth.

Limitations

Our findings from KQs 1 and 2 are not necessarily generalizable to treatment networks for other medications or diseases. Some of the findings may be a consequence of the underlying body of evidence (e.g., the number of studies for each comparison, the variation of findings in studies making the same comparison) and all of its inherent biases (e.g., selective outcome reporting, publication bias) rather than the network pattern, per se.

In addition, our selection of various sub-component network patterns was based on maximizing the amount of data available for each sub-network. Due to limited time and resources, we could not evaluate every possible network pattern within each dataset. Other approaches to choosing network patterns could also have been informative (e.g., based on entry of drugs into the market, or chronologically by publication date). But, for the purposes of this report we felt that an approach based on maximizing the amount of data available was reasonable. Due to the nature of the sub-component network patterns (i.e., that they include only some of the available evidence from a body of literature), the findings for comparisons of sub-components of the full networks are for exploration of the various types of analyses only and should not be used to inform clinical decisions.

In KQ 1, we reported several measures to compare the findings of the Bayesian MTC method with various frequentist methods. The measures have some limitations. For example, regarding the *number of comparisons* for which each frequentist method was *unable to produce a result*, our results likely underestimate the numbers that some analysts might determine if running these

analyses for a CER—because some guidance would suggest not even attempting some analyses without a certain minimum number of studies. For example, we conducted meta-regression for all comparisons that the statistical program was able to calculate a result, but it has been suggested that a certain minimal number of studies per covariate is needed for the meta-regression to produce a reliable result.[33, 45] In other words, the gap would actually be larger between the Bayesian MTC and meta-regression for the number of comparisons the method was unable to produce a result.

Another measure we used was the percent agreement between results of the Bayesian MTC method and those of each frequentist method. We considered results to agree if both methods produced a non-statistically significant (for frequentist analyses) or an unimportant (for Bayesian analyses) result for the comparison or if both analyses found a statistically significant or important result favoring the same treatment. This is an oversimplification of how results would be interpreted in a CER. For example, we did not determine whether the ultimate conclusions about comparative effectiveness or the strength of evidence grades would agree.

Our choice of analytic methods to compare was based on our judgment regarding the methods most commonly considered by analysts conducting CERs. In addition, we selected frequentist methods with some evidence to support their validity. However, we did not compare findings with the frequentist network meta-analysis method (i.e., Lumley method).[23] Our experience indicates that it is much more rarely used than the other methods, and it is generally felt that the Bayesian MTC approach has several advantages over frequentist network meta-analysis.

For Bayesian MTC analyses, we did not explore sensitivity analyses with various uninformative priors or with informative priors. However, we do not believe that the current state of the literature (i.e., the literature on SGAs and biologic DMARDs for RA) would support using informative priors. In addition, we did not include inconsistency models to test model assumptions.

For KQ 2, both of our meta-regressions rely on averages taken over patients in the trials. As such, ecological bias is a potential limitation.

Our simulations for KQ 3 were limited by the assumptions that we made to develop the scenarios. Our simulations did not examine many scenarios observed in real-world networks. For example, we set each scenario to include an equal number of studies for each drug-drug comparison, we set the sample size at 100 subjects for each study, and we did not include any multi-arm studies. In addition, we used a non-random network structure, and we did not simulate incoherence. Findings might differ if one investigated network patterns with varying numbers of studies for different comparisons or with larger or smaller (or varying) sample sizes, or if one incorporated additional complexity in the simulations.

For our simulations in KQ 3, we chose to output the probability that each drug was the best (i.e., most efficacious) for ease of presentation and to use an outcome recognizable to analysts familiar with Bayesian MTC. Being able to produce probability rankings is an advantage of the Bayesian MTC approach. However, one could argue that using mean treatment response would have been more appropriate, because it would have allowed outputs that were directly comparable with the underlying truths that we specified.

Finally, data were too sparse to include some analyses we had aimed to include at the outset. We wanted to run analyses for a continuous outcome from the RA dataset (mean change in Health Assessment Questionnaire score), as we had done for the SGA dataset, but too few studies reported sufficient data.

Future Research

We identified several issues that future research could address (Table 21). Many of these issues are related to further exploring the findings and limitations mentioned above.

Table 21. Possible targets for future research, by Key Question (KQ)

KQ	Potential Future Research
1	Our findings were based on just two real-world datasets. To determine how well our findings hold up for various real-world situations, it will be important to conduct similar analyses for other existing networks. Many (perhaps 50) publications have used Bayesian MTC meta-analyses; those could be explored.
1	For all of our Bayesian MTC meta-analyses, we ran a certain number of simulations (20,000) that were discarded and then used an additional 100,000 simulations in estimating the posterior probabilities. The impact of varying the number of simulations on the resulting findings is uncertain. Future research could address this impact to inform the most appropriate number of simulations.
1	There were numerous options of various network patterns we could have selected. Future research could use an approach similar to cumulative meta-analysis and explore how the evidence evolved over time thus, choosing network patterns based on the chronology of study publication.
1	Our analyses compared the results of various analytic approaches for both continuous and dichotomous outcomes, but we did not explore competing risk outcomes.[47] Future research could include competing risk outcomes.
2	We found a trend toward lack of efficacy or less efficacy for all of the SGAs in older adults. However, this finding could possibly be due to case mix, as studies of older adults may enroll subjects with less severe depression at baseline. Future analyses could explore this.
2	We found a trend toward greater efficacy for those with greater mean disease duration. However, this finding could possibly be due to case mix, or other factors that we did not consider. Future analyses could explore this further.
2	For our meta-regression models exploring (1) subgroup effects with a binary covariate and (2) interaction effects with a continuous covariate, we used models with the same interaction effect for all treatments. Other models have been described with independent, treatment-specific interactions and exchangeable, related treatment-specific interactions.[5] Future research could explore whether these other models would significantly alter findings.
3	For our simulations, we assumed a sample size of 100 subjects per study. Findings might differ if one investigated network patterns with larger or smaller (or varying) sample sizes.
3	For our simulations, we assumed that each network had an equal number of studies available for each comparison of interest. Findings might differ if one investigated network patterns with varying numbers of studies for different comparisons.
3	For our simulations, we used four network patterns (star, loop, one closed loop, and ladder). Many additional variations of these patterns could be assessed to determine if findings differ. In addition, complex networks that resemble real-world datasets could be developed to attempt to validate Bayesian MTC methods for specific treatments.[a]
3	Our results raised some concerns about the validity of the Bayesian MTC method for one closed loop networks, and possibly for loop patterns (but less concern). Future simulations could include additional variation of the underlying data and complexity of network patterns to determine whether these findings are consistent.
3	Our findings for one closed loop networks raise questions about the impact of a single "dangling" treatment on the results. Future simulation studies could explore this further to determine whether such "dangling" treatments should or should not be routinely included in Bayesian MTC meta-analyses. For our superior efficacy scenario, the more efficacious drug was the "dangling" treatment and future simulations could explore whether its position in the network would alter the findings.
3	For our simulations, we chose to output the probability that each drug was the best. Future research could explore whether using different outputs (e.g. mean treatment response) would either alter conclusions or uncover additional findings.

SGAs = second-generation antidepressants

[a]For example, we could develop a simulated dataset that resemble the second-generation antidepressants or the biologic DMARDs real-world data—including setting the pre-determined truth to match our real-world findings and setting the number and sample sizes of included studies to match the real-world data. Then, we could run thousands of simulations to determine the validity of Bayesian MTC methods for a simulation that very closely matches a real-world dataset for a particular treatment and health condition.

Conclusions

Bayesian MTC methods allow investigators to calculate results for many more comparisons of interest for some network patterns, including ladders and complex networks, than frequentist meta-regression or the Bucher method, in the manner that they are typically applied. When Bayesian MTC methods and various frequentist methods are each able to calculate results (as typically applied), the findings are usually in general agreement. However, findings differ for a small proportion (less than 10%) of comparisons, which could lead to differences in conclusions.

Our simulations support the validity of Bayesian MTC methods for star and ladder network patterns, but raise some concerns about one closed loop network patterns, and possibly about loop patterns (but less concern). Simulations generally found similar results for scenarios when 1 study was available for each comparison and those when more studies (2, 3, 5, or 10) were available.

Further research is needed to explore additional real-world datasets and simulated data to determine if these findings are reproducible or generalizable and to better understand the accuracy and validity of Bayesian MTC methods for various scenarios. We hope this research will inform Evidence-based Practice Centers and others conducting CERs about Bayesian MTC methods and help to inform additional research and guidance for the use of these methods.

References

1. Agency for Healthcare Research and Quality. What Is Comparative Effectiveness Research. Rockville, MD. http://www.effectivehealthcare.ahrq.gov/index.cfm/what-is-comparative-effectiveness-research1/. Accessed May 3, 2012.
2. Bucher HC, Guyatt GH, Griffith LE, et al. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol. 1997 Jun;50(6):683-91. PMID: 9250266.
3. Glenny AM, Altman DG, Song F, et al. Indirect comparisons of competing interventions. Health Technol Assess. 2005 Jul;9(26):1-134, iii-iv. PMID: 16014203.
4. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986 Sep;7(3):177-88. PMID: 3802833.
5. Dias S, Sutton AJ, Welton NJ, et al. NICE DSU Technical Support Document 3: Heterogeneity: subgroups, meta-regression, bias and bias-adjustment. Report by the Decision Support Unit. 2011. http://www.nicedsu.org.uk/TSD3%20Heterogeneity_final%20report_docx.pdf. Accessed May 4, 2012.
6. Dias S, Welton NJ, Sutton AJ, et al. A Generalised Linear Modelling Framework for Pairwise and Network Meta-analysis of Randomised Controlled Trials. Report by the Decision Support Unit. Updated August 2011. http://www.nicedsu.org.uk/TSD2%20ES%20-%2005_05_11_FINAL_updated%20Aug2011.pdf. Accessed May 4, 2012.
7. Jansen JP, Crawford B, Bergman G, et al. Bayesian meta-analysis of multiple treatment comparisons: an introduction to mixed treatment comparisons. Value Health. 2008 Sep-Oct;11(5):956-64. PMID: 18489499.
8. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med. 2004 Oct 30;23(20):3105-24. PMID: 15449338.
9. Cooper NJ, Peters J, Lai MC, et al. How valuable are multiple treatment comparison methods in evidence-based health-care evaluation? Value Health. 2011 Mar-Apr;14(2):371-80. PMID: 21296599.
10. Mills EJ, Bansback N, Ghement I, et al. Multiple treatment comparison meta-analyses: a step forward into complexity. Clin Epidemiol. 2011;3:193-202. PMID: 21750628.
11. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. Lancet. 2009 Feb 28;373(9665):746-58. PMID: 19185342.
12. Li T, Puhan MA, Vedula SS, et al. Network meta-analysis-highly attractive but more methodological research is needed. BMC Med. 2011;9:79.
13. Padwal R, Klarenbach S, Wiebe N, et al. Bariatric surgery: a systematic review and network meta-analysis of randomized trials. Obes Rev. 2011 Aug;12(8):602-21. PMID: 21438991.
14. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on indirect treatment comparisons good research practices: part 1. Value Health. 2011:417-28.
15. Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. Stat Med. 1996 Dec 30;15(24):2733-49. PMID: 8981683.
16. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. Journal of the American Statistical Association. 2006;101(474):447-59.
17. Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on indirect treatment comparisons good research practices: part 2. Value Health. 2011 Jun;14(4):429-37. PMID: 21669367.

18. Higgins JPT, Jackson D, Barrett JK, et al. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. Research Synthesis Methods. 2012;3(2):98-110.

19. Gartlehner G, Hansen RA, Morgan LC, et al. Second-Generation Antidepressants in the Pharmacologic Treatment of Adult Depression: An Update of the 2007 Comparative Effectiveness Review. (Prepared by the RTI International–University of North Carolina Evidence-based Practice Center, Contract No. 290-2007-10056-I.) AHRQ Publication No. 12-EHC012-EF Agency for Healthcare Research and Quality. Rockville, MD: December 2011. www.effectivehealthcare.ahrq.gov/reports/final.cfm

20. Donahue KE, Jonas DE, Hansen RA, et al. Drug Therapy for Rheumatoid Arthritis in Adults: An Update. Comparative Effectiveness Review No. 55. (Prepared by RTI-UNC Evidence-based Practice Center under Contract No. 290-02-0016-I.) AHRQ Publication No. 12-EHC025-EF. Rockville, MD: Agency for Healthcare Research and Quality; April 2012. www.effectivehealthcare.ahrq.gov/reports/final.cfm

21. Salanti G, Kavvoura FK, Ioannidis JP. Exploring the geometry of treatment networks. Ann Intern Med. 2008 Apr 1;148(7):544-53. PMID: 18378949.

22. Canadian Agency for Drugs and Technologies in Health (CADTH). Biological response modifier agents for adults with rheumatoid arthritis. Ottawa, ON; July 2010. http://cadth.ca/media/pdf/Biologics_for_RA_TRP_Final_Recommendations_e.pdf. Accessed May 4, 2012.

23. Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med. 2002 Aug 30;21(16):2313-24. PMID: 12210616.

24. Kasper S, de Swart H, Friis Andersen H. Escitalopram in the treatment of depressed elderly patients. Am J Geriatr Psychiatry. 2005 Oct;13(10):884-91. PMID: 16223967.

25. Roose SP, Sackeim HA, Krishnan KR, et al. Antidepressant pharmacotherapy in the treatment of depression in the very old: a randomized, placebo-controlled trial. Am J Psychiatry. 2004 Nov;161(11):2050-9. PMID: 15514406.

26. Schneider LS, Nelson JC, Clary CM, et al. An 8-week multicenter, parallel-group, double-blind, placebo-controlled study of sertraline in elderly outpatients with major depression. Am J Psychiatry. 2003 Jul;160(7):1277-85. PMID: 12832242.

27. Tollefson GD, Bosomworth JC, Heiligenstein JH, et al. A double-blind, placebo-controlled clinical trial of fluoxetine in geriatric patients with major depression. The Fluoxetine Collaborative Study Group. Int Psychogeriatr. 1995 Spring;7(1):89-104. PMID: 7579025.

28. Aletaha D, Strand V, Smolen JS, et al. Treatment-related improvement in physical function varies with duration of rheumatoid arthritis: a pooled analysis of clinical trial results. Ann Rheum Dis. 2008 Feb;67(2):238-43. PMID: 17644550.

29. Aletaha D, Ward MM. Duration of rheumatoid arthritis influences the degree of functional improvement in clinical trials. Ann Rheum Dis. 2006 Feb;65(2):227-33. PMID: 15975967.

30. Anderson JJ, Wells G, Verhoeven AC, et al. Factors predicting response to treatment in rheumatoid arthritis: the importance of disease duration. Arthritis Rheum. 2000 Jan;43(1):22-9. PMID: 10643696.

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

32. Follmann D, Elliott P, Suh I, et al. Variance imputation for overviews of clinical trials with continuous response. J Clin Epidemiol. 1992 Jul;45(7):769-73. PMID: 1619456.

33. Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]: The Cochrane Collaboration; 2011.

34. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20(1):37-46.

35. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74.

36. Borenstein M, Hedges LV, Higgins JPT, et al. Introduction to Meta-Analysis. Cambridge, UK: Wiley; 2009.

37. McCullagh P, Nelder JA. Generalized Linear Models, volume 37 of (Monographs on Statistics and Applied Probability. 2nd ed. London: Chapman & Hall; 1989.

38. Spiegelhalter DJ, Abrams KR, Myles J. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Chichester: Wiley; 2004.

39. Parmigiani G. Modeling in medical decision making: A Bayesian approach. Chichester: Wiley; 2002.

40. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Statistical Science. 1992;7(4):457-511.

41. Spiegelhalter DJ, Best NG, Carlin BP, et al. Bayesian measures of model complexity and fit. J Royal Statistical Society. 2002;64(4):583-639.

42. Carlin BP, Louis TA. Bayesian Methods for Data Analysis. 3rd ed. Boca Raton: Chapman & Hall; 2008.

43. Harbord RM, Higgins JPT. Meta-regression in Stata. Stata J. 2008;8(4):493-519.

44. Cooper NJ, Sutton AJ, Morris D, et al. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. Stat Med. 2009;28(14):1861-81.

45. Morton SC, Adams JL, Suttorp MJ, et al. Meta-regression Approaches: What, Why, When, and How? Technical Review 8 (Prepared by Southern California–RAND Evidence-based Practice Center, under Contract No 290-97-0001). AHRQ Publication No. 04-0033. Rockville, MD: Agency for Healthcare Research and Quality; March 2004.

46. Dias S, Welton NJ, Sutton AJ, et al. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials. last updated August 2011; 2011. http://www.nicedsu.org.uk. Accessed May 4, 2012.

47. Ades AE, Mavranezouli I, Dias S, et al. Network meta-analysis with competing risk outcomes. Value in Health. 2010;13(8).

Appendix A. WinBUGS Code Used in Bayesian Mixed Treatment Comparisons Meta-Analysis

The WinBUGS code used to conduct Bayesian MTC meta-analyses is given below. WinBUGS Version 1.4.3 was used for all analyses. The code was adapted from code developed for the NICE Evidence Synthesis Technical Series Documents 2 and 3.[1, 2]

KQ1: Random Effects Model for Dichotomous Data

```
# Binomial likelihood, logit link
# Random effects model for multi-arm trials

model{                          # *** PROGRAM STARTS

for(i in 1:ns){                 # LOOP THROUGH STUDIES
   w[i,1] <- 0    # adjustment for multi-arm trials is zero for control arm
   delta[i,1] <- 0           # treatment effect is zero for control arm
   mu[i] ~ dnorm(0,.0001)          # vague priors for all trial baselines
   for (k in 1:na[i]) {            # LOOP THROUGH ARMS
      r[i,k] ~ dbin(p[i,k],n[i,k]) # binomial likelihood
      logit(p[i,k]) <- mu[i] + delta[i,k]  # model for linear predictor
      rhat[i,k] <- p[i,k] * n[i,k] # expected value of the numerators

#Deviance contribution
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
         +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))        }

#  summed residual deviance contribution for this trial
   resdev[i] <- sum(dev[i,1:na[i]])
   for (k in 2:na[i]) {            # LOOP THROUGH ARMS
# trial-specific LOR distributions
      delta[i,k] ~ dnorm(md[i,k],taud[i,k])
# mean of LOR distributions (with multi-arm trial correction)
      md[i,k] <-  d[t[i,k]] - d[t[i,1]] + sw[i,k]
# precision of LOR distributions (with multi-arm trial correction)
      taud[i,k] <- tau *2*(k-1)/k
# adjustment for multi-arm RCTs
      w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])
# cumulative adjustment for multi-arm trials
      sw[i,k] <- sum(w[i,1:k-1])/(k-1)
    }
  }
totresdev <- sum(resdev[])          # Total Residual Deviance
d[1]<-0       # treatment effect is zero for reference treatment
# vague priors for treatment effects
for (k in 2:nt){  d[k] ~ dnorm(0,.0001) }
sd ~ dunif(0,5)     # vague prior for between-trial SD
tau <- pow(sd,-2)   # between-trial precision = (1/between-trial variance)

# pairwise ORs and LORs for all possible pair-wise comparisons, if nt>2
```

```
for (c in 1:(nt-1)) {
for (k in (c+1):nt) {
or[c,k] <- exp(d[k] - d[c])
lor[c,k] <- (d[k]-d[c])
}
}
# ranking on relative scale
for (k in 1:nt) {
rk[k] <- nt+1-rank(d[],k) # assumes events are “good”
# rk[k] <- rank(d[],k) # assumes events are “bad”
best[k] <- equals(rk[k],1) #calculate probability that treat k is best
}
}

# *** PROGRAM ENDS
```

KQ 1: WinBUGS Dataset for SGA Dichotomous

#Full Network (star, closed loop, loop plus one, and ladder are subsets)

#Description of data inputs

#ns = Number of studies

#nt = Number of treatments (including placebo)

#t[,x] = Treatment indicator

#r[,x] = Number achieving response on HAM-D (50% improvement of scores from baseline)

#n[,x]= Number of all randomized patients (ITT)

#na[] = Number of arms in study

list(ns=64, nt=14)

t[,1]	t[,2]	t[,3]	r[,1]	r[,2]	r[,3]	n[,1]	n[,2]	n[,3]	na[]
1	2	7	73	76	83	152	150	154	3
1	2	12	66	78	66	124	122	118	3
1	2	12	55	77	79	121	120	119	3
1	4	NA	40	205	NA	161	324	NA	2
1	4	NA	39	52	NA	122	125	NA	2
1	4	NA	48	142	NA	126	249	NA	2
1	4	NA	36	46	NA	121	123	NA	2
1	4	5	61	132	74	164	315	159	3
1	5	NA	54	55	NA	141	141	NA	2
1	5	NA	49	64	NA	139	128	NA	2
1	5	NA	26	54	NA	122	123	NA	2
1	5	6	44	117	112	137	273	274	3
1	5	7	24	32	15	70	70	33	3
1	5	11	51	129	59	99	196	97	3
1	5	11	41	126	63	93	188	86	3
1	7	NA	18	132	NA	78	285	NA	2
1	7	11	10	31	32	19	54	55	3
1	7	14	37	45	51	102	104	102	3
1	7	14	41	52	54	98	103	100	3
1	8	NA	5	9	NA	18	18	NA	2
1	10	NA	15	25	NA	42	39	NA	2
1	10	NA	14	41	NA	45	90	NA	2
1	11	NA	12	24	NA	56	55	NA	2
1	12	NA	45	70	NA	129	129	NA	2
1	12	NA	16	19	NA	49	49	NA	2
1	12	NA	49	77	NA	150	149	NA	2
1	12	NA	43	65	NA	129	132	NA	2
1	12	NA	13	26	NA	116	111	NA	2
1	14	NA	29	53	NA	102	95	NA	2
2	7	NA	37	35	NA	61	62	NA	2
2	12	NA	81	93	NA	122	126	NA	2
2	13	NA	33	21	NA	63	61	NA	2

```
3	6	NA	87	83	NA	120	120	NA	2
3	8	NA	33	31	NA	108	109	NA	2
5	6	NA	66	83	NA	138	140	NA	2
5	6	NA	81	94	NA	151	144	NA	2
5	11	NA	144	157	NA	238	240	NA	2
6	7	NA	94	89	NA	123	117	NA	2
6	11	NA	175	146	NA	232	227	NA	2
6	12	NA	75	74	NA	107	108	NA	2
6	14	NA	59	47	NA	98	100	NA	2
7	9	NA	30	35	NA	66	66	NA	2
7	10	NA	27	29	NA	61	64	NA	2
7	11	NA	67	67	NA	101	102	NA	2
7	11	NA	27	30	NA	45	45	NA	2
7	11	NA	26	25	NA	50	50	NA	2
7	11	12	57	64	70	92	96	96	3
7	12	NA	35	48	NA	120	118	NA	2
7	12	NA	63	73	NA	144	142	NA	2
7	14	NA	31	36	NA	54	55	NA	2
7	14	NA	35	35	NA	47	40	NA	2
7	14	NA	98	81	NA	170	171	NA	2
7	14	NA	153	170	NA	186	196	NA	2
7	14	NA	95	107	NA	161	153	NA	2
7	14	NA	34	48	NA	73	73	NA	2
9	11	NA	74	66	NA	139	136	NA	2
9	13	NA	61	51	NA	100	100	NA	2
10	11	NA	11	16	NA	20	20	NA	2
10	12	NA	42	41	NA	78	82	NA	2
11	13	NA	48	48	NA	53	55	NA	2
12	13	NA	37	46	NA	60	62	NA	2
12	14	NA	41	49	NA	72	75	NA	2
12	14	NA	56	56	NA	79	84	NA	2
12	14	NA	45	49	NA	82	78	NA	2
END

#Set Initial Values
#chain 1
list(d=c( NA, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0), sd=1,
mu=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0))
#chain 2
list(d=c( NA, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1), sd=4,
mu=c(-3, -3, -3, -3, -3, -3, -3, -3, -3, -3,
        -3, -3, -3, -3, -3, -3, -3, -3, -3, -3,
        -3, -3, -3, -3, -3, -3, -3, -3, -3, -3,
        -3, -3, -3, -3, -3, -3, -3, -3, -3, -3,
        -3, -3, -3, -3, -3, -3, -3, -3, -3, -3,
```

```
        -3, -3, -3, -3, -3, -3, -3, -3, -3, -3,
         -3, -3, -3, -3))
#chain 3
list(d=c( NA, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2), sd=2,
mu=c(-3, 5, -1, -3, 7, -3, -4, -3, -3, 0,
        -3, -3, 0, 3, 5, -3, -3, -1, -3, -7,
        -3, -3, 5, -1, 7, 0, 1,-4, 5, 0,
        -3, 5, -1, -3, 7, -3, -4, -3, -3, 0,
        -3, -3, 0, 3, 5, -3, -3, -1, -3, -7,
        -3, -3, 5, -1, 7, 0, 1,-4, 5, 0,
        1, -1, 5, -4))
```

KQ1: WinBUGS Dataset for RA Dichotomous

#Full Network (star, closed loop, loop plus one, and ladder are subsets)

#Description of data inputs

#ns = Number of studies

#nt = Number of treatments (including placebo)

#t[,x] = Treatment indicator

#r[,x] = Number achieving response on ACR50

#n[,x]= Number of all randomized patients (ITT)

#na[] = Number of arms in study

list(ns=31, nt=9)

t[,1]	t[,2]	t[,3]	r[,1]	r[,2]	r[,3]	n[,1]	n[,2]	n[,3]	na[]
1	2	NA	36	169	NA	219	433	NA	2
1	2	NA	14	42	NA	119	115	NA	2
1	2	7	22	63	61	110	156	165	3
1	3	NA	9	84	NA	110	328	NA	2
1	3	NA	1	36	NA	70	142	NA	2
1	3	NA	5	37	NA	62	67	NA	2
1	3	NA	2	12	NA	12	35	NA	2
1	3	NA	36	92	NA	318	318	NA	2
1	3	NA	9	28	NA	63	65	NA	2
1	3	NA	19	166	NA	200	419	NA	2
1	4	NA	9	33	NA	121	232	NA	2
1	4	NA	2	22	NA	48	105	NA	2
1	4	NA	20	43	NA	253	253	NA	2
1	4	NA	0	16	NA	12	42	NA	2
1	5	NA	1	23	NA	30	59	NA	2
1	5	NA	3	25	NA	44	44	NA	2
1	5	NA	3	19	NA	29	29	NA	2
1	5	NA	4	31	NA	80	78	NA	2
1	6	NA	13	31	NA	133	89	NA	2
1	6	NA	2	13	NA	35	35	NA	2
1	7	NA	4	33	NA	47	100	NA	2
1	7	NA	33	229	NA	363	721	NA	2
1	7	NA	22	38	NA	86	87	NA	2
1	7	NA	1	3	NA	7	14	NA	2
1	7	NA	4	94	NA	88	340	NA	2
1	8	NA	5	17	NA	40	40	NA	2
1	8	NA	16	88	NA	172	340	NA	2
1	9	NA	14	45	NA	49	99	NA	2
1	9	NA	37	302	NA	415	805	NA	2
1	9	NA	22	157	NA	204	419	NA	2
1	9	NA	38	228	NA	394	802	NA	2

END

#Initial Values
#chain 1
list(d=c(NA, 0, 0, 0, 0, 0, 0, 0, 0), sd=1, mu=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0))
#chain 2
list(d=c(NA, -1, -1, -1, -1, -1, -1, -1, -1), sd=4, mu=c(-3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3 , -3, -3, -3 , -3, -3, -3 , -3, -3, -3))
#chain 3

list(d=c(NA, 2, 2, 2, 2, 2, 2, 2, 2), sd=2, mu=c(-3, 5, -1, -3, 7, -3, -4, -3, -3, 0, -3, -3,0, 3, 5, -3, -3, -1, -3, -7, -3, -3, 5, -1,7, 0, 1,-4, 5, 0, -1))

KQ1: Random Effects Model for Continuous Data

```
# Normal likelihood, identity link
# Random effects model for multi-arm trials

model{                          # *** PROGRAM STARTS

for(i in 1:ns){                 #   LOOP THROUGH STUDIES
   w[i,1] <- 0    # adjustment for multi-arm trials is zero for control arm
   delta[i,1] <- 0            # treatment effect is zero for control arm
   mu[i] ~ dnorm(0,.0001)          # vague priors for all trial baselines
   for (k in 1:na[i]) {            #  LOOP THROUGH ARMS
      var[i,k] <- pow(se[i,k],2)   # calculate variances
      prec[i,k] <- 1/var[i,k]      # set precisions
      y[i,k] ~ dnorm(theta[i,k],prec[i,k]) # binomial likelihood
      theta[i,k] <- mu[i] + delta[i,k]  # model for linear predictor
#Deviance contribution
      dev[i,k] <- (y[i,k]-theta[i,k])*(y[i,k]-theta[i,k])*prec[i,k]
    }
#  summed residual deviance contribution for this trial
   resdev[i] <- sum(dev[i,1:na[i]])
   for (k in 2:na[i]) {            # LOOP THROUGH ARMS
# trial-specific LOR distributions
      delta[i,k] ~ dnorm(md[i,k],taud[i,k])
# mean of LOR distributions, with multi-arm trial correction
      md[i,k] <-  d[t[i,k]] - d[t[i,1]] + sw[i,k]
# precision of LOR distributions (with multi-arm trial correction)
      taud[i,k] <- tau *2*(k-1)/k
# adjustment, multi-arm RCTs
      w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])
# cumulative adjustment for multi-arm trials
      sw[i,k] <- sum(w[i,1:k-1])/(k-1)
    }
 }
totresdev <- sum(resdev[])           #Total Residual Deviance
d[1]<-0      # treatment effect is zero for control arm
# vague priors for treatment effects
for (k in 2:nt){  d[k] ~ dnorm(0,.0001) }
sd ~ dunif(0,5)     # vague prior for between-trial SD
tau <- pow(sd,-2)   # between-trial precision = (1/between-trial variance)

# All pairwise comparisons
for (c in 1:(nt-1)) {  for (k in (c+1):nt)  { diff[c,k] <- (d[c] - d[k] )}}
for (k in 1:nt) {
rk[k] <- nt+1-rank(d[],k) # assumes events are “good”
#rk[k] <- rank(d[],k) # assumes events are “bad”
best[k] <- equals(rk[k],1) #calculate probability that treat k is best
}

}                               # *** PROGRAM ENDS
```

KQ 1: WinBUGS Dataset for SGA Continuous

#Full Network (star, closed loop, loop plus one, and ladder are subsets)

#Description of data inputs

#ns = Number of studies

#nt = Number of treatments (including placebo)

#t[,x] = Treatment indicator

#y[,x] = Mean change from baseline in HAM-D score

#se[,x]= Standard error of mean change from baseline in HAM-D score

#na[] = Number of arms in study

list(ns=40, nt=14)

t[,1]	t[,2]	t[,3]	y[,1]	y[,2]	y[,3]	se[,1]	se[,2]	se[,3]	na[]		
1	4	NA	-10.7	-13.2	NA	0.84	0.82	NA	2		
1	4	NA	-9.3	-12.6	NA	0.74	0.75	NA	2		
1	4	NA	-7.5	-9.1	NA	0.66	0.67	NA	2		
1	5	6	-5.97	-7.61	-7.22	0.58	0.42	0.4	3		
1	5	11	-8.8	-12.1	-11.7	0.5	0.5	0.5	3		
1	5	11	-10.8	-12.1	-11.9	0.5	0.5	0.5	3		
1	5	NA	-10.27		-10.85		NA	0.67	0.69	NA	2
1	7	11	-11.6	-10.8	-11.1	2.04	1.29	1.27	3		
1	7	NA	-7.9	-10.3	NA	0.44	0.49	NA	2		
1	7	NA	-7	-11.2	NA	0.98	0.86	NA	2		
1	11	NA	-7.6	-7.8	NA	1.2	1.13	NA	2		
1	12	NA	-8.8	-11.1	NA	0.65	0.63	NA	2		
1	12	NA	-6.1	-6.1	NA	1.04	0.96	NA	2		
1	12	NA	-9.2	-10.53		NA	0.67	0.72	NA	2	
1	12	NA	-8.16	-11.66		NA	0.66	0.69	NA	2	
1	14	NA	-7.3	-11.7	NA	0.71	0.75	NA	2		
2	12	NA	-15.5	-16.3	NA	0.66	0.69	NA	2		
3	6	NA	-13.8	-14.7	NA	0.69	0.76	NA	2		
6	5	NA	-11.1	-9.6	NA	0.59	0.68	NA	2		
6	7	NA	-15.8	-14.7	NA	0.59	0.74	NA	2		
6	12	NA	-16.9	-16.1	NA	0.7	0.8	NA	2		
6	14	NA	-14.9	-12.9	NA	0.91	0.92	NA	2		
7	10	NA	-12.2	-11.4	NA	0.79	0.83	NA	2		
7	11	NA	-14.78		-13.92		NA	0.11	0.11	NA	2
7	11	NA	-15	-17.3	NA	1.61	1.56	NA	2		
7	13	NA	-12.2	-13.9	NA	0.91	1.02	NA	2		
7	13	NA	-14.8	-17.1	NA	1.74	1.02	NA	2		
7	14	NA	-14.4	-16.1	NA	0.88	0.7	NA	2		
7	14	NA	-10.4	-14.4	NA	1.05	0.95	NA	2		
8	11	NA	-15.9	-13.9	NA	0.66	0.66	NA	2		
8	11	NA	-13.45		-12.86		NA	1.25	1.27	NA	2
8	12	NA	-23.7	-18	NA	1.72	1.5	NA	2		
8	12	NA	-10.61		-10.98		NA	1.11	0.9	NA	2
9	11	NA	-9.22	-7.29	NA	0.51	0.54	NA	2		

```
11      13      NA      -15     -14.6   NA      0.68    0.66    NA      2
11      14      NA      -14.1   -14.7   NA      0.6     0.57    NA      2
11      14      NA      -15.5   -15.7   NA      0.95    0.9     NA      2
12      13      NA      -11.5   -12.9   NA      1.08    1.15    NA      2
12      14      NA      -11.3   -12.7   NA      0.61    0.64    NA      2
12      14      NA      -15.9   -14.3   NA      0.95    0.94    NA      2
END

#Initial Values
#chain 1
list(d=c( NA, 0,0,0,0,0,0,0,0,0,0,0,0,0), sd=1, mu=c(0, 0, 0, 0, 0, 0, 0, 0 , 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0))
#chain 2
list(d=c( NA, -1,-3,-1,1,3,-1,-3, -1,1,3,-1,-3, 1), sd=4, mu=c(-3, -3, -3, -3, -3, -3, -3,-3,-3,-3,-3,
-3, -3, -3, -3, -3, -3,-3,-3,-3,-3,
-3, -3, -3, -3, -3, -3,-3,-3,-3,-3,
-3, -3, -3, -3, -3, -3,-3,-3,-3,-3))
#chain 3
list(d=c( NA, 2,2,2,2,2,2,2,2,2,2,2,2,2), sd=2, mu=c(-3, 5, -1, -3, 7,  -3, -4, -3, 5, -1,
-3, 5, -1, -3, 7, -3, -4, -3, 5, -1,
-3, 5, -1, -3, 7, -3, -4, -3, 5, -1,

-3, 5, -1, -3, 7, -3, -4, -3, 5, -1))
```

KQ2: Random Effects Meta-Regression Model With a Subgroup Indicator Covariate

```
# Binomial likelihood, logit link, subgroup
# Random effects model for multi-arm trials

model{                          # *** PROGRAM STARTS
for(i in 1:ns){                 # LOOP THROUGH STUDIES
   w[i,1] <- 0    # adjustment for multi-arm trials is zero for control arm
   delta[i,1] <- 0            # treatment effect is zero for control arm
   mu[i] ~ dnorm(0,.0001)          # vague priors for all trial baselines
   for (k in 1:na[i]) {            # LOOP THROUGH ARMS
      r[i,k] ~ dbin(p[i,k],n[i,k]) # binomial likelihood
# model for linear predictor, covariate effect relative to treat in arm 1
      logit(p[i,k]) <- mu[i] + delta[i,k]
                  + (beta[t[i,k]]-beta[t[i,1]]) * x[i]
      rhat[i,k] <- p[i,k] * n[i,k] # expected value of the numerators
#Deviance contribution
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
         +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))        }
#  summed residual deviance contribution for this trial
   resdev[i] <- sum(dev[i,1:na[i]])
   for (k in 2:na[i]) {            # LOOP THROUGH ARMS
# trial-specific LOR distributions
      delta[i,k] ~ dnorm(md[i,k],taud[i,k])
# mean of LOR distributions (with multi-arm trial correction)
      md[i,k] <-  d[t[i,k]] - d[t[i,1]] + sw[i,k]
# precision of LOR distributions (with multi-arm trial correction)
      taud[i,k] <- tau *2*(k-1)/k
# adjustment for multi-arm RCTs
      w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])
# cumulative adjustment for multi-arm trials
      sw[i,k] <- sum(w[i,1:k-1])/(k-1)
     }
  }
totresdev <- sum(resdev[])          # Total Residual Deviance
d[1]<-0       # treatment effect is zero for reference treatment
beta[1] <- 0  # covariate effect is zero for reference treatment
for (k in 2:nt){  # LOOP THROUGH TREATMENTS
   d[k] ~ dnorm(0,.0001) # vague priors for treatment effects
   beta[k] <- B # common covariate effect
  }
B ~ dnorm(0,.0001) # vague prior for covariate effect
sd ~ dunif(0,5)     # vague prior for between-trial SD
tau <- pow(sd,-2)   # between-trial precision = (1/between-trial variance)
# treatment effect when covariate = z[j]
for (k in 1:nt){  # LOOP THROUGH TREATMENTS
   for (j in 1:nz) { dz[j,k] <- d[k] + (beta[k]-beta[1])*z[j] }
  }
# pairwise ORs and LORs for all possible pair-wise comparisons, if nt>2
for (c in 1:(nt-1)) {
```

```
    for (k in (c+1):nt)  {
# when covariate is zero
        or[c,k] <- exp(d[k] - d[c])
        lor[c,k] <- (d[k]-d[c])
# at covariate=z[j]
        for (j in 1:nz) {
            orz[j,c,k] <- exp(dz[j,k] - dz[j,c])
            lorz[j,c,k] <- (dz[j,k]-dz[j,c])
          }
     }
 }
}                                       # *** PROGRAM ENDS
```

KQ 2: WinBUGS Dataset for Meta-Regression With a Subgroup Indicator Covariate

# ns= number of studies; nt=number of treatments;
# z=values of covariate at which to calculate treatment effects; nz=length of z
list(ns=72, nt=14, z=c(1), nz=1)

t[,1]	t[,2]	t[,3]	r[,1]	r[,2]	r[,3]	n[,1]	n[,2]	n[,3]	x[]	na[]
1	2	7	73	76	83	152	150	154	0 3	
1	2	12	66	78	66	124	122	118	0 3	
1	2	12	55	77	79	121	120	119	0 3	
1	3	NA	34	34	NA	90	84	NA	1 2	
1	4	NA	40	205	NA	161	324	NA	0 2	
1	4	NA	39	52	NA	122	125	NA	0 2	
1	4	NA	48	142	NA	126	249	NA	0 2	
1	4	NA	36	46	NA	121	123	NA	0 2	
1	4	5	61	132	74	164	315	159	0 3	
1	5	NA	54	55	NA	141	141	NA	0 2	
1	5	NA	49	64	NA	139	128	NA	0 2	
1	5	NA	26	54	NA	122	123	NA	0 2	
1	5	6	44	117	112	137	273	274	0 3	
1	5	7	24	32	15	70	70	33	0 3	
1	5	11	51	129	59	99	196	97	0 3	
1	5	11	41	126	63	93	188	86	0 3	
1	7	NA	92	125	NA	336	335	NA	1 2	
1	7	NA	18	132	NA	78	285	NA	0 2	
1	7	11	10	31	32	19	54	55	0 3	
1	7	14	37	45	51	102	104	102	0 3	
1	7	14	41	52	54	98	103	100	0 3	
1	8	NA	5	9	NA	18	18	NA	0 2	
1	9	13	25	20	17	50	50	50	1 3	
1	10	NA	15	25	NA	42	39	NA	0 2	
1	10	NA	14	41	NA	45	90	NA	0 2	
1	11	NA	12	24	NA	56	55	NA	0 2	
1	11	11	72	85	100	180	168	177	1 3	
1	12	NA	45	70	NA	129	129	NA	0 2	
1	12	NA	16	19	NA	49	49	NA	0 2	
1	12	NA	49	77	NA	150	149	NA	0 2	
1	12	NA	43	65	NA	129	132	NA	0 2	
1	12	NA	96	126	NA	376	371	NA	1 2	
1	12	NA	13	26	NA	116	111	NA	0 2	
1	14	NA	29	53	NA	102	95	NA	0 2	
2	7	NA	37	35	NA	61	62	NA	0 2	
2	11	NA	34	40	NA	48	52	NA	1 2	
2	12	NA	81	93	NA	122	126	NA	0 2	
2	13	NA	33	21	NA	63	61	NA	0 2	
3	6	NA	87	83	NA	120	120	NA	0 2	
3	8	NA	33	31	NA	108	109	NA	0 2	
5	6	NA	66	83	NA	138	140	NA	0 2	
5	6	NA	81	94	NA	151	144	NA	0 2	
5	11	NA	144	157	NA	238	240	NA	0 2	

```
6	7	NA	94	89	NA	123	117	NA	0 2
6	11	NA	175	146	NA	232	227	NA	0 2
6	12	NA	75	74	NA	107	108	NA	0 2
6	14	NA	59	47	NA	98	100	NA	0 2
7	9	NA	30	35	NA	66	66	NA	0 2
7	10	NA	27	29	NA	61	64	NA	0 2
7	11	NA	67	67	NA	101	102	NA	0 2
7	11	NA	27	30	NA	45	45	NA	0 2
7	11	NA	26	25	NA	50	50	NA	0 2
7	11	12	57	64	70	92	96	96	0 3
7	12	NA	84	85	NA	119	117	NA	1 2
7	12	NA	35	48	NA	120	118	NA	0 2
7	12	NA	63	73	NA	144	142	NA	0 2
7	14	NA	31	36	NA	54	55	NA	0 2
7	14	NA	35	35	NA	47	40	NA	0 2
7	14	NA	98	81	NA	170	171	NA	0 2
7	14	NA	153	170	NA	186	196	NA	0 2
7	14	NA	95	107	NA	161	153	NA	0 2
7	14	NA	34	48	NA	73	73	NA	0 2
8	12	NA	28	25	NA	40	48	NA	1 2
9	11	NA	74	66	NA	139	136	NA	0 2
9	13	NA	61	51	NA	100	100	NA	0 2
10	11	NA	11	16	NA	20	20	NA	0 2
10	12	NA	42	41	NA	78	82	NA	0 2
11	13	NA	48	48	NA	53	55	NA	0 2
12	13	NA	37	46	NA	60	62	NA	0 2
12	14	NA	41	49	NA	72	75	NA	0 2
12	14	NA	56	56	NA	79	84	NA	0 2
12	14	NA	45	49	NA	82	78	NA	0 2
END

#Initial Values
#chain 1
list(d=c( NA, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0), mu=c(0,0,0,0,0,   0,0,0,0,0,   0,0,0,0,0,   0,0,0,0,0,
0,0,0,0,0,  0,0,0,0,0,  0,0,0,0,0,  0,0,0,0,0,  0,0,0,0,0,  0,0,0,0,0,  0,0,0,0,0,  0,0,0,0,0,
0,0,0,0,0, 0,0,0,0,0,   0,0), B=0, sd=1)
#chain 2
list(d=c( NA, -1, -1, -2, -1, -1, -2, -1, -1, -2, -1, -1, -2, 1), mu=c(-3,-3, 3,-3, 3, -3, 3,-3, 3,-3,
-3,-3, 3, 3, -3, 3, -3, -3, 3, -3, 3,-3,3,-3, -3, 3,-3,3,-3, -3, 3,-3,3,-3,-3, 3,-3,3,-3, -3, 3,-3,3,-3,-3, 3,-
3,3,-3, -3, 3,-3,3,-3,-3, 3,-3,3,-3,-3, 3,-3,3,-3,-3, 3,-3,3,-3, 3,-3  ), B=-1, sd=3)
```

KQ2: Random Effects Continuous Covariate Meta-Regression Model

```
# Binomial likelihood, logit link
# Random effects model for multi-arm trials
model{                          # *** PROGRAM STARTS
for(i in 1:ns){                 # LOOP THROUGH STUDIES
   w[i,1] <- 0    # adjustment for multi-arm trials is zero for control arm
   delta[i,1] <- 0           # treatment effect is zero for control arm
   mu[i] ~ dnorm(0,.0001)          # vague priors for all trial baselines
   for (k in 1:na[i]) {            # LOOP THROUGH ARMS
      r[i,k] ~ dbin(p[i,k],n[i,k]) # binomial likelihood
# model for linear predictor
      logit(p[i,k]) <- mu[i] + delta[i,k]
                + (beta[t[i,k]]-beta[t[i,1]]) * (x[i]-mx)
      rhat[i,k] <- p[i,k] * n[i,k] # expected value of the numerators
#Deviance contribution
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
         +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))        }
#  summed residual deviance contribution for this trial
   resdev[i] <- sum(dev[i,1:na[i]])
   for (k in 2:na[i]) {            # LOOP THROUGH ARMS
# trial-specific LOR distributions
      delta[i,k] ~ dnorm(md[i,k],taud[i,k])
# mean of LOR distributions (with multi-arm trial correction)
# covariate effect relative to treat in arm 1
      md[i,k] <-  d[t[i,k]] - d[t[i,1]] + sw[i,k]
# precision of LOR distributions (with multi-arm trial correction)
      taud[i,k] <- tau *2*(k-1)/k
# adjustment for multi-arm RCTs
      w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])
# cumulative adjustment for multi-arm trials
      sw[i,k] <- sum(w[i,1:k-1])/(k-1)
     }
  }
totresdev <- sum(resdev[])          # Total Residual Deviance
d[1]<-0       # treatment effect is zero for reference treatment
beta[1] <- 0  # covariate effect is zero for reference treatment
# vague priors for treatment effects
for (k in 2:nt){
   d[k] ~ dnorm(0,.0001)
   beta[k] <- B # common covariate effect
  }
B ~ dnorm(0,.0001) # vague prior for covariate effect
sd ~ dunif(0,5)     # vague prior for between-trial SD
tau <- pow(sd,-2)   # between-trial precision = (1/between-trial variance)
# treatment effect when covariate = z[j] (un-centring treatment effects)
for (k in 1:nt){
   for (j in 1:nz) { dz[j,k] <- d[k] - (beta[k]-beta[1])*(mx-z[j]) }
  }
# pairwise ORs and LORs for all possible pair-wise comparisons, if nt>2
```

```
for (c in 1:(nt-1)) {
   for (k in (c+1):nt)  {
# at mean value of covariate
       or[c,k] <- exp(d[k] - d[c])
       lor[c,k] <- (d[k]-d[c])
# at covariate=z[j]
       for (j in 1:nz) {
          orz[j,c,k] <- exp(dz[j,k] - dz[j,c])
          lorz[j,c,k] <- (dz[j,k]-dz[j,c])
         }
    }
 }
# ranking on relative scale
for (k in 1:nt) {
rk[k] <- nt+1-rank(d[],k) # assumes events are “good”
#rk[k] <- rank(d[],k) # assumes events are “bad”
best[k] <- equals(rk[k],1) #calculate probability that treat k is best
}                                    # *** PROGRAM ENDS
```

KQ 2: WinBUGS Dataset for Continuous Covariate Meta-Regression

# Data (RA with ACR50 – covariate x[] is disease duration)
list(ns=28, nt=9, mx=8.50, z=c(1,2,10), nz=3)

t[,1]	t[,2]	t[,3]	na[]	n[,1]	n[,2]	n[,3]	r[,1]	r[,2]	r[,3]	x[]
1	2	NA	2	219	433	NA	36	169	NA	8.7
1	2	NA	2	119	115	NA	14	42	NA	9.3
1	2	7	3	110	156	165	22	63	61	7.87
1	3	NA	2	318	318	NA	36	92	NA	10.4
1	3	NA	2	70	142	NA	1	36	NA	10
1	3	NA	2	110	328	NA	9	84	NA	11.2
1	3	NA	2	200	419	NA	19	166	NA	10.9
1	3	NA	2	62	67	NA	5	37	NA	12.3
1	3	NA	2	63	65	NA	9	28	NA	6.85
1	3	NA	2	12	35	NA	2	12	NA	7.25
1	4	NA	2	121	232	NA	9	33	NA	3.9
1	4	NA	2	253	253	NA	20	43	NA	10.5
1	4	NA	2	48	105	NA	2	22	NA	7.55
1	5	NA	2	80	78	NA	4	31	NA	12
1	5	NA	2	30	59	NA	1	23	NA	13
1	6	NA	2	133	89	NA	13	31	NA	5.5
1	6	NA	2	35	35	NA	2	13	NA	7.8
1	7	NA	2	47	100	NA	4	33	NA	7.8
1	7	NA	2	7	14	NA	1	3	NA	6.2
1	7	NA	2	363	721	NA	33	229	NA	7.5
1	7	NA	2	86	87	NA	22	38	NA	7.55
1	7	NA	2	88	340	NA	4	94	NA	8.6
1	8	NA	2	172	340	NA	16	88	NA	7
1	8	NA	2	40	40	NA	5	17	NA	11.5
1	9	NA	2	204	419	NA	22	157	NA	7.65
1	9	NA	2	49	99	NA	14	45	NA	0.865
1	9	NA	2	415	805	NA	37	302	NA	9.8
1	9	NA	2	394	802	NA	38	228	NA	9

END

Initial values for RE model
#chain 1
list(d=c(NA, 0,0,0,0,0,0, 0, 0), mu=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0), sd=1, B=0)
#chain 2
list(d=c(NA, -1,1,-1,1,-1,1, -1, 1), mu=c(-3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3), sd=0.5, B=-1)
#chain 3
list(d=c(NA, 2,-2,2,-2,2,-2, 2, -2), mu=c(-3, 5, -1, -3, 7, -3, -4, -3, -3, 0, 5, 0, -3, 5, -1, -3, 7, -3, -4, -3, -3, 0, 5, 0, 1,-1, 3,4), sd=3, B=5)

KQ 3: Simulation Code Related to Key Question 3

Initially, we created two master datasets generated from random binomial draws from which to sample sub-datasets for each of the different network pattern and sample size criteria. Each dataset reflected a different scenario we wanted to introduce to the Bayesian MTC models. The datasets were created from binomial distributions of varying probability of response and sample size. The SAS version 9.2 code is presented below.

```
/* master_even: dataset where drugs have equal efficacy and drug1 low efficacy */
data master_even (drop=i);
	do i=1 to 10000;
		r1 = rand('BINOM',0.1, 100);
		r2 = rand('BINOM',0.5, 100);
		r3 = rand('BINOM',0.5, 100);
		r4 = rand('BINOM',0.5, 100);
		output;
	end;
run;

/* master_winnner: dataset where one drug has a high efficacy and the others low efficacy */
data master_winner (drop=i);
n=100;
	do i=1 to 10000;
		r1 = rand('BINOM',0.2, 100);
		r2 = rand('BINOM',0.2, 100);
		r3 = rand('BINOM',0.2, 100);
		r4 = rand('BINOM',0.8, 100);
		output;
	end;
run;
```

KQ 3: WinBUGS Datasets

From these master datasets, sub-datasets for WinBUGS were created for each of the network patterns and sample size criteria. A total of 40,000 datasets were created by taking successive studies in each of the master datasets. In this way, studies were sampled without replacement for each of the datasets. An example dataset is shown below.

We used macros created in SAS to invoke repeated calls of openBUGS for each model run (for a total of 40,000 model runs). OpenBUGS was utilized instead of winBUGS since the simulations were performed on UNC's linux cluster. The output, in this case, the 'best' statistic, was parsed and processed in SAS for post-data management and panel plot creation.

```
#Star network, study sample size = 2
#Description of data inputs
#ns = Number of studies
#nt = Number of treatments (including placebo)
#t[,x] = Treatment indicator
#r[,x] = Number achieving response
#n[,x]= Number of all randomized patients (ITT)
#na[] = Number of arms in study

list(ns=6, nt=4)

t[,1]   t[,2]   t[,3]   r[,1]   r[,2]   r[,3]   n[,1]   n[,2]   n[,3]   na[]
1       2       7       73      76      83      152     150     154     2
1       2       12      66      78      66      124     122     118     2
1       2       12      55      77      79      121     120     119     2
1       4       NA      40      205     NA      161     324     NA      2
1       4       NA      39      52      NA      122     125     NA      2
1       4       NA      48      142     NA      126     249     NA      2
END
#Set Initial Values
#chain 1
list(d=c( NA, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0), sd=1,
mu=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0))
```

Random Effects Model for Dichotomous Data

```
# Binomial likelihood, logit link
# Random effects model for multi-arm trials

model{                          # *** PROGRAM STARTS

for(i in 1:ns){                 # LOOP THROUGH STUDIES
   w[i,1] <- 0    # adjustment for multi-arm trials is zero for control arm
   delta[i,1] <- 0            # treatment effect is zero for control arm
   mu[i] ~ dnorm(0,.0001)          # vague priors for all trial baselines
   for (k in 1:na[i]) {            # LOOP THROUGH ARMS
      r[i,k] ~ dbin(p[i,k],n[i,k]) # binomial likelihood
      logit(p[i,k]) <- mu[i] + delta[i,k]  # model for linear predictor
      rhat[i,k] <- p[i,k] * n[i,k] # expected value of the numerators

#Deviance contribution
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
         +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))        }

#  summed residual deviance contribution for this trial
   resdev[i] <- sum(dev[i,1:na[i]])
   for (k in 2:na[i]) {            # LOOP THROUGH ARMS
# trial-specific LOR distributions
      delta[i,k] ~ dnorm(md[i,k],taud[i,k])
# mean of LOR distributions (with multi-arm trial correction)
      md[i,k] <-  d[t[i,k]] - d[t[i,1]] + sw[i,k]
# precision of LOR distributions (with multi-arm trial correction)
      taud[i,k] <- tau *2*(k-1)/k
# adjustment for multi-arm RCTs
      w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])
# cumulative adjustment for multi-arm trials
      sw[i,k] <- sum(w[i,1:k-1])/(k-1)
    }
  }
totresdev <- sum(resdev[])          # Total Residual Deviance
d[1]<-0      # treatment effect is zero for reference treatment
# vague priors for treatment effects
for (k in 2:nt){  d[k] ~ dnorm(0,.0001) }
sd ~ dunif(0,5)     # vague prior for between-trial SD
tau <- pow(sd,-2)   # between-trial precision = (1/between-trial variance)
```

```
# ranking on relative scale
for (k in 1:nt) {
rk[k] <- nt+1-rank(d[],k) # assumes events are “good”
# rk[k] <- rank(d[],k) # assumes events are “bad”
best[k] <- equals(rk[k],1) #calculate probability that treat k is best
}
}
# *** PROGRAM ENDS
```

References

1. Dias S, Sutton AJ, Welton NJ, et al. NICE DSU Technical Support Document 3: Heterogeneity: subgroups, meta-regression, bias and bias-adjustment. Report by the Decision Support Unit. 2011.. http://www.nicedsu.org.uk/TSD3%20Heterogeneity.final%20report.08.05.12.pdf. Accessed May 4, 2012.

2. Dias S, Welton NJ, Sutton AJ, et al. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials. Last updated August 2011; 2011. http://www.nicedsu.org.uk.

Appendix B. Summary of Studies Included in Analyses

Table B-1. Summary of studies included in analyses of response for second-generation antidepressants

Study	N	Duration	Comparison and Dose (mg/day)	Response[a]	Population
Feighner, 1991[1]	123	6 weeks	Bupropion 225-450 Fluoxetine 20-80	37/61 35/62	Adults
Coleman, 2001[2]	456	8 weeks	Bupropion 150-400 Fluoxetine 50-200 Placebo	76/150 83/154 73/152	Adults
Weihs, 2000[3]	100	6 weeks	Bupropion 100-300 Paroxetine 10-40	34/48 40/52	Older adults
Rush, 2001[4]	248	16 weeks	Bupropion 100-300 Sertraline 50-200	81/122 93/126	Adults
Coleman, 1999[5]	364	8 weeks	Bupropion 150-400 Sertraline 50-200 Placebo	78/122 66/118 66/124	Adults
Croft, 1999[6]	360	8 weeks	Bupropion 150-400 Sertraline 50-200 Placebo	77/120 79/119 55/121	Adults
Weisler, 1994[7]	124	6 weeks	Bupropion 225-450 Trazodone 150-400	33/63 21/61	Adults
Ou, 2011[8]	240	6 weeks	Citalopram 20-40 Escitalopram 10-20	87/120 83/120	Adults
Haffmans, 1996[9]	217	6 weeks	Citalopram 20-40 Fluvoxamine 100-200	33/108 31/109	Adults
Roose, 2004[10]	174	8 weeks	Citalopram 10-40 Placebo	34/84 34/90	Older adults
Tourian, 2009[11]	638	8 weeks	Desvenlafaxine 50-100 Duloxetine 60 Placebo	132/315 74/159 61/164	Adults
Feiger, 2009[12]	244	8 weeks	Desvenlafaxine 200-400 Placebo	46/123 36/121	Adults
Boyer, 2008[13]	485	8 weeks	Desvenlafaxine 50-100 Placebo	205/324 40/161	Adults
Liebowitz, 2007[14]	247	8 weeks	Desvenlafaxine 100-200 Placebo	52/125 39/122	Adults
Septien-Velez, 2007[15]	375	8 weeks	Desvenlafaxine 200-400 Placebo	142/249 48/126	Adults
Khan, 2007[16]	278	8 weeks	Duloxetine 60 Escitalopram 10-20	66/138 83/140	Adults
Wade, 2007[17]	295	24 weeks	Duloxetine 60 Escitalopram 20	81/151 94/144	Adults
Nierenberg,2007.[18]	684	8 weeks	Duloxetine 40-60 Escitalopram 10-20 Placebo	117/273 112/274 44/137	Adults
Goldstein, 2002[19]	173	8 weeks	Duloxetine 40-120 Fluoxetine 20 Placebo	32/70 15/33 24/70	Adults
Lee, 2007[20]	478	8 weeks	Duloxetine 60 Paroxetine 20	144/238 157/240	Adults

Table B-1. Summary of studies included in analyses of response for second-generation antidepressants (continued)

Study	N	Duration	Comparison and Dose (mg/day)	Response[a]	Population
Detke, 2004[21]	367	8 weeks	Duloxetine 80-120 Paroxetine 20 Placebo	126/188 63/86 41/93	Adults
Perahia, 2006[22]	393	8 weeks	Duloxetine 80-120 Paroxetine 20 Placebo	129/196 59/97 51/99	Adults
Brannan, 2005[23]	282	7 weeks	Duloxetine 60 Placebo	55/141 54/141	Adults
Detke, 2002a[24]	267	9 weeks	Duloxetine 60 Placebo	64/128 49/139	Adults
Detke, 2002b[25]	245	9 weeks	Duloxetine 60 Placebo	54/123 26/122	Adults
Mao, 2008[26]	240	8 weeks	Escitalopram 10 Fluoxetine 20	94/123 89/117	Adults
Boulenger, 2006[27]	459	24 weeks	Escitalopram 10-20 Paroxetine 20-40	175/232 146/227	Adults
Ventura, 2007[28]	215	8 weeks	Escitalopram 10 Sertraline 50-200	75/107 74/108	Adults
Bielski, 2004[29]	198	8 weeks	Escitalopram 20 Venlafaxine 225	59/98 47/100	Adults
Hong, 2003[30]	132	6 weeks	Fluoxetine 20-40 Mirtazapine 15-45	30/66 35/66	Adults
Rush, 1998[31]	125	8 weeks	Fluoxetine 20-40 Nefazodone 100-500	27/61 29/64	Adults
De Wilde, 1993[32]	100	6 weeks	Fluoxetine 20-60 Paroxetine 20-40	26/50 25/50	Adults
Gagiano, 1993[33]	90	6 weeks	Fluoxetine 20-60 Paroxetine 20-40	27/45 30/45	Adults
Chouinard, 1999[34]	203	12 weeks	Fluoxetine 20-80 Paroxetine 20-50	67/101 67/102	Adults
Fava, 1998[35]	128	12 weeks	Fluoxetine 20-80 Paroxetine 20-50 Placebo	31/54 32/55 10/19	Adults
Fava, 2002[36]	284	10-16 weeks	Fluoxetine 20-60 Paroxetine 20-60 Sertraline 50-200	57/92 64/96 70/96	Adults
Wernicke, 1988[37]	363	6 weeks	Fluoxetine 5-40 Placebo	132/285 18/78	Adults
Tollefson, 1993[38]	671	6 weeks	Fluoxetine 20 Placebo	125/335 92/336	Older adults
Bennie, 1995[39]	286	6 weeks	Fluoxetine 20-40 Sertraline 50-100	63/144 73/142	Adults
Sechter, 1999[40]	238	24 weeks	Fluoxetine 20-60 Sertraline 50-150	35/120 48/118	Adults
Newhouse, 2005[41]	236	12 weeks	Fluoxetine 20-40 Sertraline 50-100	84/119 85/117	Older adults
Dierick, 1996[42]	314	8 weeks	Fluoxetine 20 Venlafaxine 75-150	95/161 107/153	Adults
Tylee, 1997[43]	341	12 weeks	Fluoxetine 20 Venlafaxine 75	98/170 81/171	Adults
Costa e Silva, 1998[44]	382	8 weeks	Fluoxetine 20-40 Venlafaxine 75-150	153/186 170/196	Adults
Alves, 1999[45]	87	12 weeks	Fluoxetine 20-40 Venlafaxine 75-150	35/47 35/40	Adults

Table B-1. Summary of studies included in analyses of response for second-generation antidepressants (continued)

Study	N	Duration	Comparison and Dose (mg/day)	Response[a]	Population
Tzanakaki, 2000[46]	109	6 weeks	Fluoxetine 20-60 Venlafaxine 75-250	31/54 36/55	Adults
De Nayer, 2002[47]	146	12 weeks	Fluoxetine 20 Venlafaxine 75	34/73 48/73	Adults
Rudolph, 1999[48]	301	8 weeks	Fluoxetine 20-60 Venlafaxine 75-225 Placebo	52/103 54/100 41/98	Adults
Nemeroff, 2007[49]	308	6 weeks	Fluoxetine 20-60 Venlafaxine 75-225 Placebo	45/104 51/102 37/102	Adults
Lydiard, 1989[50]	36	6 weeks	Fluvoxamine 100-300 Placebo	9/18 5/18	Adults
Rossini, 2005[51]	88	7 weeks	Fluvoxamine 200 Sertraline 150	28/40 25/48	Older adults
Benkert, 2000[52]	275	6 weeks	Mirtazapine 20-40 Paroxetine 15-45	74/139 66/136	Adults
Halikas, 1995[53]	150	6 weeks	Mirtazapine 5-35 Trazodone 40-280 Placebo	20/50 17/50 25/50	Older adults
van Moffaert, 1995[54]	200	6 weeks	Mirtazapine 24-72 Trazodone 150-450	61/100 51/100	Adults
Hicks, 2002[55]	40	8 weeks	Nefazodone 400-600 Paroxetine 20-40	11/20 16/20	Adults
Cohn, 1996[56]	81	8 weeks	Nefazodone 200-600 Placebo	25/39 15/42	Adults
Fontaine, 1994[57]	135	6 weeks	Nefazodone 100-500 Placebo	41/90 14/45	Adults
Feiger, 1996[58]	160	6 weeks	Nefazodone 100-600 Sertraline 50-200	42/78 41/82	Adults
Rickels, 1989[59]	111	6 weeks	Paroxetine 10-50 Placebo	24/55 12/56	Adults
Rapaport, 2009[60]	525	10 weeks	Paroxetine 12.5-25 Placebo	185/345 72/180	Older adults
Kasper, 2005[61]	108	6 weeks	Paroxetine 20-40 Trazodone 150-450	48/53 48/55	Adults
Reimherr, 1990[62]	299	8 weeks	Sertraline 20-200 Placebo	77/149 49/150	Adults
Lydiard, 1997[63]	261	8 weeks	Sertraline 50-200 Placebo	65/132 43/129	Adults
Olie, 1997[64]	258	6 weeks	Sertraline 50-200 Placebo	70/129 45/129	Adults
Hypericum Group, 2002[65]	227	8 weeks	Sertraline 50-100 Placebo	26/111 13/116	Adults
Schneider, 2003[66]	747	8 weeks	Sertraline 50-100 Placebo	126/371 96/376	Older adults
Blumenthal, 2007[67]	98	16 weeks	Sertraline 50-200 Placebo	19/49 16/49	Adults
Munizza, 2006[68]	122	6 weeks	Sertraline 50-100 Trazodone 150-450	37/60 46/62	Adults
Mehtonen, 2000[69]	147	8 weeks	Sertraline 50-100 Venlafaxine 75-150	41/72 49/75	Adults
Sir, 2005[70]	163	8 weeks	Sertraline 50-150 Venlafaxine 75-225	56/79 56/84	Adults

Table B-1. Summary of studies included in analyses of response for second-generation antidepressants (continued)

Study	N	Duration	Comparison and Dose (mg/day)	Response[a]	Population
Shelton, 2006[71]	160	8 weeks	Sertraline 50-150 Venlafaxine 75-225	45/82 49/78	Adults
Thase, 1997[72]	197	8 weeks	Venlafaxine 75-225 Placebo	53/95 29/102	Adults

Note: Studies are listed alphabetically by drug and chronologically within each set of drugs

[a] Response, defined as a 50% improvement on the Hamilton Depression Rating Scale (HAM-D), is presented as the number of responders out of the number in the group.

Abbreviations: mg = milligram

Table B-2. Summary of studies included in analyses of mean change in HAM-D for second-generation antidepressants

Study	N	Duration	Comparison and Dose (mg/day)	Mean change in HAM-D (SE)
Rush, 2001[4]	248	16 weeks	Bupropion 100-300 Sertraline 50-200	-15.5 (0.66) -16.3 (0.69)
Ou, 2011[8]	232	6 weeks	Citalopram 20-40 Escitalopram 10-20	-13.8 (0.69) -14.7 (0.76)
Septien-Velez, 2007[15]	245	8 weeks	Desvenlafaxine 200-400 Placebo	-12.6 (0.75) -9.3 (0.74)
Boyer, 2008[13]	325	8 weeks	Desvenlafaxine 50-100 Placebo	-13.2 (0.82) -10.7 (0.84)
Feiger, 2009[12]	235	8 weeks	Desvenlafaxine 200-400 Placebo	-9.1 (0.67) -7.5 (0.66)
Khan, 2007[16]	262	8 weeks	Duloxetine 60 Escitalopram10-20	-9.6 (0.68) -11.1 (0.59)
Nierenberg, 2007[18]	684	8 weeks	Duloxetine 60 Escitalopram 10 Placebo	-7.61 (0.42) -7.22 (0.40) -5.97 (0.58)
Detke, 2004[21]	272	8 weeks	Duloxetine 80-120 Paroxetine 20 Placebo	-12.1 (0.50) -11.7 (0.50) -8.8 (0.50)
Perahia, 2006[22]	289	8 weeks	Duloxetine 80-120 Paroxetine 20 Placebo	-12.1 (0.50) -11.7 (0.50) -8.8 (0.50)
Brannan, 2005[23]	268	7 weeks	Duloxetine 60 Placebo	-10.85 (0.69) -10.27 (0.67)
Mao, 2008[26]	231	8 weeks	Escitalopram 10 Fluoxetine 20	-15.8 (0.59) -14.7 (0.74)
Ventura, 2007[28]	211	8 weeks	Escitalopram 10 Sertraline 50-200	-16.9 (0.70) -16.1 (0.80)
Bielski, 2004[29]	195	8 weeks	Escitalopram 20 Venlafaxine 225	-14.9 (0.91) -12.9 (0.92)
Rush, 1998[31]	122	8 weeks	Fluoxetine 20-40 Nefazodone 200-500	-12.2 (0.79) -11.4 (0.83)
De Wilde, 1993[32]	78	6 weeks	Fluoxetine 20-60 Paroxetine 20-40	-15.0 (1.61) -17.3 (1.56)
Chouinard, 1999[34]	198	12 weeks	Fluoxetine 20-80 Paroxetine 20-50	-14.8 (0.11) -13.9 (0.11)
Fava, 1998[35]	128	12 weeks	Fluoxetine 20-80 Paroxetine 20-50 Placebo	-10.8 (1.29) -11.1 (1.27) -11.6 (2.04)
Wernicke, 1988[37]	169	6 weeks	Fluoxetine 5-40 Placebo	-11.2 (0.86) -7.0 (0.98)
Tollefson, 1993[38]	534	6 weeks	Fluoxetine 20 Placebo	-10.3 (0.49) -7.9 (0.44)
Perry, 1989[73]	40	6 weeks	Fluoxetine 21-50 Trazodone 241-337	-14.8 (1.74) -17.1 (1.02)
Beasley, 1991[74]	120	6 weeks	Fluoxetine 20-40 Trazodone 50-400	-12.2 (0.91) -13.9 (1.02)
Dierick, 1996[42]	314	8 weeks	Fluoxetine 20 Venlafaxine 75-150	-14.4 (0.88) -16.1 (0.70)
De Nayer, 2002[47]	131	12 weeks	Fluoxetine 20 Venlafaxine 75	-10.4 (1.05) -14.4 (0.95)
Kiev, 1997[75]	58	7 weeks	Fluvoxamine 50 Paroxetine 20	-13.5 (1.3) -12.9 (1.3)
Ushiroyama, 2004[76]	105	12 weeks	Fluvoxamine 50 Paroxetine 20	-15.9 (0.66) -13.9 (0.66)

Table B-2. Summary of studies included in analyses of mean change in HAM-D for second-generation antidepressants (continued)

Study	N	Duration	Comparison and Dose (mg/day)	Mean change in HAM-D (SE)
Nemeroff, 1995[77]	92	7 weeks	Fluvoxamine 50-150 Sertraline 50-200	-10.6 (1.11) -11.0 (0.90)
Rossini, 2005[51]	84	7 weeks	Fluvoxamine 50-200 Sertraline 25-150	-23.7 (1.72) -18.0 (1.50)
Schatzberg, 2002[78]	246	8 weeks	Mirtazapine 15-45 Paroxetine 20-40	-9.2 (0.51) -7.3 (0.54)
Ehde, 2008[79]	42	12 weeks	Paroxetine 10-40 Placebo	-7.8 (1.13) -7.6 (1.20)
Kasper, 2005[61]	108	6 weeks	Paroxetine 20-40 Trazodone 150-450	-15.0 (0.68) -14.6 (0.66)
McPartlin, 1998[80]	336	12 weeks	Paroxetine 20 Venlafaxine 75	-14.1 (0.60) -14.7 (0.57)
Ballus, 2000[81]	84	24 weeks	Paroxetine 20-40 Venlafaxine 75-150	-15.5 (0.95) -15.7 (0.90)
Reimherr, 1990[62]	283	8 weeks	Sertraline 50-200 Placebo	-11.7 (0.69) -8.2 (0.66)
Lydiard, 1997[63]	234	8 weeks	Sertraline 50-200 Placebo	-11.1 (0.63) -8.8 (0.65)
Hypericum, 2002[65]	225	8 weeks	Sertraline 50-100 Placebo	-10.5 (0.72) -9.2 (0.67)
Blumenthal, 2007[67]	98	16 weeks	Sertraline 50-200 Placebo	-6.1 (0.96) -6.1 (1.04)
Munizza, 2006[68]	121	6 weeks	Sertraline 50-100 Trazodone 150-450	-11.5 (1.08) -12.9 (1.15)
Sir, 2005[70]	158	8 weeks	Sertraline 50-150 Venlafaxine 75-225	-15.9 (0.95) -14.3 (0.94)
Shelton, 2006[71]	158	8 weeks	Sertraline 50-150 Venlafaxine 75-225	-11.3 (0.61) -12.7 (0.64)
Thase, 1997[72]	191	8 weeks	Venlafaxine 75-225 Placebo	-11.7 (0.75) -7.3 (0.71)

Note: studies are listed alphabetically by drug and chronologically within each set of drugs

Abbreviations: mg = milligram; SE = standard error

Table B-3. Summary of studies included in analyses of ACR50 response for biologic DMARDs

Author, Year Study Name	N	Comparison and Dose	Disease Duration (yrs)	Response[a]
Kremer, 2003[82] NR	234	Abatacept 10 mg/kg Placebo	9.7 8.9	42/115 14/119
Kremer, 2006[83] AIM	652	Abatacept 10 mg/kg Placebo	8.5 8.9	169/433 36/219
Schiff, 2008[84] ATTEST	431	Abatacept 10 mg/kg Infliximab 3 mg/kg Placebo	7.9 7.3 8.4	63/156 61/165 22/110
Furst, 2003[85] STAR	636	Adalimumab 40 mg Placebo	9.3 11.5	92/318 36/318
van de Putte, 2003[86] NR	212	Adalimumab 20 mg, 40 mg Placebo	10.2 9.4	36/142 1/70
Weinblatt, 2003[87] ARMADA	129	Adalimumab 40 mg Placebo	12.2 11.1	37/67 5/62
Keystone, 2004[88] NR	619	Adalimumab 20 mg, 40 mg Placebo	11.0 10.9	166/419 19/200
van de Putte, 2004[89] NR	438	Adalimumab 20 mg, 40 mg Placebo	10.8 11.6	84/328 9/110
Kim, 2007[90] NR	128	Adalimumab 40 mg Placebo	6.8 6.9	28/65 9/63
Chen, 2009[91] NR	47	Adalimumab 40 mg Placebo	6.2 8.3	12/35 2/12
Bresnihan, 1998[92] NR	353	Anakinra 75 mg, 150 mg Placebo	4.1 3.7	33/232 9/121
Cohen, 2002[93] NR	153	Anakinra 1 mg/kg, 2 mg/kg Placebo	7.3 7.8	22/105 2/48
Cohen, 2004[94] NR	506	Anakinra 100 mg Placebo	11.0 10.0	43/253 20/253
Moreland, 1997[95] NR	88	Etanercept 16 m/m sq. Placebo	NR NR	25/44 3/44
Moreland, 1999[96] NR	158	Etanercept 25 mg Placebo	12 12	31/78 4/80
Weinblatt, 1999[97] NR	89	Etanercept 25 mg Placebo	13 13	23/59 1/30
Lan, 2004[98] NR	58	Etanercept 25 mg Placebo	NR NR	19/29 3/29
Kay, 2008[99] NR	70	Golimumab 50 mg Placebo	8.2 5.6	13/35 2/35
Keystone, 2009[100] GO-FORWARD	222	Golimumab 50 mg Placebo	4.5 6.5	31/89 13/133
Maini, 1999[101] ATTRACT	428	Infliximab 3 mg/kg, 10 mg/kg Placebo	8.3 8.9	94/340 4/88
Kavanaugh, 2000[102] NR	21	Infliximab 5 mg/kg, 10 mg/kg Placebo	6.6 4.9	3/14 1/7
Abe, 2006[103] NR	147	Infliximab 3 mg/kg, 10 mg/kg Placebo	8.1 7.5	33/100 4/47
Westhovens, 2006[104] START	1,084	Infliximab 3 mg/kg, 10 mg/kg Placebo	7.1 8.4	229/721 33/363
Zhang, 2006[105] NR	173	Infliximab 3 mg/kg Placebo	7.1 8.0	38/87 22/86
Edwards, 2004[106] NR	80	Rituximab 2x1,000 mg Placebo	12 11	17/40 5/40
Emery, 2010[107] SERENE	275	Rituximab 2x500 mg, 2x1,000 mg Placebo	6.9 7.5	88/340 16/172
Maini, 2006[108] CHARISMA	148	Tocilizumab 4 mg/kg, 8 mg/kg Placebo	0.79 0.94	45/99 14/49

Table B-3. Summary of studies included in analyses of ACR50 response for biologic DMARDs (continued)

Author, Year Study Name	N	Comparison and Dose	Disease Duration (yrs)	Response[a]
Genovese, 2008[109] TOWARD	1,220	Tocilizumab 8 mg/kg Placebo	9.8 9.8	302/805 37/415
Smolen, 2008[110] OPTION	623	Tocilizumab 4 mg/kg, 8 mg/kg Placebo	7.5 7.8	157/419 22/204
Kremer, 2011[111] LITHE	1,196	Tocilizumab 4 mg/kg, 8 mg/kg Placebo	9.4 9.0	228/802 38/394

Note: studies are listed alphabetically by drug and chronologically within each set of drugs

[a] Response is presented as the number of responders/total number in the group.

Abbreviations: kg = kilogram; mg = milligram; NR = not reported; yrs = years

References

1. Feighner JP, Gardner EA, Johnston JA, et al. Double-blind comparison of bupropion and fluoxetine in depressed outpatients. J Clin Psychiatry. 1991 Aug;52(8):329-35. PMID: 1907963.

2. Coleman CC, King BR, Bolden-Watson C, et al. A placebo-controlled comparison of the effects on sexual functioning of bupropion sustained release and fluoxetine. Clin Ther. 2001 Jul;23(7):1040-58. PMID: 11519769.

3. Weihs KL, Settle EC, Jr., Batey SR, et al. Bupropion sustained release versus paroxetine for the treatment of depression in the elderly. J Clin Psychiatry. 2000 Mar;61(3):196-202. PMID: 10817105.

4. Rush AJ, Trivedi MH, Carmody TJ, et al. Response in relation to baseline anxiety levels in major depressive disorder treated with bupropion sustained release or sertraline. Neuropsychopharmacology. 2001 Jul;25(1):131-8. PMID: 11377926.

5. Coleman CC, Cunningham LA, Foster VJ, et al. Sexual dysfunction associated with the treatment of depression: a placebo-controlled comparison of bupropion sustained release and sertraline treatment. Ann Clin Psychiatry. 1999 Dec;11(4):205-15. PMID: 10596735.

6. Croft H, Settle E, Jr., Houser T, et al. A placebo-controlled comparison of the antidepressant efficacy and effects on sexual functioning of sustained-release bupropion and sertraline. Clin Ther. 1999 Apr;21(4):643-58. PMID: 10363731.

7. Weisler RH, Johnston JA, Lineberry CG, et al. Comparison of bupropion and trazodone for the treatment of major depression. J Clin Psychopharmacol. 1994 Jun;14(3):170-9. PMID: 8027413.

8. Ou JJ, Xun GL, Wu RR, et al. Efficacy and safety of escitalopram versus citalopram in major depressive disorder: a 6-week, multicenter, randomized, double-blind, flexible-dose study. Psychopharmacology (Berl). 2011 Feb;213(2-3):639-46. PMID: 20340011.

9. Haffmans PM, Timmerman L, Hoogduin CA. Efficacy and tolerability of citalopram in comparison with fluvoxamine in depressed outpatients: a double-blind, multicentre study. The LUCIFER Group. Int Clin Psychopharmacol. 1996 Sep;11(3):157-64. PMID: 8923094.

10. Roose SP, Sackeim HA, Krishnan KR, et al. Antidepressant pharmacotherapy in the treatment of depression in the very old: a randomized, placebo-controlled trial. Am J Psychiatry. 2004 Nov;161(11):2050-9. PMID: 15514406.

11. Tourian KA, Padmanabhan SK, Groark J, et al. Desvenlafaxine 50 and 100 mg/d in the treatment of major depressive disorder: an 8-week, phase III, multicenter, randomized, double-blind, placebo-controlled, parallel-group trial and a post hoc pooled analysis of three studies. Clin Ther; 2009. p. 1405-23.

12. Feiger AD, Tourian KA, Rosas GR, et al. A placebo-controlled study evaluating the efficacy and safety of flexible-dose desvenlafaxine treatment in outpatients with major depressive disorder. CNS Spectr; 2009. p. 41-50.

13. Boyer P, Montgomery S, Lepola U, et al. Efficacy, safety, and tolerability of fixed-dose desvenlafaxine 50 and 100 mg/day for major depressive disorder in a placebo-controlled trial. Int Clin Psychopharmacol; 2008. p. 243-53.

14. Liebowitz MR, Yeung PP, Entsuah R. A randomized, double-blind, placebo-controlled trial of desvenlafaxine succinate in adult outpatients with major depressive disorder. J Clin Psychiatry; 2007. p. 1663-72.

15. Septien-Velez L, Pitrosky B, Padmanabhan SK, et al. A randomized, double-blind, placebo-controlled trial of desvenlafaxine succinate in the treatment of major depressive disorder. Int Clin Psychopharmacol; 2007. p. 338-47.

16. Khan A, Bose A, Alexopoulos GS, et al. Double-blind comparison of escitalopram and duloxetine in the acute treatment of major depressive disorder. Clin Drug Investig; 2007. p. 481-92.

17. Wade A, Gembert K, Florea I. A comparative study of the efficacy of acute and continuation treatment with escitalopram versus duloxetine in patients with major depressive disorder. Curr Med Res Opin; 2007. p. 1605-14.

18. Nierenberg AA, Greist JH, Mallinckrodt CH, et al. Duloxetine versus escitalopram and placebo in the treatment of patients with major depressive disorder: onset of antidepressant action, a non-inferiority study. Curr Med Res Opin; 2007. p. 401-16.

19. Goldstein DJ, Mallinckrodt C, Lu Y, et al. Duloxetine in the treatment of major depressive disorder: a double-blind clinical trial. J Clin Psychiatry. 2002 Mar;63(3):225-31. PMID: 11926722.

20. Lee P, Shu L, Xu X, et al. Once-daily duloxetine 60 mg in the treatment of major depressive disorder: multicenter, double-blind, randomized, paroxetine-controlled, non-inferiority trial in China, Korea, Taiwan and Brazil. Psychiatry Clin Neurosci; 2007. p. 295-307.

21. Detke MJ, Wiltse CG, Mallinckrodt CH, et al. Duloxetine in the acute and long-term treatment of major depressive disorder: a placebo- and paroxetine-controlled trial. Eur Neuropsychopharmacol. 2004 Dec;14(6):457-70. PMID: 15589385.

22. Perahia DG, Wang F, Mallinckrodt CH, et al. Duloxetine in the treatment of major depressive disorder: a placebo- and paroxetine-controlled trial. Eur Psychiatry; 2006. p. 367-78.

23. Brannan SK, Mallinckrodt CH, Brown EB, et al. Duloxetine 60 mg once-daily in the treatment of painful physical symptoms in patients with major depressive disorder. J Psychiatr Res. 2005 Jan;39(1):43-53. PMID: 15504423.

24. Detke MJ, Lu Y, Goldstein DJ, et al. Duloxetine 60 mg once daily dosing versus placebo in the acute treatment of major depression. J Psychiatr Res. 2002 Nov-Dec;36(6):383-90. PMID: 12393307.

25. Detke MJ, Lu Y, Goldstein DJ, et al. Duloxetine, 60 mg once daily, for major depressive disorder: a randomized double-blind placebo-controlled trial. J Clin Psychiatry. 2002 Apr;63(4):308-15. PMID: 12000204.

26. Mao PX, Tang YL, Jiang F, et al. Escitalopram in major depressive disorder: a multicenter, randomized, double-blind, fixed-dose, parallel trial in a Chinese population. Depress Anxiety; 2008. p. 46-54.

27. Boulenger JP, Huusom AK, Florea I, et al. A comparative study of the efficacy of long-term treatment with escitalopram and paroxetine in severely depressed patients. Curr Med Res Opin; 2006. p. 1331-41.

28. Ventura D, Armstrong EP, Skrepnek GH, et al. Escitalopram versus sertraline in the treatment of major depressive disorder: a randomized clinical trial. Curr Med Res Opin; 2007. p. 245-50.

29. Bielski RJ, Ventura D, Chang CC. A double-blind comparison of escitalopram and venlafaxine extended release in the treatment of major depressive disorder. J Clin Psychiatry. 2004 Sep;65(9):1190-6. PMID: 15367045.

30. Hong CJ, Hu WH, Chen CC, et al. A double-blind, randomized, group-comparative study of the tolerability and efficacy of 6 weeks' treatment with mirtazapine or fluoxetine in depressed Chinese patients. J Clin Psychiatry. 2003 Aug;64(8):921-6. PMID: 12927007.

31. Rush AJ, Armitage R, Gillin JC, et al. Comparative effects of nefazodone and fluoxetine on sleep in outpatients with major depressive disorder. Biol Psychiatry. 1998 Jul 1;44(1):3-14. PMID: 9646878.

32. De Wilde J, Spiers R, Mertens C, et al. A double-blind, comparative, multicentre study comparing paroxetine with fluoxetine in depressed patients. Acta Psychiatr Scand. 1993 Feb;87(2):141-5. PMID: 8447241.

33. Gagiano CA. A double blind comparison of paroxetine and fluoxetine in patients with major depression. Br J Clin Res. 1993;4:145-52.

34. Chouinard G, Saxena B, Belanger MC, et al. A Canadian multicenter, double-blind study of paroxetine and fluoxetine in major depressive disorder. J Affect Disord. 1999 Jul;54(1-2):39-48. PMID: 10403145.

35. Fava M, Amsterdam JD, Deltito JA, et al. A double-blind study of paroxetine, fluoxetine, and placebo in outpatients with major depression. Ann Clin Psychiatry. 1998 Dec;10(4):145-50. PMID: 9988054.

36. Fava M, Hoog SL, Judge RA, et al. Acute efficacy of fluoxetine versus sertraline and paroxetine in major depressive disorder including effects of baseline insomnia. J Clin Psychopharmacol. 2002 Apr;22(2):137-47. PMID: 11910258.

37. Wernicke JF, Dunlop SR, Dornseif BE, et al. Low-dose fluoxetine therapy for depression. Psychopharmacol Bull. 1988;24(1):183-8. PMID: 3290940.

38. Tollefson GD, Holman SL. Analysis of the Hamilton Depression Rating Scale factors from a double-blind, placebo-controlled trial of fluoxetine in geriatric major depression. Int Clin Psychopharmacol. 1993 Winter;8(4):253-9. PMID: 8277144.

39. Bennie EH, Mullin JM, Martindale JJ. A double-blind multicenter trial comparing sertraline and fluoxetine in outpatients with major depression. J Clin Psychiatry. 1995 Jun;56(6):229-37. PMID: 7775364.

40. Sechter D, Troy S, Paternetti S, et al. A double-blind comparison of sertraline and fluoxetine in the treatment of major depressive episode in outpatients. Eur Psychiatry. 1999 Mar;14(1):41-8. PMID: 10572324.

41. Newhouse PA, Krishnan KR, Doraiswamy PM, et al. A double-blind comparison of sertraline and fluoxetine in depressed elderly outpatients. J Clin Psychiatry. 2000 Aug;61(8):559-68. PMID: 10982198.

42. Dierick M, Ravizza L, Realini R, et al. A double-blind comparison of venlafaxine and fluoxetine for treatment of major depression in outpatients. Prog Neuropsychopharmacol Biol Psychiatry. 1996 Jan;20(1):57-71. PMID: 8861177.

43. Tylee A, Beaumont G, Bowden MW, et al. A double-blind, randomized, 12-week comparison study of the safety and efficacy of venlafaxine and fluoxetine in moderate to severe major depression in general practice. Primary Care Psychiatry. 1997;3(1):51-8.

44. Costa e Silva J. Randomized, double-blind comparison of venlafaxine and fluoxetine in outpatients with major depression. J Clin Psychiatry. 1998 Jul;59(7):352-7. PMID: 9714263.

45. Alves C, Cachola I, Brandao J. Efficacy and tolerability of venlafaxine and fluoxetine in outpatients with major depression. Primary Care Psychiatry. 1999;5(2):57-63.

46. Tzanakaki M, Guazzelli M, Nimatoudis I, et al. Increased remission rates with venlafaxine compared with fluoxetine in hospitalized patients with major depression and melancholia. Int Clin Psychopharmacol; 2000. p. 29-34.

47. De Nayer A, Geerts S, Ruelens L, et al. Venlafaxine compared with fluoxetine in outpatients with depression and concomitant anxiety. Int J Neuropsychopharmacol. 2002 Jun;5(2):115-20. PMID: 12135535.

48. Rudolph RL, Feiger AD. A double-blind, randomized, placebo-controlled trial of once-daily venlafaxine extended release (XR) and fluoxetine for the treatment of depression. J Affect Disord. 1999 Dec;56(2-3):171-81. PMID: 10701474.

49. Nemeroff CB, Thase ME. A double-blind, placebo-controlled comparison of venlafaxine and fluoxetine treatment in depressed outpatients. J Psychiatr Res; 2007. p. 351-9. Epub 2005 Sep 12.

50. Lydiard RB, Laird LK, Morton WA, Jr., et al. Fluvoxamine, imipramine, and placebo in the treatment of depressed outpatients: effects on depression. Psychopharmacol Bull. 1989;25(1):68-70. PMID: 2505304.

51. Rossini D, Serretti A, Franchini L, et al. Sertraline versus fluvoxamine in the treatment of elderly patients with major depression: a double-blind, randomized trial. J Clin Psychopharmacol. 2005 Oct;25(5):471-5. PMID: 16160624.

52. Benkert O, Szegedi A, Kohnen R. Mirtazapine compared with paroxetine in major depression. J Clin Psychiatry. 2000 Sep;61(9):656-63. PMID: 11030486.

53. Halikas JA. Org 3770 (mirtazapine) versus trazodone: A placebo controlled trial in depressed elderly patients. Hum Psychopharmacol. 1995;10(Suppl 2):S125-S33.

54. van Moffaert M, de Wilde J, Vereecken A, et al. Mirtazapine is more effective than trazodone: a double-blind controlled study in hospitalized patients with major depression. Int Clin Psychopharmacol. 1995 Mar;10(1):3-9. PMID: 7622801.

55. Hicks JA, Argyropoulos SV, Rich AS, et al. Randomised controlled study of sleep after nefazodone or paroxetine treatment in out-patients with depression. Br J Psychiatry. 2002 Jun;180:528-35. PMID: 12042232.

56. Cohn CK, Robinson DS, Roberts DL, et al. Responders to antidepressant drug treatment: a study comparing nefazodone, imipramine, and placebo in patients with major depression. J Clin Psychiatry. 1996;57 Suppl 2:15-8. PMID: 8626358.

57. Fontaine R, Ontiveros A, Elie R, et al. A double-blind comparison of nefazodone, imipramine, and placebo in major depression. J Clin Psychiatry. 1994 Jun;55(6):234-41. PMID: 8071277.

58. Feiger A, Kiev A, Shrivastava RK, et al. Nefazodone versus sertraline in outpatients with major depression: focus on efficacy, tolerability, and effects on sexual function and satisfaction. J Clin Psychiatry. 1996;57 Suppl 2:53-62. PMID: 8626364.

59. Rickels K, Amsterdam J, Clary C, et al. A placebo-controlled, double-blind, clinical trial of paroxetine in depressed outpatients. Acta Psychiatr Scand Suppl. 1989;350:117-23. PMID: 2530761.

60. Rapaport MH, Lydiard RB, Pitts CD, et al. Low doses of controlled-release paroxetine in the treatment of late-life depression: a randomized, placebo-controlled trial. J Clin Psychiatry; 2009. p. 46-57.

61. Kasper S, Olivieri L, Di Loreto G, et al. A comparative, randomised, double-blind study of trazodone prolonged-release and paroxetine in the treatment of patients with major depressive disorder. Curr Med Res Opin. 2005 Aug;21(8):1139-46. PMID: 16083521.

62. Reimherr FW, Chouinard G, Cohn CK, et al. Antidepressant efficacy of sertraline: a double-blind, placebo- and amitriptyline-controlled, multicenter comparison study in outpatients with major depression. J Clin Psychiatry. 1990 Dec;51 Suppl B:18-27. PMID: 2258378.

63. Lydiard RB, Stahl SM, Hertzman M, et al. A double-blind, placebo-controlled study comparing the effects of sertraline versus amitriptyline in the treatment of major depression. J Clin Psychiatry. 1997 Nov;58(11):484-91. PMID: 9413414.

64. Olie JP, Gunn KP, Katz E. A double-blind placebo-controlled multicentre study of sertraline in the acute and continuation treatment of major depression. Eur Psychiat. 1997;12(1):34-41.

65. Hypericum Depression Trial Study Group. Effect of Hypericum perforatum (St John's wort) in major depressive disorder: a randomized controlled trial. JAMA. 2002 Apr 10;287(14):1807-14. PMID: 11939866.

66. Schneider LS, Nelson JC, Clary CM, et al. An 8-week multicenter, parallel-group, double-blind, placebo-controlled study of sertraline in elderly outpatients with major depression. Am J Psychiatry. 2003 Jul;160(7):1277-85. PMID: 12832242.

67. Blumenthal JA, Babyak MA, Doraiswamy PM, et al. Exercise and pharmacotherapy in the treatment of major depressive disorder. Psychosom Med; 2007. p. 587-96.

68. Munizza C, Olivieri L, Di Loreto G, et al. A comparative, randomized, double-blind study of trazodone prolonged-release and sertraline in the treatment of major depressive disorder. Curr Med Res Opin; 2006. p. 1703-13.

69. Mehtonen OP, Sogaard J, Roponen P, et al. Randomized, double-blind comparison of venlafaxine and sertraline in outpatients with major depressive disorder. Venlafaxine 631 Study Group. J Clin Psychiatry. 2000 Feb;61(2):95-100. PMID: 10732656.

70. Sir A, D'Souza RF, Uguz S, et al. Randomized trial of sertraline versus venlafaxine XR in major depression: efficacy and discontinuation symptoms. J Clin Psychiatry. 2005 Oct;66(10):1312-20.

71. Shelton RC, Haman KL, Rapaport MH, et al. A randomized, double-blind, active-control study of sertraline versus venlafaxine XR in major depressive disorder. J Clin Psychiatry; 2006. p. 1674-81.

72. Thase ME. Efficacy and tolerability of once-daily venlafaxine extended release (XR) in outpatients with major depression. The Venlafaxine XR 209 Study Group. J Clin Psychiatry. 1997 Sep;58(9):393-8. PMID: 9378690.

73. Perry PJ, Garvey MJ, Kelly MW, et al. A comparative trial of fluoxetine versus trazodone in outpatients with major depression. J Clin Psychiatry. 1989 Aug;50(8):290-4. PMID: 2668259.

74. Beasley CM, Jr., Dornseif BE, Pultz JA, et al. Fluoxetine versus trazodone: efficacy and activating-sedating effects. J Clin Psychiatry. 1991 Jul;52(7):294-9. PMID: 2071559.

75. Kiev A, Feiger A. A double-blind comparison of fluvoxamine and paroxetine in the treatment of depressed outpatients. J Clin Psychiatry. 1997 Apr;58(4):146-52. PMID: 9164424.

76. Ushiroyama T, Ikeda A, Ueki M. Evaluation of double-blind comparison of fluvoxamine and paroxetine in the treatment of depressed outpatients in menopause transition. J Med; 2004. p. 151-62.

77. Nemeroff CB, Ninan PT, Ballenger J, et al. Double-blind multicenter comparison of fluvoxamine versus sertraline in the treatment of depressed outpatients. Depression. 1995;3(4):163-9.

78. Schatzberg AF, Kremer C, Rodrigues HE, et al. Double-blind, randomized comparison of mirtazapine and paroxetine in elderly depressed patients. Am J Geriatr Psychiatry. 2002 Sep-Oct;10(5):541-50. PMID: 12213688.

79. Ehde DM, Kraft GH, Chwastiak L, et al. Efficacy of paroxetine in treating major depressive disorder in persons with multiple sclerosis. Gen Hosp Psychiatry; 2008. p. 40-8.

80. McPartlin GM, Reynolds A, Anderson C, et al. A comparison of once-daily venlafaxine XR and paroxetine in depressed outpatients treated in general practice. Primary Care Psychiatry. 1998 1998;4(3):127-32.

81. Ballus C, Quiros G, De Flores T, et al. The efficacy and tolerability of venlafaxine and paroxetine in outpatients with depressive disorder or dysthymia. Int Clin Psychopharmacol. 2000 Jan;15(1):43-8. PMID: 10836286.

82. Kremer JM, Westhovens R, Leon M, et al. Treatment of rheumatoid arthritis by selective inhibition of T-cell activation with fusion protein CTLA4Ig. N Engl J Med. 2003 Nov 13;349(20):1907-15. PMID: 14614165.

83. Kremer JM, Genant HK, Moreland LW, et al. Effects of abatacept in patients with methotrexate-resistant active rheumatoid arthritis: a randomized trial. Ann Intern Med. 2006 Jun 20;144(12):865-76. PMID: 16785475.

84. Schiff M, Keiserman M, Codding C, et al. Efficacy and safety of abatacept or infliximab vs placebo in ATTEST: a phase III, multi-centre, randomised, double-blind, placebo-controlled study in patients with rheumatoid arthritis and an inadequate response to methotrexate. Ann Rheum Dis. 2008 Aug;67(8):1096-103. PMID: 18055472.

85. Furst DE, Schiff MH, Fleischmann RM, et al. Adalimumab, a fully human anti tumor necrosis factor-alpha monoclonal antibody, and concomitant standard antirheumatic therapy for the treatment of rheumatoid arthritis: results of STAR (Safety Trial of Adalimumab in Rheumatoid Arthritis). J Rheumatol. 2003 Dec;30(12):2563-71. PMID: 14719195.

86. van de Putte LB, Rau R, Breedveld FC, et al. Efficacy and safety of the fully human anti-tumour necrosis factor alpha monoclonal antibody adalimumab (D2E7) in DMARD refractory patients with rheumatoid arthritis: a 12 week, phase II study. Ann Rheum Dis. 2003 Dec;62(12):1168-77. PMID: 14644854.

87. Weinblatt ME, Keystone EC, Furst DE, et al. Adalimumab, a fully human anti-tumor necrosis factor alpha monoclonal antibody, for the treatment of rheumatoid arthritis in patients taking concomitant methotrexate: the ARMADA trial. Arthritis Rheum. 2003 Jan;48(1):35-45. PMID: 12528101.

88. Keystone EC, Kavanaugh AF, Sharp JT, et al. Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. Arthritis Rheum. 2004 May;50(5):1400-11. PMID: 15146409.

89. van de Putte LB, Atkins C, Malaise M, et al. Efficacy and safety of adalimumab as monotherapy in patients with rheumatoid arthritis for whom previous disease modifying antirheumatic drug treatment has failed. Ann Rheum Dis. 2004 May;63(5):508-16. PMID: 15082480.

90. Kim HY, Lee SK, Song YW, et al. A randomized, double-blind, placebo-controlled, phase III study of the human anti-tumor necrosis factor antibody adalimumab administered as subcutaneous injections in Korean rheumatoid arthritis patients treated with methotrexate. APLAR J Rheumatol. 2007;10(1):9-16.

91. Chen DY, Chou SJ, Hsieh TY, et al. Randomized, double-blind, placebo-controlled, comparative study of human anti-TNF antibody adalimumab in combination with methotrexate and methotrexate alone in Taiwanese patients with active rheumatoid arthritis. J Formos Med Assoc. 2009 Apr;108(4):310-9. PMID: 19369178.

92. Bresnihan B, Alvaro-Gracia JM, Cobby M, et al. Treatment of rheumatoid arthritis with recombinant human interleukin-1 receptor antagonist. Arthritis Rheum. 1998 Dec;41(12):2196-204. PMID: 9870876.

93. Cohen S, Hurd E, Cush J, et al. Treatment of rheumatoid arthritis with anakinra, a recombinant human interleukin-1 receptor antagonist, in combination with methotrexate: results of a twenty-four-week, multicenter, randomized, double-blind, placebo-controlled trial. Arthritis Rheum. 2002 Mar;46(3):614-24. PMID: 11920396.

94. Cohen SB, Moreland LW, Cush JJ, et al. A multicentre, double blind, randomised, placebo controlled trial of anakinra (Kineret), a recombinant interleukin 1 receptor antagonist, in patients with rheumatoid arthritis treated with background methotrexate. Ann Rheum Dis. 2004 Sep;63(9):1062-8. PMID: 15082469.

95. Moreland LW, Baumgartner SW, Schiff MH, et al. Treatment of rheumatoid arthritis with a recombinant human tumor necrosis factor receptor (p75)-Fc fusion protein. N Engl J Med. 1997 Jul 17;337(3):141-7. PMID: 9219699.

96. Moreland LW, Schiff MH, Baumgartner SW, et al. Etanercept therapy in rheumatoid arthritis. A randomized, controlled trial. Ann Intern Med. 1999 Mar 16;130(6):478-86. PMID: 10075615.

97. Weinblatt ME, Kremer JM, Bankhurst AD, et al. A trial of etanercept, a recombinant tumor necrosis factor receptor:Fc fusion protein, in patients with rheumatoid arthritis receiving methotrexate. N Engl J Med. 1999 Jan 28;340(4):253-9. PMID: 9920948.

98. Lan JL, Chou SJ, Chen DY, et al. A comparative study of etanercept plus methotrexate and methotrexate alone in Taiwanese patients with active rheumatoid arthritis: a 12-week, double-blind, randomized, placebo-controlled study. J Formos Med Assoc. 2004 Aug;103(8):618-23. PMID: 15340661.

99. Kay J, Matteson EL, Dasgupta B, et al. Golimumab in patients with active rheumatoid arthritis despite treatment with methotrexate: a randomized, double-blind, placebo-controlled, dose-ranging study. Arthritis Rheum. 2008 Apr;58(4):964-75. PMID: 18383539.

100. Keystone EC, Genovese MC, Klareskog L, et al. Golimumab, a human antibody to tumour necrosis factor {alpha} given by monthly subcutaneous injections, in active rheumatoid arthritis despite methotrexate therapy: the GO-FORWARD Study. Ann Rheum Dis. 2009 Jun;68(6):789-96. PMID: 19066176.

101. Maini R, St Clair EW, Breedveld F, et al. Infliximab (chimeric anti-tumour necrosis factor alpha monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomised phase III trial. ATTRACT Study Group. Lancet. 1999 Dec 4;354(9194):1932-9. PMID: 10622295.

102. Kavanaugh A, St Clair EW, McCune WJ, et al. Chimeric anti-tumor necrosis factor-alpha monoclonal antibody treatment of patients with rheumatoid arthritis receiving methotrexate therapy. J Rheumatol. 2000 Apr;27(4):841-50. PMID: 10782805.

103. Abe T, Takeuchi T, Miyasaka N, et al. A multicenter, double-blind, randomized, placebo controlled trial of infliximab combined with low dose methotrexate in Japanese patients with rheumatoid arthritis. J Rheumatol. 2006 Jan;33(1):37-44. PMID: 16395748.

104. Westhovens R, Yocum D, Han J, et al. The safety of infliximab, combined with background treatments, among patients with rheumatoid arthritis and various comorbidities: a large, randomized, placebo-controlled trial. Arthritis Rheum. 2006 Apr;54(4):1075-86. PMID: 16572442.

105. Zhang FC, Hou Y, Huang F, et al. Infliximab versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a preliminary study from China. APLAR J Rheumatol. 2006;9(2):127-30.

106. Edwards JC, Szczepanski L, Szechinski J, et al. Efficacy of B-cell-targeted therapy with rituximab in patients with rheumatoid arthritis. N Engl J Med. 2004 Jun 17;350(25):2572-81. PMID: 15201414.

107. Emery P, Deodhar A, Rigby WF, et al. Efficacy and safety of different doses and retreatment of rituximab: a randomised, placebo-controlled trial in patients who are biological naive with active rheumatoid arthritis and an inadequate response to methotrexate (Study Evaluating Rituximab's Efficacy in MTX iNadequate rEsponders (SERENE)). Ann Rheum Dis. 2010 Sep;69(9):1629-35. PMID: 20488885.

108. Maini RN, Taylor PC, Szechinski J, et al. Double-blind randomized controlled clinical trial of the interleukin-6 receptor antagonist, tocilizumab, in European patients with rheumatoid arthritis who had an incomplete response to methotrexate. Arthritis Rheum. 2006 Sep;54(9):2817-29. PMID: 16947782.

109. Genovese MC, McKay JD, Nasonov EL, et al. Interleukin-6 receptor inhibition with tocilizumab reduces disease activity in rheumatoid arthritis with inadequate response to disease-modifying antirheumatic drugs: the tocilizumab in combination with traditional disease-modifying antirheumatic drug therapy study. Arthritis Rheum. 2008 Oct;58(10):2968-80. PMID: 18821691.

110. Smolen JS, Beaulieu A, Rubbert-Roth A, et al. Effect of interleukin-6 receptor inhibition with tocilizumab in patients with rheumatoid arthritis (OPTION study): a double-blind, placebo-controlled, randomised trial. Lancet. 2008 Mar 22;371(9617):987-97. PMID: 18358926.

111. Kremer JM, Blanco R, Brzosko M, et al. Tocilizumab inhibits structural joint damage in rheumatoid arthritis patients with inadequate responses to methotrexate: results from the double-blind treatment phase of a randomized placebo-controlled trial of tocilizumab safety and prevention of structural joint damage at one year. Arthritis Rheum. 2011 Mar;63(3):609-21. PMID: 21360490.

Appendix C. Evidence Networks

Figure C-1. Response for second-generation antidepressants: Full network

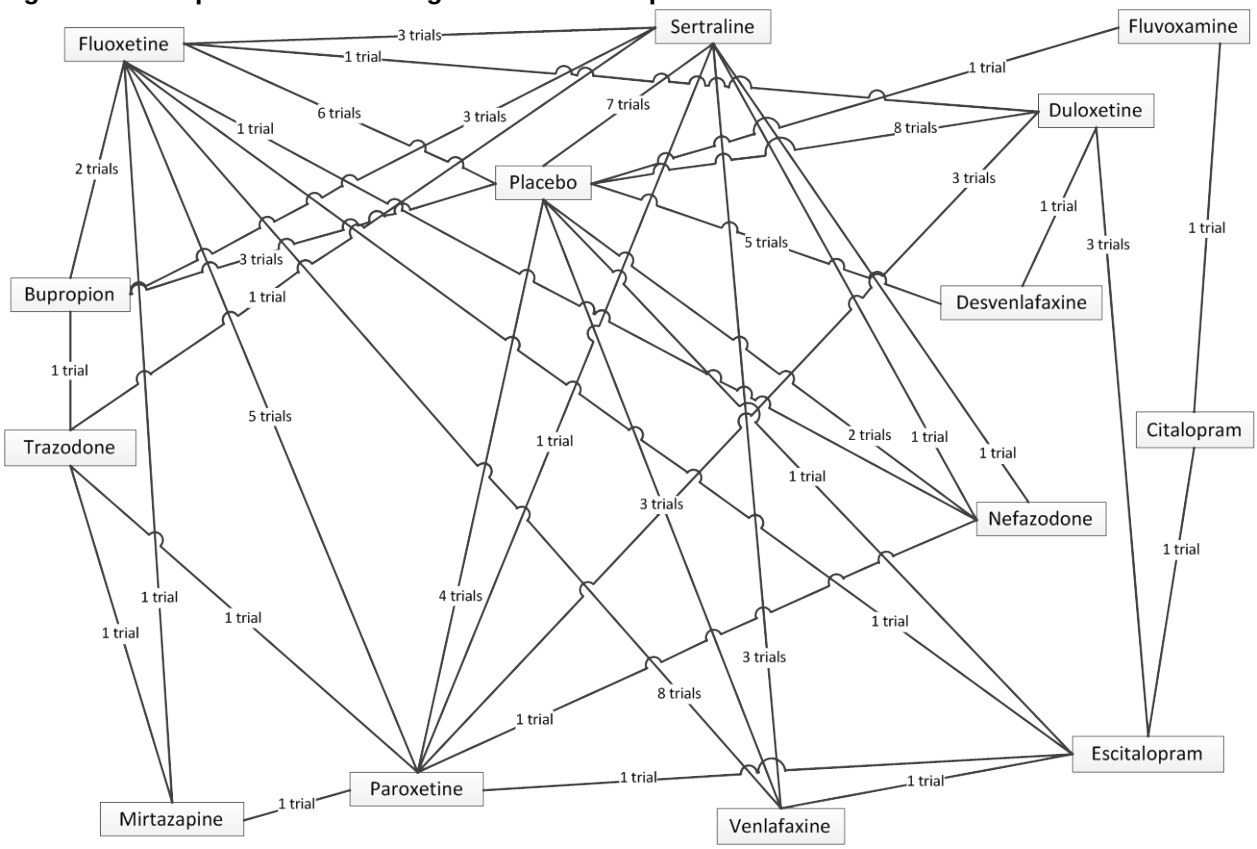


Figure C-2. Response for second-generation antidepressants: Star sub-network

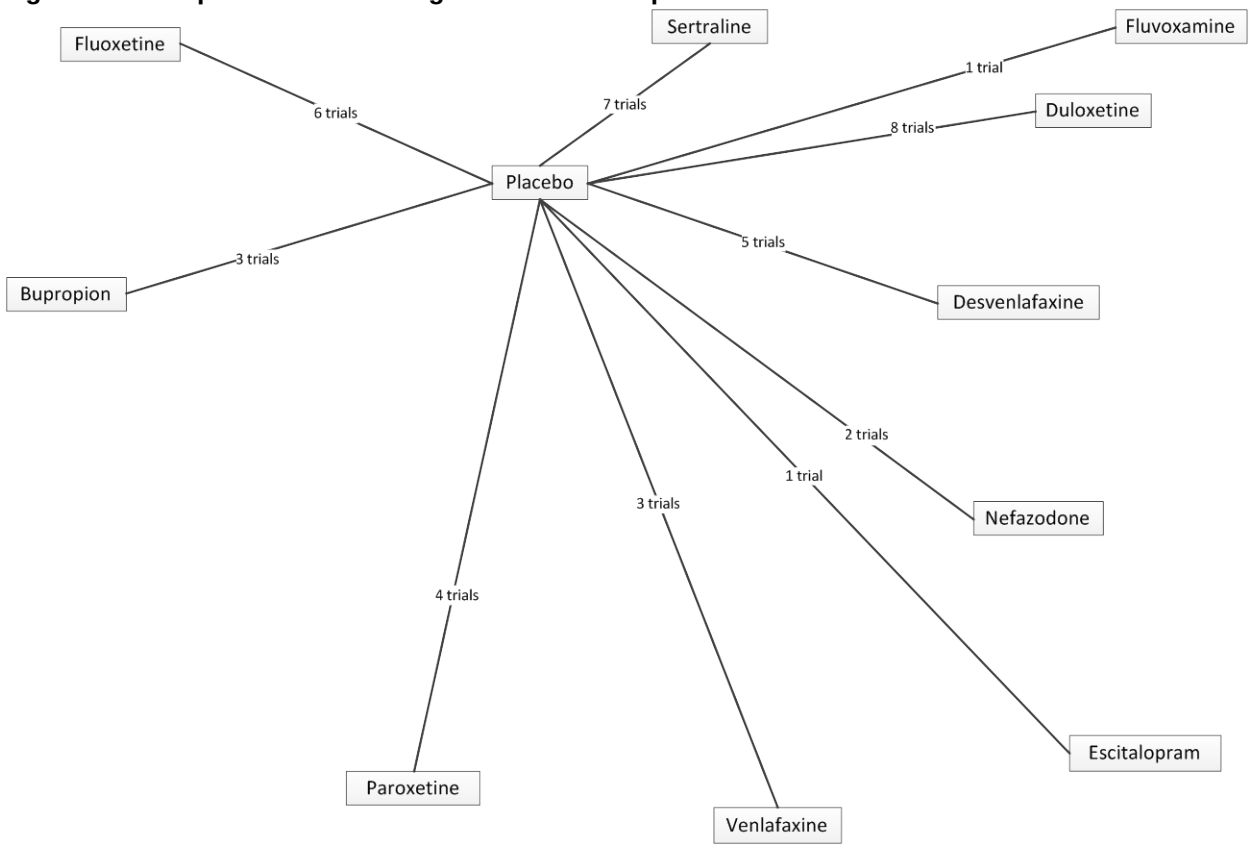


Figure C-3. Response for second-generation antidepressants: Loop sub-network

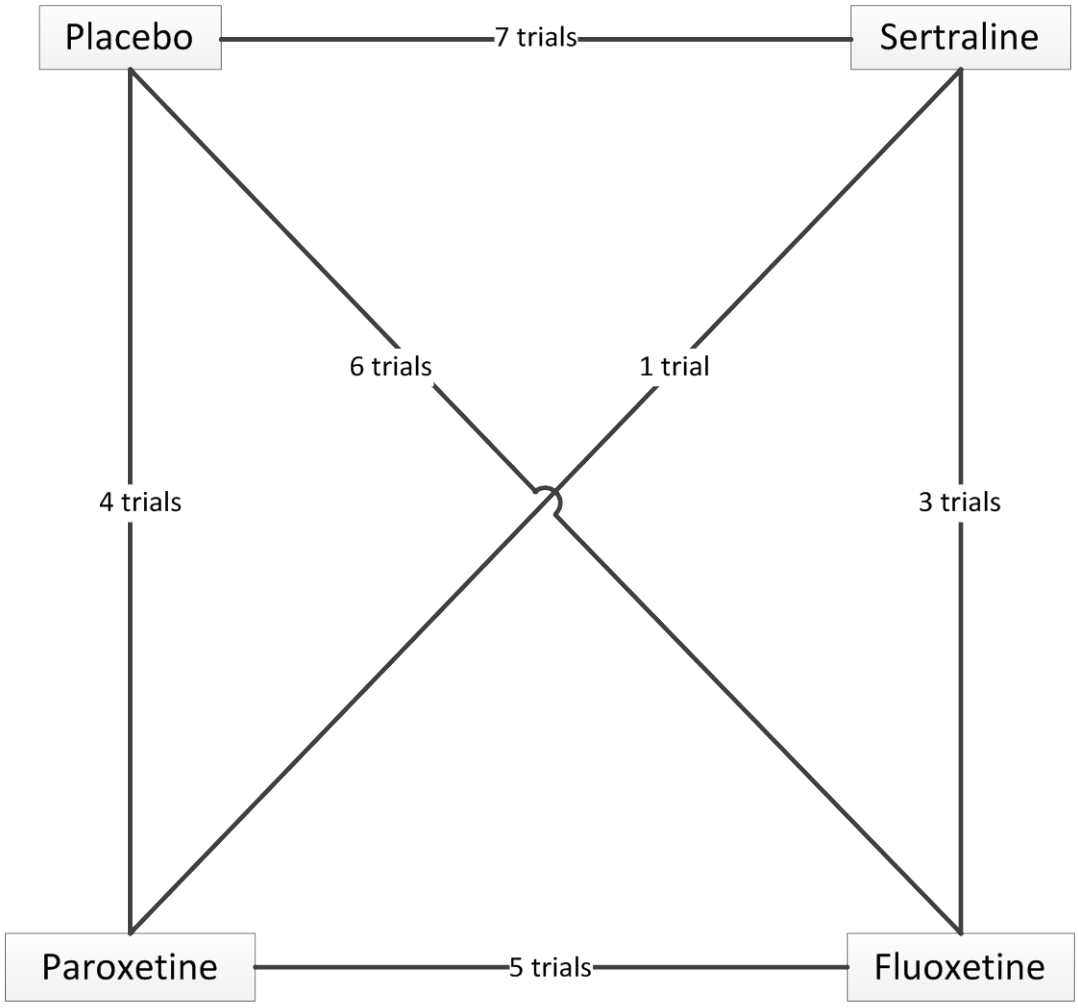


Figure C-4. Response for second-generation antidepressants: One closed loop sub-network

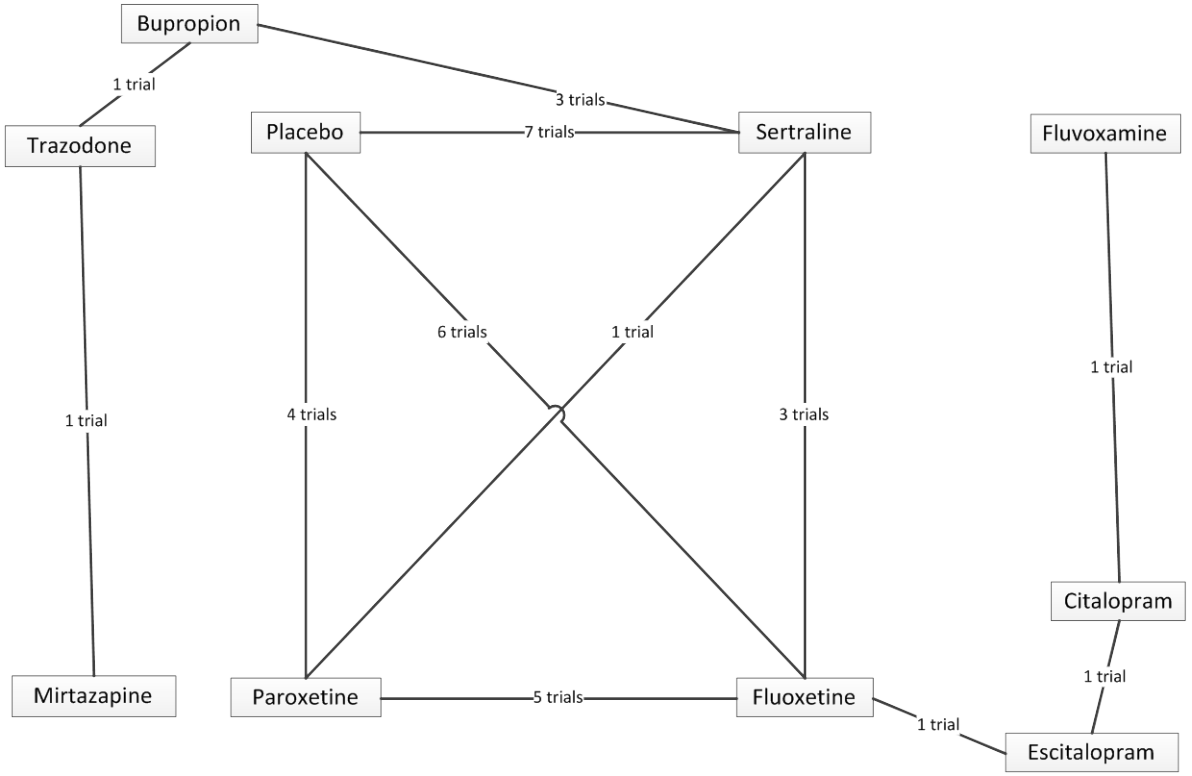


Figure C-5. Response for second-generation antidepressants: Ladder sub-network

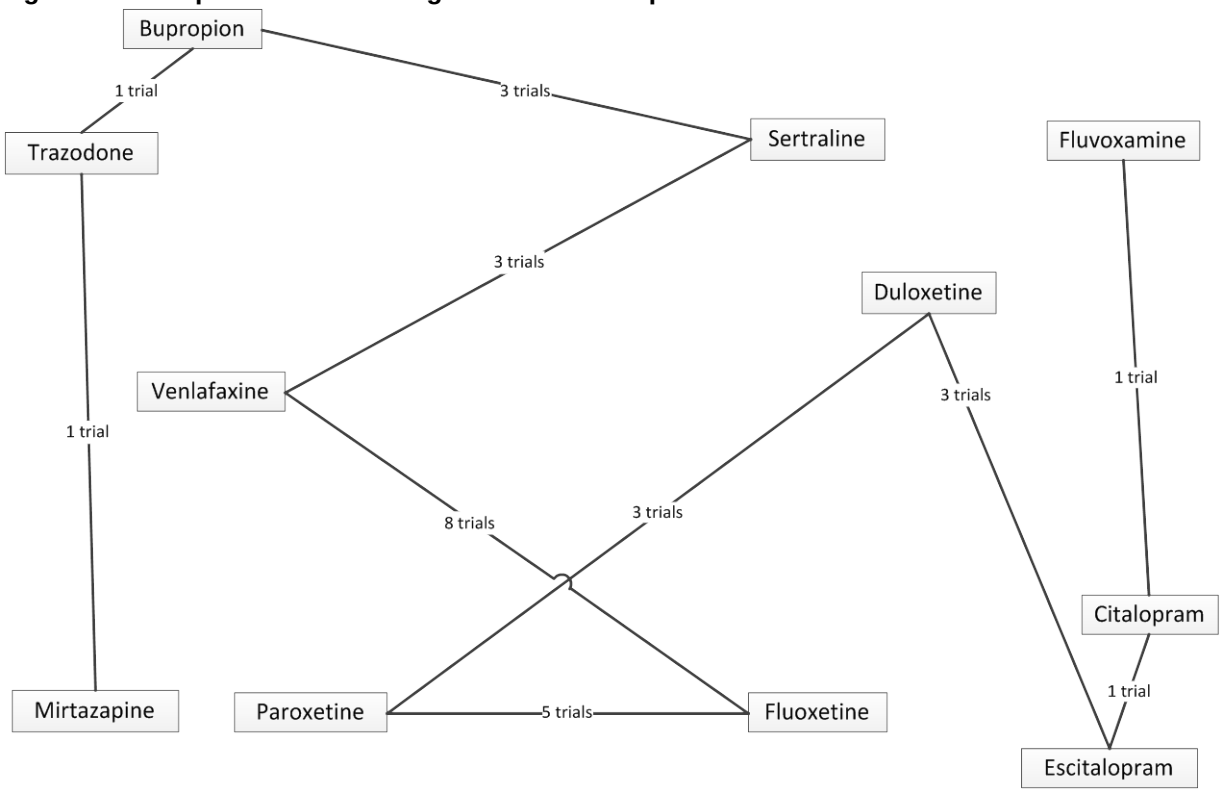


Figure C-6. Response for second-generation antidepressants: Full network, including studies in older adults (meta-regression for Key Question 2)

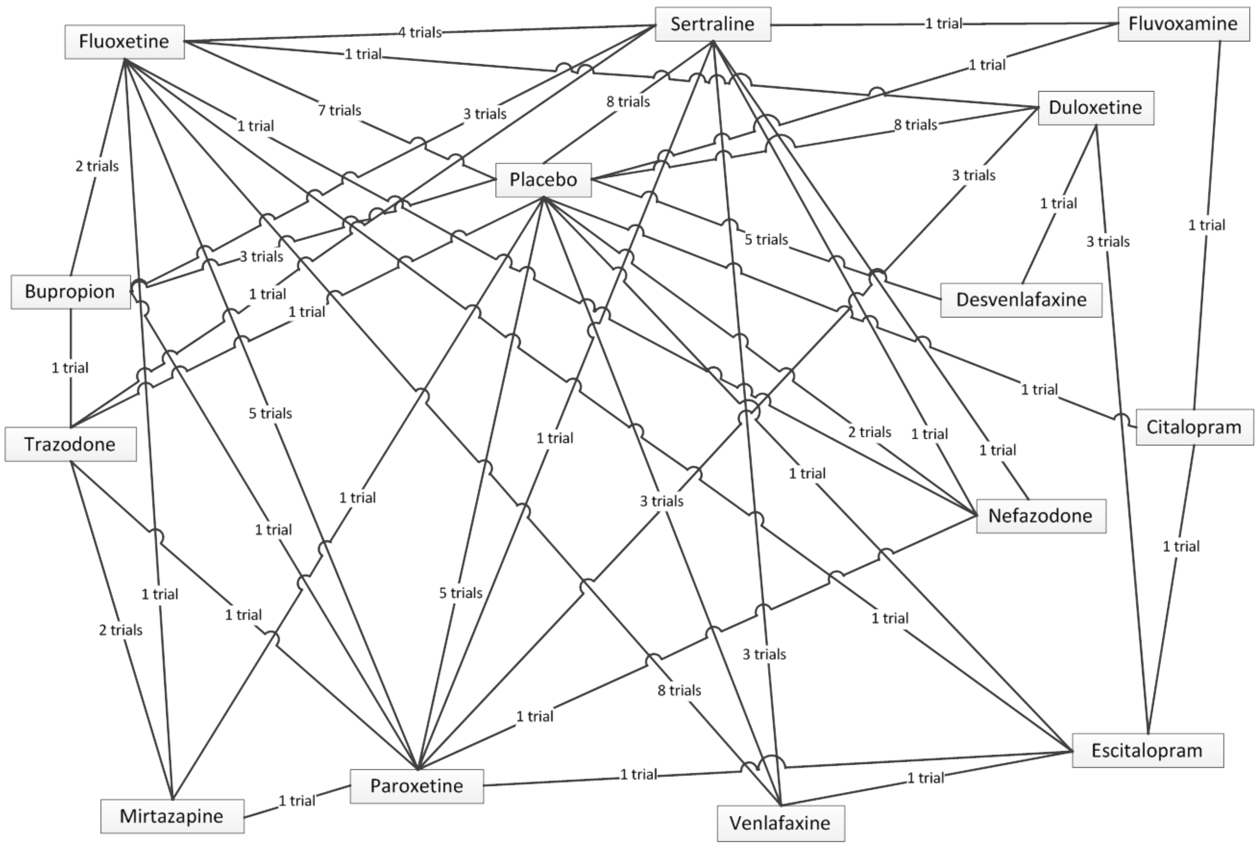


Figure C-7. Mean change in HAM-D for second-generation antidepressants: Full network

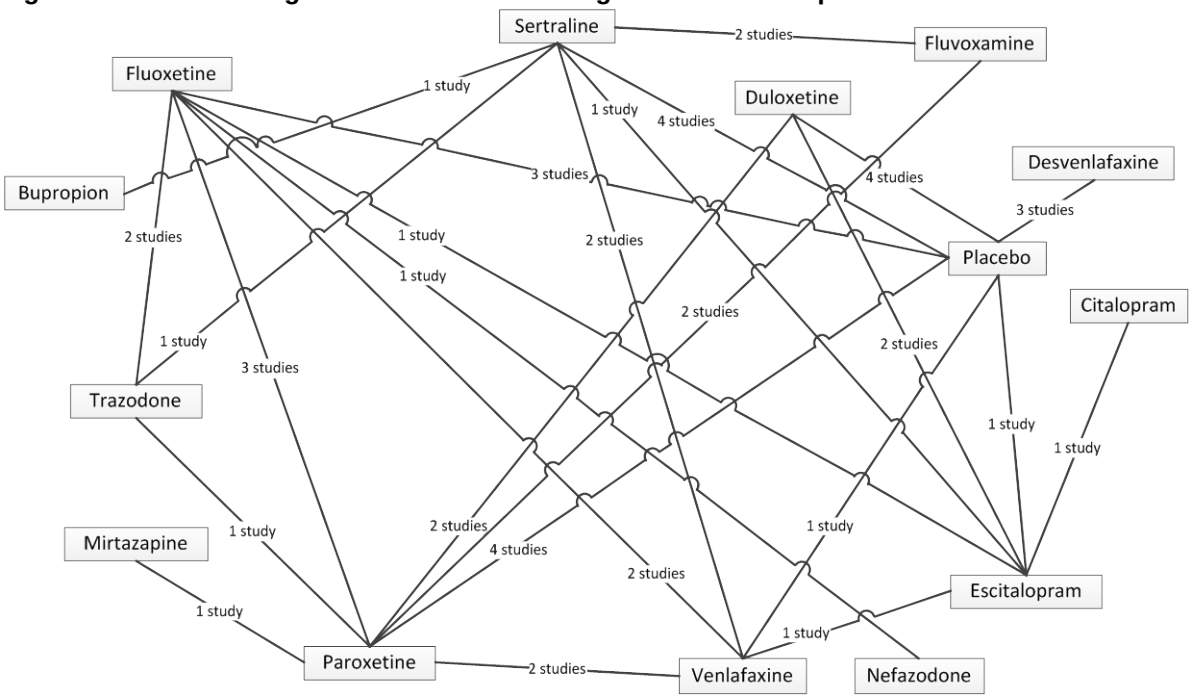


Figure C-8. Mean change in HAM-D for second-generation antidepressants: Star sub-network

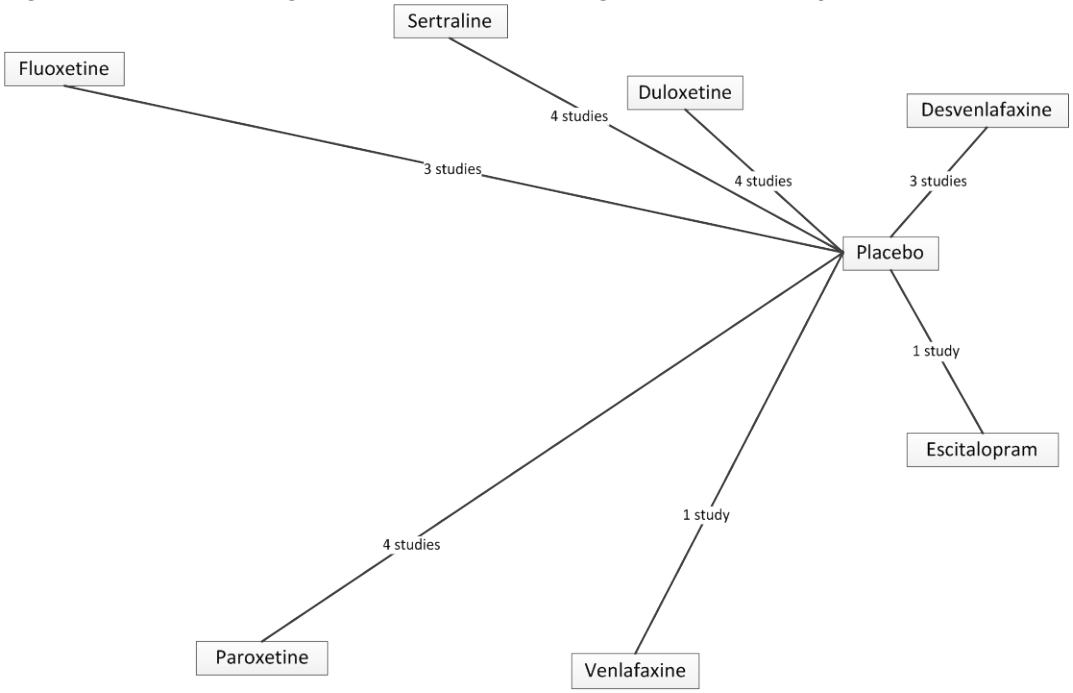


Figure C-9. Mean change in HAM-D for second-generation antidepressants: Loop sub-network

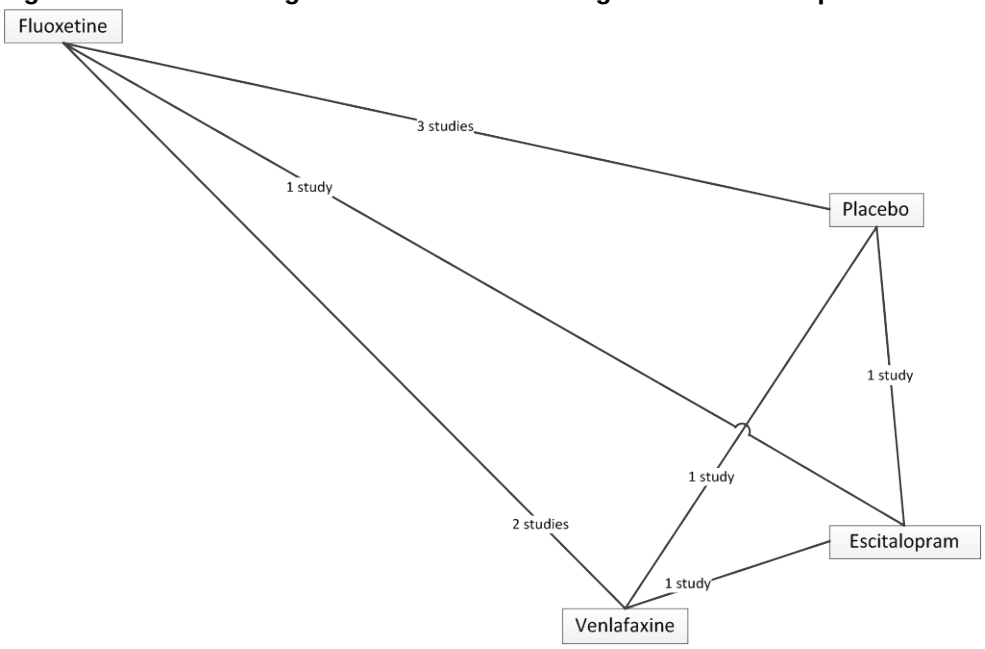


Figure C-10. Mean change in HAM-D for second-generation antidepressants: One closed loop sub-network

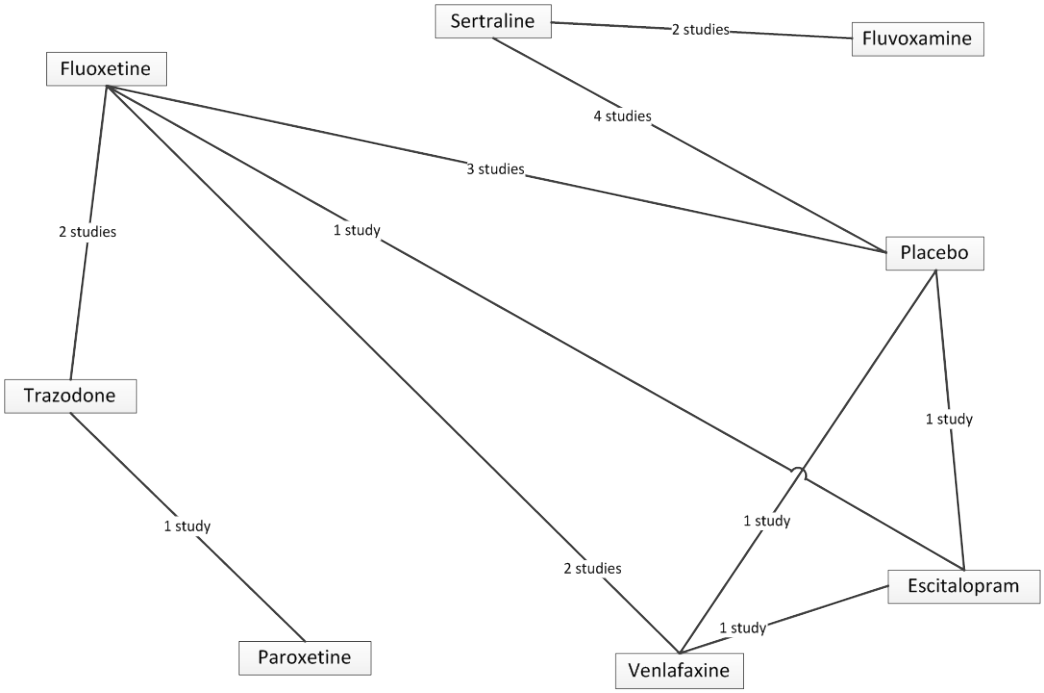


Figure C-11. Mean change in HAM-D for second-generation Antidepressants: Ladder sub-network

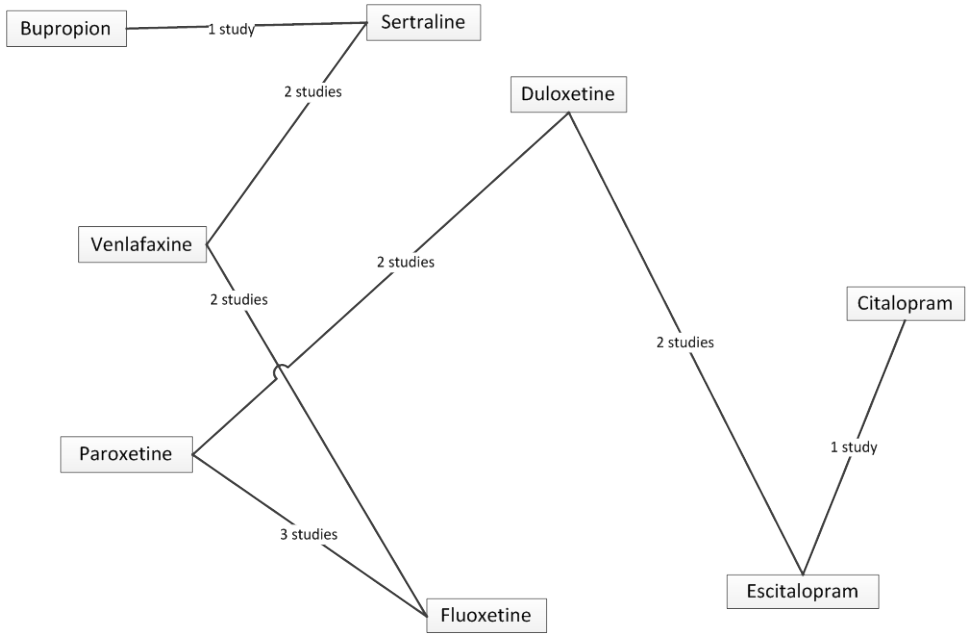


Figure C-12. ACR50 response for biologic DMARDs: Full network

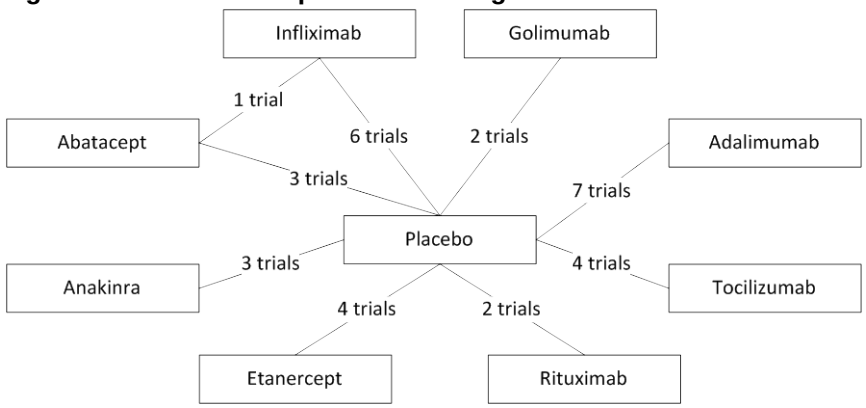


Figure C-13. ACR50 response for biologic DMARDs: Star sub-network

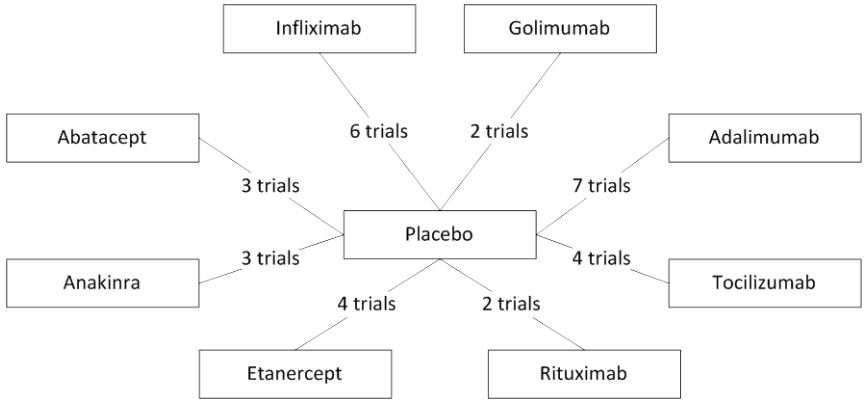


Figure C-14. ACR50 response for biologic DMARDs: Loop sub-network

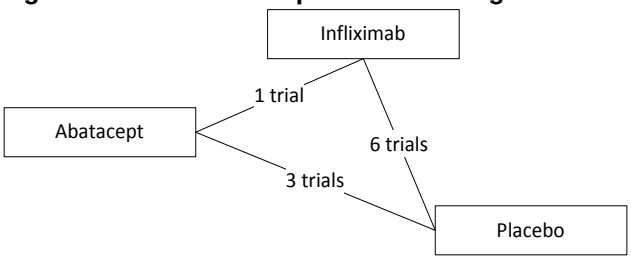


Figure C-15. ACR50 response for biologic DMARDs: One closed loop sub-network using Adalimumab

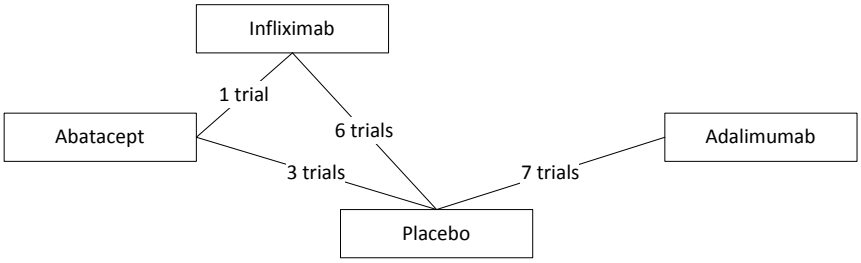


Figure C-16. ACR50 response for biologic DMARDs: One closed loop sub-network using Etanercept

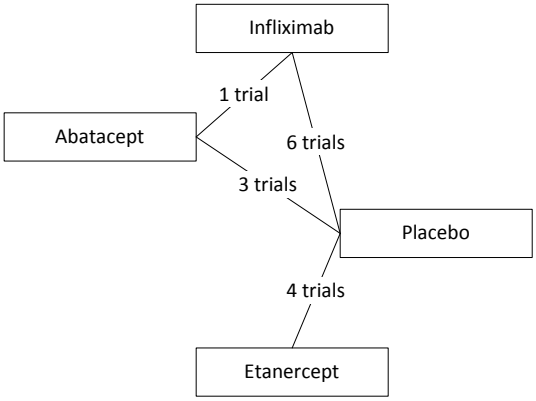


Figure C-17. ACR50 response for biologic DMARDs: Ladder sub-network using Adalimumab

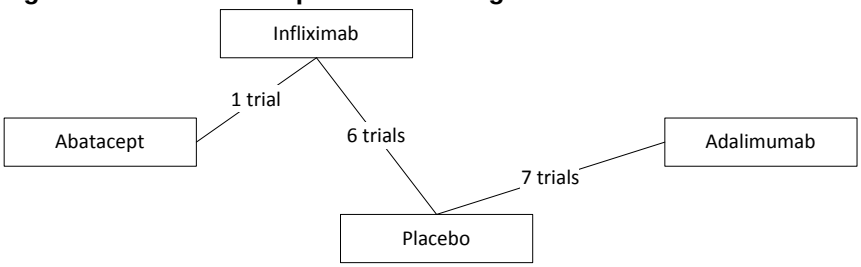


Figure C-18. ACR50 response for biologic DMARDs: Ladder sub-network using Etanercept

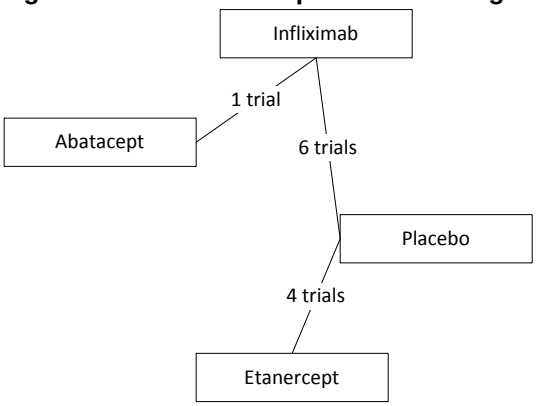
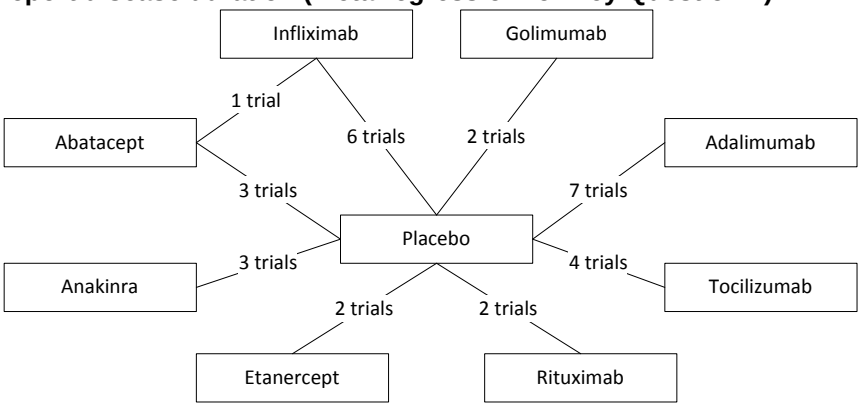


Figure C-19. ACR50 response for biologic DMARDs: Full network excluding studies that did not report disease duration (meta-regression for Key Question 2)



Appendix D. Results of Analyses

Table D-1. Comparison of response for second-generation antidepressants (full network), by method of analysis: Odds ratios (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
BUP vs. CIT	0.58 (0.29 to 1.39)	NA	NA	0.81 (0.48 to 1.35)
BUP vs. DES	0.88 (0.59 to 1.40)	0.78 (0.30 to 2.01)	0.77 (0.40 to 1.50)	0.93 (0.70 to 1.23)
BUP vs. DUL	0.99 (0.69 to 1.46)	0.85 (0.50 to 1.45)	0.87 (0.56 to 1.36)	1.00 (0.78 to 1.29)
BUP vs. ESC	0.73 (0.50 to 1.12)	1.04 (0.21 to 5.11)	1.04 (0.59 to 1.85)	0.76 (0.58 to 1.00)
BUP vs. FLUO	1.07 (0.78 to 1.52)	0.98 (0.56 to 1.71)	0.98 (0.62 to 1.56)	1.05 (0.83 to 1.32)
BUP vs. FLUV	0.62 (0.26 to 1.79)	0.59 (0.02 to 15.8)	0.58 (0.14 to 2.46)	1.03 (0.54 to 1.94)
BUP vs. MIR	0.75 (0.45 to 1.35)	0.71 (0.0 to 113.5)	IS	0.76 (0.51 to 1.12)
BUP vs. NEF	0.90 (0.55 to 1.58)	0.65 (0.19 to 2.18)	0.82 (0.35 to 1.91)	0.94 (0.64 to 1.39)
BUP vs. PAR	0.92 (0.64 to 1.38)	0.94 (0.45 to 1.99)	0.71 (0.38 to 1.33)	0.92 (0.70 to 1.19)
BUP vs. SER	0.90 (0.67 to 1.26)	0.80 (0.49 to 1.30)	0.80 (0.52 to 1.24)	0.91 (0.73 to 1.14)
BUP vs. TRA	1.09 (0.67 to 1.93)	0.54 (0.06 to 4.55)	IS	1.01 (0.69 to 1.50)
BUP vs. VEN	0.83 (0.58 to 1.22)	0.74 (0.33 to 1.62)	0.73 (0.42 to 1.28)	0.83 (0.64 to 1.07)
CIT vs. DES	1.32 (0.66 to 3.14)	NA	NA	1.15 (0.69 to 1.91)
CIT vs. DUL	1.48 (0.76 to 3.34)	NA	NA	1.24 (0.77 to 2.01)
CIT vs. ESC	1.11 (0.59 to 2.38)	NA	NA	0.94 (0.59 to 1.49)
CIT vs. FLUO	1.60 (0.82 to 3.63)	IS	IS	1.30 (0.80 to 2.11)
CIT vs. FLUV	1.03 (0.53 to 2.23)	NA	NA	1.27 (0.75 to 2.14)
CIT vs. MIR	1.11 (0.51 to 2.93)	NA	NA	0.94 (0.53 to 1.67)
CIT vs. NEF	1.34 (0.63 to 3.46)	NA	NA	1.17 (0.66 to 2.07)
CIT vs. PAR	1.38 (0.70 to 3.15)	NA	NA	1.13 (0.69 to 1.85)
CIT vs. SER	1.34 (0.69 to 3.08)	NA	NA	1.13 (0.69 to 1.84)
CIT vs. TRA	1.61 (0.74 to 4.35)	NA	NA	1.26 (0.69 to 2.27)
CIT vs. VEN	1.23 (0.62 to 2.84)	NA	NA	1.03 (0.63 to 1.69)
DES vs. DUL	1.10 (0.78 to 1.59)	1.10 (0.59 to 2.03)	1.14 (0.63 to 2.05)	1.08 (0.87 to 1.34)
DES vs. ESC	0.81 (0.55 to 1.24)	1.35 (0.22 to 8.18)	1.35 (0.67 to 2.71)	0.82 (0.63 to 1.06)
DES vs. FLUO	1.18 (0.84 to 1.73)	1.27 (0.64 to 2.54)	1.27 (0.69 to 2.34)	1.13 (0.89 to 1.42)
DES vs. FLUV	0.68 (0.29 to 1.99)	0.76 (0.06 to 10.1)	0.76 (0.17 to 3.37)	1.10 (0.59 to 2.08)
DES vs. MIR	0.82 (0.48 to 1.54)	NA	NA	0.82 (0.55 to 1.20)
DES vs. NEF	1.00 (0.61 to 1.75)	0.83 (0.21 to 3.27)	1.07 (0.42 to 2.71)	1.02 (0.70 to 1.48)
DES vs. PAR	1.02 (0.70 to 1.53)	1.23 (0.51 to 2.96)	0.93 (0.44 to 1.94)	0.99 (0.77 to 1.27)
DES vs. SER	1.00 (0.71 to 1.46)	1.05 (0.57 to 1.93)	1.04 (0.58 to 1.87)	0.98 (0.78 to 1.24)
DES vs. TRA	1.20 (0.69 to 2.26)	NA	NA	1.09 (0.72 to 1.65)
DES vs. VEN	0.91 (0.63 to 1.38)	0.95 (0.36 to 2.54)	0.96 (0.48 to 1.89)	0.89 (0.69 to 1.15)
DUL vs. ESC	0.74 (0.56 to 0.99)	1.23 (0.54 to 2.82)	1.19 (0.73 to 1.94)	0.76 (0.63 to 0.91)
DUL vs. FLUO	1.07 (0.82 to 1.43)	1.12 (0.75 to 1.68)	1.12 (0.79 to 1.59)	1.05 (0.87 to 1.26)
DUL vs. FLUV	0.62 (0.27 to 1.72)	0.69 (0.11 to 4.35)	0.67 (0.16 to 2.72)	1.02 (0.55 to 1.89)
DUL vs. MIR	0.74 (0.45 to 1.30)	0.74 (0.20 to 2.69)	IS	0.76 (0.53 to 1.08)
DUL vs. NEF	0.90 (0.58 to 1.51)	0.77 (0.33 to 1.79)	0.94 (0.43 to 2.07)	0.94 (0.66 to 1.35)
DUL vs. PAR	0.93 (0.70 to 1.24)	1.08 (0.66 to 1.77)	0.82 (0.47 to 1.41)	0.92 (0.76 to 1.10)
DUL vs. SER	0.90 (0.68 to 1.22)	0.92 (0.64 to 1.31)	0.92 (0.67 to 1.25)	0.91 (0.75 to 1.11)
DUL vs. TRA	1.08 (0.65 to 1.93)	1.25 (0.07 to 21.4)	IS	1.01 (0.69 to 1.49)
DUL vs. VEN	0.82 (0.61 to 1.15)	0.92 (0.54 to 1.59)	0.84 (0.53 to 1.34)	0.83 (0.67 to 1.03)
ESC vs. FLUO	1.44 (1.06 to 1.99)	0.94 (0.42 to 2.11)	0.94 (0.57 to 1.57)	1.38 (1.12 to 1.70)
ESC vs. FLUV	0.83 (0.37 to 2.26)	IS	0.56 (0.13 to 2.40)	1.35 (0.74 to 2.47)
ESC vs. MIR	1.00 (0.60 to 1.80)	IS	IS	1.00 (0.69 to 1.44)
ESC vs. NEF	1.21 (0.75 to 2.07)	IS	0.79 (0.33 to 1.88)	1.24 (0.86 to 1.80)
ESC vs. PAR	1.24 (0.90 to 1.74)	0.69 (0.15 to 3.13)	1.15 (0.71 to 1.88)	1.21 (0.98 to 149)
ESC vs. SER	1.21 (0.88 to 1.70)	0.77 (0.39 to 1.52)	0.77 (0.47 to 1.24)	1.20 (0.97 to 1.49)
ESC vs. TRA	1.45 (0.85 to 2.68)	IS	IS	1.33 (0.90 to 1.99)
ESC vs. VEN	1.11 (0.80 to 1.58)	0.71 (0.14 to 3.46)	0.71 (0.39 to 1.28)	1.09 (0.87 to 1.37)

Table D-1. Comparison of response for second-generation antidepressants (full network), by method of analysis: Odds ratios (95% credible interval or 95% confidence interval) (continued)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
FLUO vs. FLUV	0.57 (0.25 to 1.60)	0.60 (0.09 to 4.02)	0.60 (0.15 to 2.44)	0.98 (0.53 to 1.81)
FLUO vs. MIR	0.69 (0.43 to 1.17)	0.77 (0.34 to 1.71)	IS	0.72 (0.51 to 1.02)
FLUO vs. NEF	0.84 (0.55 to 1.36)	0.66 (0.28 to 1.56)	0.84 (0.38 to 1.86)	0.90 (0.64 to 1.21)
FLUO vs. PAR	0.86 (0.66 to 1.13)	0.73 (0.39 to 1.37)	0.73 (0.41 to 1.28)	0.88 (0.72 to 1.06)
FLUO vs. SER	0.84 (0.66 to 1.06)	0.82 (0.54 to 1.23)	0.82 (0.58 to 1.15)	0.87 (0.74 to 1.03)
FLUO vs. TRA	1.00 (0.62 to 1.75)	1.30 (0.22 to 7.66)	IS	0.97 (0.66 to 1.41)
FLUO vs. VEN	0.77 (0.61 to 0.97)	0.75 (0.42 to 1.35)	0.75 (0.46 to 1.22)	0.79 (0.67 to 0.94)
FLUV vs. MIR	0.97 (0.40 to 3.17)	NA	NA	0.74 (0.37 to 1.47)
FLUV vs. NEF	1.18 (0.50 to 3.72)	IS	1.40 (0.29 to 6.80)	0.92 (0.46 to 183)
FLUV vs. PAR	1.21 (0.54 to 3.5)	1.61 (0.13 to 20.0)	IS	0.89 (0.48 to 1.66)
FLUV vs. SER	1.18 (0.52 to 3.38)	1.37 (0.23 to 8.11)	IS	0.89 (0.48 to 1.65)
FLUV vs. TRA	1.42 (0.57 to 4.70)	NA	NA	0.99 (0.49 to 1.99)
FLUV vs. VEN	1.08 (0.48 to 3.12)	1.26 (0.05 to 35.0)	1.26 (0.30 to 5.32)	0.81 (0.43 to 1.51)
MIR vs. NEF	1.14 (0.63 to 2.29)	IS	IS	1.25 (0.78 to 1.99)
MIR vs. PAR	1.18 (0.74 to 1.97)	1.25 (0.43 to 3.64)	IS	1.21 (0.86 to 1.70)
MIR vs. SER	1.14 (0.71 to 1.98)	0.89 (0.17 to 4.61)	IS	1.20 (0.84 to 1.72)
MIR vs. TRA	1.40 (0.84 to 2.54)	IS	IS	1.34 (0.89 to 2.01)
MIR vs. VEN	1.04 (0.64 to 1.85)	1.03 (0.33 to 3.20)	IS	1.10 (0.76 to 1.58)
NEF vs. PAR	0.98 (0.62 to 1.63)	1.46 (0.47 to 4.51)	0.87 (0.35 to 2.15)	0.97 (0.68 to 1.40)
NEF vs. SER	0.96 (0.63 to 1.54)	1.23 (0.56 to 2.67)	0.98 (0.44 to 2.14)	0.97 (0.69 to 1.36)
NEF vs. TRA	1.15 (0.63 to 2.34)	IS	IS	1.07 (0.66 to 1.74)
NEF vs. VEN	0.87 (0.55 to 1.45)	0.84 (0.29 to 2.41)	0.90 (0.38 to 2.11)	0.88 (0.61 to 1.26)
PAR vs. SER	0.96 (0.72 to 1.31)	1.14 (0.71 to 1.84)	1.12 (0.65 to 1.93)	0.99 (0.81 to 1.22)
PAR vs. TRA	1.16 (0.71 to 2.04)	IS	IS	1.11 (0.76 to 1.62)
PAR vs. VEN	0.88 (0.65 to 1.23)	0.78 (0.36 to 1.70)	1.03 (0.54 to 1.96)	0.91 (0.72 to 1.14)
SER vs. TRA	1.19 (0.74 to 2.06)	1.03 (0.13 to 8.31)	IS	1.11 (0.76 to 1.62)
SER vs. VEN	0.91 (0.70 to 1.21)	0.92 (0.55 to 1.54)	0.92 (0.58 to 1.45)	0.91 (0.75 to 1.11)
TRA vs. VEN	0.71 (0.43 to 1.30)	1.51 (0.23 to 9.95)	IS	0.82 (0.55 to 1.21)

Abbreviations: BUP = bupropion; CIT = citalopram; DES = desvenlafaxine; DUL = duloxetine; ESC = escitalopram; FLUO = fluoxetine; FLUV = fluvoxamine; IS = insufficient studies; MIR = mirtazapine; NA = not applicable; NEF = nefazodone; PAR = paroxetine; SER = sertraline; TRA = trazodone; VEN = velafaxine

Table D-2. Comparison of mean change in ham-d for second-generation antidepressants (full network), by method of analysis: Weighted mean difference (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method
BUP vs. CIT	0.52 (-3.50 to 4.55)	NA	NA
BUP vs. DES	1.13 (-2.10 to 4.37)	NA	NA
BUP vs. DUL	0.65 (-2.3 to 3.60)	NA	NA
BUP vs. ESC	1.42 (-1.50 to 4.43)	IS	0.00 (-2.79 to 2.79)
BUP vs. FLUO	0.54 (-2.40 to 3.43)	NA	NA
BUP vs. FLUV	2.24 (-0.96 to 5.47)	IS	0.55 (-2.47 to 3.57)
BUP vs. MIR	2.48 (-1.20 to 6.24)	NA	NA
BUP vs. NEF	-0.25 (-4.40 to 3.79)	NA	NA
BUP vs. PAR	0.55 (-2.30 to 3.48)	NA	NA
BUP vs. SER	0.79 (-1.80 to 3.39)	Direct	Direct
BUP vs. TRA	1.44 (-1.70 to 4.68)	IS	0.60 (-3.01 to 4.21)
BUP vs. VEN	1.62 (-1.30 to 4.52)	IS	-0.74 (-4.21 to 2.73)
CIT vs. DES	0.6 (-2.80 to 3.95)	NA	NA
CIT vs. DUL	0.12 (-2.90 to 3.09)	-0.95 (-32.88 to 30.98)	0.85 (-2.49 to 4.19)
CIT vs. ESC	0.9 (-1.80 to 3.61)	Direct	Direct
CIT vs. FLUO	0.01 (-3.10 to 3.03)	IS	-0.20 (-2.88 to 2.48)
CIT vs. FLUV	1.71 (-1.70 to 5.14)	NA	NA
CIT vs. MIR	1.96 (-1.90 to 5.81)	NA	NA
CIT vs. NEF	-0.78 (-5.10 to 3.38)	NA	NA
CIT vs. PAR	0.03 (-3.00 to 3.07)	NA	NA
CIT vs. SER	0.27 (-2.80 to 3.30)	IS	0.10 (-2.76 to 2.96)
CIT vs. TRA	0.92 (-2.50 to 4.31)	NA	NA
CIT vs. VEN	1.09 (-2.00 to 4.14)	IS	-1.10 (-4.30 to 2.10)
DES vs. DUL	-0.48 (-2.40 to 1.40)	-0.65 (-2.92 to 1.61)	-0.68 (-2.25 to 0.90)
DES vs. ESC	0.29 (-1.70 to 2.32)	-1.23 (-5.06 to 2.61)	-1.23 (-2.97 to 0.51)
DES vs. FLUO	-0.59 (-2.60 to 1.28)	0.72 (-4.27 to 5.71)	-0.27 (-5.19 to 4.64)
DES vs. FLUV	1.11 (-1.30 to 3.55)	NA	NA
DES vs. MIR	1.35 (-1.60 to 4.33)	NA	NA
DES vs. NEF	-1.40 (-4.90 to 2.02)	NA	NA
DES vs. PAR	-0.58 (-2.40 to 1.31)	-0.67 (-3.33 to 2.00)	-0.72 (-2.65 to 1.21)
DES vs. SER	-0.33 (-2.20 to 1.53)	-0.41 (-2.68 to 1.86)	-0.47 (-2.15 to 1.21)
DES vs. TRA	0.32 (-2.10 to 2.71)	NA	NA
DES vs. VEN	0.49 (-1.50 to 2.43)	1.92 (-3.05 to 6.89)	1.92 (-0.34 to 4.18)
DUL vs. ESC	0.77 (-0.49 to 2.10)	-0.55 (-4.68 to 3.58)	-0.55 (-2.34 to 1.24)
DUL vs. FLUO	-0.11 (-1.50 to 1.19)	1.15 (-3.45 to 5.76)	0.41 (-4.53 to 5.34)
DUL vs. FLUV	1.59 (-0.42 to 3.64)	1.27 (-5.47 to 8.00)	1.40 (-0.43 to 3.22)
DUL vs. MIR	1.83 (-0.77 to 4.50)	2.23 (-8.66 to 13.12)	-2.23 (-3.91 to -0.55)
DUL vs. NEF	-0.90 (-4.10 to 2.21)	NA	NA
DUL vs. PAR	-0.10 (-1.30 to 1.17)	-0.04 (-2.59 to 2.51)	-0.04 (-2.02 to 1.94)
DUL vs. SER	0.15 (-1.30 to 1.53)	0.20 (-2.00 to 2.39)	0.21 (-1.53 to 1.95)
DUL vs. TRA	0.80 (-1.20 to 2.79)	-0.10 (-13.31 to 13.11)	-0.70 (2.74 to 1.34)
DUL vs. VEN	0.97 (-0.48 to 2.41)	2.60 (-2.15 to 7.35)	2.60 (0.29 to 4.91)
ESC vs. FLUO	-0.88 (-2.40 to 0.48)	0.91 (-49.15 to 50.98)	0.96 (-4.03 to 5.95)
ESC vs. FLUV	0.82 (-1.30 to 2.93)	IS	0.55 (-2.59 to 3.69)
ESC vs. MIR	1.06 (-1.70 to 3.77)	NA	NA
ESC vs. NEF	-1.70 (-5.00 to 1.47)	IS	-1.90 (-4.77 to 0.97)
ESC vs. PAR	-0.87 (-2.30 to 0.54)	0.26 (-5.51 to 6.03)	0.51 (-1.60 to 2.62)
ESC vs. SER	-0.63 (-2.10 to 0.78)	0.77 (-3.44 to 4.97)	0.76 (-1.13 to 2.65)
ESC vs. TRA	0.02 (-2.00 to 2.05)	0.78 (-17.89 to 19.45)	-2.99 (-5.85 to -0.13)
ESC vs. VEN	0.20 (-1.30 to 1.63)	IS	3.15 (0.73 to 5.57)
FLUO vs. FLUV	1.70 (-0.22 to 3.75)	IS	IS
FLUO vs. MIR	1.94 (-0.53 to 4.61)	1.72 (-4.68 to 8.11)	-2.12 (-4.30 to 0.06)

Table D-2. Comparison of mean change in ham-d for second-generation antidepressants (full network), by method of analysis: Weighted mean difference (95% credible interval or 95% confidence interval) (continued)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method
FLUO vs. NEF	-0.79 (-3.70 to 2.07)	Direct	Direct
FLUO vs. PAR	0.01 (-1.00 to 1.22)	-1.08 (-7.51 to 5.36)	-0.45 (-5.51 to 4.61)
FLUO vs. SER	0.26 (-1.00 to 1.65)	-0.92 (-5.82 to 3.97)	-0.20 (-5.17 to 4.77)
FLUO vs. TRA	0.90 (-0.81 to 2.72)	NA	NA
FLUO vs. VEN	1.08 (-0.14 to 2.41)	2.24 (-48.69 to 53.16)	2.19 (-3.00 to 7.39)
FLUV vs. MIR	0.24 (-2.70 to 3.15)	NA	NA
FLUV vs. NEF	-2.50 (-6.10 to 0.96)	NA	NA
FLUV vs. PAR	-1.70 (-3.40 to 0.06)	Direct	Direct
FLUV vs. SER	-1.40 (-3.30 to 0.40)	Direct	Direct
FLUV vs. TRA	-0.80 (-3.20 to 1.59)	IS	-2.10 (-4.56 to 0.37)
FLUV vs. VEN	-0.62 (-2.60 to 1.35)	3.53 (-20.92 to 27.98)	-1.21 (-3.33 to 0.92)
MIR vs. NEF	-2.70 (-6.70 to 1.02)	NA	NA
MIR vs. PAR	-1.90 (-4.20 to 0.39)	Direct	Direct
MIR vs. SER	-1.70 (-4.40 to 0.92)	NA	NA
MIR vs. TRA	-1.00 (-3.90 to 1.84)	IS	1.53 (-0.83 to 3.89)
MIR vs. VEN	-0.86 (-3.50 to 1.71)	IS	2.42 (0.42 to 4.42)
NEF vs. PAR	0.80 (-2.20 to 3.96)	2.01 (-20.25 to 24.27)	0.61 (-2.14 to 3.36)
NEF vs. SER	1.05 (-2.10 to 4.24)	NA	NA
NEF vs. TRA	1.70 (-1.60 to 5.15)	2.68 (-17.72 to 23.08)	-1.09 (-4.23 to 2.05)
NEF vs. VEN	1.87 (-1.20 to 5.04)	IS	2.50 (-0.64 to 5.64)
PAR vs. SER	0.24 (-1.10 to 1.47)	0.24 (-2.46 to 2.93)	0.25 (-1.82 to 2.32)
PAR vs. TRA	0.89 (-0.84 to 2.61)	0.67 (-6.88 to 8.22)	-1.70 (-4.45 to 1.05)
PAR vs. VEN	1.07 (-0.17 to 2.24)	2.66 (-4.47 to 9.79)	2.64 (0.08 to 5.20)
SER vs. TRA	0.65 (-1.20 to 2.55)	Direct	Direct
SER vs. VEN	0.82 (-0.44 to 2.09)	2.38 (-2.49 to 7.26)	2.39 (0.01 to 4.77)
TRA vs. VEN	0.18 (-1.80 to 2.08)	0.79 (-5.72 to 7.29)	-0.85 (-4.00 to 2.31)

Abbreviations: BUP = bupropion; CIT = citalopram; DES = desvenlafaxine; DUL = duloxetine; ESC = escitalopram; FLUO = fluoxetine; FLUV = fluvoxamine; IS = insufficient studies; MIR = mirtazapine; NA = not applicable; NEF = nefazodone; PAR = paroxetine; SER = sertraline; TRA = trazodone; VEN = velafaxine

Table D-3. Comparison of ACR50 response for biologic DMARDs (full network), by method of analysis: Odds ratios (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
ABA vs. ADA	0.65 (0.37 to 1.20)	0.67 (0.34 to 1.33)	0.64 (0.38 to 1.08)	0.83 (0.60 to 1.14)
ABA vs. ANA	1.26 (0.63 to 2.78)	1.32 (0.63 to 2.79)	1.32 (0.78 to 2.24)	1.82 (1.20 to 2.76)
ABA vs. ETA	0.19 (0.08 to 0.50)	0.21 (0.08 to 0.57)	0.21 (0.10 to 0.45)	0.41 (0.26 to 0.66)
ABA vs. GOL	0.54 (0.23 to 1.49)	0.59 (0.18 to 1.89)	0.57 (0.28 to 1.17)	0.76 (0.43 to 1.34)
ABA vs. INF	0.94 (0.58 to 1.65)	0.92 (0.49 to 2.29)	0.80 (0.47 to 1.38)	1.01 (0.75 to 1.35)
ABA vs. RIT	0.84 (0.39 to 2.08)	0.88 (0.34 to 1.53)	0.89 (0.49 to 1.60)	1.03 (0.64 to 1.67)
ABA vs. TOC	0.80 (0.47 to 1.57)	0.76 (0.38 to 2.44)	0.77 (0.48 to 1.25)	0.82 (0.60 to 1.13)
ADA vs. ANA	1.88 (1.01 to 3.98)	1.95 (0.82 to 4.61)	2.08 (1.12 to 3.83)	2.20 (1.49 to 3.26)
ADA vs. ETA	0.29 (0.13 to 0.73)	0.32 (0.12 to 0.84)	0.33 (0.15 to 0.75)	0.50 (0.32 to 0.78)
ADA vs. GOL	0.80 (0.36 to 2.14)	0.87 (0.27 to 2.82)	0.90 (0.41 to 1.97)	0.93 (0.53 to 1.59)
ADA vs. INF	1.39 (0.86 to 2.53)	1.42 (0.72 to 2.81)	1.26 (0.67 to 2.35)	1.22 (0.91 to 1.63)
ADA vs. RIT	1.25 (0.61 to 3.04)	1.30 (0.47 to 3.60)	1.39 (0.71 to 2.72)	1.25 (0.79 to 1.98)
ADA vs. TOC	1.19 (0.74 to 2.27)	1.20 (0.60 to 2.44)	1.21 (0.68 to 2.16)	1.00 (0.75 to 1.32)
ANA vs. ETA	0.14 (0.06 to 0.41)	0.16 (0.06 to 0.47)	0.16 (0.07 to 0.36)	0.23 (0.13 to 0.38)
ANA vs. GOL	0.39 (0.16 to 1.18)	0.45 (0.12 to 1.60)	0.43 (0.20 to 0.95)	0.42 (0.23 to 0.77)
ANA vs. INF	0.68 (0.36 to 1.47)	0.71 (0.31 to 1.63)	0.61 (0.32 to 1.13)	0.55 (0.37 to 0.82)
ANA vs. RIT	0.61 (0.27 to 1.70)	0.67 (0.22 to 1.97)	0.67 (0.34 to 1.31)	0.57 (0.33 to 0.96)
ANA vs. TOC	0.58 (0.32 to 1.30)	1.00 (0.45 to 2.21)	0.59 (0.33 to 1.04)	0.45 (0.31 to 0.67)
ETA vs. GOL	2.39 (0.92 to 8.64)	2.77 (0.72 to 10.62)	2.70 (1.05 to 6.96)	1.85 (0.98 to 3.51)
ETA vs. INF	4.17 (2.00 to 11.20)	4.34 (1.62 to 11.60)	3.78 (1.66 to 8.58)	2.45 (1.56 to 3.82)
ETA vs. RIT	3.76 (1.51 to 12.70)	4.15 (1.24 to 13.88)	4.18 (1.78 to 9.82)	2.50 (1.41 to 4.43)
ETA vs. TOC	3.59 (1.73 to 9.80)	3.63 (1.32 to 9.98)	3.65 (1.67 to 7.98)	2.00 (1.28 to 3.12)
GOL vs. INF	1.47 (0.68 to 4.05)	1.63 (0.51 to 5.20)	1.40 (0.63 to 3.08)	1.32 (0.77 to 2.28)
GOL vs. RIT	1.33 (0.53 to 4.54)	1.50 (0.24 to 4.93)	1.55 (0.68 to 3.53)	1.35 (0.70 to 2.60)
GOL vs. TOC	1.27 (0.59 to 3.57)	1.37 (0.38 to 1.97)	1.35 (0.64 to 2.86)	1.08 (0.63 to 1.86)
INF vs. RIT	0.87 (0.41 to 2.07)	0.93 (0.34 to 2.51)	1.11 (0.56 to 2.19)	1.02 (0.64 to 1.62)
INF vs. TOC	0.83 (0.51 to 1.52)	0.72 (0.38 to 1.38)	0.97 (0.54 to 1.74)	0.82 (0.62 to 1.08)
RIT vs. TOC	0.84 (0.42 to 2.12)	0.84 (0.26 to 2.77)	0.87 (0.46 to 1.65)	0.80 (0.51 to 1.26)

Abbreviations: ABA = abatacept; ADA = adalimumab; ANA = anakinra; ETA = etanercept; GOL = golimumab; INF = infliximab; RIT = rituximab; TOC = tocilizumab

Table D-4. Comparison of response for second-generation antidepressants for placebo star sub-network, by method of analysis: Odds ratios (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
BUP vs. CIT	NA	NA	NA	NA
BUP vs. DES	0.73 (0.43 to 1.35)	0.78 (0.30 to 2.01)	0.77 (0.40 to 1.50)	0.87 (0.64 to 1.18)
BUP vs. DUL	0.84 (0.51 to 1.49)	0.85 (0.50 to 1.45)	0.87 (0.56 to 1.36)	0.94 (0.70 to 1.26)
BUP vs. ESC	0.94 (0.43 to 2.56)	1.04 (0.21 to 5.11)	1.04 (0.59 to 1.85)	1.08 (0.72 to 1.63)
BUP vs. FLUO	0.94 (0.56 to 1.72)	0.98 (0.56 to 1.71)	0.98 (0.62 to 1.56)	0.99 (0.73 to 1.33)
BUP vs. FLUV	0.39 (0.11 to 2.63)	0.59 (0.02 to 15.83)	0.58 (0.14 to 2.46)	0.82 (0.26 to 2.54)
BUP vs. MIR	NA	NA	NA	NA
BUP vs. NEF	0.58 (0.27 to 1.49)	0.65 (0.19 to 2.18)	0.82 (0.35 to 1.91)	1.00 (0.51 to 1.98)
BUP vs. PAR	0.67 (0.37 to 1.39)	0.94 (0.45 to 1.99)	0.71 (0.38 to 1.33)	0.84 (0.57 to 1.23)
BUP vs. SER	0.78 (0.47 to 1.41)	0.80 (0.49 to 1.30)	0.80 (0.52 to 1.24)	0.96 (0.72 to 1.27)
BUP vs. TRA	NA	NA	NA	NA
BUP vs. VEN	0.69 (0.38 to 1.43)	0.74 (0.33 to 1.62)	0.73 (0.42 to 1.28)	0.80 (0.55 to 1.17)
CIT vs. DES	NA	NA	NA	NA
CIT vs. DUL	NA	NA	NA	NA
CIT vs. ESC	NA	NA	NA	NA
CIT vs. FLUO	NA	NA	NA	NA
CIT vs. FLUV	NA	NA	NA	NA
CIT vs. MIR	NA	NA	NA	NA
CIT vs. NEF	NA	NA	NA	NA
CIT vs. PAR	NA	NA	NA	NA
CIT vs. SER	NA	NA	NA	NA
CIT vs. TRA	NA	NA	NA	NA
CIT vs. VEN	NA	NA	NA	NA
DES vs. DUL	1.11 (0.73 to 1.79)	1.10 (0.59 to 2.03)	1.14 (0.63 to 2.05)	1.08 (0.86 to 1.34)
DES vs. ESC	1.25 (0.59 to 3.13)	1.35 (0.22 to 8.18)	1.35 (0.67 to 2.71)	1.24 (0.86 to 1.79)
DES vs. FLUO	1.25 (0.81 to 2.04)	1.27 (0.64 to 2.54)	1.27 (0.69 to 2.34)	1.14 (0.86 to 1.50)
DES vs. FLUV	0.51 (0.14 to 3.45)	0.76 (0.06 to 10.14)	0.76 (0.17 to 3.37)	0.94 (0.31 to 2.88)
DES vs. MIR	NA	NA	NA	NA
DES vs. NEF	0.77 (0.37 to 1.85)	0.83 (0.21 to 3.27)	1.07 (0.42 to 2.71)	1.15 (0.60 to 2.23)
DES vs. PAR	0.89 (0.52 to 1.69)	1.23 (0.51 to 2.96)	0.93 (0.44 to 1.94)	0.97 (0.69 to 1.35)
DES vs. SER	1.03 (0.66 to 1.69)	1.05 (0.57 to 1.93)	1.04 (0.58 to 1.87)	1.10 (0.85 to 1.43)
DES vs. TRA	NA	NA	NA	NA
DES vs. VEN	0.92 (0.53 to 1.75)	0.95 (0.36 to 2.54)	0.96 (0.48 to 1.89)	0.92 (0.65 to 1.30)
DUL vs. ESC	1.10 (0.53 to 2.70)	1.23 (0.54 to 2.82)	1.19 (0.73 to 1.94)	1.16 (0.85 to 1.58)
DUL vs. FLUO	1.10 (0.73 to 1.72)	1.12 (0.75 to 1.68)	1.12 (0.79 to 1.59)	1.06 (0.81 to 1.37)
DUL vs. FLUV	0.45 (0.12 to 3.03)	0.69 (0.11 to 4.35)	0.67 (0.16 to 2.72)	0.88 (0.29 to 2.67)
DUL vs. MIR	NA	NA	NA	NA
DUL vs. NEF	0.68 (0.33 to 1.61)	0.77 (0.33 to 1.79)	0.94 (0.43 to 2.07)	1.07 (0.56 to 2.06)
DUL vs. PAR	0.78 (0.47 to 1.45)	1.08 (0.66 to 1.77)	0.82 (0.47 to 1.41)	0.90 (0.67 to 1.21)
DUL vs. SER	0.91 (0.61 to 1.43)	0.92 (0.64 to 1.31)	0.92 (0.67 to 1.25)	1.02 (0.80 to 1.31)
DUL vs. TRA	NA	NA	NA	NA
DUL vs. VEN	0.81 (0.48 to 1.49)	0.92 (0.54 to 1.59)	0.84 (0.53 to 1.34)	0.85 (0.61 to 1.19)
ESC vs. FLUO	0.86 (0.41 to 2.17)	0.94 (0.42 to 2.11)	0.94 (0.57 to 1.57)	0.91 (0.62 to 1.34)
ESC vs. FLUV	0.35 (0.09 to 2.86)	IS	0.56 (0.13 to 2.40)	0.76 (0.24 to 2.39)
ESC vs. MIR	NA	NA	NA	NA
ESC vs. NEF	0.53 (0.21 to 1.79)	IS	0.79 (0.33 to 1.88)	0.93 (0.45 to 1.89)
ESC vs. PAR	0.61 (0.27 to 1.69)	0.69 (0.15 to 3.13)	1.15 (0.71 to 1.88)	0.78 (0.51 to 1.18)
ESC vs. SER	0.71 (0.34 to 1.79)	0.77 (0.39 to 1.52)	0.77 (0.47 to 1.24)	0.88 (0.61 to 1.29)
ESC vs. TRA	NA	NA	NA	NA
ESC vs. VEN	0.63 (0.29 to 1.72)	0.71 (0.14 to 3.46)	0.71 (0.39 to 1.28)	0.74 (0.48 to 1.14)
FLUO vs. FLUV	0.40 (0.11 to 2.63)	0.60 (0.09 to 4.02)	0.60 (0.15 to 2.44)	0.83 (0.27 to 2.55)
FLUO vs. MIR	NA	NA	NA	NA
FLUO vs. NEF	0.60 (0.29 to 1.45)	0.66 (0.28 to 1.56)	0.84 (0.38 to 1.86)	1.01 (0.52 to 1.98)

Table D-4. Comparison of response for second-generation antidepressants for placebo star sub-network, by method of analysis: Odds ratios (95% credible interval or 95% confidence interval) (continued)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
FLUO vs. PAR	0.69 (0.40 to 1.30)	0.73 (0.39 to 1.37)	0.73 (0.41 to 1.28)	0.85 (0.60 to 1.20)
FLUO vs. SER	1.23 (0.78 to 1.89)	0.82 (0.54 to 1.23)	0.82 (0.58 to 1.15)	0.97 (0.73 to 1.28)
FLUO vs. TRA	NA	NA	NA	NA
FLUO vs. VEN	0.72 (0.42 to 1.33)	0.75 (0.42 to 1.35)	0.75 (0.46 to 1.22)	0.81 (0.59 to 1.11)
FLUV vs. MIR	NA	NA	NA	NA
FLUV vs. NEF	0.79 (0.21 to 6.67)	IS	1.40 (0.29 to 6.80)	1.22 (0.34 to 4.35)
FLUV vs. PAR	0.93 (0.26 to 6.67)	1.61 (0.13 to 20.02)	IS	1.02 (0.33 to 3.20)
FLUV vs. SER	1.08 (0.31 to 7.69)	1.37 (0.23 to 8.11)	IS	1.17 (0.38 to 3.57)
FLUV vs. TRA	NA	NA	NA	NA
FLUV vs. VEN	0.96 (0.27 to 7.14)	1.26 (0.05 to 35.00)	1.26 (0.30 to 5.32)	0.98 (0.31 to 3.05)
MIR vs. NEF	NA	NA	NA	NA
MIR vs. PAR	NA	NA	NA	NA
MIR vs. SER	NA	NA	NA	NA
MIR vs. TRA	NA	NA	NA	NA
MIR vs. VEN	NA	NA	NA	NA
NEF vs. PAR	1.01 (0.47 to 2.70)	1.46 (0.47 to 4.51)	0.87 (0.35 to 2.15)	0.84 (0.42 to 1.68)
NEF vs. SER	1.16 (0.57 to 2.86)	1.23 (0.56 to 2.67)	0.98 (0.44 to 2.14)	0.95 (0.49 to 1.85)
NEF vs. TRA	NA	NA	NA	NA
NEF vs. VEN	1.04 (0.48 to 2.78)	0.84 (0.29 to 2.41)	0.90 (0.38 to 2.11)	0.80 (0.40 to 1.61)
PAR vs. SER	1.09 (0.63 to 2.04)	1.14 (0.71 to 1.84)	1.12 (0.65 to 1.93)	1.14 (0.81 to 1.61)
PAR vs. TRA	NA	NA	NA	NA
PAR vs. VEN	0.97 (0.52 to 2.04)	0.78 (0.36 to 1.70)	1.03 (0.54 to 1.96)	0.95 (0.63 to 1.43)
SER vs. TRA	NA	NA	NA	NA
SER vs. VEN	0.87 (0.51 to 1.61)	0.92 (0.55 to 1.54)	0.92 (0.58 to 1.45)	0.84 (0.59 to 1.18)
TRA vs. VEN	NA	NA	NA	NA

Abbreviations: BUP = bupropion; CIT = citalopram; DES = desvenlafaxine; DUL = duloxetine; ESC = escitalopram; FLUO = fluoxetine; FLUV = fluvoxamine; IS = insufficient studies; MIR = mirtazapine; NA = not applicable; NEF = nefazodone; PAR = paroxetine; SER = sertraline; TRA = trazodone; VEN = velafaxine

Table D-5. Comparison of response for second-generation antidepressants for loop sub-network, by method of analysis: Odds ratios (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
FLUO vs. PAR via SER	0.83 (0.64 to 1.11)	0.88 (0.19 to 4.02)	0.88 (0.44 to 1.76)	0.80 (0.63 to 1.02)
FLUO vs. PAR via placebo	NA/same	0.73 (0.39 to 1.37)	0.73 (0.41 to 1.28)	NA/same
FLUO vs. SER via PAR	0.76 (0.61 to 0.96)	0.69 (0.26 to 1.83)	0.69 (0.34 to 1.37)	0.78 (0.64 to 0.96)
FLUO vs. SER via placebo	NA/same	0.82 (0.54 to 1.23)	0.82 (0.58 to 1.15)	NA/same
PAR vs. SER via FLUO	0.90 (0.68 to 1.25)	0.71 (0.41 to 1.22)	0.71 (0.46 to 1.10)	0.97 (0.74 to 1.26)
PAR vs. SER via placebo	NA/same	1.14 (0.71 to 1.84)	1.12 (0.65 to 1.93)	NA/same

Abbreviations: FLUO = fluoxetine; NA = not applicable; PAR = paroxetine; SER = sertraline

Table D-6. Comparison of response for second-generation antidepressants for one closed loop sub-network, by method of analysis: Odds ratios (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
BUP vs. CIT	0.74 (0.30 to 2.27)	NA	NA	0.94 (0.49 to 1.80)
BUP vs. ESC	0.91 (0.46 to 2.13)	NA	NA	0.95 (0.55 to 1.64)
BUP vs. FLUO	0.99 (0.65 to 1.61)	IS	IS	1.07 (0.81 to 1.40)
BUP vs. FLUV	0.78 (0.28 to 3.13)	NA	NA	1.18 (0.53 to 2.65)
BUP vs. MIR	1.32 (0.62 to 3.85)	IS	IS	1.10 (0.51 to 2.35)
BUP vs. PAR	0.83 (0.51 to 1.43)	IS	IS	0.89 (0.64 to 1.23)
BUP vs. SER	0.77 (0.52 to 1.18)	Direct	Direct	0.87 (0.67 to 1.13)
BUP vs. TRA	2.04 (1.02 to 4.76)	Direct	Direct	1.75 (0.96 to 3.18)
CIT vs. ESC	1.11 (0.63 to 2.17)	Direct	Direct	1.01 (0.61 to 1.68)
CIT vs. FLUO	1.09 (0.50 to 2.94)	IS	IS	1.14 (0.62 to 2.09)
CIT vs. FLUV	1.05 (0.58 to 2.08)	NA	NA	1.26 (0.71 to 2.21)
CIT vs. MIR	1.37 (0.44 to 7.14)	NA	NA	1.17 (0.45 to 3.07)
CIT vs. PAR	0.91 (0.40 to 2.56)	NA	NA	0.95 (0.50 to 1.79)
CIT vs. SER	0.84 (0.37 to 2.33)	NA	NA	0.93 (0.50 to 1.73)
CIT vs. TRA	2.08 (0.70 to 10.00)	NA	NA	1.86 (0.80 to 4.35)
ESC vs. FLUO	0.98 (0.53 to 1.92)	Direct	Direct	1.12 (0.69 to 1.82)
ESC vs. FLUV	0.85 (0.39 to 2.27)	IS	IS	1.24 (0.60 to 2.56)
ESC vs. MIR	1.22 (0.42 to 5.56)	NA	NA	1.15 (0.47 to 2.83)
ESC vs. PAR	0.81 (0.42 to 1.67)	IS	IS	0.93 (0.55 to 1.58)
ESC vs. SER	0.75 (0.40 to 1.54)	IS	IS	0.91 (0.55 to 1.52)
ESC vs. TRA	1.89 (0.70 to 7.14)	NA	NA	1.84 (0.85 to 3.98)
FLUO vs. FLUV	0.78 (0.31 to 2.86)	NA	NA	1.11 (0.51 to 2.40)
FLUO vs. MIR	1.27 (0.52 to 4.35)	NA	NA	1.03 (0.47 to 2.23)
FLUO vs. PAR via SER	0.83 (0.65 to 1.10)	IS	IS	0.83 (0.65 to 1.06)
FLUO vs. PAR via placebo	NA/same	0.73 (0.39 to 1.37)	0.73 (0.41 to 1.28)	NA/same
FLUO vs. SER	0.77 (0.63 to 0.96)	0.82 (0.54 to 1.23)	0.82 (0.58 to 1.15)	0.81 (0.67 to 1.00)
FLUO vs. TRA	1.96 (0.87 to 5.88)	NA	NA	1.64 (0.88 to 3.05)
MIR vs. PAR	0.50 (0.19 to 1.61)	NA	NA	0.81 (0.37 to 1.80)
MIR vs. SER	0.47 (0.18 to 1.43)	NA	NA	0.79 (0.36 to 1.73)
MIR vs. TRA	1.45 (0.91 to 2.50)	Direct	Direct	1.59 (0.92 to 2.76)
PAR vs. SER via FLUO	0.91 (0.68 to 1.27)	IS	IS	0.98 (0.75 to 1.28)
PAR vs. SER via placebo	NA/same	1.14 (0.71 to 1.84)	1.12 (0.65 to 1.93	NA/same
PAR vs. TRA	2.33 (1.02 to 7.14)	NA	NA	1.97 (1.03 to 3.76)
SER vs. TRA	2.56 (1.16 to 7.14)	IS	IS	2.01 (1.08 to 3.75)

Abbreviations: BUP = bupropion; CIT = citalopram; ESC = escitalopram; FLUO = fluoxetine; FLUV = fluvoxamine; IS = insufficient studies; MIR = mirtazapine; NA = not applicable; PAR = paroxetine; SER = sertraline; TRA = trazodone

Table D-7. Comparison of response for second-generation antidepressants for ladder sub-network, by method of analysis: Odds ratios (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
BUP vs. CIT	0.75 (0.28 to 2.56)	NA	NA	1.36 (0.72 to 2.58)
BUP vs. DUL	1.20 (0.58 to 2.94)	NA	NA	1.43 (0.90 to 2.28)
BUP vs. ESC	0.94 (0.41 to 2.44)	NA	NA	1.24 (0.76 to 2.03)
BUP vs. FLUO	1.20 (0.68 to 2.44)	NA	NA	1.25 (0.83 to 1.90)
BUP vs. FLUV	0.79 (0.26 to 3.57)	NA	NA	1.82 (0.82 to 4.03)
BUP vs. MIR	1.30 (0.56 to 3.57)	IS	0.67 (0.26 to 1.69)	1.38 (0.65 to 2.91)
BUP vs. PAR	1.10 (0.56 to 2.38)	NA	NA	1.22 (0.79 to 1.90)
BUP vs. SER	0.95 (0.68 to 1.39)	Direct	Direct	1.08 (0.77 to 1.51)
BUP vs. TRA	2.00 (0.99 to 5.00)	Direct	Direct	2.14 (1.18 to 3.86)
BUP vs. VEN	0.79 (0.47 to 1.45)	0.82 (0.38 to 1.79)	0.82 (0.51 to 1.34)	0.98 (0.65 to 1.48)
CIT vs. DUL	1.39 (0.74 to 3.13)	1.52 (0.23 to 10.02)	1.44 (0.78 to 2.64)	1.05 (0.64 to 1.74)
CIT vs. ESC	1.11 (0.63 to 2.27)	Direct	Direct	0.91 (0.56 to 1.47)
CIT vs. FLUO	1.32 (0.63 to 3.57)	NA	NA	0.92 (0.53 to 1.59)
CIT vs. FLUV	1.05 (0.58 to 2.17)	Direct	Direct	1.33 (0.76 to 2.33)
CIT vs. MIR	1.27 (0.41 to 7.14)	NA	NA	1.01 (0.41 to 2.47)
CIT vs. PAR	1.23 (0.61 to 3.03)	NA	NA	0.90 (0.53 to 1.53)
CIT vs. SER	0.95 (0.41 to 3.23)	NA	NA	0.79 (0.44 to 1.41)
CIT vs. TRA	1.96 (0.68 to 10.00)	NA	NA	1.57 (0.72 to 3.43)
CIT vs. VEN	0.83 (0.37 to 2.38)	NA	NA	0.72 (0.41 to 1.25)
DUL vs. ESC	0.79 (0.57 to 1.06)	Direct	Direct	0.86 (0.69 to 1.08)
DUL vs. FLUO	0.94 (0.61 to 1.56)	0.96 (0.58 to 1.60)	0.97 (0.64 to 1.45)	0.87 (0.64 to 1.20)
DUL vs. FLUV	0.66 (0.28 to 1.96)	NA	NA	1.27 (0.63 to 2.56)
DUL vs. MIR	0.92 (0.33 to 3.85)	NA	NA	0.96 (0.44 to 2.10)
DUL vs. PAR	0.88 (0.65 to 1.25)	Direct	Direct	0.85 (0.67 to 1.08)
DUL vs. SER	0.69 (0.37 to 1.52)	NA	NA	0.75 (0.52 to 1.09)
DUL vs. TRA	1.43 (0.57 to 5.26)	NA	NA	1.49 (0.78 to 2.86)
DUL vs. VEN	0.60 (0.36 to 1.11)	NA	NA	0.68 (0.49 to 0.96)
ESC vs. FLUO	1.18 (0.71 to 2.22)	NA	NA	1.01 (0.71 to 1.45)
ESC vs. FLUV	0.85 (0.37 to 2.38)	IS	0.94 (0.42 to 2.11)	1.47 (0.73 to 2.93)
ESC vs. MIR	1.15 (0.41 to 5.00)	NA	NA	1.11 (0.50 to 2.48)
ESC vs. PAR	1.10 (0.72 to 1.85)	1.16 (0.57 to 2.36)	1.09 (0.77 to 1.56)	0.99 (0.72 to 1.35)
ESC vs. SER	0.86 (0.44 to 2.13)	NA	NA	0.87 (0.58 to 1.31)
ESC vs. TRA	1.79 (0.69 to 7.14)	NA	NA	1.73 (0.88 to 3.38)
ESC vs. VEN	0.75 (0.42 to 1.56)	NA	NA	0.79 (0.54 to 1.15)
FLUO vs. FLUV	0.66 (0.25 to 2.22)	NA	NA	1.45 (0.70 to 2.99)
FLUO vs. MIR	0.97 (0.38 to 3.57)	NA	NA	1.10 (0.51 to 2.35)
FLUO vs. PAR	0.91 (0.66 to 1.30)	Direct	Direct	0.98 (0.75 to 1.27)
FLUO vs. SER	0.73 (0.45 to 1.32)	0.90 (0.47 to 1.71)	0.95 (0.62 to 1.45)	0.86 (0.64 to 1.16)
FLUO vs. TRA	1.49 (0.65 to 5.00)	NA	NA	1.71 (0.91 to 3.18)
FLUO vs. VEN	0.64 (0.47 to 0.87)	Direct	Direct	0.78 (0.65 to 0.94)
FLUV vs. MIR	1.08 (0.33 to 7.69)	NA	NA	0.76 (0.28 to 2.09)
FLUV vs. PAR	1.04 (0.43 to 3.45)	NA	NA	0.67 (0.33 to 1.38)
FLUV vs. SER	0.81 (0.29 to 3.57)	NA	NA	0.59 (0.28 to 1.26)
FLUV vs. TRA	1.67 (0.53 to 11.11)	NA	NA	1.18 (0.47 to 2.94)
FLUV vs. VEN	0.71 (0.27 to 2.70)	NA	NA	0.54 (0.26 to 1.12)
MIR vs. PAR	0.66 (0.24 to 2.56)	NA	NA	0.89 (0.41 to 1.92)
MIR vs. SER	0.57 (0.25 to 1.89)	NA	NA	0.78 (0.37 to 1.66)
MIR vs. TRA	1.45 (0.92 to 2.50)	Direct	Direct	1.55 (0.90 to 2.68)
MIR vs. VEN	0.48 (0.20 to 1.69)	NA	NA	0.71 (0.33 to 1.52)
PAR vs. SER	0.78 (0.45 to 1.59)	NA	NA	0.88 (0.63 to 1.24)
PAR vs. TRA	1.61 (0.67 to 5.88)	NA	NA	1.75 (0.92 to 3.31)
PAR vs. VEN	0.68 (0.44 to 1.11)	0.83 (0.50 to 1.38)	0.87 (0.60 to 1.25)	0.80 (0.60 to 1.08)

Table D-7. Comparison of response for second-generation antidepressants for ladder sub-network, by method of analysis: Odds ratios (95% credible interval or 95% confidence interval) (continued)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
SER vs. TRA	2.04 (0.93 to 5.56)	IS	1.03 (0.46 to2.30)	1.99 (1.08 to 3.67)
SER vs. VEN	0.83 (0.55 to 1.30)	Direct	Direct	0.91 (0.68 to 1.21)
TRA vs. VEN	0.33 (0.14 to 0.99)	NA	NA	0.46 (0.25 to 0.86)

Abbreviations: BUP = bupropion; CIT = citalopram; DUL = duloxetine; ESC = escitalopram; FLUO = fluoxetine; FLUV = fluvoxamine; IS = insufficient studies; MIR = mirtazapine; NA = not applicable; PAR = paroxetine; SER = sertraline; TRA = trazodone; VEN = velafaxine

Table D-8. Comparison of mean change in HAM-D for second-generation antidepressants for placebo star sub-network, by method of analysis: Weighted mean difference (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method
DES vs. DUL	-0.65 (-2.80 to 1.38)	-0.65 (-2.92 to 1.61)	-0.68 (-2.25 to 0.90)
DES vs. ESC	-1.20 (-4.10 to 1.75)	-1.23 (-5.06 to 2.61)	-1.23 (-2.97 to 0.51)
DES vs. FLUO	0.12 (-2.30 to 2.50)	0.72 (-4.27 to 5.71)	-0.27 (-5.19 to 4.64)
DES vs. PAR	-0.88 (-3.20 to 1.24)	-0.67 (-3.33 to 2.00)	-0.72 (-2.65 to 1.21)
DES vs. SER	-0.43 (-2.60 to 1.70)	-0.41 (-2.68 to 1.86)	-0.47 (-2.15 to 1.21)
DES vs. VEN	1.94 (-1.30 to 5.21)	1.92 (-3.05 to 6.89)	1.92 (-0.34 to 4.18)
DUL vs. ESC	-0.53 (-3.30 to 2.20)	-0.55 (-4.68 to 3.58)	-0.55 (-2.34 to 1.24)
DUL vs. FLUO	0.77 (-1.40 to 2.92)	1.15 (-3.45 to 5.76)	0.41 (-4.53 to 5.34)
DUL vs. PAR	-0.24 (-2.20 to 1.63)	-0.04 (-2.59 to 2.51)	-0.04 (-2.02 to 1.94)
DUL vs. SER	0.21 (-1.70 to 2.08)	0.20 (-2.00 to 2.39)	0.21 (-1.53 to 1.95)
DUL vs. VEN	2.59 (-0.50 to 5.69)	2.60 (-2.15 to 7.35)	2.60 (0.29 to 4.91)
ESC vs. FLUO	1.30 (-1.70 to 4.28)	0.91 (-49.15 to 50.98)	0.96 (-4.03 to 5.95)
ESC vs. PAR	0.30 (-2.60 to 3.04)	0.49 (-5.90 to 6.88)	0.51 (-1.60 to 2.62)
ESC vs. SER	0.75 (-2.10 to 3.54)	0.77 (-3.44 to 4.97)	0.76 (-1.13 to 2.65)
ESC vs. VEN	3.12 (-0.61 to 6.85)	IS	3.15 (0.73 to 5.57)
FLUO vs. PAR	-1.00 (-3.40 to 1.23)	-1.08 (-7.51 to 5.36)	-0.45 (-5.51 to 4.61)
FLUO vs. SER	-0.55 (-2.80 to 1.72)	-0.92 (-5.82 to 3.97)	-0.20 (-5.17 to 4.77)
FLUO vs. VEN	1.82 (-1.50 to 5.20)	2.24 (-48.69 to 53.16)	2.19 (-3.00 to 7.39)
PAR vs. SER	0.45 (-1.50 to 2.54)	0.24 (-2.46 to 2.93)	0.25 (-1.82 to 2.32)
PAR vs. VEN	2.83 (-0.29 to 6.11)	2.66 (-4.47 to 9.79)	2.64 (0.08 to 5.20)
SER vs. VEN	2.37 (-0.75 to 5.56)	2.38 (-2.49 to 7.26)	2.39 (0.01 to 4.77)

Abbreviations: DES = desvenlafaxine; DUL = duloxetine; ESC = escitalopram; FLUO = fluoxetine; IS = insufficient studies; PAR = paroxetine; SER = sertraline; VEN = venlafaxine

Table D-9. Comparison of mean change in HAM-D for second-generation antidepressants for loop sub-network, by method of analysis: Weighted mean difference (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Via
ESC vs. FLUO	-1.20 (-4.00 to 1.41)	0.91 (-49.15 to 50.98)	0.96 (-4.03 to 5.95)	Placebo
ESC vs. FLUO	-1.20 (-4.00 to 1.41)	IS	-4.74 (-8.11 to -1.36)	VEN
ESC vs. VEN	0.65 (-2.90 to 3.94)	IS	0.60 (-2.27 to 3.47)	FLUO
FLUO vs. VEN	1.87 (-0.90 to 4.77)	IS	-0.90 (-4.00 to 2.20)	ESC

Abbreviations: ESC = escitalopram; FLUO = fluoxetine; IS = insufficient studies; VEN = venlafaxine

Table D-10. Comparison of mean change in HAM-D for second-generation antidepressants for one closed loop sub-network, by method of analysis: Weighted mean difference (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Via
ESC vs. FLUO	-1.20 (-4.60 to 1.80)	0.91 (-49.15 to 50.98)	0.96 (-4.03 to 5.95)	placebo
ESC vs. FLUO	-1.20 (-4.00 to 1.41)	IS	-4.74 (-8.11 to -1.36)	VEN
ESC vs. FLUV	0.74 (-4.20 to 5.66)	NA	NA	NA
ESC vs. PAR	-1.10 (-5.30 to 2.93)	NA	NA	NA
ESC vs. SER	-1.10 (-4.70 to 2.18)	0.77 (-3.44 to 4.97)	0.76 (-1.13 to 2.65)	placebo
ESC vs. TRA	-0.26 (-4.20 to 3.69)	NA	NA	NA
ESC vs. VEN	0.67 (-2.30 to 3.45)	IS	0.60 (-2.27 to 3.47)	FLUO
FLUO vs. FLUV	1.92 (-2.70 to 6.71)	NA	NA	NA
FLUO vs. PAR	0.04 (-2.90 to 3.24)	IS	-1.49 (-4.38 to 1.40)	TRA
FLUO vs. SER	0.04 (-3.00 to 3.06)	-0.92 (-5.82 to 3.97)	-0.20 (-5.17 to 4.77)	placebo
FLUO vs. TRA	0.92 (-1.90 to 3.96)	NA	NA	NA
FLUO vs. VEN	1.85 (-0.50 to 4.31)	IS	-0.90 (-4.00 to 2.20)	ESC
FLUO vs. VEN	1.85 (-0.50 to 4.31)	2.24 (-48.69 to 53.16)	2.19 (-3.00 to 7.39)	placebo
FLUV vs. PAR	-1.90 (-7.40 to 3.69)	NA	NA	NA
FLUV vs. SER	-1.90 (-5.60 to 1.63)	NA	NA	NA
FLUV vs. TRA	-1.00 (-6.50 to 4.47)	NA	NA	NA
FLUV vs. VEN	-0.07 (-5.10 to 4.71)	NA	NA	NA
PAR vs. SER	0.00 (-4.40 to 4.13)	NA	NA	NA
PAR vs. TRA	0.88 (-2.30 to 4.06)	NA	NA	NA
PAR vs. VEN	1.81 (-2.20 to 5.56)	NA	NA	NA
SER vs. TRA	0.88 (-3.20 to 5.21)	NA	NA	NA
SER vs. VEN	1.81 (-1.50 to 5.21)	NA	NA	NA
TRA vs. VEN	0.93 (-2.90 to 4.61)	IS	3.59 (0.45 to 6.73)	FLUO

Abbreviations: ESC = escitalopram; FLUO = fluoxetine; FLUV = fluvoxamine; IS = insufficient studies; NA = not applicable; PAR = paroxetine; SER = sertraline; TRA = trazodone; VEN = venlafaxine

Table D-11. Comparison of mean change in HAM-D for second-generation antidepressants for ladder sub-network, by method of analysis: Weighted mean difference (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Via
BUP vs. CIT	-2.20 (-9.40 to 5.43)	NA	NA	NA
BUP vs. DUL	-1.70 (-7.60 to 4.43)	NA	NA	NA
BUP vs. ESC	-1.30 (-7.60 to 5.39)	NA	NA	NA
BUP vs. FLUO	-1.70 (-7.00 to 3.25)	NA	NA	NA
BUP vs. PAR	-2.00 (-7.50 to 3.68)	NA	NA	NA
BUP vs. SER	0.83 (-2.70 to 4.30)	NA	NA	NA
BUP vs. VEN	1.04 (-3.50 to 5.14)	IS	-0.74 (-4.21 to 2.73)	SER
CIT vs. DUL	0.54 (-3.90 to 4.71)	IS	0.85 (-2.49 to 4.19)	ESC
CIT vs. ESC	0.91 (-2.70 to 4.51)	NA	NA	NA
CIT vs. FLUO	0.52 (-5.30 to 5.59)	NA	NA	NA
CIT vs. PAR	0.24 (-4.80 to 5.09)	NA	NA	NA
CIT vs. SER	3.02 (-3.80 to 9.39)	NA	NA	NA
CIT vs. VEN	3.23 (-3.20 to 8.95)	NA	NA	NA
DUL vs. ESC	0.36 (-1.90 to 2.82)	NA	NA	NA
DUL vs. FLUO	-0.02 (-3.70 to 2.90)	NA	NA	NA
DUL vs. PAR	-0.30 (-2.60 to 2.06)	NA	NA	NA
DUL vs. SER	2.48 (-2.60 to 7.27)	NA	NA	NA
DUL vs. VEN	2.69 (-1.70 to 6.62)	NA	NA	NA
ESC vs. FLUO	-0.38 (-4.90 to 3.26)	NA	NA	NA
ESC vs. PAR	-0.67 (-4.10 to 2.62)	0.26 (-5.51 to 6.03)	-0.35 (-3.17 to 2.47)	DUL
ESC vs. SER	2.11 (-3.60 to 7.35)	NA	NA	NA
ESC vs. VEN	2.32 (-2.90 to 6.80)	NA	NA	NA
FLUO vs. PAR	-0.28 (-2.20 to 2.36)	NA	NA	NA
FLUO vs. SER	2.50 (-1.10 to 6.53)	1.66 (-31.65 to 34.98)	2.68 (-1.01 to 6.36)	VEN
FLUO vs. VEN	2.71 (0.03 to 5.54)	NA	NA	NA
PAR vs. SER	2.78 (-1.70 to 7.05)	NA	NA	NA
PAR vs. VEN	2.99 (-0.74 to 6.23)	-4.13 (-12.41 to 4.15)	1.89 (-0.86 to 4.64)	FLUO
SER vs. VEN	0.21 (-2.50 to 2.65)	NA	NA	NA

Abbreviations: BUP = bupropion; CIT = citalopram; DUL = duloxetine; ESC = escitalopram; FLUO = fluoxetine; IS = insufficient studies; NA = not applicable; PAR = paroxetine; SER = sertraline; VEN = venlafaxine

Table D-12. Comparison of ACR50 response for biologic DMARDs for placebo star sub-network, by method of analysis: Odds ratios (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
ABA vs. ADA	0.59 (0.33 to 1.14)	0.67 (0.34 to 1.33)	0.64 (0.38 to 1.08)	0.83 (0.60 to 1.14)
ABA vs. ANA	1.15 (0.57 to 2.63)	1.32 (0.63 to 2.79)	1.32 (0.78 to 2.24)	1.82 (1.20 to 2.76)
ABA vs. ETA	0.17 (0.08 to 0.49)	0.21 (0.08 to 0.57)	0.21 (0.10 to 0.45)	0.41 (0.26 to 0.66)
ABA vs. GOL	0.49 (0.21 to 1.39)	0.59 (0.18 to 1.89)	0.57 (0.28 to 1.17)	0.76 (0.43 to 1.34)
ABA vs. INF	0.85 (0.48 to 1.72)	0.92 (0.49 to 2.29)	0.80 (0.47 to 1.38)	1.01 (0.75 to 1.35)
ABA vs. RIT	0.78 (0.36 to 2.00)	0.88 (0.34 to 1.53)	0.89 (0.49 to 1.60)	1.03 (0.64 to 1.67)
ABA vs. TOC	0.74 (0.42 to 1.49)	0.76 (0.38 to 2.44)	0.77 (0.48 to 1.25)	0.82 (0.60 to 1.13)
ADA vs. ANA	1.89 (0.99 to 4.00)	1.95 (0.82 to 4.61)	2.08 (1.12 to 3.83)	2.20 (1.49 to 3.26)
ADA vs. ETA	0.28 (0.13 to 0.77)	0.32 (0.12 to 0.84)	0.33 (0.15 to 0.75)	0.50 (0.32 to 0.78)
ADA vs. GOL	0.81 (0.36 to 2.13)	0.87 (0.27 to 2.82)	0.90 (0.41 to 1.97)	0.93 (0.53 to 1.59)
ADA vs. INF	1.39 (0.83 to 2.56)	1.42 (0.72 to 2.81)	1.26 (0.67 to 2.35)	1.22 (0.91 to 1.63)
ADA vs. RIT	1.27 (0.62 to 3.13)	1.30 (0.47 to 3.60)	1.39 (0.71 to 2.72)	1.25 (0.79 to 1.98)
ADA vs. TOC	1.20 (0.74 to 2.33)	1.20 (0.60 to 2.44)	1.21 (0.68 to 2.16)	1.00 (0.75 to 1.32)
ANA vs. ETA	0.14 (0.06 to 0.41)	0.16 (0.06 to 0.47)	0.16 (0.07 to 0.36)	0.23 (0.13 to 0.38)
ANA vs. GOL	0.39 (0.16 to 1.18)	0.45 (0.12 to 1.60)	0.43 (0.20 to 0.95)	0.42 (0.23 to 0.77)
ANA vs. INF	0.68 (0.36 to 1.49)	0.71 (0.31 to 1.63)	0.61 (0.32 to 1.13)	0.55 (0.37 to 0.82)
ANA vs. RIT	0.62 (0.27 to 1.72)	0.67 (0.22 to 1.97)	0.67 (0.34 to 1.31)	0.57 (0.33 to 0.96)
ANA vs. TOC	0.58 (0.32 to 1.32)	1.00 (0.45 to 2.21)	0.59 (0.33 to 1.04)	0.45 (0.31 to 0.67)
ETA vs. GOL	2.38 (0.90 to 8.33)	2.77 (0.72 to 10.62)	2.70 (1.05 to 6.96)	1.85 (0.98 to 3.51)
ETA vs. INF	4.17 (1.96 to 11.11)	4.34 (1.62 to 11.60)	3.78 (1.66 to 8.58)	2.45 (1.56 to 3.82)
ETA vs. RIT	3.85 (1.52 to 12.50)	4.15 (1.24 to 13.88)	4.18 (1.78 to 9.82)	2.50 (1.14 to 4.43)
ETA vs. TOC	3.57 (1.69 to 10.00)	3.63 (1.32 to 9.98)	3.65 (1.67 to 7.98)	2.00 (1.28 to 3.12)
GOL vs. INF	1.45 (0.68 to 4.17)	1.63 (0.51 to 5.20)	1.40 (0.63 to 3.08)	1.32 (0.77 to 2.28)
GOL vs. RIT	1.33 (0.53 to 4.76)	1.50 (0.24 to 4.93)	1.55 (0.68 to 3.53)	1.35 (0.70 to 2.60)
GOL vs. TOC	1.27 (0.59 to 3.70)	1.37 (0.38 to 1.97)	1.35 (0.64 to 2.86)	1.08 (0.63 to 1.86)
INF vs. RIT	0.88 (0.41 to 2.13)	0.93 (0.34 to 2.51)	1.11 (0.56 to 2.19)	1.02 (0.64 to 1.62)
INF vs. TOC	0.83 (0.50 to 1.56)	0.72 (0.38 to 1.38)	0.97 (0.54 to 1.74)	0.82 (0.62 to 1.08)
RIT vs. TOC	0.83 (0.41 to 2.13)	0.84 (0.26 to 2.77)	0.87 (0.46 to 1.65)	0.80 (0.51 to 1.26)

Abbreviations: ABA = abatacept; ADA = adalimumab; ANA = anakinra; ETA = etanercept; GOL = golimumab; INF = infliximab; RIT = rituximab; TOC = tocilizumab

Table D-13. Comparison of ACR50 response for biologic DMARDs for loop sub-network, by method of analysis: Odds ratios (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
ABA vs. INF	0.92 (0.50 to 1.79)	0.90 (0.33 to 2.44)	0.77 (0.48 to 1.25)	1.04 (0.78 to 1.38)

Abbreviations: ABA = abatacept; INF = infliximab

Table D-14. Comparison of ACR50 response for biologic DMARDs for one closed loop sub-network using adalimumab, by method of analysis: Odds ratios (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
ABA vs. ADA	0.62 (0.31 to 1.33)	0.67 (0.34 to 1.33)	0.64 (0.38 to 1.08)	0.83 (0.60 to 1.15)
ABA vs. INF	0.93 (0.51 to 1.89)	0.92 (0.49 to 2.29)	0.80 (0.47 to 1.38)	1.01 (0.75 to 1.36)
ADA vs. INF	1.39 (0.78 to 2.94)	1.42 (0.72 to 2.81)	1.26 (0.67 to 2.35)	1.21 (0.91 to 1.63)

Abbreviations: ABA = abatacept; ADA = adalimumab; INF = infliximab

Table D-15. Comparison of ACR50 response for biologic DMARDs for one closed loop sub-network using etanercept, by method of analysis: Odds ratios (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
ABA vs. ETA	0.19 (0.08 to 0.52)	0.21 (0.08 to 0.57)	0.21 (0.10 to 0.45)	0.41 (0.25 to 0.67)
ABA vs. INF	0.94 (0.56 to 1.67)	0.88 (0.34 to 2.29)	0.80 (0.47 to 1.38)	1.00 (0.75 to 1.35)
ETA vs. INF	4.17 (1.96 to 11.11)	4.34 (1.62 to 11.60)	3.78 (1.66 to 8.58)	2.46 (1.54 to 3.93)

Abbreviations: ABA = abatacept; ETA = etanercept; INF = infliximab

Table D-16. Comparison of ACR50 response for biologic DMARDs for ladder sub-network, by method of analysis: Odds ratios (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
ABA vs. ETA	0.19 (0.05 to 1.05)	NA	NA	0.79 (0.50 to 1.26)
ABA vs. INF	0.98 (0.38 to 3.45)	Direct	Direct	0.98 (0.65 to 1.47)
ETA vs. INF	4.00 (1.72 to 12.50)	4.34 (1.62 to 11.60)	3.78 (1.66 to 8.58)	1.24 (0.91 to 1.67)

Abbreviations: ABA = abatacept; ETA = etanercept; INF = infliximab

Appendix E. Comparison of Precision

For each set of results in Appendix D, the tables below provide a comparison of the precision (determined by width of the 95% credible interval or confidence interval) for the Bayesian MTC meta-analysis and the frequentist methods used. The darkest shading indicates the most precise result and the lightest indicates the least precise. The drug-drug comparisons for each of the tables below are limited to those where at least one of the frequentist methods could produce a result.

Table E-1. Comparison of response for second-generation antidepressants (full network), by method of analysis: Odds ratios (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
BUP vs. CIT	0.58 (0.29 to 1.39)	NA	NA	0.81 (0.48 to 1.35)
BUP vs. DES	0.88 (0.59 to 1.40)	0.78 (0.30 to 2.01)	0.77 (0.40 to 1.50)	0.93 (0.70 to 1.23)
BUP vs. DUL	0.99 (0.69 to 1.46)	0.85 (0.50 to 1.45)	0.87 (0.56 to 1.36)	1.00 (0.78 to 1.29)
BUP vs. ESC	0.73 (0.50 to 1.12)	1.04 (0.21 to 5.11)	1.04 (0.59 to 1.85)	0.76 (0.58 to 1.00)
BUP vs. FLUO	1.07 (0.78 to 1.52)	0.98 (0.56 to 1.71)	0.98 (0.62 to 1.56)	1.05 (0.83 to 1.32)
BUP vs. FLUV	0.62 (0.26 to 1.79)	0.59 (0.02 to 15.8)	0.58 (0.14 to 2.46)	1.03 (0.54 to 1.94)
BUP vs. MIR	0.75 (0.45 to 1.35)	0.71 (0.0 to 113.5)	IS	0.76 (0.51 to 1.12)
BUP vs. NEF	0.90 (0.55 to 1.58)	0.65 (0.19 to 2.18)	0.82 (0.35 to 1.91)	0.94 (0.64 to 1.39)
BUP vs. PAR	0.92 (0.64 to 1.38)	0.94 (0.45 to 1.99)	0.71 (0.38 to 1.33)	0.92 (0.70 to 1.19)
BUP vs. SER	0.90 (0.67 to 1.26)	0.80 (0.49 to 1.30)	0.80 (0.52 to 1.24)	0.91 (0.73 to 1.14)
BUP vs. TRA	1.09 (0.67 to 1.93)	0.54 (0.06 to 4.55)	IS	1.01 (0.69 to 1.50)
BUP vs. VEN	0.83 (0.58 to 1.22)	0.74 (0.33 to 1.62)	0.73 (0.42 to 1.28)	0.83 (0.64 to 1.07)
CIT vs. DES	1.32 (0.66 to 3.14)	NA	NA	1.15 (0.69 to 1.91)
CIT vs. DUL	1.48 (0.76 to 3.34)	NA	NA	1.24 (0.77 to 2.01)
CIT vs. ESC	1.11 (0.59 to 2.38)	NA	NA	0.94 (0.59 to 1.49)
CIT vs. FLUO	1.60 (0.82 to 3.63)	IS	IS	1.30 (0.80 to 2.11)
CIT vs. FLUV	1.03 (0.53 to 2.23)	NA	NA	1.27 (0.75 to 2.14)
CIT vs. MIR	1.11 (0.51 to 2.93)	NA	NA	0.94 (0.53, to1.67)
CIT vs. NEF	1.34 (0.63 to 3.46)	NA	NA	1.17 (0.66 to 2.07)
CIT vs. PAR	1.38 (0.70 to 3.15)	NA	NA	1.13 (0.69 to1.85)
CIT vs. SER	1.34 (0.69 to 3.08)	NA	NA	1.13 (0.69 to 1.84)
CIT vs. TRA	1.61 (0.74 to 4.35)	NA	NA	1.26 (0.69 to 2.27)
CIT vs. VEN	1.23 (0.62 to 2.84)	NA	NA	1.03 (0.63 to 1.69)
DES vs. DUL	1.10 (0.78 to 1.59)	1.10 (0.59 to 2.03)	1.14 (0.63 to 2.05)	1.08 (0.87 to 1.34)
DES vs. ESC	0.81 (0.55 to 1.24)	1.35 (0.22 to 8.18)	1.35 (0.67 to 2.71)	0.82 (0.63 to 1.06)
DES vs. FLUO	1.18 (0.84 to 1.73)	1.27 (0.64 to 2.54)	1.27 (0.69 to 2.34)	1.13 (0.89 to 1.42)
DES vs. FLUV	0.68 (0.29 to 1.99)	0.76 (0.06 to 10.1)	0.76 (0.17 to 3.37)	1.10 (0.59 to 2.08)
DES vs. MIR	0.82 (0.48 to 1.54)	NA	NA	0.82 (0.55 to 1.20)
DES vs. NEF	1.00 (0.61 to 1.75)	0.83 (0.21 to 3.27)	1.07 (0.42 to 2.71)	1.02 (0.70 to 1.48)
DES vs. PAR	1.02 (0.70 to 1.53)	1.23 (0.51 to 2.96)	0.93 (0.44 to 1.94)	0.99 (0.77 to 1.27)
DES vs. SER	1.00 (0.71 to 1.46)	1.05 (0.57 to 1.93)	1.04 (0.58 to 1.87)	0.98 (0.78 to 1.24)
DES vs. TRA	1.20 (0.69 to 2.26)	NA	NA	1.09 (0.72 to 1.65)
DES vs. VEN	0.91 (0.63 to 1.38)	0.95 (0.36 to 2.54)	0.96 (0.48 to 1.89)	0.89 (0.69 to 1.15)
DUL vs. ESC	0.74 (0.56 to 0.99)	1.23 (0.54 to 2.82)	1.19 (0.73 to 1.94)	0.76 (0.63 to 0.91)
DUL vs. FLUO	1.07 (0.82 to 1.43)	1.12 (0.75 to 1.68)	1.12 (0.79 to 1.59)	1.05 (0.87 to 1.26)
DUL vs. FLUV	0.62 (0.27 to 1.72)	0.69 (0.11 to 4.35)	0.67 (0.16 to 2.72)	1.02 (0.55 to 1.89)
DUL vs. MIR	0.74 (0.45 to 1.30)	0.74 (0.20 to 2.69)	IS	0.76 (0.53 to 1.08)
DUL vs. NEF	0.90 (0.58 to 1.51)	0.77 (0.33 to 1.79)	0.94 (0.43 to 2.07)	0.94 (0.66 to 1.35)
DUL vs. PAR	0.93 (0.70 to 1.24)	1.08 (0.66 to 1.77)	0.82 (0.47 to 1.41)	0.92 (0.76 to 1.10)
DUL vs. SER	0.90 (0.68 to 1.22)	0.92 (0.64 to 1.31)	0.92 (0.67 to 1.25)	0.91 (0.75 to 1.11)
DUL vs. TRA	1.08 (0.65 to 1.93)	1.25 (0.07 to 21.4)	IS	1.01 (0.69 to 1.49)
DUL vs. VEN	0.82 (0.61 to 1.15)	0.92 (0.54 to 1.59)	0.84 (0.53 to 1.34)	0.83 (0.67 to 1.03)
ESC vs. FLUO	1.44 (1.06 to 1.99)	0.94 (0.42 to 2.11)	0.94 (0.57 to 1.57)	1.38 (1.12 to 1.70)
ESC vs. FLUV	0.83 (0.37 to 2.26)	IS	0.56 (0.13 to 2.40)	1.35 (0.74 to 2.47)

Table E-1. Comparison of response for second-generation antidepressants (full network), by method of analysis: Odds ratios (95% credible interval or 95% confidence interval) (continued)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
ESC vs. MIR	1.00 (0.60 to 1.80)	IS	IS	1.00 (0.69 to 1.44)
ESC vs. NEF	1.21 (0.75 to 2.07)	IS	0.79 (0.33 to 1.88)	1.24 (0.86 to 1.80)
ESC vs. PAR	1.24 (0.90 to 1.74)	0.69 (0.15 to 3.13)	1.15 (0.71 to 1.88)	1.21 (0.98 to 1.49)
ESC vs. SER	1.21 (0.88 to 1.70)	0.77 (0.39 to 1.52)	0.77 (0.47 to 1.24)	1.20 (0.97 to 1.49)
ESC vs. TRA	1.45 (0.85 to 2.68)	IS	IS	1.33 (0.90 to 1.99)
ESC vs. VEN	1.11 (0.80 to 1.58)	0.71 (0.14 to 3.46)	0.71 (0.39 to 1.28)	1.09 (0.87 to 1.37)
FLUO vs. FLUV	0.57 (0.25 to 1.60)	0.60 (0.09 to 4.02)	0.60 (0.15 to 2.44)	0.98 (0.53 to 1.81)
FLUO vs. MIR	0.69 (0.43 to 1.17)	0.77 (0.34 to 1.71)	IS	0.72 (0.51 to 1.02)
FLUO vs. NEF	0.84 (0.55 to 1.36)	0.66 (0.28 to 1.56)	0.84 (0.38 to 1.86)	0.90 (0.64 to 1.21)
FLUO vs. PAR	0.86 (0.66 to 1.13)	0.73 (0.39 to 1.37)	0.73 (0.41 to 1.28)	0.88 (0.72 to 1.06)
FLUO vs. SER	0.84 (0.66 to 1.06)	0.82 (0.54 to 1.23)	0.82 (0.58 to 1.15)	0.87 (0.74 to 1.03)
FLUO vs. TRA	1.00 (0.62 to 1.75)	1.30 (0.22 to 7.66)	IS	0.97 (0.66 to 1.41)
FLUO vs. VEN	0.77 (0.61 to 0.97)	0.75 (0.42 to 1.35)	0.75 (0.46 to 1.22)	0.79 (0.67 to 0.94)
FLUV vs. MIR	0.97 (0.40 to 3.17)	NA	NA	0.74 (0.37 to 1.47)
FLUV vs. NEF	1.18 (0.50 to 3.72)	IS	1.40 (0.29 to 6.80)	0.92 (0.46 to 1.83)
FLUV vs. PAR	1.21 (0.54 to 3.5)	1.61 (0.13 to 20.0)	IS	0.89 (0.48 to 1.66)
FLUV vs. SER	1.18 (0.52 to 3.38)	1.37 (0.23 to 8.11)	IS	0.89 (0.48 to 1.65)
FLUV vs. TRA	1.42 (0.57 to 4.70)	NA	NA	0.99 (0.49 to 1.99)
FLUV vs. VEN	1.08 (0.48 to 3.12)	1.26 (0.05 to 35.0)	1.26 (0.30 to 5.32)	0.81 (0.43 to 1.51)
MIR vs. NEF	1.14 (0.63 to 2.29)	IS	IS	1.25 (0.78 to 1.99)
MIR vs. PAR	1.18 (0.74 to 1.97)	1.25 (0.43 to 3.64)	IS	1.21 (0.86 to 1.70)
MIR vs. SER	1.14 (0.71 to 1.98)	0.89 (0.17 to 4.61)	IS	1.20 (0.84 to 1.72)
MIR vs. TRA	1.40 (0.84 to 2.54)	IS	IS	1.34 (0.89 to 2.01)
MIR vs. VEN	1.04 (0.64 to 1.85)	1.03 (0.33 to 3.20)	IS	1.10 (0.76 to 1.58)
NEF vs. PAR	0.98 (0.62 to 1.63)	1.46 (0.47 to 4.51)	0.87 (0.35 to 2.15)	0.97 (0.68 to 1.40)
NEF vs. SER	0.96 (0.63 to 1.54)	1.23 (0.56 to 2.67)	0.98 (0.44 to 2.14)	0.97 (0.69 to 1.36)
NEF vs. TRA	1.15 (0.63 to 2.34)	IS	IS	1.07 (0.66 to 1.74)
NEF vs. VEN	0.87 (0.55 to 1.45)	0.84 (0.29 to 2.41)	0.90 (0.38 to 2.11)	0.88 (0.61 to 1.26)
PAR vs. SER	0.96 (0.72 to 1.31)	1.14 (0.71 to 1.84)	1.12 (0.65 to 1.93)	0.99 (0.81 to 1.22)
PAR vs. TRA	1.16 (0.71 to 2.04)	IS	IS	1.11 (0.76 to 1.62)
PAR vs. VEN	0.88 (0.65 to 1.23)	0.78 (0.36 to 1.70)	1.03 (0.54 to 1.96)	0.91 (0.72 to 1.14)
SER vs. TRA	1.19 (0.74 to 2.06)	1.03 (0.13 to 8.31)	IS	1.11 (0.76 to 1.62)
SER vs. VEN	0.91 (0.70 to 1.21)	0.92 (0.55 to 1.54)	0.92 (0.58 to 1.45)	0.91 (0.75 to 1.11)
TRA vs. VEN	0.71 (0.43 to 1.30)	1.51 (0.23 to 9.95)	IS	0.82 (0.55 to 1.21)

Abbreviations: BUP = bupropion; CIT = citalopram; DES = desvenlafaxine; DUL = duloxetine; ESC = escitalopram; FLUO = fluoxetine; FLUV = fluvoxamine; IS = insufficient studies; MIR = mirtazapine; NA = not applicable; NEF = nefazodone; PAR = paroxetine; SER = sertraline; TRA = trazodone; VEN = velafaxine

Table E-2. Comparison of mean change in HAM-D for second-generation antidepressants (full network), by method of analysis: Weighted mean difference (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method
BUP vs. ESC	1.42 (-1.50 to 4.43)	IS	0.00 (-2.79 to 2.79)
BUP vs. FLUV	2.24 (-0.96 to 5.47)	IS	0.55 (-2.47 to 3.57)
BUP vs. TRA	1.44 (-1.70 to 4.68)	IS	0.60 (-3.01 to 4.21)
BUP vs. VEN	1.62 (-1.30 to 4.52)	IS	-0.74 (-4.21 to 2.73)
CIT vs. DUL	0.12 (-2.90 to 3.09)	-0.95 (-32.88 to 30.98)	0.85 (-2.49 to 4.19)
CIT vs. FLUO	0.01 (-3.10 to 3.03)	IS	-0.20 (-2.88 to 2.48)
CIT vs. SER	0.27 (-2.80 to 3.30)	IS	0.10 (-2.76 to 2.96)
CIT vs. VEN	1.09 (-2.00 to 4.14)	IS	-1.10 (-4.30 to 2.10)
DES vs. DUL	-0.48 (-2.40 to 1.40)	-0.65 (-2.92 to 1.61)	-0.68 (-2.25 to 0.90)
DES vs. ESC	0.29 (-1.70 to 2.32)	-1.23 (-5.06 to 2.61)	-1.23 (-2.97 to 0.51)
DES vs. FLUO	-0.59 (-2.60 to 1.28)	0.72 (-4.27 to 5.71)	-0.27 (-5.19 to 4.64)
DES vs. PAR	-0.58 (-2.40 to 1.31)	-0.67 (-3.33 to 2.00)	-0.72 (-2.65 to 1.21)
DES vs. SER	-0.33 (-2.20 to 1.53)	-0.41 (-2.68 to 1.86)	-0.47 (-2.15 to 1.21)
DES vs. VEN	0.49 (-1.50 to 2.43)	1.92 (-3.05 to 6.89)	1.92 (-0.34 to 4.18)
DUL vs. ESC	0.77 (-0.49 to 2.10)	-0.55 (-4.68 to 3.58)	-0.55 (-2.34 to 1.24)
DUL vs. FLUO	-0.11 (-1.50 to 1.19)	1.15 (-3.45 to 5.76)	0.41 (-4.53 to 5.34)
DUL vs. FLUV	1.59 (-0.42 to 3.64)	1.27 (-5.47 to 8.00)	1.40 (-0.43 to 3.22)
DUL vs. MIR	1.83 (-0.77 to 4.50)	2.23 (-8.66 to 13.12)	-2.23 (-3.91 to -0.55)
DUL vs. PAR	-0.10 (-1.30 to 1.17)	-0.04 (-2.59 to 2.51)	-0.04 (-2.02 to 1.94)
DUL vs. SER	0.15 (-1.30 to 1.53)	0.20 (-2.00 to 2.39)	0.21 (-1.53 to 1.95)
DUL vs. TRA	0.80 (-1.20 to 2.79)	-0.10 (-13.31 to 13.11)	-0.70 (-2.74 to 1.34)
DUL vs. VEN	0.97 (-0.48 to 2.41)	2.60 (-2.15 to 7.35)	2.60 (0.29 to 4.91)
ESC vs. FLUO	-0.88 (-2.40 to 0.48)	0.91 (-49.15 to 50.98)	0.96 (-4.03 to 5.95)
ESC vs. FLUV	0.82 (-1.30 to 2.93)	IS	0.55 (-2.59 to 3.69)
ESC vs. NEF	-1.70 (-5.00 to 1.47)	IS	-1.90 (-4.77 to 0.97)
ESC vs. PAR	-0.87 (-2.30 to 0.54)	0.26 (-5.51 to 6.03)	0.51 (-1.60 to 2.62)
ESC vs. SER	-0.63 (-2.10 to 0.78)	0.77 (-3.44 to 4.97)	0.76 (-1.13 to 2.65)
ESC vs. TRA	0.02 (-2.00 to 2.05)	0.78 (-17.89 to 19.45)	-2.99 (-5.85 to -0.13)
ESC vs. VEN	0.20 (-1.30 to 1.63)	IS	3.15 (0.73 to 5.57)
FLUO vs. MIR	1.94 (-0.53 to 4.61)	1.72 (-4.68 to 8.11)	-2.12 (-4.30 to 0.06)
FLUO vs. PAR	0.01 (-1.00 to 1.22)	-1.08 (-7.51 to 5.36)	-0.45 (-5.51 to 4.61)
FLUO vs. SER	0.26 (-1.00 to 1.65)	-0.92 (-5.82 to 3.97)	-0.20 (-5.17 to 4.77)
FLUO vs. VEN	1.08 (-0.14 to 2.41)	2.24 (-48.69 to 53.16)	2.19 (-3.00 to 7.39)
FLUV vs. TRA	-0.80 (-3.20 to 1.59)	IS	-2.10 (-4.56 to 0.37)
FLUV vs. VEN	-0.62 (-2.60 to 1.35)	3.53 (-20.92 to 27.98)	-1.21 (-3.33 to 0.92)
MIR vs. TRA	-1.00 (-3.90 to 1.84)	IS	1.53 (-0.83 to 3.89)
MIR vs. VEN	-0.86 (-3.50 to 1.71)	IS	2.42 (0.42 to 4.42)
NEF vs. PAR	0.80 (-2.20 to 3.96)	2.01 (-20.25 to 24.27)	0.61 (-2.14 to 3.36)
NEF vs. TRA	1.70 (-1.60 to 5.15)	2.68 (-17.72 to 23.08)	-1.09 (-4.23 to 2.05)
NEF vs. VEN	1.87 (-1.20 to 5.04)	IS	2.50 (-0.64 to 5.64)
PAR vs. SER	0.24 (-1.10 to 1.47)	0.24 (-2.46 to 2.93)	0.25 (-1.82 to 2.32)
PAR vs. TRA	0.89 (-0.84 to 2.61)	0.67 (-6.88 to 8.22)	-1.70 (-4.45 to 1.05)
PAR vs. VEN	1.07 (-0.17 to 2.24)	2.66 (-4.47 to 9.79)	2.64 (0.08 to 5.20)
SER vs. VEN	0.82 (-0.44 to 2.09)	2.38 (-2.49 to 7.26)	2.39 (0.01 to 4.77)
TRA vs. VEN	0.18 (-1.80 to 2.08)	0.79 (-5.72 to 7.29)	-0.85 (-4.00 to 2.31)

Abbreviations: BUP = bupropion; CIT = citalopram; DES = desvenlafaxine; DUL = duloxetine; ESC = escitalopram; FLUO = fluoxetine; FLUV = fluvoxamine; IS = insufficient studies; MIR = mirtazapine; NA = not applicable; NEF = nefazodone; PAR = paroxetine; SER = sertraline; TRA = trazodone; VEN = velafaxine

Table E-3. Comparison of ACR50 response for biologic DMARDs (full network), by method of analysis: Odds ratios (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
ABA vs. ADA	0.65 (0.37 to 1.20)	0.67 (0.34 to 1.33)	0.64 (0.38 to 1.08)	0.83 (0.60 to 1.14)
ABA vs. ANA	1.26 (0.63 to 2.78)	1.32 (0.63 to 2.79)	1.32 (0.78 to 2.24)	1.82 (1.20 to 2.76)
ABA vs. ETA	0.19 (0.08 to 0.50)	0.21 (0.08 to 0.57)	0.21 (0.10 to 0.45)	0.41 (0.26 to 0.66)
ABA vs. GOL	0.54 (0.23 to 1.49)	0.59 (0.18 to 1.89)	0.57 (0.28 to 1.17)	0.76 (0.43 to 1.34)
ABA vs. INF	0.94 (0.58 to 1.65)	0.92 (0.49 to 2.29)	0.80 (0.47 to 1.38)	1.01 (0.75 to 1.35)
ABA vs. RIT	0.84 (0.39 to 2.08)	0.88 (0.34 to 1.53)	0.89 (0.49 to 1.60)	1.03 (0.64 to 1.67)
ABA vs. TOC	0.80 (0.47 to 1.57)	0.76 (0.38 to 2.44)	0.77 (0.48 to 1.25)	0.82 (0.60 to 1.13)
ADA vs. ANA	1.88 (1.01 to 3.98)	1.95 (0.82 to 4.61)	2.08 (1.12 to 3.83)	2.20 (1.49 to 3.26)
ADA vs. ETA[a]	0.29 (0.13 to 0.73)	0.32 (0.12 to 0.84)	0.33 (0.15 to 0.75)	0.50 (0.32 to 0.78)
ADA vs. GOL	0.80 (0.36 to 2.14)	0.87 (0.27 to 2.82)	0.90 (0.41 to 1.97)	0.93 (0.53 to 1.59)
ADA vs. INF	1.39 (0.86 to 2.53)	1.42 (0.72 to 2.81)	1.26 (0.67 to 2.35)	1.22 (0.91 to 1.63)
ADA vs. RIT	1.25 (0.61 to 3.04)	1.30 (0.47 to 3.60)	1.39 (0.71 to 2.72)	1.25 (0.79 to 1.98)
ADA vs. TOC	1.19 (0.74 to 2.27)	1.20 (0.60 to 2.44)	1.21 (0.68 to 2.16)	1.00 (0.75 to 1.32)
ANA vs. ETA	0.14 (0.06 to 0.41)	0.16 (0.06 to 0.47)	0.16 (0.07 to 0.36)	0.23 (0.13 to 0.38)
ANA vs. GOL	0.39 (0.16 to 1.18)	0.45 (0.12 to 1.60)	0.43 (0.20 to 0.95)	0.42 (0.23 to 0.77)
ANA vs. INF	0.68 (0.36 to 1.47)	0.71 (0.31 to 1.63)	0.61 (0.32 to 1.13)	0.55 (0.37 to 0.82)
ANA vs. RIT	0.61 (0.27 to 1.70)	0.67 (0.22 to 1.97)	0.67 (0.34 to 1.31)	0.57 (0.33 to 0.96)
ANA vs. TOC	0.58 (0.32 to 1.30)	1.00 (0.45 to 2.21)	0.59 (0.33 to 1.04)	0.45 (0.31 to 0.67)
ETA vs. GOL	2.39 (0.92 to 8.64)	2.77 (0.72 to 10.62)	2.70 (1.05 to 6.96)	1.85 (0.98 to 3.51)
ETA vs. INF	4.17 (2.00 to 11.20)	4.34 (1.62 to 11.60)	3.78 (1.66 to 8.58)	2.45 (1.56 to 3.82)
ETA vs. RIT	3.76 (1.51 to 12.70)	4.15 (1.24 to 13.88)	4.18 (1.78 to 9.82)	2.50 (1.41 to 4.43)
ETA vs. TOC	3.59 (1.73 to 9.80)	3.63 (1.32 to 9.98)	3.65 (1.67 to 7.98)	2.00 (1.28 to 3.12)
GOL vs. INF	1.47 (0.68 to 4.05)	1.63 (0.51 to 5.20)	1.40 (0.63 to 3.08)	1.32 (0.77 to 2.28)
GOL vs. RIT	1.33 (0.53 to 4.54)	1.50 (0.24 to 4.93)	1.55 (0.68 to 3.53)	1.35 (0.70 to 2.60)
GOL vs. TOC	1.27 (0.59 to 3.57)	1.37 (0.38 to 1.97)	1.35 (0.64 to 2.86)	1.08 (0.63 to 1.86)
INF vs. RIT	0.87 (0.41 to 2.07)	0.93 (0.34 to 2.51)	1.11 (0.56 to 2.19)	1.02 (0.64 to 1.62)
INF vs. TOC	0.83 (0.51 to 1.52)	0.72 (0.38 to 1.38)	0.97 (0.54 to 1.74)	0.82 (0.62 to 1.08)
RIT vs. TOC	0.84 (0.42 to 2.12)	0.84 (0.26 to 2.77)	0.87 (0.46 to 1.65)	0.80 (0.51 to 1.26)

[a]The Bayesian MTC method and the Bucher method produced the same precision and therefore are shaded the same (i.e., they tied for second most precise for this comparison)

Abbreviations: ABA = abatacept; ADA = adalimumab; ANA = anakinra; ETA = etanercept; GOL = golimumab; INF = infliximab; RIT = rituximab; TOC = tocilizumab

Table E-4. Comparison of response for second-generation antidepressants for placebo star sub-network, by method of analysis: Odds ratios (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
BUP vs. DES	0.73 (0.43 to 1.35)	0.78 (0.30 to 2.01)	0.77 (0.40 to 1.50)	0.87 (0.64 to 1.18)
BUP vs. DUL	0.84 (0.51 to 1.49)	0.85 (0.50 to 1.45)	0.87 (0.56 to 1.36)	0.94 (0.70 to 1.26)
BUP vs. ESC	0.94 (0.43 to 2.56)	1.04 (0.21 to 5.11)	1.04 (0.59 to 1.85)	1.08 (0.72 to 1.63)
BUP vs. FLUO	0.94 (0.56 to 1.72)	0.98 (0.56 to 1.71)	0.98 (0.62 to 1.56)	0.99 (0.73 to 1.33)
BUP vs. FLUV	0.39 (0.11 to 2.63)	0.59 (0.02 to 15.83)	0.58 (0.14 to 2.46)	0.82 (0.26 to 2.54)
BUP vs. NEF	0.58 (0.27 to 1.49)	0.65 (0.19 to 2.18)	0.82 (0.35 to 1.91)	1.00 (0.51 to 1.98)
BUP vs. PAR	0.67 (0.37 to 1.39)	0.94 (0.45 to 1.99)	0.71 (0.38 to 1.33)	0.84 (0.57 to 1.23)
BUP vs. SER	0.78 (0.47 to 1.41)	0.80 (0.49 to 1.30)	0.80 (0.52 to 1.24)	0.96 (0.72 to 1.27)
BUP vs. VEN	0.69 (0.38 to 1.43)	0.74 (0.33 to 1.62)	0.73 (0.42 to 1.28)	0.80 (0.55 to 1.17)
DES vs. DUL	1.11 (0.73 to 1.79)	1.10 (0.59 to 2.03)	1.14 (0.63 to 2.05)	1.08 (0.86 to 1.34)
DES vs. ESC	1.25 (0.59 to 3.13)	1.35 (0.22 to 8.18)	1.35 (0.67 to 2.71)	1.24 (0.86 to 1.79)
DES vs. FLUO	1.25 (0.81 to 2.04)	1.27 (0.64 to 2.54)	1.27 (0.69 to 2.34)	1.14 (0.86 to 1.50)
DES vs. FLUV	0.51 (0.14 to 3.45)	0.76 (0.06 to 10.14)	0.76 (0.17 to 3.37)	0.94 (0.31 to 2.88)
DES vs. NEF	0.77 (0.37 to 1.85)	0.83 (0.21 to 3.27)	1.07 (0.42 to 2.71)	1.15 (0.60 to 2.23)
DES vs. PAR	0.89 (0.52 to 1.69)	1.23 (0.51 to 2.96)	0.93 (0.44 to 1.94)	0.97 (0.69 to 1.35)
DES vs. SER	1.03 (0.66 to 1.69)	1.05 (0.57 to 1.93)	1.04 (0.58 to 1.87)	1.10 (0.85 to 1.43)
DES vs. VEN	0.92 (0.53 to 1.75)	0.95 (0.36 to 2.54)	0.96 (0.48 to 1.89)	0.92 (0.65 to 1.30)
DUL vs. ESC	1.10 (0.53 to 2.70)	1.23 (0.54 to 2.82)	1.19 (0.73 to 1.94)	1.16 (0.85 to 1.58)
DUL vs. FLUO	1.10 (0.73 to 1.72)	1.12 (0.75 to 1.68)	1.12 (0.79 to 1.59)	1.06 (0.81 to 1.37)
DUL vs. FLUV	0.45 (0.12 to 3.03)	0.69 (0.11 to 4.35)	0.67 (0.16 to 2.72)	0.88 (0.29 to 2.67)
DUL vs. NEF	0.68 (0.33 to 1.61)	0.77 (0.33 to 1.79)	0.94 (0.43 to 2.07)	1.07 (0.56 to 2.06)
DUL vs. PAR	0.78 (0.47 to 1.45)	1.08 (0.66 to 1.77)	0.82 (0.47 to 1.41)	0.90 (0.67 to 1.21)
DUL vs. SER	0.91 (0.61 to 1.43)	0.92 (0.64 to 1.31)	0.92 (0.67 to 1.25)	1.02 (0.80 to 1.31)
DUL vs. VEN	0.81 (0.48 to 1.49)	0.92 (0.54 to 1.59)	0.84 (0.53 to 1.34)	0.85 (0.61 to 1.19)
ESC vs. FLUO	0.86 (0.41 to 2.17)	0.94 (0.42 to 2.11)	0.94 (0.57 to 1.57)	0.91 (0.62 to 1.34)
ESC vs. FLUV	0.35 (0.09 to 2.86)	IS	0.56 (0.13 to 2.40)	0.76 (0.24 to 2.39)
ESC vs. NEF	0.53 (0.21 to 1.79)	IS	0.79 (0.33 to 1.88)	0.93 (0.45 to 1.89)
ESC vs. PAR	0.61 (0.27 to 1.69)	0.69 (0.15 to 3.13)	1.15 (0.71 to 1.88)	0.78 (0.51 to 1.18)
ESC vs. SER	0.71 (0.34 to 1.79)	0.77 (0.39 to 1.52)	0.77 (0.47 to 1.24)	0.88 (0.61 to 1.29)
ESC vs. VEN	0.63 (0.29 to 1.72)	0.71 (0.14 to 3.46)	0.71 (0.39 to 1.28)	0.74 (0.48 to 1.14)
FLUO vs. FLUV	0.40 (0.11 to 2.63)	0.60 (0.09 to 4.02)	0.60 (0.15 to 2.44)	0.83 (0.27 to 2.55)
FLUO vs. NEF	0.60 (0.29 to 1.45)	0.66 (0.28 to 1.56)	0.84 (0.38 to 1.86)	1.01 (0.52 to 1.98)
FLUO vs. PAR	0.69 (0.40 to 1.30)	0.73 (0.39 to 1.37)	0.73 (0.41 to 1.28)	0.85 (0.60 to 1.20)
FLUO vs. SER	0.81 (0.53 to 1.28)	0.82 (0.54 to 1.23)	0.82 (0.58 to 1.15)	0.97 (0.73 to 1.28)
FLUO vs. VEN	0.72 (0.42 to 1.33)	0.75 (0.42 to 1.35)	0.75 (0.46 to 1.22)	0.81 (0.59 to 1.11)
FLUV vs. NEF	0.79 (0.21 to 6.67)	IS	1.40 (0.29 to 6.80)	1.22 (0.34 to 4.35)
FLUV vs. PAR	0.93 (0.26 to 6.67)	1.61 (0.13 to 20.02)	IS	1.02 (0.33 to 3.20)
FLUV vs. SER	1.08 (0.31 to 7.69)	1.37 (0.23 to 8.11)	IS	1.17 (0.38 to 3.57)
FLUV vs. VEN	0.96 (0.27 to 7.14)	1.26 (0.05 to 35.00)	1.26 (0.30 to 5.32)	0.98 (0.31 to 3.05)
NEF vs. PAR	1.01 (0.47 to 2.70)	1.46 (0.47 to 4.51)	0.87 (0.35 to 2.15)	0.84 (0.42 to 1.68)
NEF vs. SER	1.16 (0.57 to 2.86)	1.23 (0.56 to 2.67)	0.98 (0.44 to 2.14)	0.95 (0.49 to 1.85)
NEF vs. VEN	1.04 (0.48 to 2.78)	0.84 (0.29 to 2.41)	0.90 (0.38 to 2.11)	0.80 (0.40 to 1.61)
PAR vs. SER	1.09 (0.63 to 2.04)	1.14 (0.71 to 1.84)	1.12 (0.65 to 1.93)	1.14 (0.81 to 1.61)
PAR vs. VEN	0.97 (0.52 to 2.04)	0.78 (0.36 to 1.70)	1.03 (0.54 to 1.96)	0.95 (0.63 to 1.43)
SER vs. VEN	0.87 (0.51 to 1.61)	0.92 (0.55 to 1.54)	0.92 (0.58 to 1.45)	0.84 (0.59 to 1.18)

Abbreviations: BUP = bupropion; CIT = citalopram; DES = desvenlafaxine; DUL = duloxetine; ESC = escitalopram; FLUO = fluoxetine; FLUV = fluvoxamine; IS = insufficient studies; MIR = mirtazapine; NA = not applicable; NEF = nefazodone; PAR = paroxetine; SER = sertraline; TRA = trazodone; VEN = velafaxine

Table E-5. Comparison of response for second-generation antidepressants for loop sub-network, by method of analysis: Odds ratios (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
FLUO vs. PAR via SER	0.83 (0.64 to 1.11)	0.88 (0.19 to 4.02)	0.88 (0.44 to 1.76)	0.73 (0.51 to 1.05)
FLUO vs. PAR via placebo	Same as above	0.73 (0.39 to 1.37)	0.73 (0.41 to 1.28)	0.85 (0.60 to 1.20)
FLUO vs. SER via PAR	0.76 (0.61 to 0.96)	0.69 (0.26 to 1.83)	0.69 (0.34 to 1.37)	0.86 (0.64 to 1.16)
FLUO vs. SER via placebo	Same as above	0.82 (0.54 to 1.23)	0.82 (0.58 to 1.15)	0.97 (0.73 to 1.28)
PAR vs. SER via FLUO	0.90 (0.68 to 1.25)	0.71 (0.41 to 1.22)	0.71 (0.46 to 1.10)	0.73 (0.51 to 1.05)
PAR vs. SER via placebo	Same as above	1.14 (0.71 to 1.84)	1.12 (0.65 to 1.93)	1.14 (0.81 to 1.61)

Abbreviations: FLUO = fluoxetine; NA = not applicable; PAR = paroxetine; SER = sertraline

Table E-6. Comparison of response for second-generation antidepressants for one closed loop sub-network, by method of analysis: Odds ratios (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-regression	Bucher method	Logistic regression
BUP vs. CIT	0.74 (0.30 to 2.27)	NA	NA	0.94 (0.49 to 1.80)
BUP vs. ESC	0.91 (0.46 to 2.13)	NA	NA	0.95 (0.55 to 1.64)
BUP vs. FLUO	0.99 (0.65 to 1.61)	IS	IS	1.07 (0.81 to 1.40)
BUP vs. FLUV	0.78 (0.28 to 3.13)	NA	NA	1.18 (0.53 to 2.65)
BUP vs. MIR	1.32 (0.62 to 3.85)	IS	IS	1.10 (0.51 to 2.35)
BUP vs. PAR	0.83 (0.51 to 1.43)	IS	IS	0.89 (0.64 to 1.23)
BUP vs. SER	0.77 (0.52 to 1.18)	Direct	Direct	0.87 (0.67 to 1.13)
BUP vs. TRA	2.04 (1.02 to 4.76)	Direct	Direct	1.75 (0.96 to 3.18)
CIT vs. ESC	1.11 (0.63 to 2.17)	Direct	Direct	1.01 (0.61 to 1.68)
CIT vs. FLUO	1.09 (0.50 to 2.94)	IS	IS	1.14 (0.62 to 2.09)
CIT vs. FLUV	1.05 (0.58 to 2.08)	NA	NA	1.26 (0.71 to 2.21)
CIT vs. MIR	1.37 (0.44 to 7.14)	NA	NA	1.17 (0.45 to 3.07)
CIT vs. PAR	0.91 (0.40 to 2.56)	NA	NA	0.95 (0.50 to 1.79)
CIT vs. SER	0.84 (0.37 to 2.33)	NA	NA	0.93 (0.50 to 1.73)
CIT vs. TRA	2.08 (0.70 to 10.00)	NA	NA	1.86 (0.80 to 4.35)
ESC vs. FLUO	0.98 (0.53 to 1.92)	Direct	Direct	1.12 (0.69 to 1.82)
ESC vs. FLUV	0.85 (0.39 to 2.27)	IS	IS	1.24 (0.60 to 2.56)
ESC vs. MIR	1.22 (0.42 to 5.56)	NA	NA	1.15 (0.47 to 2.83)
ESC vs. PAR	0.81 (0.42 to 1.67)	IS	IS	0.93 (0.55 to 1.58)
ESC vs. SER	0.75 (0.40 to 1.54)	IS	IS	0.91 (0.55 to 1.52)
ESC vs. TRA	1.89 (0.70 to 7.14)	NA	NA	1.84 (0.85 to 3.98)
FLUO vs. FLUV	0.78 (0.31 to 2.86)	NA	NA	1.11 (0.51 to 2.40)
FLUO vs. MIR	1.27 (0.52 to 4.35)	NA	NA	1.03 (0.47 to 2.23)
FLUO vs. PAR via SER	0.83 (0.65 to 1.10)	IS	IS	0.83 (0.65 to 1.06)
FLUO vs. PAR via placebo	NA/same	0.73 (0.39 to 1.37)	0.73 (0.41 to 1.28)	NA/same
FLUO vs. SER	0.77 (0.63 to 0.96)	0.82 (0.54 to 1.23)	0.82 (0.58 to 1.15)	0.81 (0.67 to 1.00)
FLUO vs. TRA	1.96 (0.87 to 5.88)	NA	NA	1.64 (0.88 to 3.05)
MIR vs. PAR	0.50 (0.19 to 1.61)	NA	NA	0.81 (0.37 to 1.80)
MIR vs. SER	0.47 (0.18 to 1.43)	NA	NA	0.79 (0.36 to 1.73)
MIR vs. TRA	1.45 (0.91 to 2.50)	Direct	Direct	1.59 (0.92 to 2.76)
PAR vs. SER via FLUO	0.91 (0.68 to 1.27)	IS	IS	0.98 (0.75 to 1.28)
PAR vs. SER via placebo	NA/same	1.14 (0.71 to 1.84)	1.12 (0.65 to 1.93)	NA/same
PAR vs. TRA	2.33 (1.02 to 7.14)	NA	NA	1.97 (1.03 to 3.76)
SER vs. TRA	2.56 (1.16 to 7.14)	IS	IS	2.01 (1.08 to 3.75)

Abbreviations: BUP = bupropion; CIT = citalopram; ESC = escitalopram; FLUO = fluoxetine; FLUV = fluvoxamine; IS = insufficient studies; MIR = mirtazapine; NA = not applicable; PAR = paroxetine; SER = sertraline; TRA = trazodone

Table E-7. Comparison of response for second-generation antidepressants for ladder sub-network, by method of analysis: Odds ratios (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
BUP vs. CIT	0.75 (0.28 to 2.56)	NA	NA	1.36 (0.72 to 2.58)
BUP vs. DUL	1.20 (0.58 to 2.94)	NA	NA	1.43 (0.90 to 2.28)
BUP vs. ESC	0.94 (0.41 to 2.44)	NA	NA	1.24 (0.76 to 2.03)
BUP vs. FLUO	1.20 (0.68 to 2.44)	NA	NA	1.25 (0.83 to 1.90)
BUP vs. FLUV	0.79 (0.26 to 3.57)	NA	NA	1.82 (0.82 to 4.03)
BUP vs. MIR	1.30 (0.56 to 3.57)	IS	0.67 (0.26 to 1.69)	1.38 (0.65 to 2.91)
BUP vs. PAR	1.10 (0.56 to 2.38)	NA	NA	1.22 (0.79 to 1.90)
BUP vs. SER	0.95 (0.68 to 1.39)	Direct	Direct	1.08 (0.77 to 1.51)
BUP vs. TRA	2.00 (0.99 to 5.00)	Direct	Direct	2.14 (1.18 to 3.86)
BUP vs. VEN	0.79 (0.47 to 1.45)	0.82 (0.38 to 1.79)	0.82 (0.51 to 1.34)	0.98 (0.65 to 1.48)
CIT vs. DUL	1.39 (0.74 to 3.13)	1.52 (0.23 to 10.02)	1.44 (0.78 to 2.64)	1.05 (0.64 to 1.74)
CIT vs. ESC	1.11 (0.63 to 2.27)	Direct	Direct	0.91 (0.56 to 1.47)
CIT vs. FLUO	1.32 (0.63 to 3.57)	NA	NA	0.92 (0.53 to 1.59)
CIT vs. FLUV	1.05 (0.58 to 2.17)	Direct	Direct	1.33 (0.76 to 2.33)
CIT vs. MIR	1.27 (0.41 to 7.14)	NA	NA	1.01 (0.41 to 2.47)
CIT vs. PAR	1.23 (0.61 to 3.03)	NA	NA	0.90 (0.53 to 1.53)
CIT vs. SER	0.95 (0.41 to 3.23)	NA	NA	0.79 (0.44 to 1.41)
CIT vs. TRA	1.96 (0.68 to 10.00)	NA	NA	1.57 (0.72 to 3.43)
CIT vs. VEN	0.83 (0.37 to 2.38)	NA	NA	0.72 (0.41 to 1.25)
DUL vs. ESC	0.79 (0.57 to 1.06)	Direct	Direct	0.86 (0.69 to 1.08)
DUL vs. FLUO	0.94 (0.61 to 1.56)	0.96 (0.58 to 1.60)	0.97 (0.64 to 1.45)	0.87 (0.64 to 1.20)
DUL vs. FLUV	0.66 (0.28 to 1.96)	NA	NA	1.27 (0.63 to 2.56)
DUL vs. MIR	0.92 (0.33 to 3.85)	NA	NA	0.96 (0.44 to 2.10)
DUL vs. PAR	0.88 (0.65 to 1.25)	Direct	Direct	0.85 (0.67 to 1.08)
DUL vs. SER	0.69 (0.37 to 1.52)	NA	NA	0.75 (0.52 to 1.09)
DUL vs. TRA	1.43 (0.57 to 5.26)	NA	NA	1.49 (0.78 to 2.86)
DUL vs. VEN	0.60 (0.36 to 1.11)	NA	NA	0.68 (0.49 to 0.96)
ESC vs. FLUO	1.18 (0.71 to 2.22)	NA	NA	1.01 (0.71 to 1.45)
ESC vs. FLUV	0.85 (0.37 to 2.38)	IS	0.94 (0.42 to 2.11)	1.47 (0.73 to 2.93)
ESC vs. MIR	1.15 (0.41 to 5.00)	NA	NA	1.11 (0.50 to 2.48)
ESC vs. PAR	1.10 (0.72 to 1.85)	1.16 (0.57 to 2.36)	1.09 (0.77 to 1.56)	0.99 (0.72 to 1.35)
ESC vs. SER	0.86 (0.44 to 2.13)	NA	NA	0.87 (0.58 to 1.31)
ESC vs. TRA	1.79 (0.69 to 7.14)	NA	NA	1.73 (0.88 to 3.38)
ESC vs. VEN	0.75 (0.42 to 1.56)	NA	NA	0.79 (0.54 to 1.15)
FLUO vs. FLUV	0.66 (0.25 to 2.22)	NA	NA	1.45 (0.70 to 2.99)
FLUO vs. MIR	0.97 (0.38 to 3.57)	NA	NA	1.10 (0.51 to 2.35)
FLUO vs. PAR	0.91 (0.66 to 1.30)	Direct	Direct	0.98 (0.75 to 1.27)
FLUO vs. SER	0.73 (0.45 to 1.32)	0.90 (0.47 to 1.71)	0.95 (0.62 to 1.45)	0.86 (0.64 to 1.16)
FLUO vs. TRA	1.49 (0.65 to 5.00)	NA	NA	1.71 (0.91 to 3.18)
FLUO vs. VEN	0.64 (0.47 to 0.87)	Direct	Direct	0.78 (0.65 to 0.94)
FLUV vs. MIR	1.08 (0.33 to 7.69)	NA	NA	0.76 (0.28 to 2.09)
FLUV vs. PAR	1.04 (0.43 to 3.45)	NA	NA	0.67 (0.33 to 1.38)
FLUV vs. SER	0.81 (0.29 to 3.57)	NA	NA	0.59 (0.28 to 1.26)
FLUV vs. TRA	1.67 (0.53 to 11.11)	NA	NA	1.18 (0.47 to 2.94)
FLUV vs. VEN	0.71 (0.27 to 2.70)	NA	NA	0.54 (0.26 to 1.12)
MIR vs. PAR	0.66 (0.24 to 2.56)	NA	NA	0.89 (0.41 to 1.92)
MIR vs. SER	0.57 (0.25 to 1.89)	NA	NA	0.78 (0.37 to 1.66)
MIR vs. TRA	1.45 (0.92 to 2.50)	Direct	Direct	1.55 (0.90 to 2.68)
MIR vs. VEN	0.48 (0.20 to 1.69)	NA	NA	0.71 (0.33 to 1.52)
PAR vs. SER	0.78 (0.45 to 1.59)	NA	NA	0.88 (0.63 to 1.24)
PAR vs. TRA	1.61 (0.67 to 5.88)	NA	NA	1.75 (0.92 to 3.31)
PAR vs. VEN	0.68 (0.44 to 1.11)	0.83 (0.50 to 1.38)	0.87 (0.60 to 1.25)	0.80 (0.60 to 1.08)
SER vs. TRA	2.04 (0.93 to 5.56)	IS	1.03 (0.46 to2.30)	1.99 (1.08 to 3.67)

Table E-7. Comparison of response for second-generation antidepressants for ladder sub-network, by method of analysis: Odds ratios (95% credible interval or 95% confidence interval) (continued)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
SER vs. VEN	0.83 (0.55 to 1.30)	Direct	Direct	0.91 (0.68 to 1.21)
TRA vs. VEN	0.33 (0.14 to 0.99)	NA	NA	0.46 (0.25 to 0.86)

Abbreviations: BUP = bupropion; CIT = citalopram; DUL = duloxetine; ESC = escitalopram; FLUO = fluoxetine; FLUV = fluvoxamine; IS = insufficient studies; MIR = mirtazapine; NA = not applicable; PAR = paroxetine; SER = sertraline; TRA = trazodone; VEN = venlafaxine

Table E-8. Comparison of mean change in HAM-D for second-generation antidepressants for placebo star sub-network, by method of analysis: Weighted mean difference (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method
DES vs. DUL	-0.65 (-2.80 to 1.38)	-0.65 (-2.92 to 1.61)	-0.68 (-2.25 to 0.90)
DES vs. ESC	-1.20 (-4.10 to 1.75)	-1.23 (-5.06 to 2.61)	-1.23 (-2.97 to 0.51)
DES vs. FLUO	0.12 (-2.30 to 2.50)	0.72 (-4.27 to 5.71)	-0.27 (-5.19 to 4.64)
DES vs. PAR	-0.88 (-3.20 to 1.24)	-0.67 (-3.33 to 2.00)	-0.72 (-2.65 to 1.21)
DES vs. SER	-0.43 (-2.60 to 1.70)	-0.41 (-2.68 to 1.86)	-0.47 (-2.15 to 1.21)
DES vs. VEN	1.94 (-1.30 to 5.21)	1.92 (-3.05 to 6.89)	1.92 (-0.34 to 4.18)
DUL vs. ESC	-0.53 (-3.30 to 2.20)	-0.55 (-4.68 to 3.58)	-0.55 (-2.34 to 1.24)
DUL vs. FLUO	0.77 (-1.40 to 2.92)	1.15 (-3.45 to 5.76)	0.41 (-4.53 to 5.34)
DUL vs. PAR	-0.24 (-2.20 to 1.63)	-0.04 (-2.59 to 2.51)	-0.04 (-2.02 to 1.94)
DUL vs. SER	0.21 (-1.70 to 2.08)	0.20 (-2.00 to 2.39)	0.21 (-1.53 to 1.95)
DUL vs. VEN	2.59 (-0.50 to 5.69)	2.60 (-2.15 to 7.35)	2.60 (0.29 to 4.91)
ESC vs. FLUO	1.30 (-1.70 to 4.28)	0.91 (-49.15 to 50.98)	0.96 (-4.03 to 5.95)
ESC vs. PAR	0.30 (-2.60 to 3.04)	0.49 (-5.90 to 6.88)	0.51 (-1.60 to 2.62)
ESC vs. SER	0.75 (-2.10 to 3.54)	0.77 (-3.44 to 4.97)	0.76 (-1.13 to 2.65)
ESC vs. VEN	3.12 (-0.61 to 6.85)	IS	3.15 (0.73 to 5.57)
FLUO vs. PAR	-1.00 (-3.40 to 1.23)	-1.08 (-7.51 to 5.36)	-0.45 (-5.51 to 4.61)
FLUO vs. SER	-0.55 (-2.80 to 1.72)	-0.92 (-5.82 to 3.97)	-0.20 (-5.17 to 4.77)
FLUO vs. VEN	1.82 (-1.50 to 5.20)	2.24 (-48.69 to 53.16)	2.19 (-3.00 to 7.39)
PAR vs. SER	0.45 (-1.50 to 2.54)	0.24 (-2.46 to 2.93)	0.25 (-1.82 to 2.32)
PAR vs. VEN	2.83 (-0.29 to 6.11)	2.66 (-4.47 to 9.79)	2.64 (0.08 to 5.20)
SER vs. VEN	2.37 (-0.75 to 5.56)	2.38 (-2.49 to 7.26)	2.39 (0.01 to 4.77)

Abbreviations: DES = desvenlafaxine; DUL = duloxetine; ESC = escitalopram; FLUO = fluoxetine; IS = insufficient studies; PAR = paroxetine; SER = sertraline; VEN = venlafaxine

Table E-9. Comparison of mean change in HAM-D for second-generation antidepressants for loop sub-network, by method of analysis: Weighted mean difference (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Via
ESC vs. FLUO	-1.20 (-4.00 to 1.41)	0.91 (-49.15 to 50.98)	0.96 (-4.03 to 5.95)	Placebo
ESC vs. FLUO	-1.20 (-4.00 to 1.41)	IS	-4.74 (-8.11 to -1.36)	VEN
ESC vs. VEN	0.65 (-2.90 to 3.94)	IS	0.60 (-2.27 to 3.47)	FLUO
FLUO vs. VEN	1.87 (-0.90 to 4.77)	IS	-0.90 (-4.00 to 2.20)	ESC

Abbreviations: ESC = escitalopram; FLUO = fluoxetine; IS = insufficient studies; VEN = venlafaxine

Table E-10. Comparison of mean change in HAM-D for second-generation antidepressants for one closed loop sub-network, by method of analysis: Weighted mean difference (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Via
ESC vs. FLUO	-1.20 (-4.60 to 1.80)	0.91 (-49.15 to 50.98)	0.96 (-4.03 to 5.95)	placebo
ESC vs. FLUO	-1.20 (-4.00 to 1.41)	IS	-4.74 (-8.11 to -1.36)	VEN
ESC vs. SER	-1.10 (-4.70 to 2.18)	0.77 (-3.44 to 4.97)	0.76 (-1.13 to 2.65)	placebo
ESC vs. VEN	0.67 (-2.30 to 3.45)	IS	0.60 (-2.27 to 3.47)	FLUO
FLUO vs. PAR	0.04 (-2.90 to 3.24)	IS	-1.49 (-4.38 to 1.40)	TRA
FLUO vs. SER	0.04 (-3.00 to 3.06)	-0.92 (-5.82 to 3.97)	-0.20 (-5.17 to 4.77)	placebo
FLUO vs. VEN	1.85 (-0.50 to 4.31)	IS	-0.90 (-4.00 to 2.20)	ESC
FLUO vs. VEN	1.85 (-0.50 to 4.31)	2.24 (-48.69 to 53.16)	2.19 (-3.00 to 7.39)	placebo
TRA vs. VEN	0.93 (-2.90 to 4.61)	IS	3.59 (0.45 to 6.73)	FLUO

Abbreviations: ESC = escitalopram; FLUO = fluoxetine; FLUV = fluvoxamine; IS = insufficient studies; NA = not applicable; PAR = paroxetine; SER = sertraline; TRA = trazodone; VEN = venlafaxine

Table E-11. Comparison of mean change in HAM-D for second-generation antidepressants for ladder sub-network, by method of analysis: Weighted mean difference (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Via
BUP vs. VEN	1.04 (-3.50 to 5.14)	IS	-0.74 (-4.21 to 2.73)	SER
CIT vs. DUL	0.54 (-3.90 to 4.71)	IS	0.85 (-2.49 to 4.19)	ESC
ESC vs. PAR	-0.67 (-4.10 to 2.62)	0.26 (-5.51 to 6.03)	-0.35 (-3.17 to 2.47)	DUL
FLUO vs. SER	2.50 (-1.10 to 6.53)	1.66 (-31.65 to 34.98)	2.68 (-1.01 to 6.36)	VEN
PAR vs. VEN	2.99 (-0.74 to 6.23)	-4.13 (-12.41 to 4.15)	1.89 (-0.86 to 4.64)	FLUO

Abbreviations: BUP = bupropion; CIT = citalopram; DUL = duloxetine; ESC = escitalopram; FLUO = fluoxetine; IS = insufficient studies; NA = not applicable; PAR = paroxetine; SER = sertraline; VEN = venlafaxine

Table E-12. Comparison of ACR50 response for biologic DMARDs for placebo star sub-network, by method of analysis: Odds ratios (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
ABA vs. ADA	0.59 (0.33 to 1.14)	0.67 (0.34 to 1.33)	0.64 (0.38 to 1.08)	0.83 (0.60 to 1.14)
ABA vs. ANA	1.15 (0.57 to 2.63)	1.32 (0.63 to 2.79)	1.32 (0.78 to 2.24)	1.82 (1.20 to 2.76)
ABA vs. ETA	0.17 (0.08 to 0.49)	0.21 (0.08 to 0.57)	0.21 (0.10 to 0.45)	0.41 (0.26 to 0.66)
ABA vs. GOL	0.49 (0.21 to 1.39)	0.59 (0.18 to 1.89)	0.57 (0.28 to 1.17)	0.76 (0.43 to 1.34)
ABA vs. INF	0.85 (0.48 to 1.72)	0.92 (0.49 to 2.29)	0.80 (0.47 to 1.38)	1.01 (0.75 to 1.35)
ABA vs. RIT	0.78 (0.36 to 2.00)	0.88 (0.34 to 1.53)	0.89 (0.49 to 1.60)	1.03 (0.64 to 1.67)
ABA vs. TOC	0.74 (0.42 to 1.49)	0.76 (0.38 to 2.44)	0.77 (0.48 to 1.25)	0.82 (0.60 to 1.13)
ADA vs. ANA	1.89 (0.99 to 4.00)	1.95 (0.82 to 4.61)	2.08 (1.12 to 3.83)	2.20 (1.49 to 3.26)
ADA vs. ETA	0.28 (0.13 to 0.77)	0.32 (0.12 to 0.84)	0.33 (0.15 to 0.75)	0.50 (0.32 to 0.78)
ADA vs. GOL	0.81 (0.36 to 2.13)	0.87 (0.27 to 2.82)	0.90 (0.41 to 1.97)	0.93 (0.53 to 1.59)
ADA vs. INF	1.39 (0.83 to 2.56)	1.42 (0.72 to 2.81)	1.26 (0.67 to 2.35)	1.22 (0.91 to 1.63)
ADA vs. RIT	1.27 (0.62 to 3.13)	1.30 (0.47 to 3.60)	1.39 (0.71 to 2.72)	1.25 (0.79 to 1.98)
ADA vs. TOC	1.20 (0.74 to 2.33)	1.20 (0.60 to 2.44)	1.21 (0.68 to 2.16)	1.00 (0.75 to 1.32)
ANA vs. ETA	0.14 (0.06 to 0.41)	0.16 (0.06 to 0.47)	0.16 (0.07 to 0.36)	0.23 (0.13 to 0.38)
ANA vs. GOL	0.39 (0.16 to 1.18)	0.45 (0.12 to 1.60)	0.43 (0.20 to 0.95)	0.42 (0.23 to 0.77)
ANA vs. INF	0.68 (0.36 to 1.49)	0.71 (0.31 to 1.63)	0.61 (0.32 to 1.13)	0.55 (0.37 to 0.82)
ANA vs. RIT	0.62 (0.27 to 1.72)	0.67 (0.22 to 1.97)	0.67 (0.34 to 1.31)	0.57 (0.33 to 0.96)
ANA vs. TOC	0.58 (0.32 to 1.32)	1.00 (0.45 to 2.21)	0.59 (0.33 to 1.04)	0.45 (0.31 to 0.67)
ETA vs. GOL	2.38 (0.90 to 8.33)	2.77 (0.72 to 10.62)	2.70 (1.05 to 6.96)	1.85 (0.98 to 3.51)
ETA vs. INF	4.17 (1.96 to 11.11)	4.34 (1.62 to 11.60)	3.78 (1.66 to 8.58)	2.45 (1.56 to 3.82)
ETA vs. RIT	3.85 (1.52 to 12.50)	4.15 (1.24 to 13.88)	4.18 (1.78 to 9.82)	2.50 (1.14 to 4.43)
ETA vs. TOC	3.57 (1.69 to 10.00)	3.63 (1.32 to 9.98)	3.65 (1.67 to 7.98)	2.00 (1.28 to 3.12)
GOL vs. INF	1.45 (0.68 to 4.17)	1.63 (0.51 to 5.20)	1.40 (0.63 to 3.08)	1.32 (0.77 to 2.28)
GOL vs. RIT	1.33 (0.53 to 4.76)	1.50 (0.24 to 4.93)	1.55 (0.68 to 3.53)	1.35 (0.70 to 2.60)
GOL vs. TOC	1.27 (0.59 to 3.70)	1.37 (0.38 to 1.97)	1.35 (0.64 to 2.86)	1.08 (0.63 to 1.86)
INF vs. RIT	0.88 (0.41 to 2.13)	0.93 (0.34 to 2.51)	1.11 (0.56 to 2.19)	1.02 (0.64 to 1.62)
INF vs. TOC	0.83 (0.50 to 1.56)	0.72 (0.38 to 1.38)	0.97 (0.54 to 1.74)	0.82 (0.62 to 1.08)
RIT vs. TOC	0.83 (0.41 to 2.13)	0.84 (0.26 to 2.77)	0.87 (0.46 to 1.65)	0.80 (0.51 to 1.26)

Abbreviations: ABA = abatacept; ADA = adalimumab; ANA = anakinra; ETA = etanercept; GOL = golimumab; INF = infliximab; RIT = rituximab; TOC = tocilizumab

Table E-13. Comparison of ACR50 response for biologic DMARDs for loop sub-network, by method of analysis: Odds ratios (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
ABA vs. INF	0.92 (0.50 to 1.79)	0.90 (0.33 to 2.44)	0.77 (0.48 to 1.25)	0.82(0.60 to 1.13)

Abbreviations: ABA = abatacept; INF = infliximab

Table E-14. Comparison of ACR50 response for biologic DMARDs for one closed loop sub-network using Adalimumab, by method of analysis: Odds ratios (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
ABA vs. ADA	0.62 (0.31 to 1.33)	0.67 (0.34 to 1.33)	0.64 (0.38 to 1.08)	0.83 (0.60 to 1.14)
ABA vs. INF	0.93 (0.51 to 1.89)	0.92 (0.49 to 2.29)	0.80 (0.47 to 1.38)	1.01 (0.75 to 1.35)
ADA vs. INF	1.39 (0.78 to 2.94)	1.42 (0.72 to 2.81)	1.26 (0.67 to 2.35)	1.22 (0.91 to 1.63)

Abbreviations: ABA = abatacept; ADA = adalimumab; INF = infliximab

Table E-15. Comparison of ACR50 response for biologic DMARDs for one closed loop sub-network using Etanercept, by method of analysis: Odds ratios (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
ABA vs. ETA	0.19 (0.08 to 0.52)	0.21 (0.08 to 0.57)	0.21 (0.10 to 0.45)	0.41 (0.26 to 0.66)
ABA vs. INF	0.94 (0.56 to 1.67)	0.88 (0.34 to 2.29)	0.80 (0.47 to 1.38)	1.01 (0.75 to 1.35)
ETA vs. INF	4.17 (1.96 to 11.11)	4.34 (1.62 to 11.60)	3.78 (1.66 to 8.58)	2.45 (1.56 to 3.82)

Abbreviations: ABA = abatacept; ETA = etanercept; INF = infliximab

Table E-16. Comparison of ACR50 response for biologic DMARDs for ladder sub-network, by method of analysis: Odds ratios (95% credible interval or 95% confidence interval)

Comparison	Bayesian MTC	Meta-Regression	Bucher Method	Logistic Regression
ABA vs. ETA	0.19 (0.05 to 1.05)	NA	NA	0.79 (0.50 to 1.26)
ABA vs. INF	0.98 (0.38 to 3.45)	Direct	Direct	0.98 (0.65 to 1.47)
ETA vs. INF	4.00 (1.72 to 12.50)	4.34 (1.62 to 11.60)	3.78 (1.66 to 8.58)	1.24 (0.91 to 1.67)

Abbreviations: ABA = abatacept; ETA = etanercept; INF = infliximab

Appendix F. Distribution Plots From Simulation Study

For each of the scenarios in KQ3, the histograms below provide a graphical representation of the distribution of the probability of best treatment statistic.

Figure F-1. Distribution of best treatment probability: Star pattern, equivalent efficacy scenario

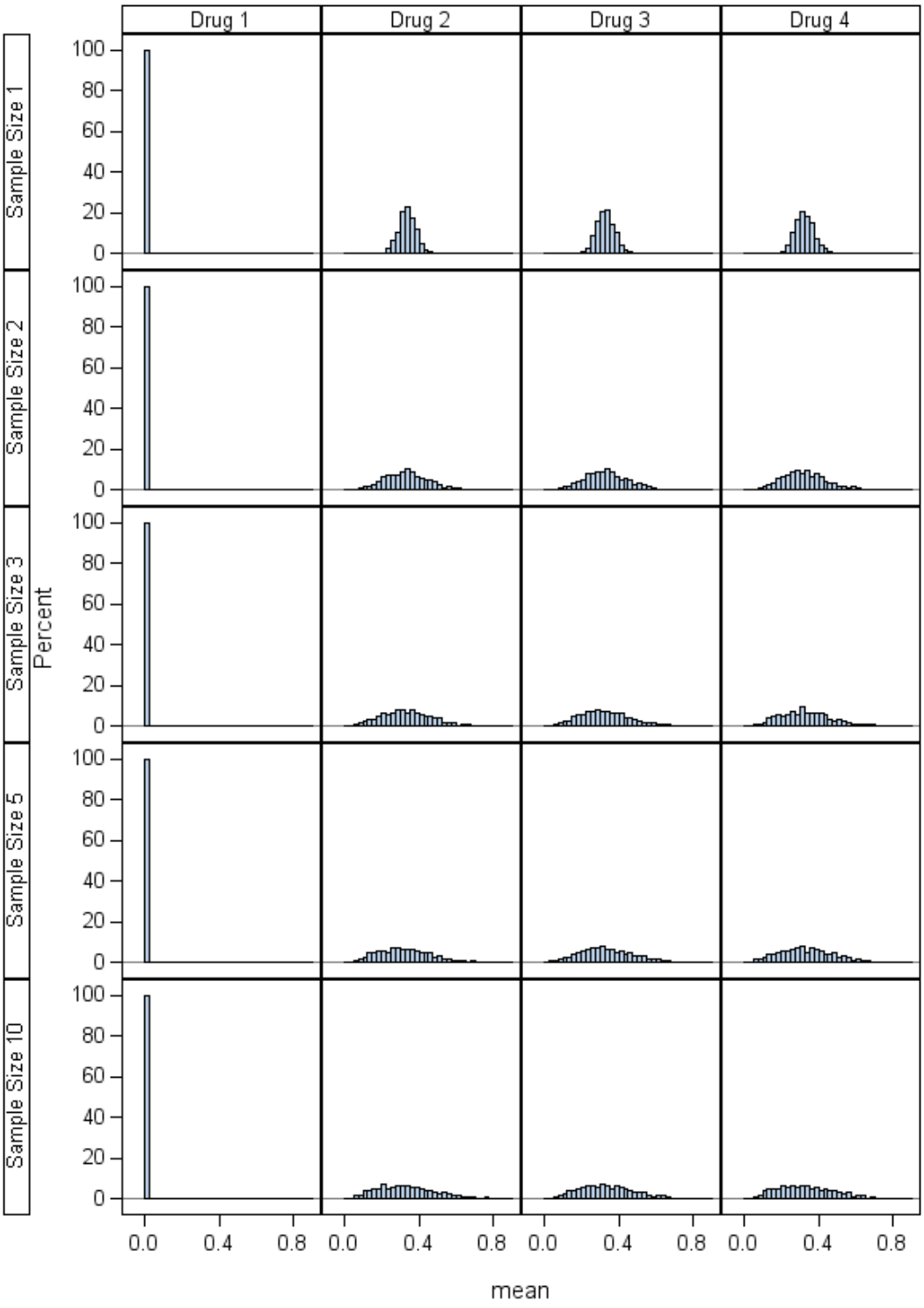


Figure F-2. Distribution of best treatment probability: Loop pattern, equivalent efficacy scenario

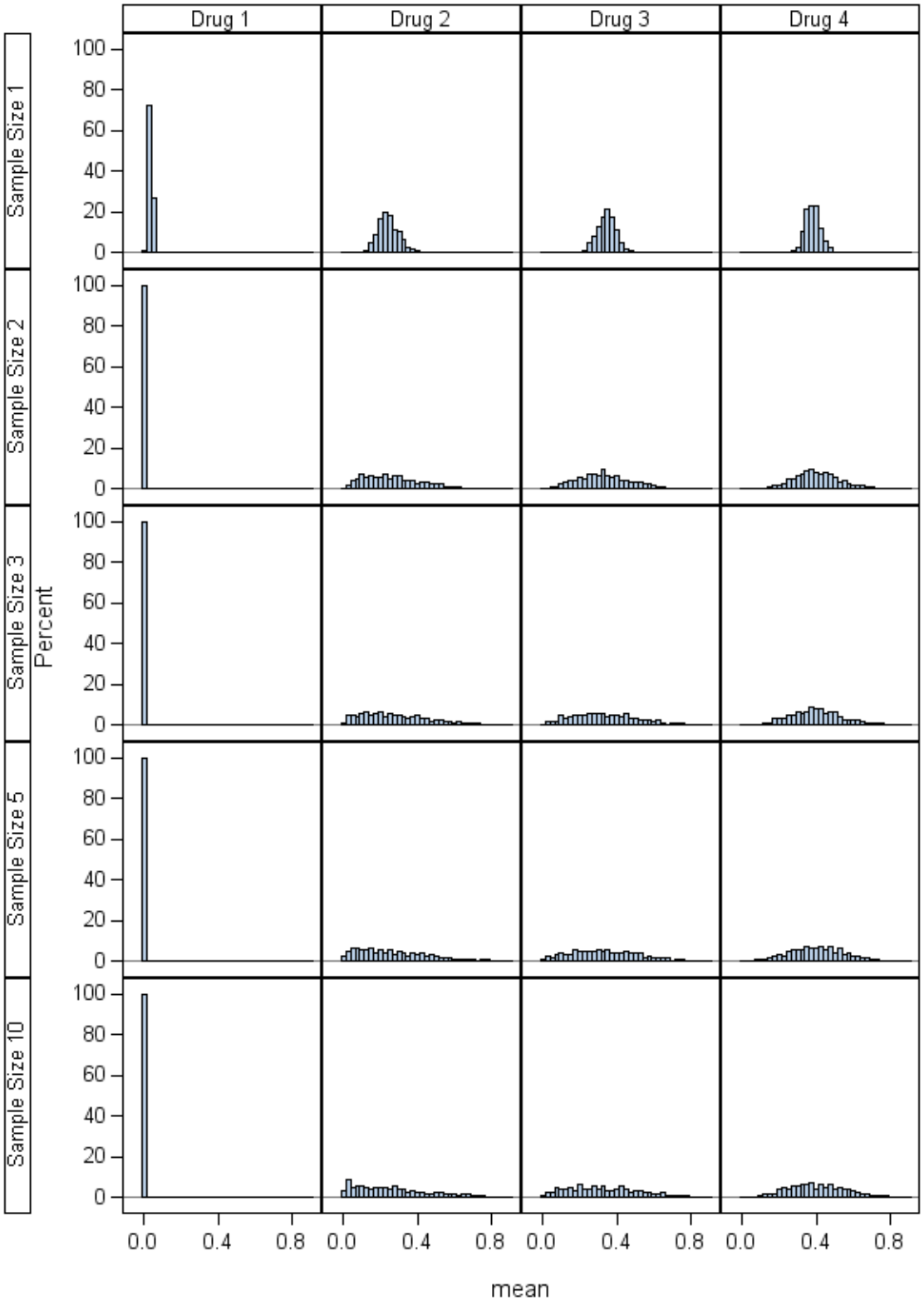


Figure F-3. Distribution of best treatment probability: One closed loop pattern, equivalent efficacy scenario

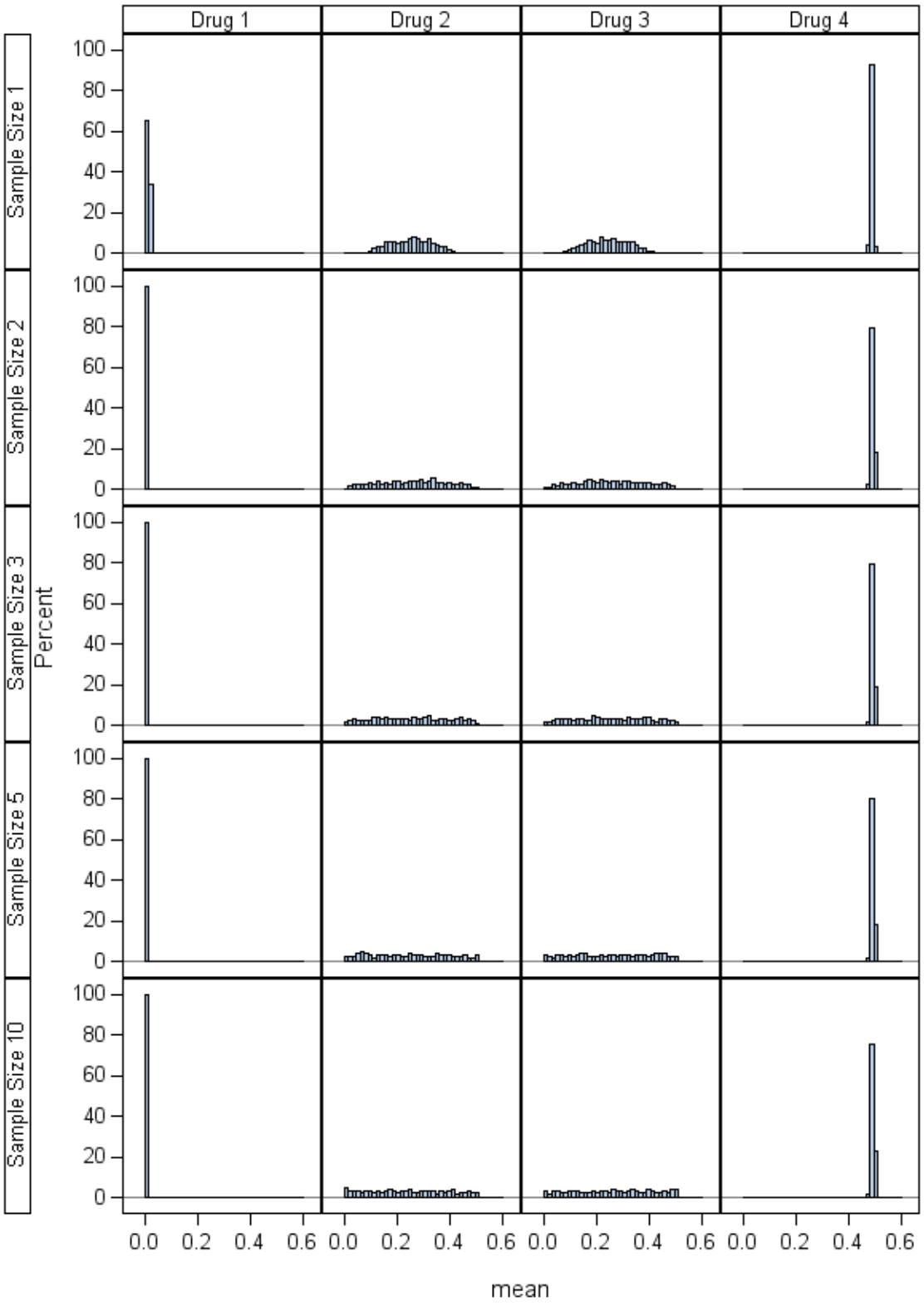


Figure F-4. Distribution of best treatment probability: Ladder pattern, equivalent efficacy scenario

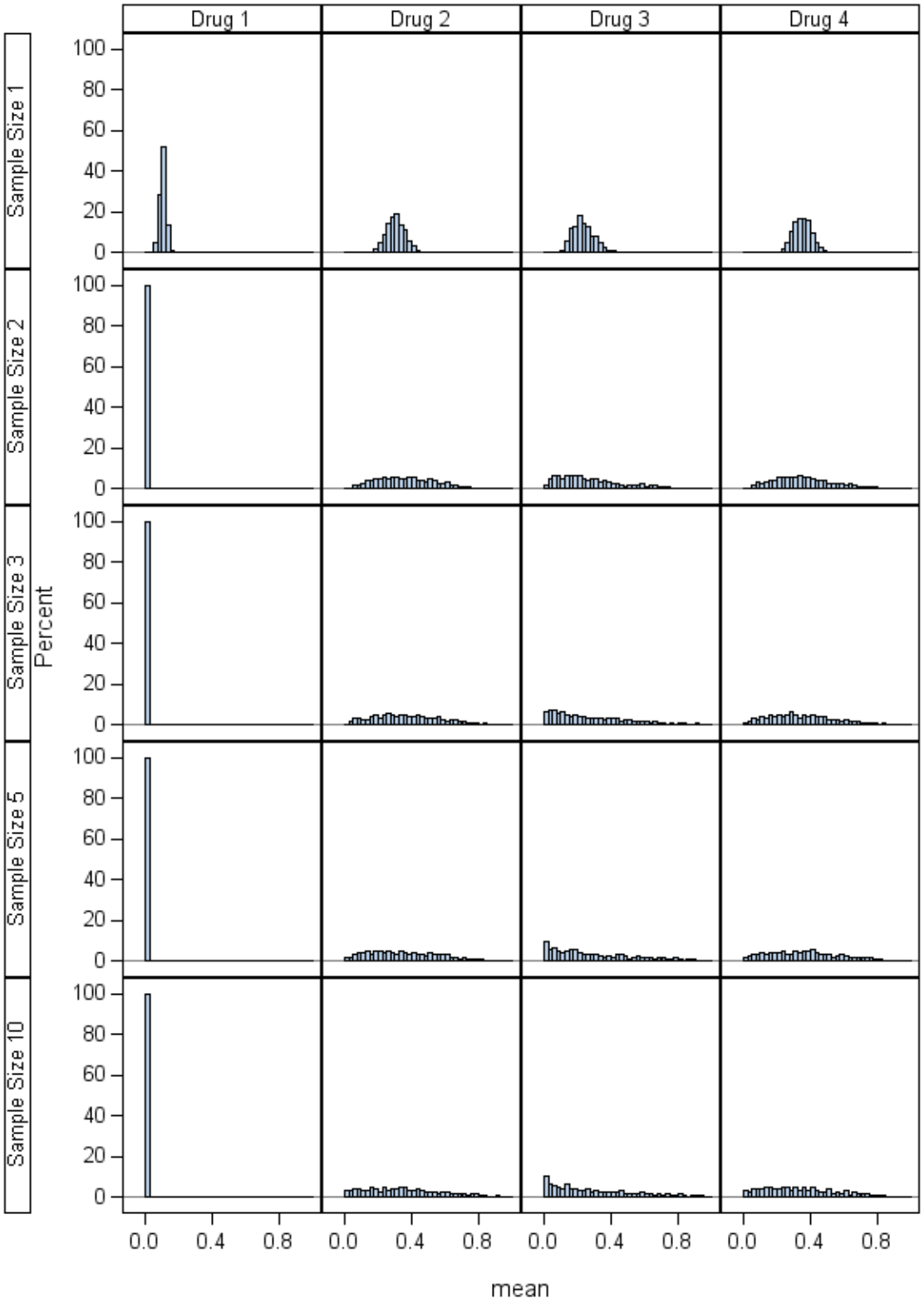


Figure F-5. Distribution of best treatment probability: Star pattern, superior efficacy scenario

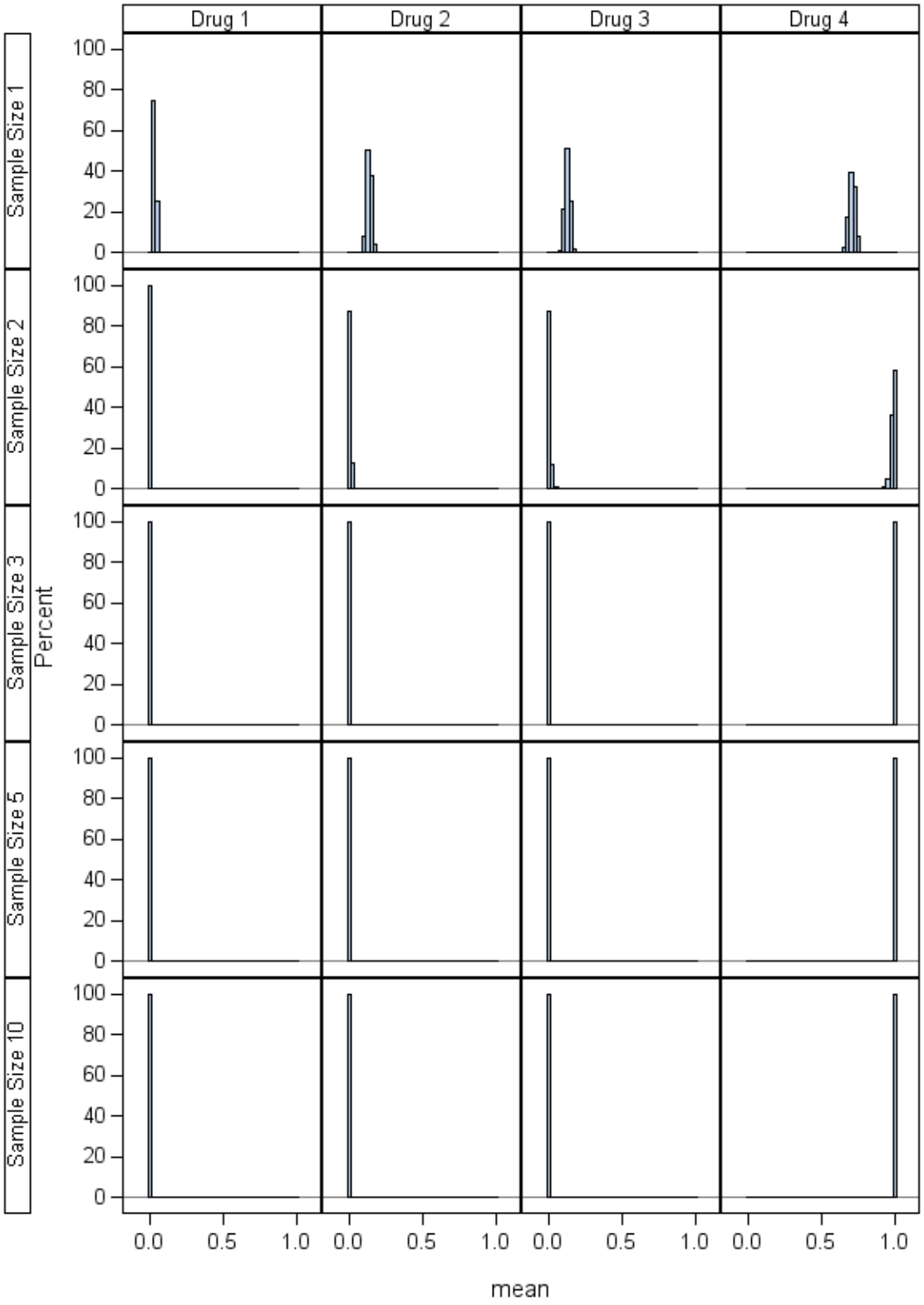


Figure F-6. Distribution of best treatment probability: Loop pattern, superior efficacy scenario

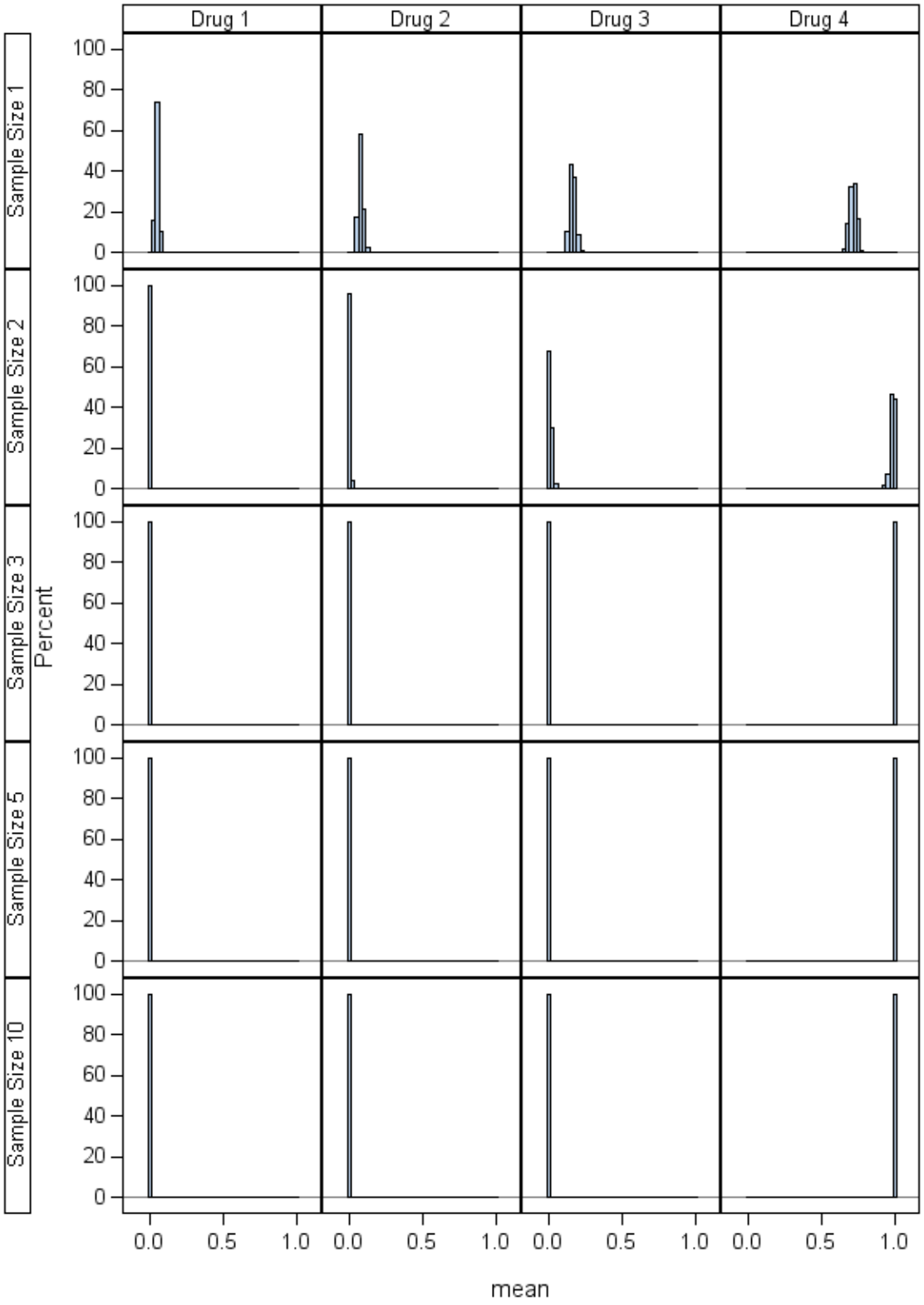


Figure F-7. Distribution of best treatment probability: One closed loop pattern, superior efficacy scenario

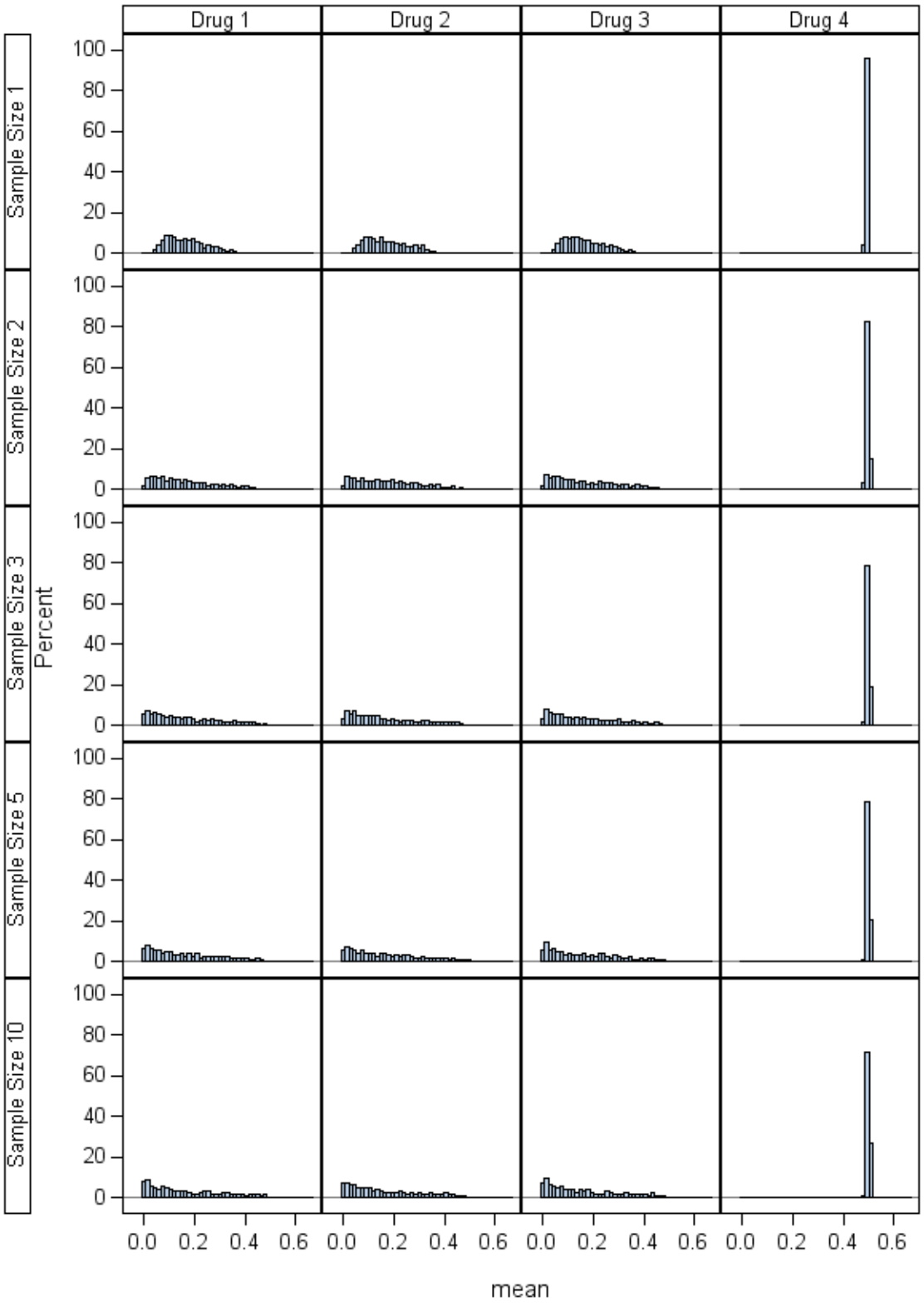


Figure F-8. Distribution of best treatment probability: Ladder pattern, superior efficacy scenario