Hamza Tasnim (Orcid ID: 0000-0002-4700-6990)
Chalkou Konstantina (Orcid ID: 0000-0001-9718-021X)
Benkert Pascal (Orcid ID: 0000-0001-6525-8174)
Lorscheider Johannes (Orcid ID: 0000-0003-1100-2506)
zecca chiara (Orcid ID: 0000-0002-9990-3431)
Iglesias Urrutia Cynthia Paola (Orcid ID: 0000-0002-3426-0930)
Manca Andrea (Orcid ID: 0000-0001-8342-8421)
Furukawa Toshi A (Orcid ID: 0000-0003-2159-3776)
Cipriani Andrea (Orcid ID: 0000-0001-5179-8321)

Synthesizing cross-design evidence and cross-format data using network meta-regression

Tasnim Hamza[1,2], Konstantina Chalkou[1,2], Fabio Pellegrini[3], Jens Kuhle[4], Pascal Benkert[5], Johannes Lorscheider[4], Chiara Zecca[6,7], Cynthia P Iglesias-Urrutia [8], Andrea Manca[9], Toshi A. Furukawa[10], Andrea Cipriani[11], Georgia Salanti[1].

[1]Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland.
[2]Graduate School for Health Sciences, University of Bern, Switzerland.
[3]Biogen Digital Health, Biogen Spain, Madrid, Spain.
[4]Department of Neurology, University Hospital Basel and Departments of Biomedicine and Clinical Research, University Hospital Basel, University of Basel, Basel, Switzerland.
[5]Department of Clinical Research, University of Basel, Basel, Switzerland.
[6]Multiple Sclerosis Center, Neurocenter of Southern Switzerland, EOC, Lugano, Switzerland
[7]Faculty of biomedical Sciences, Università della Svizzera Italiana, Lugano, Switzerland.
[8]Department of Health Sciences, University of York, York, United Kingdom.
[9]Centre for Health Economics, University of York, York, United Kingdom.
[10]Department of Health Promotion and Human Behavior, and Department of Clinical Epidemiology, Graduate School of Medicine/School of Public Health, Kyoto University, Kyoto, Japan.
[11]Department of Psychiatry, University of Oxford, Oxford, United Kingdom.
[12]Oxford Health NHS Foundation Trust, Warneford Hospital, Oxford, United Kingdom.

Funding

TH, KH, CI, AM and GS are funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 825162. AC is supported by the National Institute for Health Research (NIHR) Oxford Cognitive Health Clinical Research Facility, by an NIHR Research Professorship (grant RP-2017-08-ST2-006), by the NIHR Oxford and Thames Valley Applied Research Collaboration, and by the NIHR Oxford Health Biomedical Research Centre (grant BRC-1215-20005). The views expressed are those of the authors and not necessarily those of the UK National Health Service, the NIHR, or the UK Department of Health. CZ holds a grant from Ente Ospedaliero Cantonale for senior researchers.

Conflict of interest

FP is an employee of and holds stocks/stock options in Biogen. JL has received research support from Innosuisse – a Swiss innovation agency, Biogen. and Novartis and received speaker honoraria and/or compensation for serving on advisory boards from Roche, Teva, and Novartis. Ente Ospedaliero Cantonale (employer) received compensation for Chiara Zecca's speaking activities, consulting fees, or research grants from Abbvie, Almirall, Biogen Idec, Bristol Meyer Squibb, Genzyme, Lundbeck, Merck, Novartis, Teva Pharma, and Roche. TAF reports personal fees from DT Axis, Kyoto University Original, MSD and SONY, and a grant from Shionogi, outside the submitted work; In addition, TAF has patents 2020-548587 and 2022-082495 pending, and intellectual properties for Kokoro-app licensed to Mitsubishi-Tanabe. Other authors have no conflicts of interest to declare relevant to the content of this article.

Acknowledgment

We are very grateful to Suvitha Subramaniam for her great help extracting data from the Swiss Multiple Sclerosis Cohort registry for this project.

Abstract (250 words)

In network meta-analysis (NMA), we synthesize all relevant evidence about health outcomes with competing treatments. The evidence may come from randomized clinical trials (RCT) or non-randomized studies (NRS) as individual participant data (IPD) or as aggregate data (AD). We present a suite of Bayesian NMA and network meta-regression (NMR) models allowing for cross-design and cross-format synthesis. The models integrate a three-level hierarchical model for synthesizing IPD and AD into four approaches. The four approaches account for differences in the design and risk of bias (RoB) in the RCT and NRS evidence. These four approaches variously ignoring differences in RoB, using NRS to construct penalized treatment effect priors and bias-adjustment models that control the contribution of information from high RoB studies in two different ways. We illustrate the methods in a network of three pharmacological interventions and placebo for patients with relapsing-remitting multiple sclerosis. The estimated relative treatment effects do not change much when we accounted for differences in design and RoB. Conducting network meta-regression showed that intervention efficacy decreases with increasing participant age. We also re-analysed a network of 431 RCT comparing 21 antidepressants, and we did not observe material changes in intervention efficacy when adjusting for studies' high RoB. We re-analysed both case studies accounting for different study RoB. In summary, the described suite of NMA/NMR models enables inclusion of all relevant evidence while incorporating information on the within-study bias in both observational and experimental data and enabling estimation of individualized treatment effects through the inclusion of participant characteristics.

Key words: real-world evidence, observational studies, randomised controlled trials, risk of bias

1 Introduction

Network meta-analysis (NMA) is a widely used tool to synthesise the available evidence that may vary in design and format (1–3). Evidence may come either from a randomized clinical trial (RCT) or a non-randomized study (NRS); as either individual participant data (IPD) or aggregate data (AD). As heterogeneity is a common attribute of evidence synthesis, many published comparative effectiveness reviews account for covariates that modify the treatment effect in a network meta-regression (NMR) (4,5). The effect of study-level covariates can be modelled using only AD, while IPD is needed to adjust for patient-level covariates to avoid aggregation bias (6) and confounding when NRS are included. The inclusion of these participant characteristics also enables estimating individualized treatment effects.

Matching-Adjusted Indirect Comparison (6–8)(7) and simulated treatment comparison methods (8) have been used to combine evidence from IPD and AD using reweighting techniques and regression models, respectively, to adjust for effect modifiers. However, this adjustment needs to be done separately for each treatment comparison and requires IPD for at least one of each treatment comparison. The performance of these methods has been investigated in two simulation studies. Phillippo et al. found that matching-adjusted indirect comparison performs poorly when its underlying assumptions are violated (9). Remiro-Azócar et al. showed that the current use of simulated treatment comparison method yields often biased estimates (10). Jansen proposed combining IPD and AD in an NMA by integrating the underlying IPD distribution of the AD studies (11). The method was applied initially to binary outcomes and extended to other data types (12). The three-level hierarchical model extends the standard NMR model combining IPD and AD by introducing a new level differentiating between the two formats (11,13–15).

While most published NMAs only synthesize RCT evidence, there is growing interest incorporating non-randomized or real-world evidence in these analyses (16,17). The inclusion of evidence from NRS has many potential advantages, such as better reflected clinical practice realities; the data in follow-up studies are collected over relatively long time periods; and finally, NRS are essential when RCTs are less feasible (e.g., in rare conditions). While RCT evidence is considered to be of lower risk of bias when compared with NRS, a Cochrane review found little evidence that RCTs and NRSs provide different estimates of treatment effect (18). Also, many empirical studies have identified different types of bias possibly present in many RCTs. For example, Schulz et al. (19) found that RCTs with inadequate allocation concealment

or lack of blinding tended to exaggerate the estimated treatment effect and provide biased results. Similarly, Chalmers et al. (20) showed major differences between treatment and control effects in unblinded trials, as well as trials lacking proper randomisation when compared with double-blinded studies. Wood et al. (21) found that the treatment effect estimates of subjective outcomes (outcomes are dependent on judgment from an assessor or patient-reported) were exaggerated for studies with poor allocation concealment or lack of blinding.

Several methods have been proposed for combining various designs in NMA contexts. Three approaches have been proposed to synthesize RCT and NRS evidence (22,23): the first combines studies of different design ignoring their differences (we call this the naïve approach); an alternative is to use NRS evidence to construct penalized treatment effect priors; and a third approach is to add a new level to reflect differences in study designs using a three-level hierarchical model. This last approach requires the network to include several studies on each design which is not the case for most NMAs (23). Dias et al. (24) presented an NMA model that adjusts for the within-study risk of bias (RoB) of RCTs by adding a bias indicator. The bias indicator was assigned a binary value of 0 for low RoB studies; 1 for high RoB studies; and a uniform distribution for studies with unclear RoB. Verde (25) proposed to model the unadjusted and adjusted relative treatment effect simultaneously using a bimodal normal distribution. The model was developed for pairwise meta-analysis.

We extend the two RoB adjustment methods described above by accounting for the uncertainty in each RoB judgment in Dias et al.'s model and by drawing from Verde's approach into NMA (24,25). Then we build a Bayesian cross-NMA/NMR model by integrating the approaches that combine RCT and NRS evidence into the three-level hierarchical model, which combines IPD and AD. This model enables estimating treatment effects for specific subgroups of patients through the inclusion of participant characteristics. Bias-adjusted models can be used to explore the impact of the different levels of bias in RCTs. We will illustrate this by modelling the risk of bias in a network of RCTs with AD comparing various antidepressants.

This work has been done within the HTx Horizon 2020 project. HTx is supported by the European Union, lasting for 5 years from January 2019. The main aim of HTx is to create a framework for the Next Generation Health Technology Assessment (HTA) to support patient-centered, societally oriented, real-time decision-making on access to and reimbursement for health technologies throughout Europe.

2 Examples

We analysed two networks of interventions: one of pharmacological agents in relapsing-remitting multiple sclerosis (RRMS) and another of antidepressant treatments (Figure 1). In both examples, RoB judgements were formulated using the Cochrane RoB tool 1 (26).

Relapsing-Remitting Multiple Sclerosis (RRMS) drugs network: The agents to manage RRMS were compared in systematic reviews of RCTs and NMAs (27,28). We contribute to the methodological literature by analysing the IPD and AD from five RCTs (29–33) and the Swiss Multiple Sclerosis Cohort (SMSC) (34).

We defined the inclusion criteria for patients from the SMSC to be consistent with the RCTs' criteria. We only included people from the SMSC with RRMS treated with any of the three active agents shown in Table 1. Compared with available RCTs, individuals in the SMSC are followed for longer. To avoid immortal time bias, we specified the length and the start of follow-up for each individual (35,36). Since two years was the typical duration of the RCTs we included, we defined cycles of length of two years from when a patient initiated a treatment in SMSC; we recorded their outcome during these two years of follow-up.

To investigate the effectiveness of the treatments in subgroups of people, we explored whether age at the time of treatment initiation modifies the treatment effect. Individuals with RRMS have flare-ups of relapses or symptoms; between these flare-ups, they are free of symptoms (37). Our outcome of interest is relapse at two years of follow-up. We use the odds ratio (OR) to compare treatments. When OR of treatment A vs B ($OR_{AB} = odds_A/odds_B$) is less than 1, treatment A is more effective than treatment B.

Figure 1a, Table 1 and Appendix Table 1-3 summarize the data available, their format and the RoB in each study.

Antidepressants: Our dataset includes AD from 431 RCTs (263 at moderate RoB and 168 at low RoB) comparing 21 antidepressants and placebo (38). The outcome of interest is response to treatment defined as 50% reported reduction in depression symptoms. In the original article, the authors performed a sensitivity analysis by including only low RoB studies in their analysis (38). We re-analysed their dataset by controlling the impact of information from studies at different levels of RoB. (The dataset is available at https://data.mendeley.com/datasets/83rthbp8ys/2.)

3 Methods

We review existing NMR models to combine different data formats—IPD or AD—in this section. We then extend these models by combining the evidence from RCT and NRS in four different ways. Table 2 provides an overview of these four models and Table 3 summarizes the notation used. All models we later introduce are implemented in a new R package called *crossnma* available on CRAN (https://CRAN.R-project.org/package=crossnma). The R code for the analysis of both examples and the antidepressant dataset can be found at the following URL: https://github.com/htx-r/crossnma-theoretical-paper-analysis.

3.1 Synthesizing cross-format data: individual participant data (IPD) and aggregate data (AD)

To combine IPD and AD data into the three-level hierarchical model of network meta-regression, we divided the model into three parts; in the first two parts, the model is set for IPD and AD separately. Next, we present how we combined the evidence from both parts. We describe all models assuming binary outcomes; however, they can be adapted easily to other outcome types, such as time-to-event data, as described by Saramago et al. (39). The NMA models are simply the NMR models without covariate terms.

Part I: Network meta-regression (NMR) model for IPD studies

Assuming y_{ijk} is a binary outcome of participant i in study j under treatment arm k, we place a Bernoulli distribution for y_{ijk} with a probability of an event to occur p_{ijk}. This probability is then linked to the control/treatment effect via a logistic transformation. The study-specific baseline effect u_{jb} is the log-odds in the reference treatment b in that study. The treatment effect δ_{jbk} represents the log odds ratio of treatment k relative to the reference treatment b. Both effects δ_{jbk} and u_{jb} are defined when the participant and mean covariates equal zero.

To estimate subgroup-specific treatment effects, we consider the covariate effect by adding the following three parameters (i) a regression coefficient, β_{0j}, which captures the prognostic effect of the covariate in study j; (ii) a between-study regression coefficient, $\beta^{B}_{1,jbk}$, which quantifies the interaction between the relative treatment effect and the mean covariate value across studies; and (iii) a within-study regression coefficient, $\beta^{W}_{1,jbk}$, which models the treatment-covariate interaction effect at the individual level. The two coefficients β_{0j} and $\beta^{W}_{1,jbk}$ are estimated using the participant-level covariate x_{ijk}, while $\beta^{B}_{1,jbk}$ requires only the

study mean covariate ($\bar{x}_j$) that is often reported in the publication. The term $\beta^B_{1,jbk} - \beta^W_{1,jbk}$ quantifies the discrepancy among the between- and the within-covariate estimates or the aggregation bias (40). In the following, we summarize the likelihood and the parametrization of the model in IPD studies:

$$y_{ijk} \sim Bernoulli(p_{ijk})$$

$$\text{Logit}(p_{ijk}) = \begin{cases} u_{jb} + \beta_{0j} x_{ijk} & \text{if } k = b \\ u_{jb} + \delta_{jbk} + \beta_{0j}\, x_{ijk} + & \text{if } k \neq b \\ \beta^W_{1,jbk} x_{ijk} + (\beta^B_{1,jbk} - \beta^W_{1,jbk})\, \bar{x}_{j.} & \end{cases}$$

Where $j = 1, \dots, ns_{IPD}$, and ns_{IPD} is the total number of IPD studies.

Part II: NMR model for AD studies

We model the published information from each AD study next. For each treatment k in study j, we place a binomial distribution for the corresponding number of events r_{jk} with sample size n_{jk} and probabilities of the event to occur $p_{.jk}$.

$$r_{jk} \sim Bin(p_{.jk},\, n_{jk})$$

$$Logit(p_{.jk}) = \begin{cases} u_{jb} & \text{if } k = b \\ u_{jb} + \delta_{jbk} + \beta^B_{1,jbk}\, \bar{x}_j & \text{if } k \neq b. \end{cases}$$

We incorporate the study-level covariate effect by adding only $\beta^B_{1,jbk}\, \bar{x}_j$. Here, $j = 1 + ns_{IPD}, \dots, ns_{IPD} + ns_{AD}$ where ns_{AD} is the total number of AD studies.

Part III: Combine the evidence from IPD and AD

We combine the relative treatment effects and the between-study regression coefficients from IPD and AD parts via an exchangeable model

$$\delta_{jbk} \sim N(d_{Ak} - d_{Ab}\ , \tau^2\ ), \qquad \beta^B_{1,jbk} \sim N\left(B^B_{1,Ak} - B^B_{1,Ab}, \tau^2_B\right),$$

where $j = 1, \dots, ns_{IPD} + ns_{AD}$.

The within-study regression estimates from only IPD studies ($j = 1, \dots, ns_{IPD}$) are synthesized as

$$\beta^W_{1,jbk} \sim N\left(B^W_{1,Ak} - B^W_{1,Ab}, \tau^2_W\right).$$

Here, A represents the reference treatment in the whole network; therefore, $d_{AA}, B^W_{1,AA}, B^B_{1,AA} = 0$.

Alternatively, a common-effect model can be assumed

$$\delta_{jbk} = d_{Ak} - d_{Ab}, \qquad \beta^{B}_{1,jbk} = B^{B}_{1,Ak} - B^{B}_{1,Ab}, \qquad \beta^{W}_{1,jbk} = B^{W}_{1,Ak} - B^{W}_{2,Ab}.$$

We summarize the model assumptions in Table 4.

We assumed minimally informative priors for $u_{jb}, \beta_{0j} \sim \mathrm{N}(0, 10^2)$ and also for the basic parameters $B^{W}_{1,Ak}, B^{B}_{1,Ak}, d_{Ak} \sim \mathrm{N}(0, 10^2)$. For all heterogeneity parameters, we assigned a uniform distribution $\tau, \tau_B, \tau_W \sim Unif(0,2)$ which allows for difference of log-odds ratios of 2 (or 7.4 of odds ratio) across trials in the treatment and the covariate effect. This change is adequately large on the log scale; hence, the given prior can be considered sufficiently vague.

In all models we present with random treatment effects, we accounted for correlations induced by multi-arm studies using a multivariate distribution as in the standard NMA methods (2). In Appendix 2, we describe how we accounted for multi-arm studies in bias-adjusted model 2.

3.2 Synthesizing cross-design data: Randomized trials and observational data

The model we described in Section 3.1 can be applied to RCT or NRS studies, separately. Next, we describe four different approaches to combine evidence from RCTs and NRSs into the model from Section 3.1.

3.2.1 Unadjusted network meta-regression

Using the simplest approach, we integrate the NRS evidence into the RCT model without differentiation between the two designs. Technically, this means we only need to expand the index of study j to involve both study designs. For IPD, it becomes $j = 1, \ldots, ns_{IPD,RCT} + ns_{IPD,NRS}$ and in AD part, $j = ns_{IPD,RCT} + ns_{IPD,NRS} + 1$,…, $ns_{IPD,RCT} + ns_{IPD,NRS} + ns_{AD,RCT} + ns_{AD,NRS}$.

3.2.2 Using non-randomized studies (NRS) to construct priors for the treatment effects

Using NRS evidence to construct priors for the treatment effects in the RCT model is a two-step approach. In the first step, the (network) meta-regression—with only NRS data—estimates the relative treatment effects with posterior distribution of mean $\tilde{d}^{NRS}_{Ak}$ and variance V^{NRS}_{Ak}. In the second step, the posteriors of NRS results—accounting for possible confounders—are then used as priors for the corresponding basic parameters in the RCT model; $d^{RCT}_{Ak} \sim N(\tilde{d}^{NRS}_{Ak}, V^{NRS}_{Ak})$. Treatment effects not observed in NRS are given vague priors (see part III of Section 3.1). Another possibility when constructing the prior is to use the estimated between-NRS heterogeneity ($\tilde{\tau}^2_{NRS}$) instead of the posterior variance V^{NRS}_{Ak}.

Instead of performing the analysis in two steps, the RCT and NRS synthesis can be conducted simultaneously and seamlessly incorporate the information from NRS in the RCT model.

We can control the potential dominance of NRS evidence (eg, because of the large sample size) on the RCT model by either shifting the NRS means with a bias term ς or by dividing the variance in the prior distribution with a common inflation factor $w,\ 0 < w < 1$; $d_{Ak} \sim N(\tilde{d}_{Ak}^{NRS} + \varsigma, V_{Ak}^{NRS}/w)$. When $w = 1$, NRS evidence is used at face value and when $w \approx 0$, NRS evidence is ignored.

3.2.3 Bias-adjusted model 1

We incorporate judgments about study risk of bias in bias-adjusted model 1 and model 2. The indicator R_j takes binary values 0 (no bias) or 1 (bias) according to a Bernoulli distribution with probability of risk of bias π_j that relates to the study design characteristics. These characteristics are used to summarized the study risk of bias into low, high or unclear. In bias-adjusted model 1, we extend the method introduced by Dias et al. (24) by adding a treatment-specific bias term $\gamma_{2,jbk} R_j$ to the relative treatment effect for both the AD and IPD parts of the model (41). A multiplicative model can also be employed, where treatment effects are multiplied by $\gamma_{1,jbk}^{R_j}$. These bias terms penalize the high RoB studies for potential overestimation or underestimation by adjusting their relative treatment effects. Next, we extend the model from Section 3.1 to adjust for bias.

Part I: NMR model for IPD studies

We model the IPD studies from both designs simultaneously; we differentiate between the designs by including the study-level bias terms. We can add either multiplicative $\gamma_{1,jbk}$ bias effects, additive $\gamma_{2,jbk}$ bias effects, or both (in this case, δ_{jbk} should be dropped from the additive part) as

$$\text{logit}(p_{ijk}) = \begin{cases} u_{jb} + \beta_{0j} x_{ijk} & \text{if } k = b \\ u_{jb} + \overbrace{\delta_{jbk} \gamma_{1,jbk}^{R_j}}^{multiplicative} + \overbrace{\delta_{jbk} + \gamma_{2,jbk} R_j}^{additive} + & \text{if } k \neq b \\ \beta_{0j}\, x_{ijk} + \beta_{1,jbk}^{W} x_{ijk} + \left(\beta_{1,jbk}^{B} - \beta_{1,jbk}^{W}\right) \bar{x}_{j.} & \end{cases} \quad (1)$$

where $j = 1, \ldots, ns_{IPD,RCT} + ns_{IPD,NRS}$.

The bias indicator R_j follows a Bernoulli distribution with a bias probability $\pi_j = P(R_j = 1)$

$$R_j = \begin{cases} 1, & \text{if study } j \text{ has high risk of bias} \\ 0, & \text{otherwise} \end{cases}$$

$$R_j \sim \text{Bernoulli}(\pi_j).$$

Then based on the risk of bias for each study, a different beta distribution is placed for π_j.

$$\pi_j \sim Beta(a_1, a_2).$$

The hyperparameters a_1 and a_2 should be chosen in a way that reflects the risk of bias for each study. The degree of skewness in beta distribution can be controlled by the ratio a_1/a_2 . When a_1/a_2 equals 1 (or $a_1 = a_2$), there is no skewness in the beta distribution (the distribution is reduced to a uniform distribution), which is appropriate for studies with unclear risk of bias. When a_1 is much larger than a_2, the mean of probability of bias (expected value of $\pi_j = a_1/(a_1 + a_2)$) is closer to 1 as the study will have a high bias probability, which leads to a 'major' bias adjustment.

Alternatively, we can use the study characteristics $\boldsymbol{z_j} = (z_{1,j}, z_{2,j}, \dots, z_{m,j})$ (eg, the concealment of the study) to predict π_j through a logistic transformation as follows

$$logit(\pi_j) = e + \boldsymbol{f^T z_j}.$$

where $\boldsymbol{f^T} = (f_1, \dots, f_m)$ is a vector of covariate effect on the odds ratio of bias and e is the overall odds of bias. The superscript **T** transposes the vector. A minimally informative prior is located in the regression coefficients $e, \boldsymbol{f^T} \sim N(0, 10^2)$.

We alternatively describe the logistic model with additive bias effect in equation (1) by the following parametrisation

$$\text{logit}(p_{ijk}) = \begin{cases} u_{jb} + \beta_{0j} x_{ijk} & \text{if } k = b \\ u_{jb} + (1 - R_j)\delta_{jbk} + R_j \delta_{jbk}^{bias} + & \text{if } k \neq b \\ \beta_{0j}\, x_{ijk} + \beta_{1,jbk}^{W} x_{ijk} + \left(\beta_{1,jbk}^{B} - \beta_{1,jbk}^{W}\right) \bar{x}_j. & \end{cases} \quad (2)$$

where $\delta_{jbk}^{bias} = \delta_{jbk} + \gamma_{jbk}$.

Part II: NMR model for AD studies

Similarly, we add the two bias terms to model the summary data.

$$Logit(p_{.jk}) = \begin{cases} u_{jb} & \text{if } k = b \\ u_{jb} + \overbrace{\delta_{jbk}\gamma_{1,jbk}^{R_j}}^{multiplicative} + \overbrace{\delta_{jbk} + \gamma_{2,jbk} R_j}^{additive} + \beta_{1,jbk}^{B}\, \bar{x}_j & \text{if } k \neq b \end{cases} \quad (3)$$

where $j = ns_{IPD,RCT} + ns_{IPD,NRS} + 1, \dots, ns_{IPD,RCT} + ns_{IPD,NRS} + ns_{AD,RCT} + ns_{AD,NRS}$.
Again, when multiplicative and additive parts are both considered in the model, the term δ_{jbk} needs to be removed from the additive term.

Other parametrisation of the logistic model with additive bias effect in equation (3) is

$$Logit(p_{.jk}) = \begin{cases} u_{jb} & \text{if } k = b \\ u_{jb} + (1 - R_j)\delta_{jbk} + R_j \delta_{jbk}^{bias} + \beta_{1,jbk}^{B}\, \bar{x}_j & \text{if } k \neq b. \end{cases} \quad (4)$$

Part III: Combine the evidence from IPD and AD

In addition to the covariates' effects and the treatment effects, here we also combine the multiplicative and the additive treatment-specific bias effects across studies by assuming they are either exchangeable ($\gamma_{1,jbk} \sim N\left(g_{1,bk}, \tau_{1,\gamma}^2\right)$, $\gamma_{2,jbk} \sim N(g_{2,bk}, \tau_{2,\gamma}^2)$) or common ($\gamma_{1,jbk} = g_{1,bk}$ and $\gamma_{2,jbk} = g_{2,bk}$). We set priors for the between-study standard deviation again as $\tau_{1,\gamma}, \tau_{2,\gamma} \sim \text{Unif}(0,2)$.

For the other parameterisation in equations (2) and (4), the bias-adjusted relative treatment effect δ_{jbk}^{bias} can be assumed exchangeable across studies

$$\delta_{jbk}^{bias} \sim N\left(g_{bk} + d_{Ak} - d_{Ab}, \frac{\tau^2}{q_j}\right)$$

or common as

$$\delta_{jbk}^{bias} = g_{bk} + d_{Ak} - d_{Ab}.$$

In this case, instead of assigning prior to the between-study heterogeneity in bias effect τ_γ, we model the RoB weight $q_j = \tau^2/(\tau^2 + \tau_\gamma^2)$ for each study. The quantity represents the proportion of the between-study heterogeneity that is not explained by accounting for risk of bias. These weights take values between 0 and 1, $0 < q_j < 1$, and they are either given fixed values (as Spiegelhalter and Best proposed (42)) or assigned a prior to let the data estimate them, $q_j \sim Beta(v, 1)$(as Verde assumed (25)). The values of v determine the extent studies at high risk of bias will be down-weighted on average. Setting $v = 1$ gives $E\left(q_j\right) = v/(v + 1) = 0.5$, which means that high risk of bias studies will be penalized by 50% on average.

Dias et al. (24) proposed to model the mean bias effect ($g_{1,bk}$, $g_{2,bk}$) based on the compared treatments. One approach is to assume a common mean bias for studies that compare active treatments with an inactive treatment (placebo, standard, or no treatment)

$$g_{m,bk} = \begin{cases} g_m & \text{if } b \text{ is inactive treatment} \\ 0 & \text{if } b \text{ and } k \text{ are active treatments} \end{cases}$$

where $m = \{1,2\}$.

In this case, the mean bias effect cancels out contrasts for comparing two active treatments. When exchangeable bias parameters are used, active vs active comparisons have an expected bias effect of zero with uncertainty the common bias-heterogeneity parameters $\tau_{1,\gamma}^2$, $\tau_{2,\gamma}^2$ for multiplicative and additive, respectively.

Instead of assuming zero bias in active vs active comparison, we could assume a common and fixed bias effect g_m^{act} :

$$g_{m,bk} = \begin{cases} g_m & \text{if } b \text{ is inactive treatment} \\ (-1)^{dir_{bk}}\, g_m^{act} & \text{if } b \text{ and } k \text{ are active treatments} \end{cases}$$

The direction of bias (dir_{bk}) varies by the comparison type and should be defined in the data. The bias in active vs inactive comparisons will favour the active treatment. However, the direction of bias is less clear in studies that compare active treatments with each other. The direction of bias could be linked to other types of bias, such as 'optimism bias'–a bias favouring the newest treatment. In this case, the direction of bias in each active vs active comparison is set to be either 0, meaning that bias favours b over k; or 1, meaning that k is favoured to b. We could also follow a data-driven approach and assign the bias direction a Bernoulli distribution $dir \sim \text{Bernoulli}(p_{dir})$ where the probability of b to be favoured over k, p_{dir}, is given a beta distribution $p_{dir} \sim Beta(a_3, a_4)$. The shape of this beta distribution is characterized by a_3 and a_4. When a_3 is set a value less than a_4, the study is more likely to be favouring b over k.

3.2.4 Bias-adjusted model 2

Extending the model initially introduced by Verde (25), bias-adjusted model 2 parametrises the relative treatment effect using a bimodal normal distribution that involves the bias parameters (25). We define the bias-adjusted relative treatment effect θ_{jbk} as follows in both parts of the NMR.

Part I: NMR model for IPD studies

$$\text{Logit}(p_{ijk}) = \begin{cases} u_{jb} + \beta_{0j} x_{ijk} & \text{if } k = b \\ u_{jb} + \theta_{jbk} + \beta_{0j}\, x_{ijk} + & \text{if } k \neq b \\ \beta_{1,jbk}^{W} x_{ijk} + \left(\beta_{1,jbk}^{B} - \beta_{1,jbk}^{W}\right) \bar{x}_j. & \end{cases}$$

Part II: NMR model for AD studies

We also add the bias adjustment term to AD part

$$Logit(p_{.jk}) = \begin{cases} u_{jb} & \text{if } k = b \\ u_{jb} + \theta_{jbk} + \beta_{1,jbk}^{B}\, \bar{x}_j & \text{if } k \neq b. \end{cases}$$

Part III: Combine the evidence from IPD and AD

The coefficients from the covariates effect and treatment effects are combined as in the previous models. We additionally combine the bias-adjusted relative treatment effect θ_{jbk} via exchangeable model with a mixture of two normal distributions

$$\theta_{jbk} \sim (1-\pi_j) N(d_{Ak} - d_{Ab}, \tau^2) + \pi_j N(d_{Ak} - d_{Ab} + \gamma_{jbk}, \tau^2 + \tau_\gamma^2).$$

Assuming a common-effect model we can alternatively summarize these relative effects

$$\theta_{jbk} = (1-\pi_j)(d_{Ak} - d_{Ab}) + \pi_j (d_{Ak} - d_{Ab} + \gamma_{jbk}) = d_{Ak} - d_{Ab} + \pi_j \gamma_{jbk}.$$

This model adjusts the relative treatment effect by a bias effect that is proportional to the bias probability in each study. The bias parameters γ_{jbk} across studies are assigned either the exchangeable- or common-effect model and then the mean bias effects g_{bk} are also combined across comparisons.

Following what we describe in Section 3.2.3, the between-study standard deviation τ_γ can also be modelled in two different ways. We set a prior either for $\tau_\gamma \sim \text{Unif}(0,2)$ or for $q_j \sim Beta(v, 1)$ where $q_j = \tau^2/(\tau^2 + \tau_\gamma^2)$ represents the RoB weight for each study. However, choosing the prior for q_j could be more meaningful in practice as v represents the discounting in study weight. All other syntheses are performed as outlined for bias-adjusted model 1 in Section 3.2.3.

4 Implementation of the models and results

We implemented the models in a Bayesian setting using Just Another Gibbs Sampler (43) software through R (44). For all models, we ran two chains each for 100 000 iterations, discarded the first 40 000 samples, and thinned by 1. We examined the convergence of chains on each parameter by either visually inspecting the trace plots or checking the Gelman-Rubin statistic, $\hat{R}$, which measures the agreement between the within- and between-chains of MCMC; it should be approximately 1 when the chain converges properly. We evaluated model performance using the deviance information criterion (DIC), with the preferable model the one with the lowest DIC values (45). From here onwards, point estimates refer to posterior medians. Of note, the study-specific ORs in Figure 2 were calculated within a frequentist framework, and the lines represent confidence intervals. We analyzed IPD studies with the *glm()* function in R and AD studies with the *metabin()* function (from *meta* package).

4.1 Immunomodulatory agents in RRMS

We conducted NMA and NMR assuming a common treatment effect across studies (the small number of studies did not allow efficient estimation of heterogeneity). We included age as a covariate in the NMR model which was centred around mean age 38 to improve convergence. We also assumed a common age effect across studies. For the IPD part of the models, we set the within- and between-study age effects equal: $\beta^{W}_{1,jbk} = \beta^{B}_{1,jbk}$. The little variation in mean participant age across the included studies (see Table 1) renders the estimation of $\beta^{B}_{1,jbk}$. In bias-adjusted models 1 and 2, we assigned two different informative prior distributions for the bias probability π_j: a $Beta(100,1)$ for high RoB studies and $Beta(1,100)$ for low RoB studies (see Appendix Figure 2 and Table 1). We assumed additive bias effects and combined them across studies into a common parameter. The direction of bias was assumed to favour the active treatment rather placebo in RCTs and any other treatment over glatiramer acetate in the SMSC since it is the oldest treatment. We set placebo as a network reference for all analyses except when using NRS information as a prior; in that case, natalizumab was used as the reference treatment.

We first analysed the data using the SMSC data to construct priors for the treatment effects. The posterior distributions of the logORs were $d_{DF\ vs\ N} \sim N(-0.01, 0.2)$ (dimethyl fumarate vs natalizumab) and $d_{GA\ vs\ N} \sim N(1.56, 0.33)$ (glatiramer acetate vs natalizumab). The basic parameter of placebo vs natalizumab (not observed in the cohort) was assigned an approximately uninformative prior ($d_{P\ vs\ N} \sim N(0, 10^2)$). In Appendix Figure 1, we present the results when these posteriors were used as (discounted) priors in the NMA of the RCT data assuming different values of w. Only the estimated effect of glatiramer acetate vs natalizumab changed slightly when incorporating the non-randomized evidence because the SMSC has a much smaller sample size (n =206) than all RCTs together (n =3 891).

Figure 2 and Appendix Tables 4–7 show the NMA ORs and the corresponding 95% credible intervals (CrI) using no adjustment and bias-adjusted models 1 and 2. The adjustment for the different bias effects did not materially change the estimated ORs. The small change we observed for glatiramer acetate in bias-adjusted models can be attributed to the high risk of bias in Bornstein and Johnson studies (32,33).

For bias-adjusted model 1, the bias effect $exp(g)$ was estimated 0.705 (95% CrI: 0.198–1.459). The OR of the active treatments when compared with placebo in high RoB studies are on average 0.705 times the OR in low RoB studies, yet the uncertainty is very large. In the

bias-adjusted model 2, $exp(g)$ was more precisely estimated at 0.323 (95% CrI: 0.126–0.821). This means that on average high RoB studies tend to overestimate the efficacy of the active treatments. We investigated the convergence of the model parameters in Appendix Figure 4 and Appendix Table 8. The bias parameter g estimated from the bias-adjusted model 1 has a slightly poor convergence when compared with other parameters.

We incorporated the effect of age in bias-adjusted model 1; Figure 3 presents the NMR ORs of active vs placebo for various age values. The estimated age coefficient $exp(B)$ was 0.984 (95% CrI: 0.264–1.935) suggesting that for an increase in age by one year the ORs of each treatment vs placebo decreases by 1-0.984.

Table 6 summarizes the DIC values for the unadjusted analysis and the bias-adjusted models 1 and 2. Because bias-adjusted model 2 has a lowest DIC (DIC for IPD model=90365 and for AD model =158), it is preferred over other models. The model that uses NRS evidence as a prior has DIC for IPD model 87144 and for AD model 142. This model was excluded from the comparison because it only uses RCT data.

4.2 Antidepressants for major depression

We conducted an NMA assuming a random treatment effect across studies. For bias-adjusted models 1 and 2, we used additive bias effects and combined them across studies assuming random-effects. The bias probability π_j of moderate and low RoB studies was given prior distributions $Beta(20,1)$ and $Beta(1,20)$, respectively (see Appendix Figure 3). When we set the direction of bias in studies comparing an active drug to placebo, we assumed mean bias g^P, and the antidepressant was assumed the favoured treatment; then in active vs active comparisons, we assumed bias g^{act}, and the sponsored treatment was assumed the favoured treatment. In other cases, the mean bias was set to zero. We performed a sensitivity analysis to investigate the robustness of the results with less informative prior distributions for the bias probability π_j in both bias models $Beta(10,1)$ and $Beta(1,10)$ for studies at moderate and low RoB, respectively.

Table 5 shows the estimates of bias effect parameters using the bias-adjusted models 1 and 2. The results suggest that moderate RoB studies do not provide different estimates of the effectiveness of the active interventions versus placebo, whereas the effects of sponsored treatments are overestimated on average. In bias-adjusted model 1, the OR of the active

treatment (sponsored) against active (not sponsored) in low RoB studies are on average 1.186 times the OR in high RoB studies. We also fitted the bias-adjusted models 1 and 2 by re-parametrising the heterogeneity τ_γ using the weights q_j .We set $q_j = 1$ for studies at low RoB and $q_j \sim Beta(1/3,1)$ for moderate RoB studies which reduces their weight on average by 25% or $q_j \sim Beta(4,1)$) for 80% weight reduction. The results do not materially change.

As expected, the bias indicator (R_j) was estimated to be 1 on average for studies with moderate RoB and 0 for studies with low RoB. Also, the convergence of bias parameters was good in the antidepressants example because of the large number of studies (see Appendix Figure 5 and Appendix Figure 6).

Figure 4 presents the resulting OR and 95% CrI for the adjusted and unadjusted models. Controlling for the information from the moderate RoB studies scarcely changed the effects of active drugs vs placebo. Using less informative priors for the bias probability and for between-study heterogeneity in the bias effect did not materially change these conclusions (Appendix Figure 7 and Appendix Figure 8). The estimate of between-study heterogeneity in treatment effect was 0.210 (95% CrI: 0.169–0.251) in unadjusted model, which decreased when bias-adjusted model 1 was applied to 0.176 (95% CrI: 0.089–0.236) and the estimate in bias-adjusted model 2 was 0.213 (95% CrI: 0.147–0.291). The differences in the estimates of between-study heterogeneity are minor and their CrI overlap to a large extent.

We compared the bias-adjusted models 1 and 2 and unadjusted model by calculating the DIC; their values are reported in Table 6. The bias-adjusted model 1 performs better than the unadjusted and bias-adjusted models 2.

5 Discussion

We introduced a suite of Bayesian NMA and NMR models to synthesize evidence that comes from different study designs and in different data formats. We extended the three-level hierarchical model for combining IPD and AD with four models incorporating RCT and NRS evidence. The first model ignores differences in design and RoB between studies; the second uses NRS to construct discounted treatment effect priors; and two models adjust for the risk of bias in each study. The bias effect can be multiplied or added to the relative treatment effect. The multiplicative bias is more likely to describe better cases of selective outcome reporting. In such cases, results from studies with small true effects are magnified considerably to "cross the significance line" while results from studies with large true effects are exaggerated only a bit or not at all.

We implemented the four NMA/NMR models in a dataset comparing treatments for RRMS patients. The estimated treatment effects were consistent, irrespective of the model used. When age was included as a covariate, the efficacy of active treatments relative to placebo decreased with increasing age. In other words, all active treatments become less effective for older patients, which aligns with previous findings (46,47). We also illustrated the bias-adjusted models in a network of AD from RCTs on antidepressants. The results from sponsored drug arms in head-to-head studies tended to be larger than those in non-sponsored arms. In the original analysis, Cipriani et al. (38) did not detect any impact of sponsoring in the estimated efficacy of the antidepressants. Note, however, that our bias-adjusted models estimate the interaction between risk of bias and sponsoring and hence it is possible that sponsoring plays a role in modifying the treatment effect only in studies with moderate risk of bias.

Our methods tackle the bias issue at the quantitative synthesis stage. However, there are two issues to consider when such analyses are conducted. First, empirical evidence has shown that the treatment effects are often exaggerated in high risk of bias studies (48). In these cases, one can employ diagnostics to evaluate the impact of such large study results (49–51) and then fit models that decrease the impact of those studies either by employing non-normal random effect distributions (52) or by shrinking the relative treatment effects towards equivalence (53,54). Second, the bias (for NRS, in particular) should also be mitigated at study design and when interpreting results. In their comprehensive framework, Sarri et al. (55) proposed seven steps outlining how to combine RCT and NRS data in NMA. They proposed different considerations for interpreting findings, suggesting a way that reflects the differences in evidence type. Their framework suggests a certain critical assessment of NRS, which can be used in our bias-adjusted models.

Some limitations of our proposed models need to be acknowledged. First, the bias-adjusted models require several studies at different levels of RoB. In the absence of many studies, strong assumptions can be imposed on bias parameters via informative priors. Our first example of RRMS only included six studies; we assigned highly informative beta distributions to the bias probability. We used less informative priors in the case of antidepressants' network because many studies were available. Second, the results of the analysis can be sensitive to the prior assumptions in model parameters. For this reason, sensitivity analyses should be conducted to investigate the robustness of the estimates, using different priors if possible. Sensitivity to prior distributions is particularly important for the probability of bias and the covariate effect parameters. In our examples, we found that bias-adjusted model 2 was more sensitive to the prior assigned to the bias probability when compared with bias-adjusted model.

Third, choosing down-weighting parameters for the model that uses the NRS data to construct prior information is not straightforward. However, Efthimiou et al. (23) outlined different considerations to guide this choice.

Finally, the estimated treatment effect can be influenced by the sample size of the study when reporting bias is suspected or for other reasons associated with small study effects. Hence, a study can overestimate the treatment effect for reasons related to its sample size and/or a high RoB. In a hierarchical random-effects models study-specific estimates from small studies tend to be pulled towards the overall mean, and hence overestimation of treatment effects in small studies tends to be less of a problem (53). However, we recommend that the presence of small-study effects are routinely checked before conducting any synthesis. If there is no strong evidence of small-study effects, bias-adjusted models 1 or 2 can be applied.

To implement the proposed models, there are further worthy considerations. These include performing a comprehensive systematic review to identify relevant RCTs and NRSs (following the framework introduced by Sarri et al. (55)). In our RRMS example, we included RCTs identified in a previous systematic review with available IPD (27,28) and observational data from the SMSC. For clinically-relevant results after analysis, more data needs to be included to apply our methods to an extended network of all drugs used to treat patients with RRMS, such as presented by Jenkins et al. (56). In their review, Jenkins et al. showed how including NRS data in the synthesis model increased the between-study heterogeneity and therefore the uncertainty around the effect estimates. By accounting for potential effect modifiers and differences in RoB, other studies can investigate whether our models explain large between-study heterogeneity.

Combining individual and aggregate data has two key advantages when compared with analysing aggregate data solely. First, aggregate data studies contribute only to estimating interactions between mean values of effect modifiers and treatment, yet individual data studies account for interactions at the individual patient-level, thus avoiding ecological bias. Second, individual data adjust for prognostic factors and covariates that predict the outcome and the course of the disease regardless of the assigned treatment (46). Adjusting for prognostic factors is desirable (57) in order to improve the interpretation and the external validity of the findings (58); enhance the precision of the estimated treatment effects (59); and correct potential imbalance in baselines after randomisation (60).

Incorporating NRS evidence into NMA models that traditionally only include RCTs is increasingly important in several clinical research settings, such as when conducting RCTs are less feasible for rare conditions. A recent scoping review of methods that combine RCT and

NRS in NMA (61) reveals that unadjusted synthesis is the most popular approach, probably for its ease of use. The unadjusted analysis, however, can be considered as an initial step but not the primary analysis, as it ignores the differences in design and RoB. Accounting for within-study bias in both observational and experimental data, our suite of models offers a viable alternative. Our approach also allows estimating individualized treatment effects through the inclusion of participant characteristics.

HIGHLIGHTS

What is already known?

The evidence in network meta-analysis (NMA) typically comes from randomized clinical trials (RCT) where aggregate data (AD) are extracted from published reports. Retrieving individual participant data (IPD) allows considering participant covariates to explain some of the heterogeneity/inconsistency in the network and identify effect modifiers. Additionally, evidence from non-randomized studies (NRS) reflects the reality in clinical practice and bridges the efficacy-effectiveness gap.

What is new?

This paper describes a Bayesian suite for evidence synthesis which extends and integrates four different approaches that combine RCT and NRS evidence into a three-level hierarchical model for the synthesis of IPD and AD. We call this suite a cross-NMA/NMR model since it enables cross-design and cross-format synthesis.

Potential impact for *Research Synthesis Methods* readers outside the authors' field.

By describing and demonstrating the cross-NMA/NMR suite of models, we hope to facilitate the inclusion of all relevant evidence that comes from multiple sources. Synthesis of all sources of evidence and formats of data, will increase power and relevance of NMA results.

Data Availability Statement

The models we introduce in this paper are implemented in a new R package called *crossnma*, available on CRAN (https://CRAN.R-project.org/package=crossnma). The R code for the analysis of both examples and the antidepressant dataset can be found at the following URL: https://github.com/htx-r/crossnma-theoretical-paper-analysis.

References

1. Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ. 2005 Oct 15;331(7521):897–900.

2. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med. 2004 Oct 30;23(20):3105–24.

3. Efthimiou O, Debray TPA, van Valkenhoef G, Trelle S, Panayidou K, Moons KGM, et al. GetReal in network meta-analysis: a review of the methodology. Res Synth Methods. 2016 Sep;7(3):236–63.

4. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions. 2nd ed. Chichester (UK): John Wiley & Sons; 2019.

5. Borenstein M, Hedges LV, Higgins JPT, Rothstien HR. Introduction to Meta-An alysis [Internet]. John Wiley & Sons, Ltd; 2009 [cited 2022 Feb 4]. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470743386.ch1

6. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd ed. Philadelphia: Wolters Kluwer Health; 2008.

7. Signorovitch JE, Wu EQ, Yu AP, Gerrits CM, Kantor E, Bao Y, et al. Comparative effectiveness without head-to-head trials: a method for matching-adjusted indirect comparisons applied to psoriasis treatment with adalimumab or etanercept. Pharmacoeconomics. 2010;28(10):935–45.

8. Ishak KJ, Proskorovsky I, Benedict A. Simulation and matching-based approaches for indirect comparison of treatments. Pharmacoeconomics. 2015 Jun;33(6):537–49.

9. Phillippo DM, Dias S, Ades AE, Welton NJ. Assessing the performance of population adjustment methods for anchored indirect comparisons: A simulation study. Stat Med. 2020 Dec 30;39(30):4885–911.

10. Remiro-Azócar A, Heath A, Baio G. Methods for population adjustment with limited access to individual patient data: A review and simulation study. Res Synth Methods. 2021 Nov;12(6):750–75.

11. Jansen JP. Network meta-analysis of individual and aggregate level data. Res Synth Methods. 2012 Jun;3(2):177–90.

12. Phillippo DM, Dias S, Ades AE, Belger M, Brnabic A, Schacht A, et al. Multilevel network meta-regression for population-adjusted treatment comparisons. J R Stat Soc Ser A Stat Soc. 2020 Jun;183(3):1189–210.

13. Saramago P, Sutton AJ, Cooper NJ, Manca A. Mixed treatment comparisons using aggregate and individual participant level data. Stat Med. 2012 Dec 10;31(28):3516–36.

14. Donegan S, Williamson P, D'Alessandro U, Garner P, Smith CT. Combining individual patient data and aggregate data in mixed treatment comparison meta-analysis: Individual patient data may be beneficial if only for a subset of trials. Stat Med. 2013 Mar 15;32(6):914–30.

15. Thom HHZ, Capkun G, Cerulli A, Nixon RM, Howard LS. Network meta-analysis combining individual patient and aggregate data from a mixture of study designs with an application to pulmonary arterial hypertension. BMC Med Res Methodol. 2015 Apr 12;15:34.

16. Bell H, Wailoo A, Hernandez Alava M, Grieve R, Neves De Faria RI, Gibson L, et al. The use of real world data for the estimation of treatment effects in NICE decision making. NICE Decision Support Unit; 2016.

17. Food and Drug Administration (FDA). Framework for FDA's Real-World Evidence Program. FDA. 2018;

18. Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. Cochrane Database Syst Rev. 2014 Apr 29;(4):MR000034.

19. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA. 1995 Feb 1;273(5):408–12.

20. Chalmers TC, Celano P, Sacks HS, Smith H. Bias in treatment assignment in controlled clinical trials. N Engl J Med. 1983 Dec 1;309(22):1358–61.

21. Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ. 2008 Mar 15;336(7644):601–5.

22. Schmitz S, Adams R, Walsh C. Incorporating data from various trial designs into a mixed treatment comparison model. Stat Med. 2013 Jul 30;32(17):2935–49.

23. Efthimiou O, Mavridis D, Debray TPA, Samara M, Belger M, Siontis GCM, et al. Combining randomized and non-randomized evidence in network meta-analysis. Stat Med. 2017 15;36(8):1210–26.

24. Dias S, Welton NJ, Marinho VCC, Salanti G, Higgins JPT, Ades AE. Estimation and adjustment of bias in randomized evidence by using mixed treatment comparison meta-analysis. Journal of the Royal Statistical Society. 2010;173:613–29.

25. Verde PE. A bias-corrected meta-analysis model for combining, studies of different types and quality. Biom J. 2021 Feb;63(2):406–22.

26. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011 Oct 18;343:d5928.

27. Tramacere I DGC, Salanti G DR, Filippini G. Immunomodulators and immunosuppressants for relapsing-remitting multiple sclerosis: a network meta-

analysis. Cochrane Database of Systematic Reviews 2015 [Internet]. CD011381(9). Available from: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD011381.pub2/full

28. Giovannoni G, Lang S, Wolff R, Duffy S, Hyde R, Kinter E, et al. A Systematic Review and Mixed Treatment Comparison of Pharmaceutical Interventions for Multiple Sclerosis. Neurol Ther. 2020 Dec;9(2):359–74.

29. Polman CH, O'Connor PW, Havrdova E, Hutchinson M, Kappos L, Miller DH, et al. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. N Engl J Med. 2006 Mar 2;354(9):899–910.

30. Fox RJ, Miller DH, Phillips JT, Hutchinson M, Havrdova E, Kita M, et al. Placebo-controlled phase 3 study of oral BG-12 or glatiramer in multiple sclerosis. N Engl J Med. 2012 Sep 20;367(12):1087–97.

31. Gold R, Kappos L, Arnold DL, Bar-Or A, Giovannoni G, Selmaj K, et al. Placebo-controlled phase 3 study of oral BG-12 for relapsing multiple sclerosis. N Engl J Med. 2012 Sep 20;367(12):1098–107.

32. Bornstein MB, Miller A, Slagle S, Weitzman M, Crystal H, Drexler E, et al. A pilot trial of Cop 1 in exacerbating-remitting multiple sclerosis. N Engl J Med. 1987 Aug 13;317(7):408–14.

33. Johnson KP, Brooks BR, Cohen JA, Ford CC, Goldstein J, Lisak RP, et al. Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis: results of a phase III multicenter, double-blind, placebo-controlled trial. 1995. Neurology. 1995;57(12 Suppl 5):S16-24.

34. Disanto G, Benkert P, Lorscheider J, Mueller S, Vehoff J, Zecca C, et al. The Swiss Multiple Sclerosis Cohort-Study (SMSC): A Prospective Swiss Wide Investigation of Key Phases in Disease Evolution and New Treatment Options. PLoS ONE. 2016;11(3):e0152347.

35. Suissa S. Immortal Time Bias in Pharmacoepidemiology. American Journal of Epidemiology. 2008 Feb 15;167(4):492–9.

36. Lévesque LE, Hanley JA, Kezouh A, Suissa S. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. BMJ. 2010 Mar 12;340:b5087.

37. Goldenberg MM. Multiple Sclerosis Review. P T. 2012 Mar;37(3):175–84.

38. Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. Lancet. 2018 Apr 7;391(10128):1357–66.

39. Saramago P, Chuang LH, Soares MO. Network meta-analysis of (individual patient) time to event data alongside (aggregate) count data. BMC Med Res Methodol. 2014 Sep 10;14:105.

40. Riley RD, Lambert PC, Staessen JA, Wang J, Gueyffier F, Thijs L, et al. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. Statistics in Medicine. 2008;27(11):1870–93.

41. Turner RM, Spiegelhalter DJ, Smith GCS, Thompson SG. Bias modelling in evidence synthesis. J R Stat Soc Ser A Stat Soc. 2009 Jan;172(1):21–47.

42. Spiegelhalter DJ, Best NG. Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. Stat Med. 2003 Dec 15;22(23):3687–709.

43. Plummer M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. 2003;

44. RStudio Team. RStudio: Integrated Development Environment for R. RStudio, Inc [Internet]. 2019; Available from: http://www.rstudio.com/

45. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2002;64(4):583–639.

46. Chalkou K, Steyerberg E, Egger M, Manca A, Pellegrini F, Salanti G. A two-stage prediction model for heterogeneous effects of treatments. Stat Med. 2021 Sep 10;40(20):4362–75.

47. Pellegrini F, Copetti M, Bovis F, Cheng D, Hyde R, de Moor C, et al. A proof-of-concept application of a novel scoring approach for personalized medicine in multiple sclerosis. Mult Scler. 2019 May 30;1352458519849513.

48. Savović J, Turner RM, Mawdsley D, Jones HE, Beynon R, Higgins JPT, et al. Association Between Risk-of-Bias Assessments and Results of Randomized Trials in Cochrane Reviews: The ROBES Meta-Epidemiologic Study. Am J Epidemiol. 2018 May;187(5):1113–22.

49. Petropoulou M, Salanti G, Rücker G, Schwarzer G, Moustaki I, Mavridis D. A forward search algorithm for detecting extreme study effects in network meta-analysis. Stat Med. 2021 Nov 10;40(25):5642–56.

50. Noma H, Gosho M, Ishii R, Oba K, Furukawa TA. Outlier detection and influence diagnostics in network meta-analysis. Res Synth Methods. 2020 Nov;11(6):891–902.

51. Viechtbauer W, Cheung MWL. Outlier and influence diagnostics for meta-analysis. Res Synth Methods. 2010 Apr;1(2):112–25.

52. Lee KJ, Thompson SG. Flexible parametric models for random-effects distributions. Stat Med. 2008 Feb 10;27(3):418–34.

53. Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D. The BUGS Book: A Practical Introduction to Bayesian Analysis [Internet]. Chapman and Hall/CRC; 2012. Available from: https://doi.org/10.1201/b13613

54. Efthimiou O, White IR. The dark side of the force: Multiplicity issues in network meta-analysis and how to address them. Res Synth Methods. 2020 Jan;11(1):105–22.

55. Sarri G, Patorno E, Yuan H, Guo JJ, Bennett D, Wen X, et al. Framework for the synthesis of non-randomised studies and randomised controlled trials: a guidance on conducting a systematic review and meta-analysis for healthcare decision making. BMJ Evid Based Med. 2020 Dec 9;bmjebm-2020-111493.

56. Jenkins DA, Hussein H, Martina R, Dequen-O'Byrne P, Abrams KR, Bujkiewicz S. Methods for the inclusion of real-world evidence in network meta-analysis. BMC Med Res Methodol. 2021 Oct 9;21(1):207.

57. Harrell F. RCT Analyses With Covariate Adjustment. https://www.fharrell.com/post/covadj/#disqus_thread. 2020

58. Hauck WW, Anderson S, Marcus SM. Should we adjust for covariates in nonlinear regression analyses of randomized trials? Control Clin Trials. 1998 Jun;19(3):249–56.

59. Robinson LD, Jewell N P. Some surprising results about covariate adjustment in logistic regression models. International Statistical Review. 1991;59(2):227–40.

60. Steyerberg EW, Bossuyt PM, Lee KL. Clinical trials in acute myocardial infarction: should we adjust for baseline characteristics? Am Heart J. 2000 May;139(5):745–51.

61. Zhang K, Arora P, Sati N, Béliveau A, Troke N, Veroniki AA, et al. Characteristics and methods of incorporating randomized and nonrandomized evidence in network meta-analyses: a scoping review. J Clin Epidemiol. 2019;113:1–10.

Tables

Table 1 Study characteristics and assigned priors for bias probability of the network of treatments for the relapsing-remitting multiple sclerosis in Figure 1a.

IPD; Individual Participant Data, AD; Aggregate Data, RCT; Randomized Clinical Trial,

NRS; Non-Randomised Study

Study	Treatments	Number of patients with at least one relapse in two years	Sample size	Design & data formal	Risk of bias (RoB)	Mean age	Distribution of bias probability π_j
AFFIRM (29)	Natalizumab, Placebo	359	939	RCT IPD	low	36	Beta(1,100)
CONFIRM (30)	Dimethyl fumarate, Glatiramer acetate, Placebo	451	1417	RCT IPD	low	37	Beta(1,100)
DEFINE (31)	Dimethyl fumarate, Placebo	394	1234	RCT IPD	low	39	Beta(1, 100)
Swiss Multiple Sclerosis Cohort (34)	Dimethyl fumarate, Glatiramer acetate, Natalizumab	44	206	NRS IPD	high	46	Beta(100,1)
Bornstein (32)	Glatiramer acetate, Placebo	30	50	RCT AD	high	34	Beta(100,1)
Johnson (33)	Glatiramer acetate, Placebo	186	251	RCT AD	high	30	Beta(100,1)

Table 2 Overview of the presented models allowing for cross-design and cross-format synthesis in network meta-regression

	Unadjusted analysis	Using NRS to form a prior distribution	Bias-adjusted model 1	Bias-adjusted model 2
Accounting for RoB from RCT and NRS	RoB is not considered.	The NRS evidence is shifted and/or down-weighted using the parameters ς and w, respectively. The RoB in the RCT is not considered.	For high RoB studies (NRS or RCT), the model shifts/multiplies the relative treatment effects by γ_{jbk} and/or downweighs the study contribution when the estimates are combined. The method differentiates NRS evidence from RCT by setting relatively greater bias probability (π_j) for NRS compared to RCT.	The model adjusts the relative treatment effects by γ_{jbk} where the adjustment is proportional to the bias probability of the study. It allows also to downweigh the study contribution through $\tau_\gamma{}^1$ or $q_j{}^2$. The bias probability (π_j) can be assumed greater for NRS compared to RCT.
Key model parameters	Relative treatment effect δ_{jbk}. Covariate effect β_{0j}. Within-study covariate-treatment interaction ($\beta_{1,jbk}^W$). Between-study covariate-treatment interaction ($\beta_{1,jbk}^B$).	Same as unadjusted analysis.	Same as unadjusted analysis. Bias effect; multiplicative ($\gamma_{1,jbk}$) and/or additive $\gamma_{2,jbk}$. Bias indicator R_j. Bias probability π_j.	The covariate parameters; β_{0j}, $\beta_{1,jbk}^W$ and $\beta_{1,jbk}^B$. Bias-adjusted relative treatment effect θ_{jbk} Bias effect γ_{jbk} (only additive) Bias probability π_j

Features, advantages, and challenges	Easy to implement using standard statistical software. Mostly used in practice. Recommended only as an initial analysis.	Choosing a value for ς (mean bias shift) and the inflation factor w can be challenging in practice. Should be used with a range of parameter values.	Can be used to model multiplicative bias effects. Compared to bias-adjusted model 2, an extra parameter, R_j, needs to be estimated. We recommend running a sensitivity analysis by choosing different values for a_1, a_2 (hyperparameters of the prior beta distribution assigned to π_j).	It allows for more uncertainty about our risk of bias judgment. It has slightly a better convergence for the bias effect parameters compared to bias-adjusted model 1. A sensitivity analysis for a_1, a_2 is recommended. The bias-adjusted model 2 is more sensitive to the prior assigned to π_j compared to bias-adjusted model 1, particularly when there are a few studies to synthesize.

RoB, Risk of Bias in the study; RCT, Randomised Clinical Trials; NRS, Non-Randomised Studies

[1] τ_γ is the between-study heterogeneity in bias effect

[2] $q_j = \tau^2/(\tau^2 + \tau_\gamma^2)$ *and represents the proportion of the between-study heterogeneity that is not explained by accounting for risk of bias*

Table 3 Notation for the synthesis models

Notation	Description
$i = 1, \ldots, np_j$	participant id
$j = 1, \ldots, ns$	study id
$k = 1, \ldots, K$	treatment index
$ns_{IPD}, ns_{AD}, ns_{RCT}, ns_{NRS}$ $ns_{IPD,RCT}, ns_{AD,RCT}$ $ns_{IPD,NRS}, ns_{AD,NRS}$	the number of studies. The index refers to the design or format of the study or both
y_{ijk}	binary outcome (0/1)
p_{ijk}	probability of the event to occur
r_{jk}	the number of events per arm
n_{jk}	the sample size per arm
b	the study-specific reference
u_{jb}	The treatment effect of the study-specific reference b when $x_{ijk} = \bar{x}_j = 0$
δ_{jbk}	log(OR) of treatment k relative to b
x_{ijk}	the covariate
$\bar{x}_j$	the mean covariate for study j
d_{Ak}	the basic parameters where d_{AA}=0 when A set as the reference in the network
z_j	study characteristics to estimate the bias probability π_j

Table 4 Assumptions about the model parameters

Parameter	**Assumptions**
Relative treatment effect (δ_{jbk})	Random-effects: $\delta_{jbk} \sim N(d_{Ak} - d_{Ab}, \tau^2)$
	Common-effect: $\delta_{jbk} = d_{Ak} - d_{Ab}$
Covariate effect (β_{0j})	Independent effects: $\beta_{0j} \sim N(0, 10^2)$
	Random-effects: $\beta_{0j} \sim N(B_0, \tau_0^2)$
Within-study covariate-treatment interaction ($\beta_{1,jbk}^W$)	Random-effects: $\beta_{1,jbk}^W \sim N(B_{1,Ak}^W - B_{1,Ab}^W, \tau_W^2)$
	Common-effect: $\beta_{1,jbk}^W = B_{1,Ak}^W - B_{1,Ab}^W$
Between-study covariate-treatment interaction ($\beta_{1,jbk}^B$)	Random-effects: $\beta_{1,jbk}^B \sim N(B_{1,Ak}^B - B_{1,Ab}^B, \tau_B^2)$
	Common-effect: $\beta_{1,jbk}^B = B_{1,Ak}^B - B_{1,Ab}^B$
Bias effect ($\gamma_{m,jbk}$) $m = \{1,2\}$	Random-effects: $\gamma_{m,jbk} \sim N(g_{m,bk}, \tau_{m,\gamma}^2)$
	Common-effect: $\gamma_{m,jbk} = g_{m,bk}$
Mean bias effect ($g_{m,bk}$)	$g_{m,bk} = \begin{cases} g_m & \text{if } b \text{ is inactive treatment} \\ 0 & \text{if } b \text{ and } k \text{ are active treatments} \end{cases}$
	$g_{m,bk} = \begin{cases} g_m & \text{if } b \text{ is inactive treatment} \\ (-1)^{dir_{bk}} g_m^{act} & \text{if } b \text{ and } k \text{ are active treatment} \end{cases}$
Bias indicator	$R_j \sim Bernoulli(\pi_j)$
Bias probability (π_j)	$\pi_j \sim Beta(a_1, a_2)$
	$logit(\pi_j) = e + \mathbf{f}^{\mathbf{T}} \mathbf{z_j}$

Table 5 The mean estimates and 95% credible intervals from bias-adjusted models 1 and 2 for the antidepressants network shown in Figure 1b.

g^p, the additive bias effect on log odds ratio for active-placebo comparisons; g^{act}, the additive bias effect on log odds ratio for active-active comparisons (sponsored treatment assumed to be favoured); RoB, risk of bias in the study; CrI, credible interval.

	Bias-adjusted model 1	Bias-adjusted model 2
Model assuming a prior $\tau_\gamma \sim \text{Unif}(0,2)$ for the heterogeneity in bias effects		
Primary analysis: Bias probability distribution (low RoB: $\pi_j \sim$Beta(1,20), moderate RoB: $\pi_j \sim$Beta(20,1))		
Mean bias effect: exp(g^p)	1.090 (0.975, 1.249)	1.035 (0.939, 1.143)
Mean bias effect: exp(g^{act})	1.186 (1.054, 1.335)	1.182 (1.054, 1.335)
Heterogeneity in bias effect: τ_γ (95% CrI)	0.130 (0.005, 0.261)	0.185 (0.128, 0.251)
Sensitivity analysis: Bias probability distribution (low RoB; $\pi_j \sim$Beta(1,10), moderate RoB; $\pi_j \sim$Beta(10,1))		
Mean bias effect: exp(g^p)	1.163 (0.966, 1.421)	1.035 (0.878, 1.224)
Mean bias effect: exp(g^{act})	1.257 (1.095, 1.478)	1.271 (1.094, 1.600)
Heterogeneity in bias effect: τ_γ	0.206 (0.078, 0.318)	0.210 (0.127, 0.354)
Model that re-parametrises the heterogeneity using weights $q_j \sim Beta(v,1)$ where $q_j = \tau^2/(\tau^2+\tau_\gamma^2)$		
Low RoB studies: no down-weighting; Moderate RoB studies: down-weight by 25%		
Mean bias effect: exp(g^p)	0.985 (0.786, 1.475)	0.817 (0.549, 1.112)
Mean bias effect: exp(g^{act})	1.222 (1.073, 1.476)	1.427 (1.173, 1.942)
Low RoB studies: no down-weighting; Moderate RoB studies: down-weighting by 80%		
Mean bias effect: exp(g^p)	1.012 (0.860, 1.167)	1.008 (0.851, 1.153)
Mean bias effect: exp(g^{act})	1.203 (1.067, 1.383)	1.231 (1.081, 1.470)

Table 6 Deviance information criterion of the network meta-analysis models (NMA) fitted to the network of treatments for the relapsing-remitting multiple sclerosis (RRMS) in Figure 1a and for the NMA models fitted to the antidepressants network in Figure 1b.

IPD; Individual Participant Data, AD; Aggregate Data,

	RRMS example		Antidepressant example
	IPD model	AD model	AD model
Unadjusted analysis	90492	187	2667
Bias-adjusted model 1	90508	248	2648
Bias-adjusted model 2	90365	158	2664

Figures captions

Figure 1 Network plots of (a) treatments for patients with relapsing-remitting multiple sclerosis compared in randomised controlled trials (solid, grey edges) and in the Swiss Multiple Sclerosis Cohort (dashed, black edges). The outcome is relapse in two years (b) antidepressants and placebo compared in randomised clinical trials. The outcome is response to treatment. The thickness of the edges is proportional to the number of trials comparing each pair of treatments.

Figure 2 Relapse odds ratios with 95% credible intervals of all comparisons of treatments among patients with relapsing-remitting multiple sclerosis. The estimates are computed by conducting unadjusted analysis and bias-adjusted analysis 1 and 2 in a Bayesian framework of the data in the network of Figure 1a. The study-specific estimates have been computed in frequentist framework and hence the lines represent confidence intervals. To compute these estimate, we used glm() function to analyze IPD studies and metabin() function *(from* meta *package) to analyze AD studies.*

Figure 3 The relationship between patient age (in years) and the estimated odds ratio with 95% credible intervals (the shaded areas) for active treatments vs placebo among patients with relapsing-remitting multiple sclerosis estimated with network meta-regression with bias-adjusted model 1.

Figure 4 Response odds ratio with 95% credible interval for each antidepressant vs placebo estimated from unadjusted analysis and bias-adjusted models 1 and 2 using the data presented in the network of Figure 1b. A random-effects network meta-analysis model is assumed to estimate treatment and bias effects.

Appendix Figures captions

Appendix Figure 1

Appendix Figure 2

Appendix Figure 3

Appendix Figure 4

Appendix Figure 5

Appendix Figure 6

Appendix Figure 7

Appendix Figure 8

Appendix Tables captions

Appendix Table 1

Appendix Table 2

Appendix Table 3

Appendix Table 4

Appendix Table 5

Appendix Table 6

Appendix Table 7

Appendix Table 8

Figure 1a

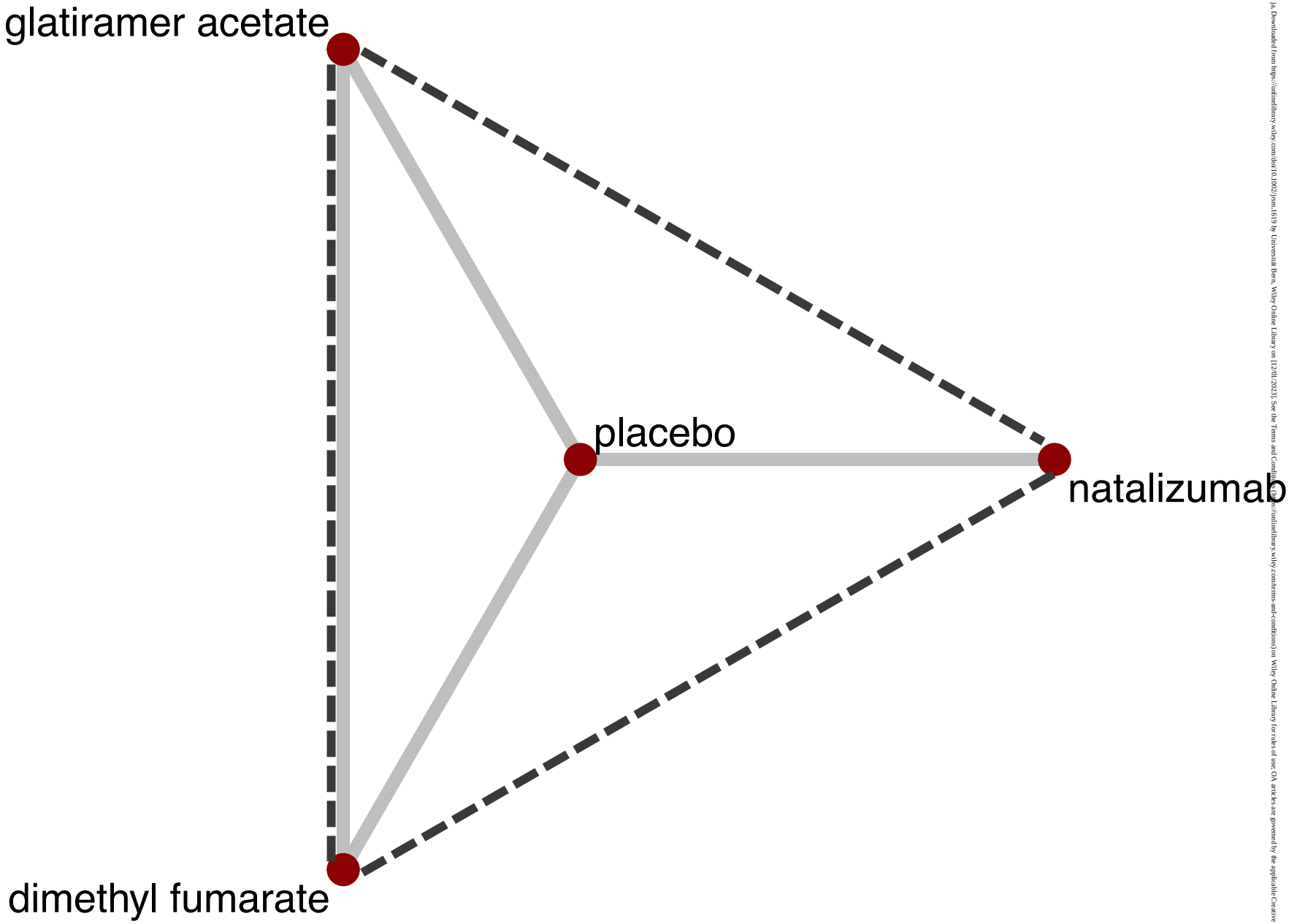
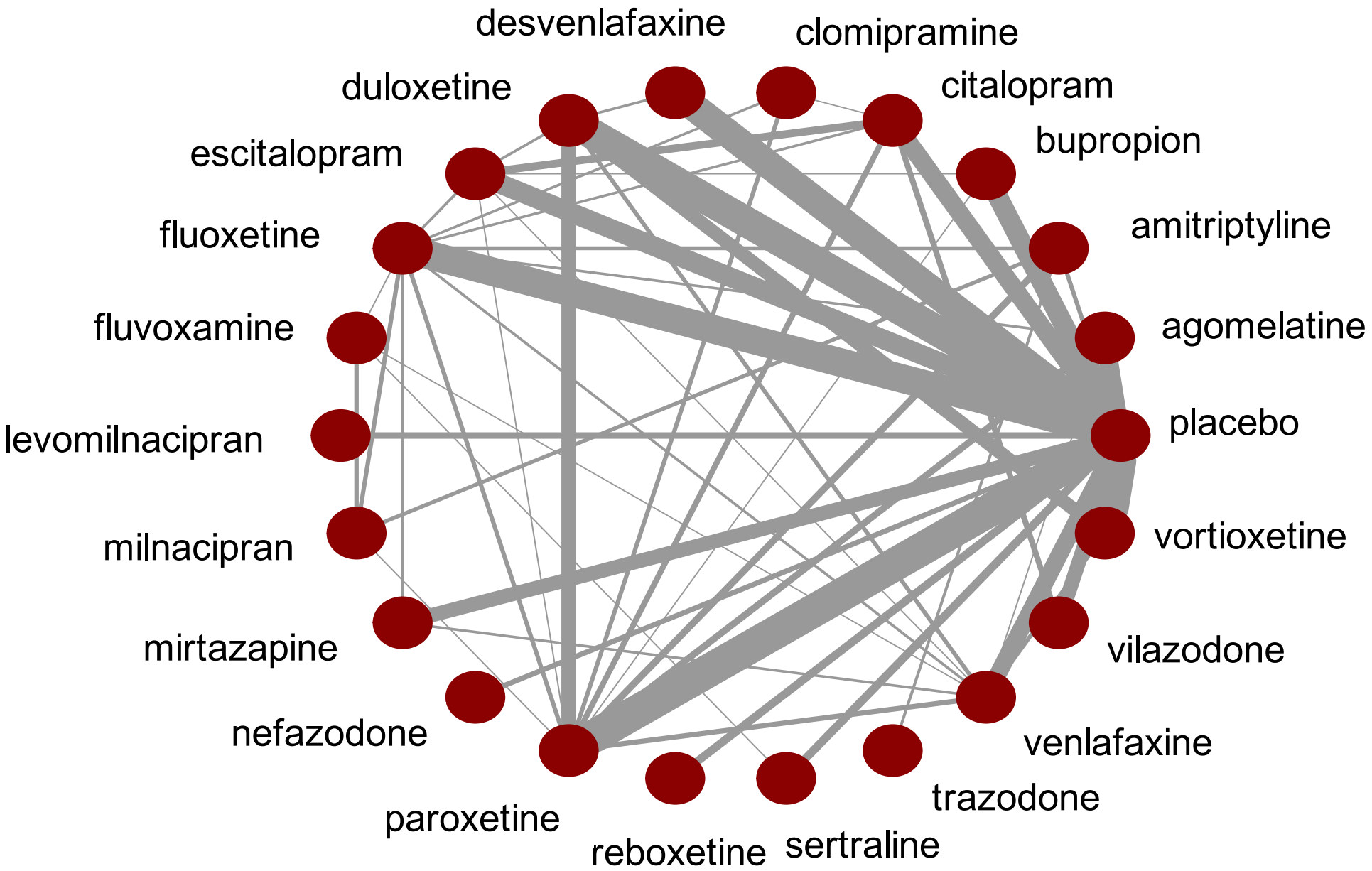


Figure 1b



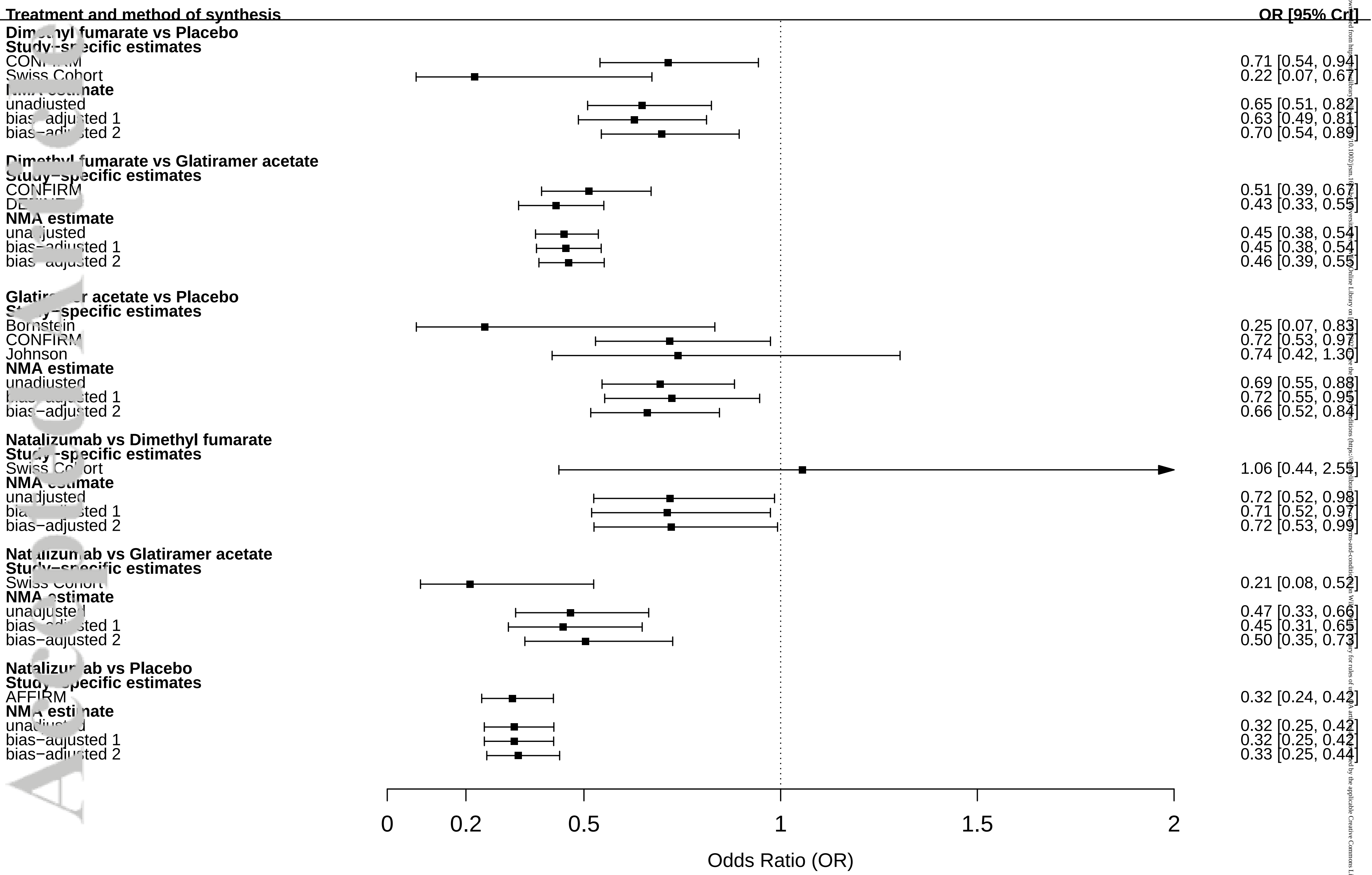


Figure 3

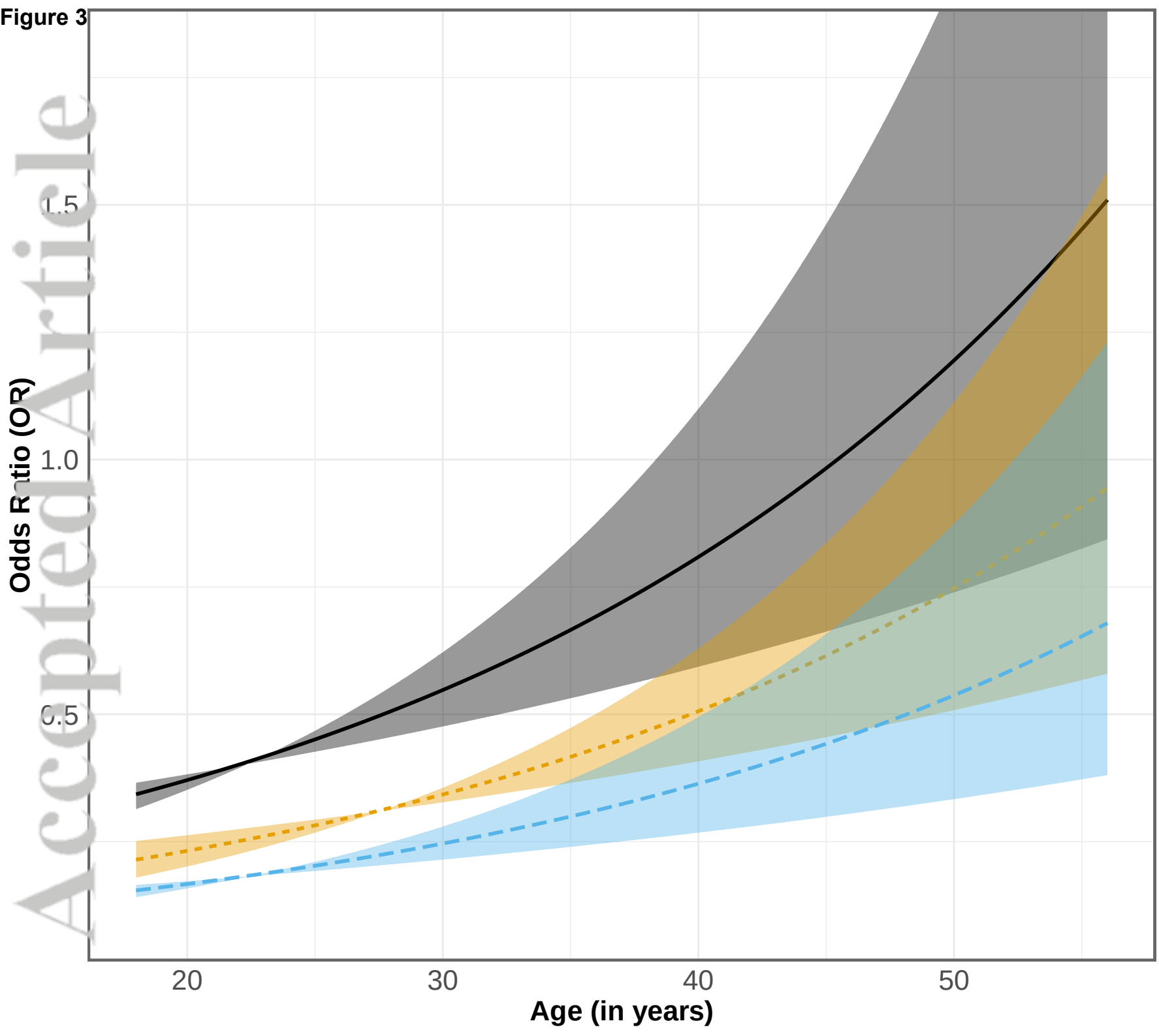


Figure 4