

RESEARCH

Open Access



Bayesian network meta-analysis methods for combining individual participant data and aggregate data from single arm trials and randomised controlled trials

Janharpreet Singh^{1*}, Sandro Gsteiger², Lorna Wheaton¹, Richard D. Riley³, Keith R. Abrams⁴, Clare L. Gillies⁵ and Sylwia Bujkiewicz¹

Abstract

Background: Increasingly in network meta-analysis (NMA), there is a need to incorporate non-randomised evidence to estimate relative treatment effects, and in particular in cases with limited randomised evidence, sometimes resulting in disconnected networks of treatments. When combining different sources of data, complex NMA methods are required to address issues associated with participant selection bias, incorporating single-arm trials (SATs), and synthesising a mixture of individual participant data (IPD) and aggregate data (AD). We develop NMA methods which synthesise data from SATs and randomised controlled trials (RCTs), using a mixture of IPD and AD, for a dichotomous outcome.

Methods: We propose methods under both contrast-based (CB) and arm-based (AB) parametrisations, and extend the methods to allow for both within- and across-trial adjustments for covariate effects. To illustrate the methods, we use an applied example investigating the effectiveness of biologic disease-modifying anti-rheumatic drugs for rheumatoid arthritis (RA). We applied the methods to a dataset obtained from a literature review consisting of 14 RCTs and an artificial dataset consisting of IPD from two SATs and AD from 12 RCTs, where the artificial dataset was created by removing the control arms from the only two trials assessing tocilizumab in the original dataset.

Results: Without adjustment for covariates, the CB method with independent baseline response parameters (*CBunadjInd*) underestimated the effectiveness of tocilizumab when applied to the artificial dataset compared to the original dataset, albeit with significant overlap in posterior distributions for treatment effect parameters. The CB method with exchangeable baseline response parameters produced effectiveness estimates in agreement with *CBunadjInd*, when the predicted baseline response estimates were similar to the observed baseline response. After adjustment for RA duration, there was a reduction in across-trial heterogeneity in baseline response but little change in treatment effect estimates.

(Continued on next page)

*Correspondence: js929@leicester.ac.uk

¹BioStatistics Research Group, Department of Health Sciences, University of Leicester, Leicester, UK

Full list of author information is available at the end of the article



© The Author(s). 2022 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

Conclusions: Our findings suggest incorporating SATs in NMA may be useful in some situations where a treatment is disconnected from a network of comparator treatments, due to a lack of comparative evidence, to estimate relative treatment effects. The reliability of effect estimates based on data from SATs may depend on adjustment for covariate effects, although further research is required to understand this in more detail.

Keywords: Evidence synthesis, Network meta-analysis, Single-arm trials, Individual participant data, Arm-based methods, Bayesian hierarchical methods, Rheumatoid arthritis, Observational evidence

Background

Traditionally, meta-analysis of randomised controlled trials (RCTs) has been used to quantitatively pool published evidence on the effectiveness of a treatment, or a network of treatments through network meta-analysis (NMA), across trials with similar characteristics (e.g. patient population, trial design and conduct) [1–3]. In the published literature, results from a trial are usually reported as aggregate data (AD) summarising the average treatment effect. Synthesising evidence from RCTs has been considered a gold standard as randomised treatment allocation minimises the risk of confounding in treatment effect estimates.

In circumstances where randomised evidence is limited, non-randomised evidence, such as data from single-arm trials (SATs), may have to be considered in determining treatment effectiveness [4]. Such cases are becoming increasingly common in health technology assessment (HTA), as new technologies receive accelerated approval from regulatory agencies based on SATs leading to disconnected networks of treatments [5]. This creates an issue for reimbursement decision-making by HTA bodies where the interest is in unbiased estimates of relative treatment effects. Consequently, there is a need for synthesis methods which combine randomised and non-randomised evidence, whilst addressing issues associated with the latter, such as susceptibility to confounding and the lack of evidence from a comparator arm in the case of SATs [6].

For a particular trial, individual participant data (IPD) may be available recording the effect of the treatment on outcomes of interest, as well as a set of covariates, for each participant. Meta-analysis incorporating IPD has been shown to overcome issues associated with detecting treatment-covariate interactions, and prognostic effects, compared to a meta-regression of AD only [7, 8]. Moreover, it accounts for variability in covariate values across participants within a trial, allowing within- and across-trial interactions to be estimated separately which mitigates the risk of aggregation bias [9, 10]. However, the availability of IPD is often subject to a data-sharing agreement with the trial sponsor which often involves a lengthy, complex process due to regulatory considerations (e.g. data protection and privacy). Furthermore, access to IPD

may be particularly challenging when data are required from a number of trials, and for trials assessing different treatments developed by different manufacturers. Consequently, IPD are unlikely to be available for all trials in the synthesis and methods have been developed which combine a mixture of IPD and AD [11–14]. Population-adjustment methods have been developed to estimate a relative treatment effect for the specific situation where IPD are available for a trial assessing a particular treatment and only AD are available for another trial assessing a comparator, which is a common occurrence in HTA [15]. Such methods make the assumption that the treatment effect is equivalent in the two study populations at each level of a set of effect modifiers [16]. More recently, multilevel network meta-regression methods have been proposed which extend the population-adjustment methods beyond the two-study case, and also avoid the issue of aggregation bias by using IPD to inform the covariate distributions in the AD model [17]. Similarly, methods have been proposed which use IPD to estimate regression coefficients for effect modifiers, and synthesise IPD and AD via either a hierarchical model or by constructing empirical prior distributions [18].

Extensions to meta-analysis, which allow the synthesis of SATs and RCTs, have been proposed which assume that baseline response parameters are exchangeable (rather than independent) across trials [19]. This approach has been developed to a NMA context using a mixture of IPD and AD and a contrast-based approach to NMA [20]. Arm-based NMA methods, which parametrise absolute treatment effects across trial arms, also allow SATs to be incorporated into a synthesis of RCTs in a pairwise meta-analysis [21] and a NMA context [22, 23]. Under a Bayesian framework, these methods allow data from SATs to provide information on the pooled treatment effect and between-study heterogeneity parameters via a prior distribution, and can be down-weighted according to commensurability with RCT data [24]. These methods have been criticised for compromising randomised treatment allocation and introducing bias in treatment effect estimates from RCTs [25, 26]. However, they may be helpful to include when the only evidence available for a treatment is from SATs, and the increase in risk of bias associated with the exchangeability assumption may have

little impact in practice [27, 28]. Alternative methods seek to match similar trial arms based on reported characteristics and either entering the matched pairs into the NMA directly or by plugging-in the baseline response estimate from a matched trial to estimate a relative treatment effect from SATs [29, 30].

In this paper, we develop methods to synthesise SATs and two-arm RCTs, using a mixture of IPD and AD, for a dichotomous outcome. We consider methods with and without adjustment for a covariate effect. We build on contrast- and arm-based NMA methods [22, 23] to incorporate SATs, and use shared-parameters to synthesise IPD and AD [13, 14]. We apply the developed methods using mixture of IPD and AD from RCTs assessing the effectiveness of biologic disease-modifying anti-rheumatic drugs (bDMARDs) as treatments for rheumatoid arthritis (RA), as an applied example. We describe the IPD and AD datasets used in the applied example in “[Applied example: assessing biologics for treating rheumatoid arthritis \(RA\)](#)” section, before detailing the methods for a Bayesian implementation in “[Methods](#)” section, and reporting the results from the application in “[Results](#)” section. In “[Discussion](#)” section, we discuss issues related to the use of the methods and suggest how these can be investigated further.

Applied example: assessing biologics for treating rheumatoid arthritis (RA)

In this paper, we consider a case-study in rheumatoid arthritis (RA); a chronic auto-immune condition causing joint inflammation in (mostly) elderly patients. The American College of Rheumatology (ACR) response criteria provide a measure of patient response to treatment, with an ACR20 outcome representing a 20% improvement in RA symptoms as defined by the criteria [31]. A number of bDMARDs (hereafter referred to as biologics) have been developed to offer a choice of treatment strategies in managing the disease [32].

Table 1 summarises the original dataset used in the applied example. Data were available from 14 RCTs, each assessing the effectiveness of a first-line biologic versus placebo as treatment for RA, with 5,821 participants in total. In 9 trials, a biologic was allocated to multiple trial arms (ranging from two to four arms) in order to assess different dosage regimens. For this applied example, these data were aggregated and considered as from a single trial arm. There was not significant variability in dosage regimens across trials assessing a particular biologic. All trial arms consisted of participants receiving methotrexate as background therapy. IPD were available for two trials of tocilizumab [33, 34], whilst only AD were available for the remaining trials [35–49]. An ACR20 response was considered as an outcome to measure treatment effectiveness, with IPD providing responder status for each participant and AD providing the number of participants achieving

a response in each trial arm. We considered RA duration (in years) as a covariate in our models, as there was evidence of an association between baseline response and RA duration present in the exploratory analysis of the IPD. Table 1 lists the mean RA duration at baseline in each trial arm.

Figure 1 illustrates the network structure in terms of the evidence available for each treatment comparison. The network consists of trials assessing four biologics; infliximab, tocilizumab, golimumab, and adalimumab. Each trial compared a biologic therapy versus placebo, and there were no trials comparing biologics head-to-head.

In this paper, we consider methods which can be applied to synthesise data from SATs and RCTs, in a NMA context. For the applied example, an artificial dataset was created by removing the placebo arms from the two tocilizumab trials. As a result, the artificial dataset included two SATs assessing tocilizumab. The aim of the synthesis was to compare the effectiveness of tocilizumab with infliximab, golimumab, and adalimumab, in terms of the ACR20 outcome.

Methods

In this section, we describe Bayesian NMA methods used to synthesise data from SATs and two-arm RCTs, under both contrast- and arm-based model parametrisations. We focus on one-stage methods which synthesise IPD and AD in a single model, as opposed to two-stage methods where IPD are first aggregated and then synthesised with AD. We assume data are available as IPD from N^{IPD} SATs followed by AD from N^{AD} two-arm RCTs, and let $j = 1, \dots, N^{IPD}, \dots, N^{IPD} + N^{AD}$ index both trial sets. We start by describing the likelihood functions used to model a dichotomous outcome using IPD and AD, then describe methods under a contrast-based parametrisation, followed by methods under an arm-based parametrisation. Under each parametrisation, we describe methods without adjustment for covariates and methods which include a covariate to adjust for its effect.

Likelihood models for a dichotomous outcome

Let Y_{ijk} represent a dichotomous outcome indicating responder status (i.e. responder 1, non-responder 0) for participant i in trial j assigned to treatment k . Then, for trials from which IPD are available, we assume that the outcome Y_{ijk} is a Bernoulli random variable,

$$Y_{ijk} \sim \text{Bern}(p_{ijk}), \quad j = 1, \dots, N^{IPD}; \quad (1)$$

where p_{ijk} represents the probability of response for each participant.

Similarly, let n_{jk} and r_{jk} represent the numbers of participants and responders, respectively, in trial j allocated

Table 1 Summary (arm-level) data from RCTs assessing biologic therapies as treatments for RA, (N - number of participants, ACR20 (%) - percent of ACR20 responders)

Trial	Lead author	Year	Data type	Intervention	N	ACR20 (%)	Mean RA duration (years)
1	Genovese	2008	IPD	Tocilizumab	803	60.8	9.8
1	Genovese	2008	IPD	Placebo	413	24.5	9.8
2	Smolen	2008	IPD	Tocilizumab	418	53.1	7.4
2	Smolen	2008	IPD	Placebo	204	26.5	7.8
3	Maini	1999	AD	Infliximab	340	57.4	8.3
3	Maini	1999	AD	Placebo	88	22.7	8.9
4	Weinblatt	2003	AD	Adalimumab	209	60.3	12.7
4	Weinblatt	2003	AD	Placebo	62	14.5	11.1
5	Keystone	2004	AD	Adalimumab	419	62.1	11.0
5	Keystone	2004	AD	Placebo	200	29.5	10.9
6	Chen	2008	AD	Adalimumab	35	54.3	6.2
6	Chen	2008	AD	Placebo	12	33.3	8.3
7	Kay	2008	AD	Golimumab	137	61.3	7.9
7	Kay	2008	AD	Placebo	35	37.1	5.6
8	Keystone	2009	AD	Golimumab	178	59.6	5.6
8	Keystone	2009	AD	Placebo	133	27.8	6.5
9	Kremer	2010	AD	Golimumab	257	43.6	8.8
9	Kremer	2010	AD	Placebo	129	24.8	7.4
10	Schiff	2011	AD	Infliximab	165	59.4	7.3
10	Schiff	2011	AD	Placebo	110	41.8	8.4
11	Emery	2013	AD	Golimumab	318	61.6	3.5
11	Emery	2013	AD	Placebo	160	49.4	2.9
12	Kim	2013	AD	Infliximab	71	50.7	9.8
12	Kim	2013	AD	Placebo	72	30.6	7.4
13	Tanaka	2013	AD	Golimumab	173	72.8	8.4
13	Tanaka	2013	AD	Placebo	88	33.0	8.7
14	Weinblatt	2013	AD	Golimumab	395	58.5	6.9
14	Weinblatt	2013	AD	Placebo	197	24.9	7.0

treatment k . Then, for trials from which only AD are available, the number of responders in each arm r_{jk} is assumed to follow a Binomial distribution,

$$r_{jk} \sim \text{Bin}(n_{jk}, p_{jk}), \quad j = N^{IPD} + 1, \dots, N^{IPD} + N^{AD}; \quad (2)$$

where p_{jk} is the probability of response in each trial arm. For both data-types, we employ the logit link function to transform response probabilities onto the linear predictor scale on which treatment effects can be assumed to be linearly additive.

The number of responders in each arm r_{jk} is the sum over the participant-level Bernoulli outcomes Y_{ijk} , which are assumed to be identically-distributed. If there are unobserved participant-level covariates which influence the individual response probabilities p_{ijk} , then the corresponding outcomes Y_{ijk} are unlikely to be identically-

distributed. Thus, the AD likelihood in Eq. (2) is the correct aggregate version of the IPD likelihood in Eq. (1) only if there are no important unobserved participant-level covariates.

Contrast-based (CB) methods

Here, we describe methods under a contrast-based parametrisation where treatment effects are expressed relative to a baseline response (i.e. the outcome observed in the control arm). We denote B as index for the baseline treatment in each trial.

Methods without covariates

We start by describing methods which only model the response data and do not include data on covariates, as an intermediate step so that it is clear how these methods can be extended to incorporate covariate adjust-

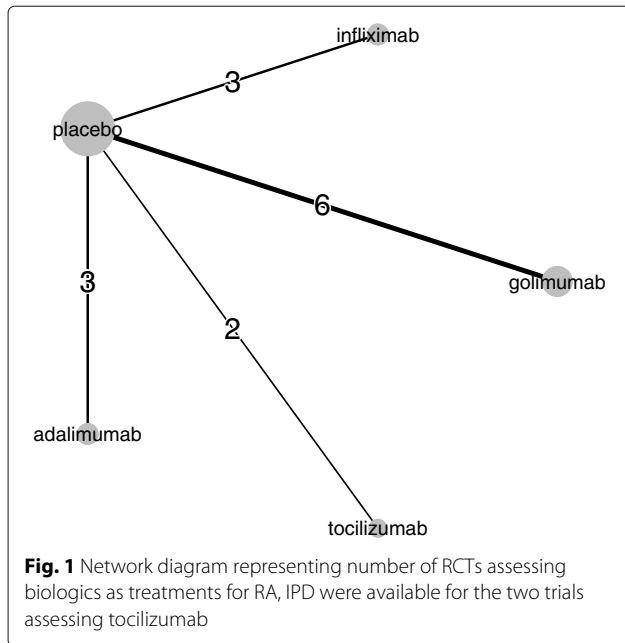


Fig. 1 Network diagram representing number of RCTs assessing biologics as treatments for RA, IPD were available for the two trials assessing tocilizumab

ment in “Methods with covariate-adjustment” section. The IPD and AD models are equivalent, but are described separately to clarify how shared-parameters allow both datasets to contribute to the estimation of the treatment effect.

For IPD, the model is given by,

$$\text{logit}(p_{ijk}) = \phi_j + \delta_{jBk} \times I_{ijk}, \quad j = 1, \dots, N^{IPD}; \quad (3)$$

where ϕ_j represents the log odds of a response on the baseline treatment B in trial j , δ_{jBk} is the log odds ratio of effect on treatment k relative to B , and I_{ijk} is a participant-level treatment indicator variable (i.e. treatment 1, baseline 0).

Similarly, for AD, the model can be written as,

$$\begin{aligned} \text{logit}(p_{jk}) &= \phi_j + \delta_{jBk} \times I_{jk}, \\ j &= N^{IPD} + 1, \dots, N^{IPD} + N^{AD}; \end{aligned} \quad (4)$$

where ϕ_j and δ_{jBk} have the same interpretation as for the IPD model (3), and I_{jk} is an arm-level treatment indicator variable.

The treatment effects δ_{jBk} are assumed to be exchangeable across trials comparing treatment k with B ,

$$\delta_{jBk} \sim N(d_{Bk}, \sigma^2), \quad j = 1, \dots, N^{IPD} + N^{AD}; \quad (5)$$

with d_{Bk} representing the pooled log odds ratio. The between-study heterogeneity in the treatment effects is quantified by the standard deviation parameter σ . Here, σ is assumed to be common across treatment comparisons as we consider methods for application in cases where data are limited.

The pooled effects d_{Bk} can be represented in terms of basic parameters by assuming consistency within the network $d_{Bk} = d_{1k} - d_{1B}$ ($d_{11} = 0$). Thus, effect estimates

can be obtained indirectly for treatment comparisons for which there are no trials providing a head-to-head estimate. We place a non-informative Normal prior distribution on each of the basic parameters $d_{1k} \sim N(0, 10^2)$ and the baseline response parameters $\phi_j \sim N(0, 10^2)$. Hong et al. place a non-informative Uniform prior distribution for the between-study heterogeneity parameter, $\sigma \sim U(0, 10)$. However, we note that this prior distribution may be weakly-informative in a synthesis with few trials [50] and $\sigma \sim U(0, 2)$ has been recommended as a more appropriate prior distribution [51] which we use here (The results of a sensitivity analysis using $\sigma \sim U(0, 10)$ are reported in Appendix E for comparison).

Traditionally, NMA methods have been applied to synthesise data from RCTs which assume independent baseline response parameters ϕ_j [3]. This accounts for differences in prognostic factors between trials, and ensures that the impact of randomised treatment allocation is reflected in the treatment effect estimates. However, such an approach is restricted to trials with two or more arms and the interest here is to incorporate data from SATs into the synthesis. Therefore, we assume exchangeable baseline response parameters as in [19],

$$\phi_j \sim N(m_\phi, \sigma_\phi^2), \quad j = 1, \dots, N^{IPD} + N^{AD}; \quad (6)$$

where m_ϕ represents the mean baseline response, and σ_ϕ quantifies between-study heterogeneity in the baseline response. The benefit of this assumption is that it allows a baseline response to be predicted, and a treatment effect to be estimated, for each SAT in the synthesis so that pooled estimates are based on all of the available evidence. The cost of this assumption is that randomisation is compromised for RCTs in the synthesis, which will bias treatment effect estimates where there is significant between-study heterogeneity (which may be due to differences in prognostic factors) in baseline response estimates. We specify a non-informative Normal prior distribution for the mean baseline response parameter $m_\phi \sim N(0, 10^2)$, and a non-informative Uniform prior distribution for the between-study heterogeneity parameter $\sigma_\phi \sim U(0, 2)$. We label the methods with independent and exchangeable baseline response parameters by *CBunadjInd* and *CBunadjEx*, respectively.

Alternatively, the *CBunadjInd* method can be implemented in two stages to predict baseline response estimates for SATs in the synthesis, whilst maintaining the effects of randomisation and limiting bias in treatment effect estimates for RCTs. In the first stage, the *CBunadjEx* method is applied to the data from the set of RCTs in the synthesis. The posterior estimates for the mean and standard deviation parameters (i.e. \hat{m}_ϕ and $\hat{\sigma}_\phi$), describing the distribution of the baseline response across RCTs, are recorded. In the second stage, the *CBunadjInd* method

is applied to synthesise SATs and RCTs, where an informative Normal prior distribution is placed on the baseline response parameters corresponding to SATs (i.e. $\phi_j \sim N(\hat{m}_\phi, \hat{\sigma}_\phi^2)$). This prior distribution is informed by the posterior estimates recorded from the first stage of the method, which is equivalent to using the posterior predictive distribution of the baseline response across RCTs. The posterior predictive distribution, as opposed to the posterior distribution, is recommended by Dias et al. to account for between-study heterogeneity [52]. Ultimately, pooled treatment effect estimates, and inference, are based only on the second stage of the method. Thus, SATs are incorporated into the synthesis and contribute to the pooled treatment effect estimates, whilst the assumption of independent baseline response parameters ensures randomisation in RCTs is not compromised.

Methods with covariate-adjustment

Study-level adjustment only Here, we consider methods which include a covariate to adjust the baseline response. We begin by extending the *CBunadjEx* method to adjust for a covariate at the trial-level, and let \bar{x}_j represent the mean covariate value summarising participants in trial j . We use the superscripts A and W to denote across- and within-trial effects, respectively. For IPD, the extended model is given by,

$$\text{logit}(p_{ijk}) = \phi_j + \alpha^A \times \bar{x}_j + \delta_{jBk} \times I_{ijk}, \quad j = 1, \dots, N^{IPD}; \quad (7)$$

where ϕ_j represents the log odds of response on baseline treatment B when \bar{x}_j is equal to zero, and α^A is the change in the log odds for a unit increase in \bar{x}_j . Thus, α^A represents a trial-level covariate effect, accounting for the association between the baseline response ϕ_j and the mean covariate \bar{x}_j , across trials. Here, there is no adjustment for the interaction between the relative treatment effect and the mean covariate, and so δ_{jBk} represents a marginal treatment effect.

Similarly, for AD, the extended model can be described by,

$$\text{logit}(p_{jk}) = \phi_j + \alpha^A \times \bar{x}_j + \delta_{jBk} \times I_{jk}, \quad j = N^{IPD} + 1, \dots, N^{IPD} + N^{AD}; \quad (8)$$

where ϕ_j and α^A have the same interpretation as for the IPD model (7). A non-informative Normal prior distribution is recommended for the across-trial covariate effect parameter $\alpha^A \sim N(0, 10^2)$, and we label this method *CBadjEx*.

Participant-level and study-level adjustment In this paper, we consider methods to synthesise a mixture of IPD and AD. IPD allows participant-level and trial-level covariate effects to be estimated separately, avoiding

aggregation bias which may occur when the assumption that these effects are equivalent does not hold [8]. We extend the *CBadjEx* method in Eq. (7) to adjust for the covariate both at the participant-level and the trial-level. We let x_{ijk} represent the covariate value for participant i in trial j allocated treatment k . For IPD, the extended model is given by,

$$\text{logit}(p_{ijk}) = \phi_j + \alpha^A \times \bar{x}_j + \alpha^W \times (x_{ijk} - \bar{x}_j) + \delta_{jBk} \times I_{ijk}, \quad j = 1, \dots, N^{IPD}; \quad (9)$$

where α^W is the change in the log odds of a response on baseline treatment B for a unit increase in $(x_{ijk} - \bar{x}_j)$. Thus, α^W represents a common (participant-level) covariate effect within a trial. A non-informative Normal prior distribution is recommended for within-trial covariate effect parameter $\alpha^W \sim N(0, 10^2)$, and we label this method *CBadjEbEx* (where *Eb* denotes that this method accounts for ecological aggregation bias - the difference between the across- and within-trial covariate effects).

Arm-based (AB) methods

Hong et al. have proposed NMA models under an arm-based parametrisation, where treatment effects are expressed as the absolute effects across trial arms, using either IPD [22] or AD [23]. Here, we extend their models to synthesise a mixture of IPD and AD.

Methods without covariates

We begin by describing models with outcome data only. For IPD, the model is given by,

$$\text{logit}(p_{ijk}) = \theta_k + v_{jk}, \quad j = 1, \dots, N^{IPD}; \quad (10)$$

where θ_k represents the pooled log odds of response on treatment k , and v_{jk} are error terms (i.e. the difference between the log odds of response on treatment k in trial j and the pooled log odds of response on treatment k). A non-informative Normal prior distribution is suggested for each pooled response parameter $\theta_k \sim N(0, 10^2)$.

Similarly, for AD, the model is described by,

$$\text{logit}(p_{jk}) = \theta_k + v_{jk}, \quad j = N^{IPD} + 1, \dots, N^{IPD} + N^{AD}; \quad (11)$$

where θ_k and v_{jk} have the same interpretation as for the IPD model (10).

The error terms v_{jk} are assumed to be exchangeable across trials,

$$v_{jk} \sim N(0, \tau^2), \quad j = 1, \dots, N^{IPD} + N^{AD}; \quad (12)$$

where τ quantifies the between-study heterogeneity and is assumed to be common. A non-informative Uniform prior distribution is recommended for the between-study heterogeneity parameter $\tau \sim U(0, 2)$.

The application in this paper considers a synthesis of SATs and two-arm RCTs, and we make an additional assumption regarding the heterogeneity observed in the latter. For each two-arm RCT, we let k_1 and k_2 denote the treatment assigned in arms 1 and 2, respectively. Then, the error terms are assumed to follow a bivariate distribution,

$$\begin{pmatrix} v_{jk_1} \\ v_{jk_2} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_{k_1 k_2} = \begin{pmatrix} \tau^2 & \rho\tau^2 \\ \rho\tau^2 & \tau^2 \end{pmatrix}\right), \quad (13)$$

$$j = N^{IPD} + 1, \dots, N^{IPD} + N^{AD};$$

with covariance matrix $\Sigma_{k_1 k_2}$. The responses observed in arms from RCTs are likely to be similar (compared to two SATs), due to the balance in prognostic factors achieved by randomised treatment allocation [53]. The similarity is represented by a common distribution, which is a bivariate distribution in the case of data from two-arm RCTs, accounting for the correlation between treatment arms. A non-informative Uniform prior distribution is suggested for the correlation parameter in the covariance matrix $\rho \sim U(-1, 1)$. We label this method as *ABunadj*.

Methods with covariate-adjustment

Study-level adjustment only Here, we extend the *ABunadj* method to allow adjustment for a covariate. We begin by adjusting the absolute treatment effect in each arm for the average value of the covariate \bar{x}_{jk} across participants in trial j and treatment arm k . For IPD, the extended model is given by,

$$\text{logit}(p_{ijk}) = \theta_k + v_{jk} + \beta^A \times \bar{x}_{jk}, \quad j = 1, \dots, N^{IPD}; \quad (14)$$

where θ_k represents the pooled log odds of response to treatment k when \bar{x}_{jk} is equal to zero, and β^A is the change in the pooled log odds for a unit increase in \bar{x}_{jk} . Thus, β^A represents the covariate effect across trial arms. Under the arm-based parametrisation, we follow the approach by Hong et al. and adjust for the arm-level mean covariate to account for differences between arms within each trial, in addition to differences between trials [22]. In some cases, data on the arm-level mean covariate \bar{x}_{jk} may not be available from the published trial report, in which case adjustment would be restricted to the reported mean trial-level covariate \bar{x}_j .

Similarly, for AD, the extended model can be written as,

$$\begin{aligned} \text{logit}(p_{jk}) &= \theta_k + v_{jk} + \beta^A \times \bar{x}_{jk}, \\ j &= N^{IPD} + 1, \dots, N^{IPD} + N^{AD}; \end{aligned} \quad (15)$$

where θ_k and β^A have the same interpretation as for the IPD model (14). We suggest a non-informative Normal prior distribution for the arm-level covariate effect parameter $\beta^A \sim N(0, 10^2)$, and we label this method *ABadj*.

Participant-level and study-level adjustment In a similar way to that described for the contrast-based methods, the covariate effect at the participant-level may differ from the effect at the arm-level. To avoid the aggregation bias which is introduced as a result of this difference, we extend the *ABadj* method in Eq. (14) to separate the across- and within-trial covariate effects. For IPD the extended model is given by,

$$\begin{aligned} \text{logit}(p_{ijk}) &= \theta_k + v_{jk} + \beta^A \times \bar{x}_{jk} + \beta^W \times (x_{ijk} - \bar{x}_{jk}), \\ j &= 1, \dots, N^{IPD}; \end{aligned} \quad (16)$$

where β^W represents the change in the pooled log odds for a unit increase in $(x_{ijk} - \bar{x}_{jk})$. We place a non-informative Normal distribution on the within-trial covariate effect parameter $\beta^W \sim N(0, 10^2)$. We label this method *ABadjEb*.

Implementation and software

All methods were applied by fitting models in Stan [54], using Markov chain Monte Carlo (MCMC) sampling to estimate Bayesian posterior distributions for model parameters. Each implementation consisted of four MCMC chains, where 1,000 iterations were used as an initial burn-in period for each chain. Posterior estimates were based on 1,000 samples per chain, and checked for sensitivity to changes in initial values. The effective sample size and \hat{R} statistics were extracted for an initial assessment to identify signs of chain non-convergence, and further checks were made by inspecting trace and autocorrelation plots [55].

Results

In this section, we report the results from applying the methods described in “Methods” section to the datasets described in “Applied example: assessing biologics for treating rheumatoid arthritis (RA)” section. All methods were applied to both the original dataset (consisting of 14 RCTs) and the artificial dataset (consisting of two SATs and 12 RCTs from the original dataset). In tables and figures, where a method was applied to the original dataset it is labelled with the suffix *1*, and where it was applied to the artificial dataset it is labelled with the suffix *2* (e.g. the contrast-based unadjusted method with independent baseline response parameters applied to the original dataset is labelled *CBunadjInd1*). Under the contrast-based parametrisation, the methods require a reference treatment to be specified for which we chose placebo. We report the results for the methods without covariates first, and then report the results from applying the methods with adjustment on the baseline response using RA duration (years) as a covariate.

Methods without covariates

Contrast-based methods

The contrast-based methods without adjustment for covariates are described in “Methods without covariates” section, and consist of the methods with independent and exchangeable baseline response parameters, *CBunadjInd* and *CBunadjEx*, respectively. Figure 2 presents the posterior median and 95% credible interval (CrI) estimates for the basic parameters d_{1k} , representing log odds ratios (ORs) comparing biologics versus placebo in terms of the ACR20 outcome, for each method when applied to the original and artificial datasets.

Considering the application to the original dataset first, the *CBunadjInd* method estimates the OR comparing tocilizumab versus placebo to be 3.94 (95% CrI: 1.99, 8.25). This indicates that, in the population represented by the set of trials in the synthesis, a random sample of participants allocated tocilizumab have on average almost four times the odds of achieving an ACR20 response compared to a similar sample of participants allocated

placebo. When the *CBunadjInd* method was applied to the artificial dataset, using a two-stage approach (see final paragraph in “Methods without covariates” section), the OR was estimated to be 3.10 (1.34, 6.82). This suggests that the effectiveness of tocilizumab is underestimated in the artificial dataset compared to the original dataset, albeit with significant overlap in CrIs. This result can be explained by the difference in the observed baseline response in the original dataset and the baseline response predicted by the *CBunadjInd* method when applied to the artificial dataset. Figure A.1 (Appendix A) presents the posterior estimates of the baseline response parameters ϕ_j representing the log odds of achieving an ACR20 response on placebo in each trial. For trials one and two, assessing tocilizumab versus placebo, the predicted baseline response overestimates the observed baseline response in the original dataset, which inversely impacts the basic parameter estimates.

Considering the application to the original dataset, the *CBunadjEx* method estimates the OR comparing

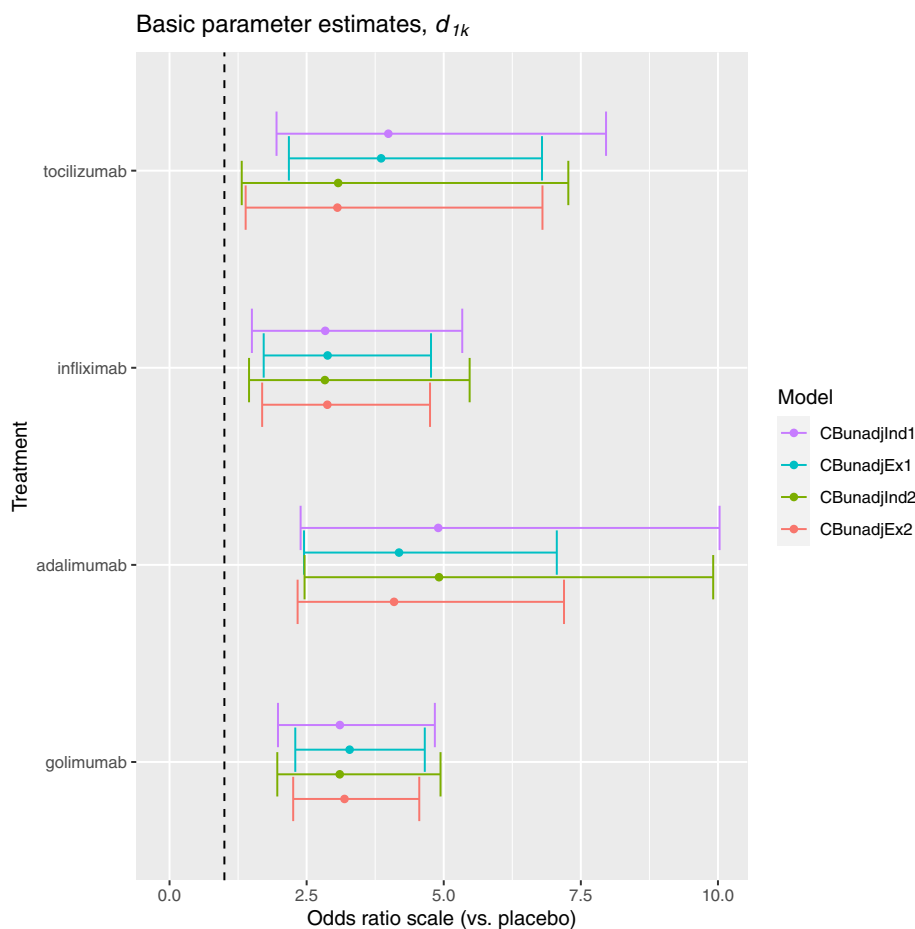


Fig. 2 Posterior median and 95% credible interval estimates for log odds ratios comparing each biologic versus placebo in terms of the ACR20 outcome, for contrast-based methods without covariates

tocilizumab versus placebo to be 3.86 (2.20, 6.89). This estimate is in agreement with the estimate from the *CBunadjInd* method applied to the original dataset, indicating that the exchangeability assumption had little impact on the baseline response estimates. Figure A.1 (Appendix A) confirms that there is little difference in baseline response estimates between *CBunadjInd* and *CBunadjEx* for trials one and two. When applied to the artificial dataset, *CBunadjEx* estimates the OR to be 3.10 (1.40, 6.96). This result is consistent with the estimate from applying *CBunadjInd* to the artificial dataset using the two-stage approach, but underestimates the OR in the original dataset. This suggests that the disagreement in the effect estimates for tocilizumab is due to the difference between the observed and predicted baseline response estimates, and that the exchangeability assumption had little impact on the results. There is significant overlap in the CrIs corresponding to the effect estimates across all applications, which indicates that the disagreement introduced by using data from SATs (i.e. removing data from an RCT control arm) is relatively limited for this example. Figure 2 also shows that uncertainty increases in treatment effect estimates corresponding to tocilizumab when the control arms are removed, but there is no change in uncertainty for the effect estimates corresponding to the other treatments.

Figure 2 suggests that the exchangeability assumption has a greater impact on effect estimates for adalimumab. Considering the application to the original dataset, the *CBunadjInd* method estimates an OR of 4.90 (2.39, 9.78) for adalimumab versus placebo. In contrast, the *CBunadjEx* method estimates an OR of 4.18 (2.44, 7.17). A similar disagreement is observed in the artificial dataset where the *CBunadjInd* method estimates an OR of 4.81 (2.34, 9.78), whilst *CBunadjEx* estimates an OR of 4.10 (2.34, 6.96). This indicates that the exchangeability assumption introduces a disagreement in the effect estimates for adalimumab in both datasets. Figure A.1 (Appendix A) shows that the baseline response estimates corresponding to trials assessing adalimumab versus placebo (in particular, trials four and six) are significantly shifted toward the mean baseline response (represented by the dashed-line). This suggests that the exchangeability assumption can lead to disagreement in effect estimates, particularly where there is significant between-study heterogeneity in baseline response estimates. It follows that adjusting the baseline response by including covariates, to account for between-study heterogeneity, could mitigate disagreement in effect estimates introduced by the exchangeability assumption. Despite evidence of disagreement in effect estimates there is significant overlap in CrIs, suggesting that the impact of assuming exchangeability is relatively limited for this example.

Arm-based methods

The arm-based method without covariate-adjustment *ABunadj* is described in “Methods without covariates” section. Figure 3 presents the posterior median and 95% CrI estimates for the pooled absolute effects θ_k representing the log odds of achieving an ACR20 outcome associated with each intervention.

Considering the application to the original dataset first, *ABunadj* estimates the pooled log odds on placebo to be -0.88 (-1.10, -0.67). This implies that, in a population represented by the trials in the synthesis, a random sample of participants allocated placebo would on average have a probability of achieving an ACR20 response of 0.29 (0.25, 0.34). For tocilizumab, the pooled log odds is estimated to be 0.33 (-0.17, 0.81), which corresponds to a response probability of 0.58 (0.46, 0.69). This suggests that tocilizumab is significantly more effective than placebo, in terms of the ACR20 outcome, and is consistent with the results from the contrast-based methods. When *ABunadj* was applied to the artificial dataset, the pooled log odds on placebo was estimated to be -0.83 (-1.09, -0.57), representing a response probability of 0.30 (0.25, 0.36). This is very similar to the estimate of the pooled response probability for *ABunadj* applied to the original dataset. The pooled log odds on tocilizumab is estimated to be 0.28 (-0.26, 0.82), corresponding to a response probability of 0.57 (0.44, 0.69). Here, the result slightly underestimates the effect estimate from the original dataset. This indicates that the removal of data on placebo response, in trials assessing tocilizumab versus placebo, impacted the pooled response estimate for tocilizumab. This may be because the *ABunadj* method assumes that response estimates are correlated at the between-study level, where the correlation ρ was estimated to be 0.29 (-0.42, 0.80) for the application to the original dataset, and 0.31 (-0.40, 0.84) for the application to the artificial dataset. There is significant overlap in CrIs for each intervention associated with each dataset, which suggests the impact of the removal of data was limited in this example.

Table B.2 (Appendix B) lists posterior median and 95% CrI estimates for the model parameters corresponding to each application of *ABunadj*. The between-study heterogeneity τ was estimated to be 0.34 (0.22, 0.53) when *ABunadj* was applied to the original dataset, which is slightly smaller than the estimate of 0.35 (0.23, 0.55) from the application to the artificial dataset.

Methods with covariate-adjustment

Contrast-based methods

In this section, we report the results from applying methods which include an adjustment of the baseline response for RA duration. We begin by considering the results from the contrast-based methods with exchangeable baseline response parameters. Figure 4 presents the posterior

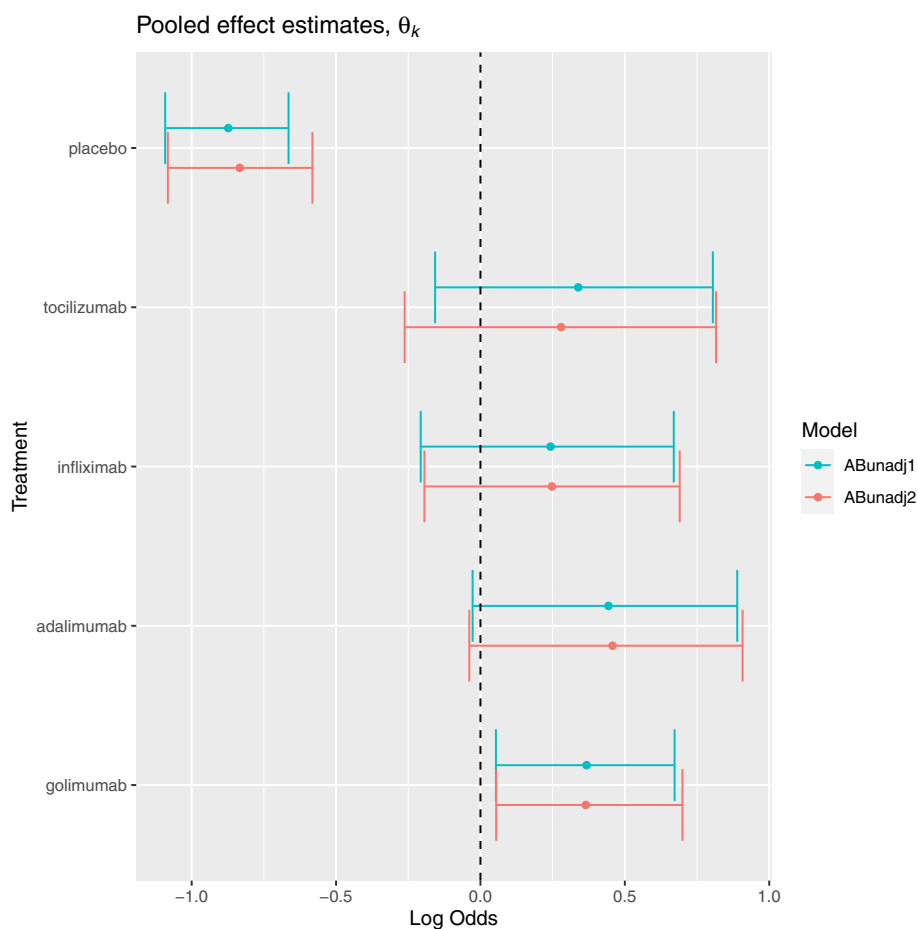


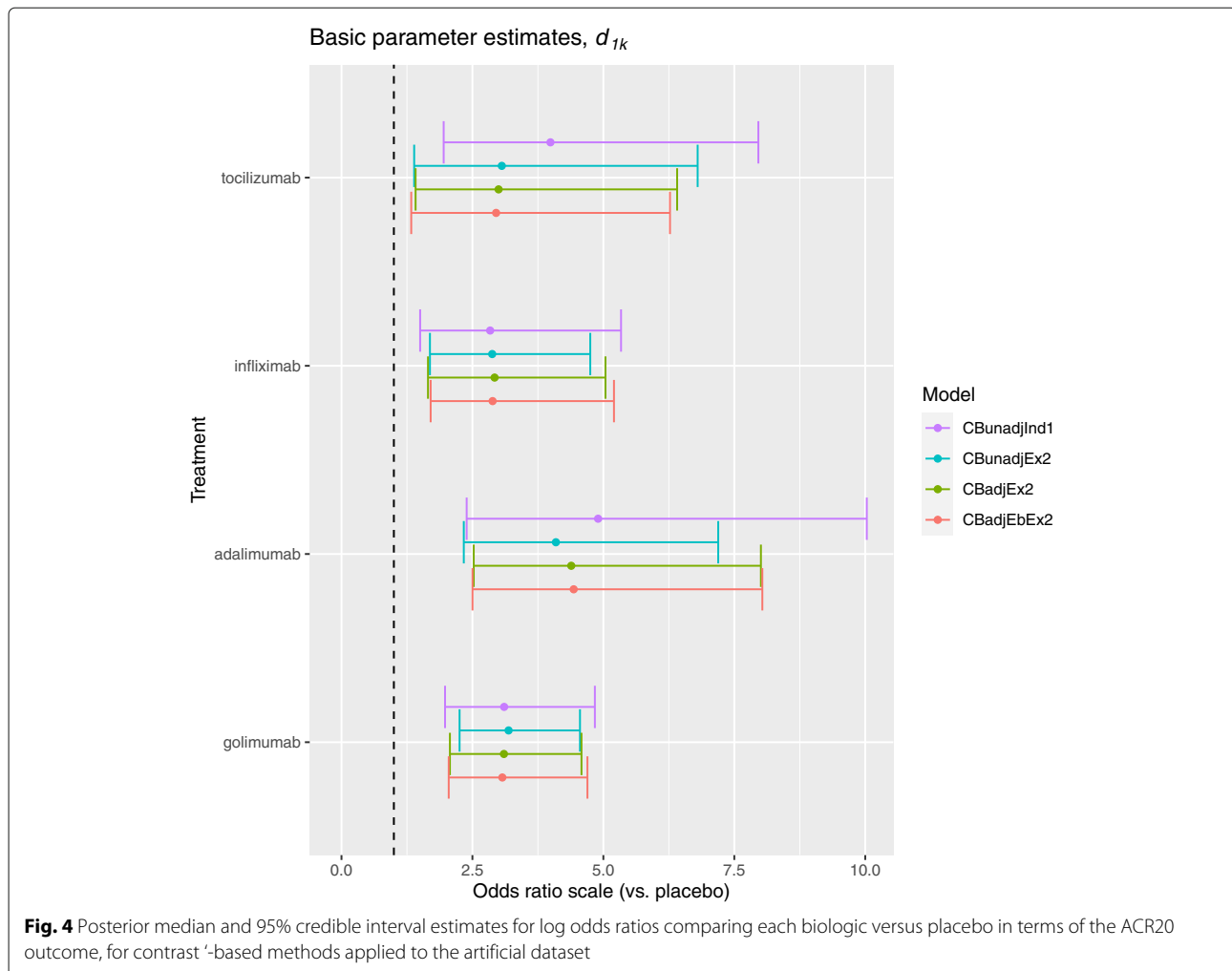
Fig. 3 Posterior median and 95% credible interval estimates for the pooled log odds on each intervention in terms of the ACR20 outcome, for arm-based methods without covariates

median and 95% CrI estimates for the unadjusted and adjusted log ORs, corresponding to the contrast-based methods applied to the artificial dataset.

For tocilizumab, the OR estimated by the *CBadjEx* method, which included adjustment for the mean RA duration, was 3.00 (1.42, 6.42) which suggests that in a trial of newly-diagnosed RA patients (i.e. mean RA duration equal to zero years), participants allocated tocilizumab have approximately three times greater odds of achieving an ACR20 response compared to a similar group of participants allocated placebo. This result underestimates the unadjusted OR 4.10 (2.34, 6.96) estimated by the *CBunadj* method, which represents the average effect of tocilizumab across varying levels of RA duration. This difference in results indicates that the effect of tocilizumab may vary with RA duration, and Table B.3 (Appendix B) confirms the presence of an across trial covariate effect (α^A) of -0.10 (-0.20, 0.00). This implies that the effect of a biologic versus placebo in a trial in which a sample of patients have had RA for an average of two years will be

approximately 10% ($100 \times (1 - \exp(-0.10))$) smaller than in a trial in which a similar sample of patients have had RA for an average of one year.

The *CBadjEbEx* method, which also includes an adjustment for the within-trial covariate effect of RA duration, estimates an OR of 2.94 (1.34, 6.30) comparing tocilizumab versus placebo. This result is in strong agreement with the *CBadjEx* method, which suggests a lack of a covariate effect at the participant-level. Table B.3 (Appendix B) lists the within-trial covariate effect (α^W) estimate as 0.00 (-0.02, 0.01). The difference between across- and within-trial covariate effect estimates indicates aggregation bias, and may be due to a difference in the distribution of RA duration within trials and the distribution of mean RA duration across trials. Figure 4 also presents the posterior estimates for the unadjusted log OR (comparing tocilizumab versus placebo) from the contrast-based method with independent baseline response parameters (*CBunadjInd*) applied to the original dataset. The unadjusted OR, estimated in the artificial



dataset, underestimates the unadjusted OR estimated in the original dataset, and this difference does not decrease after adjustment for the covariate effect of RA duration. This shows that the adjustment has not mitigated the disagreement in results between the original and artificial datasets. Figure A.2 (Appendix A) presents the posterior estimates for the baseline response corresponding to each method. For trials one and two, there is little difference in the estimates before and after adjustment. This could be due to the fact that RA duration does not explain heterogeneity in the baseline response.

For adalimumab versus placebo, the adjusted OR estimated by the *CBadjEx* method was 4.39 (2.53, 8.00). This estimate is more consistent with the unadjusted OR from the original dataset, estimated by *CBunadjInd* to be 4.90 (2.39, 10.07). In Figure A.2 (Appendix A), the adjusted baseline response estimates in trials four, five, and six (i.e. trials assessing adalimumab versus placebo) show some discrepancy with respect to the unadjusted baseline response estimates from the original dataset.

Table B.3 (Appendix B) shows that there is a reduction in between-study heterogeneity in baseline response (σ_ϕ) after adjusting for the mean RA duration, from 0.33 (0.15, 0.60) estimated by *CBunadjEx* to 0.28 (0.12, 0.56) estimated by *CBadjEx*. This indicates that the inclusion of mean RA duration is explaining some of the heterogeneity in baseline response, although there is little change in the mean baseline response m_ϕ .

Arm-based methods

Figure 5 presents the posterior median and 95% CrI estimates for the unadjusted and adjusted arm-based methods, applied to the artificial dataset. For placebo, the unadjusted pooled log odds of achieving an ACR20 response was estimated to be -0.83 (-1.08, -0.58), corresponding to a response probability of 0.30 (0.25, 0.36). This result was relatively unchanged after adjustment for mean RA duration at the arm-level, where the adjusted pooled log odds were estimated to be -0.85 (-1.09, -0.61) and the corresponding response probability was 0.30 (0.25, 0.36).

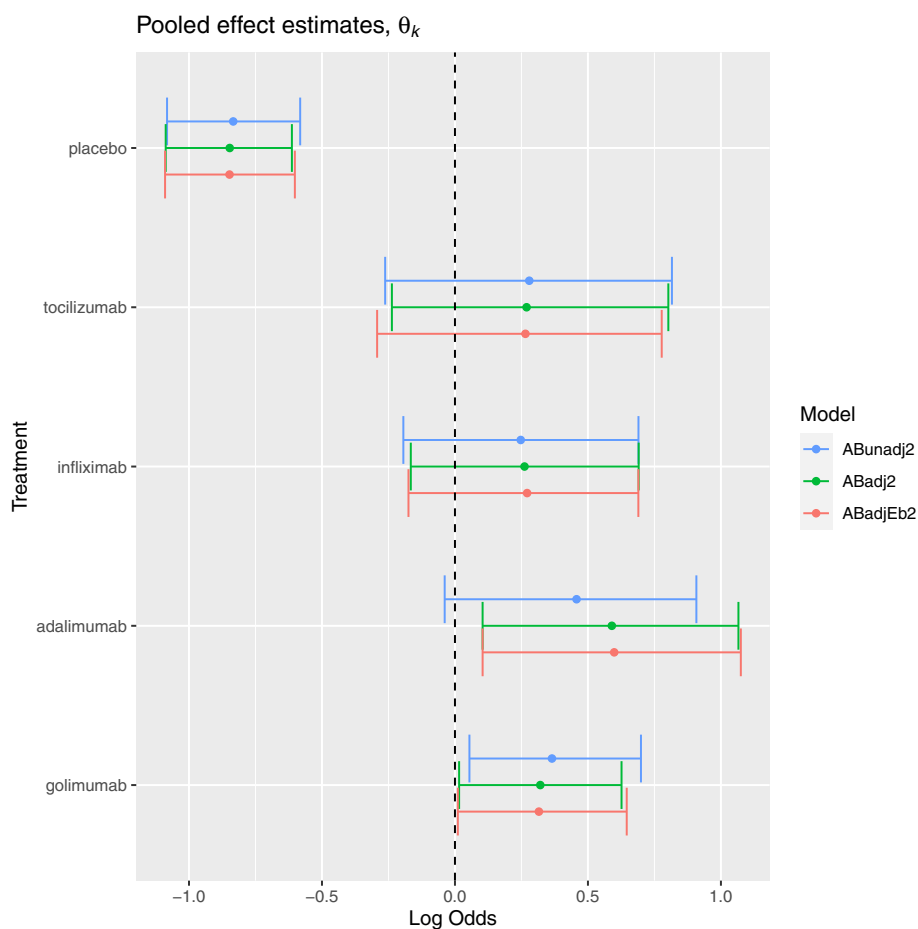


Fig. 5 Posterior median and 95% credible interval estimates for the pooled log odds on each intervention in terms of the ACR20 outcome, for arm-based methods applied to the artificial dataset

Similarly, the adjustment did not seem to have an impact on effect estimates for tocilizumab, where the unadjusted pooled log odds were estimated to be 0.28 (-0.26, 0.82) and the adjusted pooled log odds were 0.27 (-0.24, 0.80). In contrast, the effect on adalimumab was larger after adjustment indicating the presence of a covariate effect across trials. The unadjusted pooled log odds of response were 0.59 (0.10, 1.07) corresponding to a response probability of 0.64 (0.52, 0.74), whilst the adjusted pooled log odds of response were 0.46 (-0.04, 0.91) corresponding to a response probability of 0.61 (0.49, 0.71). Table B.4 (Appendix B) lists the posterior estimates for the across-trial covariate effect β^A , estimated by the *ABadj* method to be -0.06 (-0.14, 0.02). This suggests that there is a decrease of approximately 6% in the pooled log odds of achieving an ACR20 response, for each additional year a sample of patients spend diagnosed with RA. There does not appear to be evidence of a within-trial covariate effect β^W , which is estimated by the *ABadjEb* method to be 0.00 (-0.02, 0.01). Although there does not seem to

be a noticeable difference in the between-study heterogeneity (τ) before and after adjustment, the correlation ρ decreases from 0.31 (-0.40, 0.87) to 0.23 (-0.49, 0.82).

Table B.5 (Appendix B) lists deviance information criterion (DIC) statistics, a measure of model fit [56] where smaller values are preferred, for each method applied to the artificial dataset. There does not seem to be a significant difference in model fit between the arm- and contrast-based methods. The inclusion of RA duration as a covariate did not provide a significant reduction in DIC.

Discussion

In this paper, we develop contrast- and arm-based methods for NMA synthesising SATs and RCTs, using a mixture of IPD and AD, for a dichotomous outcome. We also extended these methods to adjust the baseline response for a covariate, to account for within- and across-trial covariate effects. We applied the methods to a dataset of 14 RCTs assessing biologics versus placebo for RA and an artificial dataset, including two SATs (assessing

tocilizumab), generated from the original dataset of 14 RCTs by removing the control arm in the two tocilizumab trials. Placebo was designated as the reference treatment for the network, and so the contrast-based methods assumed a fixed baseline treatment across trials. The contrast-based unadjusted methods underestimated the treatment effect for tocilizumab versus placebo, when applied to the artificial dataset compared to the original dataset. This difference appeared to be caused by an overestimate of the predicted baseline response for the SATs. The contrast-based methods with exchangeable baseline response parameters showed disagreement with the methods assuming independent parameters, in treatment effect estimates for adalimumab, in both datasets. This disagreement seemed to be caused by a shift toward the mean baseline response associated with the exchangeability assumption for the baseline response parameters. Despite these discrepancies, there was large overlap in posterior distributions for treatment effect parameters between applications to the two datasets. Under both the contrast- and arm-based parametrisation, there was some evidence of an across-trial covariate effect associated with mean RA duration, but a within-trial covariate effect was not detected. However, the adjustment appeared to make little difference to the treatment effect estimates. The arm-based methods showed greater agreement in effect estimates between the applications to the two datasets.

Our applied example enabled an assessment of the developed methods to synthesise SATs and RCTs against a synthesis of RCTs alone. A similar approach was used by Beliveau et al. [27], where connected networks of RCTs were disconnected by removing data from trials, to understand the impact of assuming exchangeable baseline response parameters to connect the disconnected networks. In the majority of cases, the exchangeability assumption seemed to make relatively little difference to the results, as posterior distributions for treatment effect parameters showed large overlap. This is consistent with the findings from our applied example, where there was also significant overlap in posterior distributions for treatment effect parameters in the artificial and original datasets, although the effect estimate for tocilizumab had greater precision in the original dataset where more data were available. Despite this, there was a shift in effect estimates associated with assuming exchangeable baseline response parameters, due to the influence of the mean baseline response (e.g. for adalimumab). To better understand this discrepancy, we also applied the contrast-based method with independent baseline response parameters to the artificial dataset using a two-stage approach. This approach is recommended by Dias et al. [52] to ensure treatment effect estimates are not influenced by assumptions regarding the baseline response. This mitigated the discrepancy associated with assuming exchangeable

baseline response parameters, and the results were more consistent with the application to the original dataset for adalimumab. However, the discrepancy remained for treatment effect estimates associated with tocilizumab, indicating that the predicted baseline response in the artificial dataset was not consistent with the observed baseline response in the original dataset. This implies that incorporating SATs into a NMA to estimate relative treatment effects depends significantly on how representative the mean baseline response across the RCTs in the synthesis is of the hypothetical baseline response for the participants in the SATs. This highlights the importance of considering whether each trial in the synthesis is informative about the target population in the SATs [52]. It also suggests that explaining between-study heterogeneity in the baseline response, by including covariates to account for differences in prognostic factors in trial participants, could improve the predicted baseline response. Thus, incorporating SATs into a synthesis may be beneficial when decision-makers require relative effect estimates for a particular treatment, and are reluctant to delay the decision until a RCT has been undertaken. A review of HTA submissions found an increasing use of data from SATs in recent years, particularly to provide evidence on an external patient cohort to determine relative treatment effectiveness [57]. Malottki et al. conducted a multiple technology appraisal of five biologics as second-line treatment for RA [58], in which they identified five RCTs and 28 SATs relevant to the decision problem. Only two biologics were compared head-to-head via an indirect comparison based on two RCTs and the SATs were not included in any quantitative synthesis, resulting in large uncertainty in effectiveness estimates.

When seeking to combine data from SATs and RCTs in a meta-analysis (or NMA), particularly in a decision-making context, it can be beneficial to implement a range of methods to understand how alternative parametrisations influence the results. A comparison of arm- and contrast-based methods requires the calculation of additional quantities as a function of model parameters. For arm-based methods, a pooled relative treatment effect can be calculated from the pooled arm-level outcomes, $d_{1k} = \theta_k - \theta_1$. For contrast-based methods, a pooled absolute treatment effect can be calculated from the mean baseline response and the pooled relative treatment effect, $\theta_k = m_\phi + d_{ik}$.

We also applied methods which included adjustment for a covariate, to account for between-study heterogeneity in baseline response, considering RA duration to be a prognostic factor associated with ACR20 response. In the artificial dataset, an across-trial covariate effect was associated with mean RA duration. The results indicate that the adjustment provided a small reduction in between-study heterogeneity, but this did not make treatment

effect estimates more consistent with those from the application to the original dataset. The residual between-study heterogeneity estimated after adjustment suggested the presence of unmeasured prognostic factors. Despite some evidence of an across-trial covariate effect, there was no evidence of a within-trial covariate effect, indicating the presence of aggregation bias. The difference between these findings may be due to the impact of unmeasured covariates, associated with both an ACR20 response and mean RA duration across trials. An exploratory analysis of the IPD, corresponding to the two tocilizumab trials for which it was available (see Table 1), was undertaken to assess the distribution of RA duration across patients within each trial. The distribution of RA duration within each trial did not differ significantly from the distribution of mean RA duration across the other 12 RCTs (listed in Table 1). Here, the goal of the adjustment for RA duration was to use data from RCTs to predict baseline response estimates more representative (compared to the unadjusted estimates) of the patient population in the SATs. Whilst we focus on adjustment on the baseline response, other authors have looked at how a treatment effect varies with respect to a trial characteristic (treatment-covariate interaction) [59], via meta-regression or network meta-regression. Population-adjustment methods, such as matching-adjustment indirect comparison and simulated treatment comparison methods, have been proposed for a similar goal where IPD are available for one trial and only AD are available for another [16]. Similarly, NMA methods have been developed which combine IPD and AD whilst accounting for a non-linear association between the treatment effect and a particular covariate to mitigate aggregation bias [60]. More recently, Hong et al. propose NMA methods using AD to define a prior distribution for pooled treatment effect and treatment-covariate interaction parameters estimated from IPD, whilst down-weighting the AD via a power term or a commensurability parameter, which also avoid introducing aggregation bias into the synthesis [61].

Due to the limited data available in the artificial dataset, we assumed a common within-trial covariate effect parameter. An assumption of independent within-trial covariate effects may have accounted for the difference in covariate distributions between the two trials for which IPD were available. Thom et al. were unable to detect across-trial covariate effects in a similar application due to limited data [20]. In our applied example, there were no trials comparing biologics head-to-head, preventing an assessment of inconsistency via the recommended methods [62]. Due to this, and the limited data, we did not consider adjustment of treatment effects for potential effect modifiers.

In this paper, we made the assumption that the intervention arms in RCTs assessing tocilizumab versus placebo

were representative of SATs assessing tocilizumab. In practice, the populations represented by the trials may differ due to factors associated with trial design (e.g. participant inclusion criteria). A better assessment of the methods would be provided by a simulation study, where data are simulated for a network of RCTs and then a subset of data corresponding to participants in control arms are removed, to provide a more accurate representation of data from SATs. This would allow the methods to be applied in a greater number of scenarios to understand how differences in trial populations can influence treatment effect estimates when incorporating SATs into a network. Furthermore, many of the trials considered in the applied example consisted of multiple intervention arms, from which data were combined to create a single intervention arm. The proposed methods can be extended for application to multi-arm RCTs, by considering within-trial correlation between treatment effects relative to a common baseline treatment [51].

Conclusion

In this paper, we propose contrast- and arm-based NMA methods which allow the incorporation of SATs into a synthesis of RCTs, to estimate relative treatment effectiveness when there is no alternative (due to a disconnected network). Further work is required to understand how adjustment for covariate effects can be incorporated to predict a more representative baseline response estimate for SATs, and treatment effect estimates more consistent with randomised evidence.

Abbreviations

NMA: Network meta-analysis; SAT: Single arm trial; IPD: Individual participant data; AD: Aggregate data; RCT: Randomised controlled trial; CB: Contrast based; AB: Arm based; RA: Rheumatoid arthritis; HTA: Health technology assessment; bDMARD: biologic disease modifying anti rheumatic drug; ACR: American College of Rheumatology; MCMC: Markov chain Monte Carlo; CrI: Credible interval; OR: Odds ratio; DIC: Deviance information criterion

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12874-022-01657-y>.

Additional file 1: Supplementary material for: bayesian network meta-analysis methods for combining individual participant data and aggregate data from single arm trials and randomised controlled trials. Includes: Figures A.1-2, Tables B.1-5, Stan programming code for models.

Acknowledgements

This research used the ALICE/SPECTRE High Performance Computing Facility at the University of Leicester. The authors would like to thank Roche for providing access to the individual participant data used in the analysis.

Authors' contributions

JS and SB conceptualised the study. JS, LW, and SG undertook the data curation. JS undertook the formal analysis and contributed the original draft of the manuscript. SB, KRA and CG provided supervision for JS. SG and RDR provided additional methodological advice. All authors contributed to manuscript revisions and approved the final manuscript.

Funding

JS was supported by UK National Institute for Health Research (NIHR) Doctoral Research Fellowship [award no. NIHR300190]. LW was supported by NIHR Methods Fellowship [award no. RM-FI-2017-08-027] and subsequently by NIHR Pre-doctoral Fellowship [NIHR301013]. KRA and SB were supported by the UK Medical Research Council [grant no. MR/R025223/1]. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The datasets analysed during the study are included in this published article (Table 1, applicable only to data extracted from published manuscripts). For eligible studies qualified researchers may request access to individual patient level clinical data through a data request platform. At the time of writing this request platform is Vivli (<https://vivli.org/ourmember/roche/>). For up to date details on Roche's Global Policy on the Sharing of Clinical Information and how to request access to related clinical study documents, see here: https://go.roche.com/data_sharing.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

JS, SG, LW, RDR, and CG do not have conflict of interest. SB has served as a paid consultant providing methodological advice to NICE, Roche and RTI Health Solutions, has received payments for educational events from Roche, and has received research funding from European Federation of Pharmaceutical Industries & Associations (EFPIA) and Johnson & Johnson. KRA has served as a paid consultant, providing methodological advice, to; Abbvie, Amaris, Allergan, Astellas, AstraZeneca, Boehringer Ingelheim, Bristol-Meyers Squibb, Creativ-Ceutical, GSK, ICON/Oxford Outcomes, Ipsen, Janssen, Eli Lilly, Merck, NICE, Novartis, NovoNordisk, Pfizer, PRMA, Roche and Takeda, and has received research funding from Association of the British Pharmaceutical Industry (ABPI), European Federation of Pharmaceutical Industries & Associations (EFPIA), Pfizer, Sanofi and Swiss Precision Diagnostics. He is a Partner and Director of Visible Analytics Limited, a healthcare consultancy company.

Author details

¹Biostatistics Research Group, Department of Health Sciences, University of Leicester, Leicester, UK. ²F. Hoffmann-La Roche Ltd, Basel, Switzerland. ³Centre for Prognosis Research, School of Medicine, University of Keele, Staffordshire, UK. ⁴Department of Statistics, University of Warwick, Coventry, UK. ⁵Leicester Real World Evidence Unit, Diabetes Research Centre, University of Leicester, Leicester, UK.

Received: 24 December 2021 Accepted: 23 May 2022

Published online: 11 July 2022

References

1. Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. *BMJ*. 1997;315(7121):1533–37.
2. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177–88.
3. Lu G, Ades A. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004;23(20):3105–24.
4. Anderson M, Naci H, Morrison D, Osipenko L, Mossialos E. A review of nice appraisals of pharmaceuticals 2000–2016 found variation in establishing comparative clinical effectiveness. *J Clin Epidemiol*. 2019;105:50–59.
5. Woolacott N, Corbett M, Jones-Diette J, Hodgson R. Methodological challenges for the evaluation of clinical effectiveness in the context of accelerated regulatory approval: an overview. *J Clin Epidemiol*. 2017;90:108–18.
6. Bell H, Wailoo A, Hernandez M, Grieve R, Faria R, Gibson L, Grimm S. The use of real world data for the estimation of treatment effects in nice decision making. Decision Support Unit, SCHARR, University of Sheffield. 2016.
7. Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *J Clin Epidemiol*. 2002;55(1):86–94.
8. Riley RD, Steyerberg EW. Meta-analysis of a binary outcome using individual participant data and aggregate data. *Res Synth Methods*. 2010;1(1):2–19.
9. Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI. Individual patient-versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Stat Med*. 2002;21(3):371–87.
10. Jackson C, Best N, Richardson S. Hierarchical related regression for combining aggregate and individual data in studies of socio-economic disease risk factors. *J R Stat Soc Ser A Stat Soc*. 2008;171(1):159–78.
11. Riley RD, Lambert PC, Staessen JA, Wang J, Gueyffier F, Thijs L, Bouillon F. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Stat Med*. 2008;27(11):1870–93.
12. Sutton AJ, Kendrick D, Coupland CA. Meta-analysis of individual-and aggregate-level data. *Stat Med*. 2008;27(5):651–69.
13. Saramago P, Sutton AJ, Cooper NJ, Manca A. Mixed treatment comparisons using aggregate and individual participant level data. *Stat Med*. 2012;31(28):3516–36.
14. Donegan S, Williamson P, D'Alessandro U, Smith CT. Assessing the consistency assumption by exploring treatment by covariate interactions in mixed treatment comparison meta-analysis: individual patient-level covariates versus aggregate trial-level covariates. *Stat Med*. 2012;31(29):3840–57.
15. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. Methods for population-adjusted indirect comparisons in health technology appraisal. *Med Dec Making*. 2018;38(2):200–11.
16. Phillippo DM, Dias S, Elstada A, Ades A, Welton NJ. Population adjustment methods for indirect comparisons: a review of national institute for health and care excellence technology appraisals. *Int J Technol Assess Health Care*. 2019;35(3):221–28.
17. Phillippo DM, Dias S, Belger M, Brnabic A, Schacht A, Saure D, Kadziola Z, Welton NJ. Multilevel network meta-regression for population-adjusted treatment comparisons. *J R Stat Soc Ser A Stat Soc*. 2020;183(3):1189–210.
18. Kanters S, Karim ME, Thorlund K, Anis AH, Zoratti M, Bansback N. Comparing the use of aggregate data and various methods of integrating individual patient data to network meta-analysis and its application to first-line art. *BMC Med Res Methodol*. 2021;21(1):1–15.
19. Begg CB, Pilote L. A model for incorporating historical controls into a meta-analysis. *Biometrics*. 1991;47(3):899–906.
20. Thom HH, 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;15(1):1–16.
21. Zhang J, Ko C-W, Nie L, Chen Y, Tiwari R. Bayesian hierarchical methods for meta-analysis combining randomized-controlled and single-arm studies. *Stat Methods Med Res*. 2019;28(5):1293–310.
22. Hong H, Fu H, Price KL, Carlin BP. Incorporation of individual-patient data in network meta-analysis for multiple continuous endpoints, with application to diabetes treatment. *Stat Med*. 2015;34(20):2794–819.
23. Hong H, Chu H, Zhang J, Carlin BP. A bayesian missing data framework for generalized multiple outcome mixed treatment comparisons. *Res Synth Methods*. 2016;7(1):6–22.
24. Wang Z, Lin L, Murray T, Hodges JS, Chu H. Bridging randomized controlled trials and single-arm trials using commensurate priors in arm-based network meta-analysis. *Ann Appl Stat*. 2021;15(4):1767–87.
25. Dias S, Ades A. Absolute or relative effects? arm-based synthesis of trial data. *Res Synth Methods*. 2016;7(1):23.
26. Hong H, Chu H, Zhang J, Carlin BP. Rejoinder to the discussion of "a bayesian missing data framework for generalized multiple outcome mixed treatment comparisons," by s. dias and ae ades. *Res Synth Methods*. 2016;7(1):29.
27. Bélièveau A, Goring S, Platt RW, Gustafson P. Network meta-analysis of disconnected networks: how dangerous are random baseline treatment effects?. *Res Synth Methods*. 2017;8(4):465–74.

28. White IR, Turner RM, Karahalios A, Salanti G. A comparison of arm-based and contrast-based models for network meta-analysis. *Stat Med*. 2019;38(27):5197–213.
29. Schmitz S, Maguire A, Morris J, Ruggeri K, Haller E, Kuhn I, Leahy J, Homer N, Khan A, Bowden J, et al. The use of single armed observational data to closing the gap in otherwise disconnected evidence networks: a network meta-analysis in multiple myeloma. *BMC Med Res Methodol*. 2018;18(1):1–18.
30. Leahy J, Thom H, Jansen JP, Gray E, O'Leary A, White A, Walsh C. Incorporating single-arm evidence into a network meta-analysis using aggregate level matching: assessing the impact. *Stat Med*. 2019;38(14):2505–23.
31. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham III CO, Birnbaum NS, Burmester GR, Bykerk VP, Cohen MD, et al. 2010 rheumatoid arthritis classification criteria: an american college of rheumatology/european league against rheumatism collaborative initiative. *Arthritis Rheum*. 2010;62(9):2569–81.
32. Aletaha D, Smolen JS. Diagnosis and management of rheumatoid arthritis: a review. *JAMA*. 2018;320(13):1360–72.
33. Genovese MC, McKay JD, Nasonov EL, Mysler EF, da Silva NA, Alecock E, Woodworth T, Gomez-Reino JJ. 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 Off J Am Coll Rheumatol*. 2008;58(10):2968–80.
34. Smolen JS, Beaulieu A, Rubbert-Roth A, Ramos-Remus C, Rovensky J, Alecock E, Woodworth T, Alten R, Investigators O, et al. Effect of interleukin-6 receptor inhibition with tocilizumab in patients with rheumatoid arthritis (optin study): a double-blind, placebo-controlled, randomised trial. *Lancet*. 2008;371(9617):987–97.
35. Maini R, St Clair EW, Breedveld F, Furst D, Kalden J, Weisman M, Smolen J, Emery P, Harriman G, Feldmann M, et al. Infliximab (chimeric anti-tumor necrosis factor α monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomised phase iii trial. *Lancet*. 1999;354(9194):1932–39.
36. Weinblatt ME, Keystone EC, Furst DE, Moreland LW, Weisman MH, Birbara CA, Teoh LA, Fischkoff SA, Chartash EK. Adalimumab, a fully human anti-tumor necrosis factor α monoclonal antibody, for the treatment of rheumatoid arthritis in patients taking concomitant methotrexate: the armada trial. *Arthritis Rheum*. 2003;48(1):35–45.
37. Keystone EC, Kavanaugh AF, Sharp JT, Tannenbaum H, Hua Y, Teoh LS, Fischkoff SA, Chartash EK. 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;50(5):1400–11.
38. Kremer J, Ritchlin C, Mendelsohn A, Baker D, Kim L, Xu Z, Han J, Taylor P. Golimumab, a new human anti-tumor necrosis factor α antibody, administered intravenously in patients with active rheumatoid arthritis: Forty-eight-week efficacy and safety results of a phase iii randomized, double-blind, placebo-controlled study. *Arthritis Rheum*. 2010;62(4):917–28.
39. Chen D-Y, Chou S-J, Hsieh T-Y, Chen Y-H, Chen H-H, Hsieh C-W, Lan J-L. 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;108(4):310–19.
40. Kay J, Matteson EL, Dasgupta B, Nash P, Durez P, Hall S, Hsia EC, Han J, Wagner C, Xu Z, 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;58(4):964–75.
41. Keystone E, Heijde DVD, Mason Jr D, Landewé R, Vollenhoven RV, Combe B, Emery P, Strand V, Mease P, Desai C, et al. Certolizumab pegol plus methotrexate is significantly more effective than placebo plus methotrexate in active rheumatoid arthritis: findings of a fifty-two-week, phase iii, multicenter, randomized, double-blind, placebo-controlled, parallel-group study. *Arthritis Rheum Off J Am Coll Rheumatol*. 2008;58(11):3319–29.
42. Smolen J, Landewé RB, Mease P, Brzezicki J, Mason D, Luijckens K, van Vollenhoven RF, Kavanaugh A, Schiff M, Burmester GR, et al. Efficacy and safety of certolizumab pegol plus methotrexate in active rheumatoid arthritis: the rapid 2 study. a randomised controlled trial. *Ann Rheum Dis*. 2009;68(6):797–804.
43. Keystone E, Genovese M, Klareskog L, Hsia E, Miranda P, et al. Golimumab, a human antibody to tnf- α given by monthly subcutaneous injections, in active rheumatoid arthritis despite methotrexate therapy: The go-forward study. *Ann Rheum Dis*. 2009;68:789–96.
44. Schiff M, Keiserman M, Codding C, Songcharoen S, Berman A, Nayiager S, Saldate C, Li T, Aranda R, Becker J, 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;67(8):1096–103.
45. Choy E, McKenna F, Vencovsky J, Valente R, Goel N, VanLunen B, Davies O, Stahl H-D, Alten R. Certolizumab pegol plus mtx administered every 4 weeks is effective in patients with ra who are partial responders to mtx. *Rheumatology*. 2012;51(7):1226–34.
46. Emery P, Fleischmann RM, Moreland LW, Hsia EC, Strusberg I, Durez P, Nash P, Amante EJB, Churchill M, Park W, et al. Golimumab, a human anti-tumor necrosis factor α monoclonal antibody, injected subcutaneously every four weeks in methotrexate-naive patients with active rheumatoid arthritis: twenty-four-week results of a phase iii, multicenter, randomized, double-blind, placebo-controlled study of golimumab before methotrexate as first-line therapy for early-onset rheumatoid arthritis. *Arthritis Rheum Off J Am Coll Rheumatol*. 2009;60(8):2272–83.
47. Kim J, Ryu H, Yoo D-H, Park S-H, Song G-G, Park W, Cho C-S, Song Y-W. A clinical trial and extension study of infliximab in korean patients with active rheumatoid arthritis despite methotrexate treatment. *J Korean Med Sci*. 2013;28(12):1716–22.
48. Tanaka Y, Harigai M, Takeuchi T, Yamanaka H, Ishiguro N, Yamamoto K, Miyasaka N, Koike T, Kanazawa M, Oba T, et al. Golimumab in combination with methotrexate in japanese patients with active rheumatoid arthritis: results of the go-forth study. *Ann Rheum Dis*. 2012;71(6):817–24.
49. Weinblatt ME, Bingham CO, Mendelsohn AM, Kim L, Mack M, Lu J, Baker D, Westhovens R. Intravenous golimumab is effective in patients with active rheumatoid arthritis despite methotrexate therapy with responses as early as week 2: results of the phase 3, randomised, multicentre, double-blind, placebo-controlled go-further trial. *Ann Rheum Dis*. 2013;72(3):381–89.
50. Röver C, Bender R, Dias S, Schmid CH, Schmidli H, Sturtz S, Weber S, Friede T. On weakly informative prior distributions for the heterogeneity parameter in bayesian random-effects meta-analysis. *Res Synth Methods*. 2021;12(4):448–74.
51. Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials. 2011. Last updated September 2016; available from <http://www.nicedsu.org.uk>.
52. Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU Technical Support Document 5: Evidence synthesis in the baseline natural history model. 2011. Last updated April 2012; available from <http://www.nicedsu.org.uk>.
53. Zhang J, Ko C-W, Nie L, Chen Y, Tiwari R. Bayesian hierarchical methods for meta-analysis combining randomized-controlled and single-arm studies. *Stat Methods Med Res*. 2019;28(5):1293–310.
54. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A. Stan: A probabilistic programming language. *J Stat Softw*. 2017;76(1):1–32.
55. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci*. 1992;7(4):457–72.
56. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Ser B Stat Methodol*. 2002;64(4):583–639.
57. Patel D, Grimson F, Mihaylova E, Wagner P, Warren J, van Engen A, Kim J. Use of external comparators for health technology assessment submissions based on single-arm trials. *Value Health*. 2021;24(8):1118–125.
58. Malottki K, Barton P, Tsourapas A, Uthman AO, Liu Z, Routh K, et al. Adalimumab, etanercept, infliximab, rituximab and abatacept for the

treatment of rheumatoid arthritis after the failure of a tumour necrosis factor inhibitor: a systematic review and economic evaluation. *Health Technol Assess.* 2011;15(14).

59. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted?. *Stat Med.* 2002;21(11):1559–73.
60. Jansen JP. Network meta-analysis of individual and aggregate level data. *Res Synth Methods.* 2012;3(2):177–90.
61. Hong H, Fu H, Carlin BP. Power and commensurate priors for synthesizing aggregate and individual patient level data in network meta-analysis. *J R Stat Soc: Ser C: Appl Stat.* 2018;67(4):1047–69.
62. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. NICE DSU Technical Support Document 4: Inconsistency in Networks of Evidence Based on Randomised Controlled Trials. 2011. Last updated April 2014; available from <http://www.nicedsu.org.uk>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

