Hierarchical network meta-analysis models to address sparsity of events and differing treatment classifications with regard to adverse outcomes

Fiona C. Warren[a*†], Keith R. Abrams[b], Alex J. Sutton[b]

[a]University of Exeter Medical School, St Luke's Campus, Magdalen Road, Exeter, EX1 2LU, UK
[b]Department of Health Sciences, University of Leicester, University Road, Leicester, LE1 7RH
*Correspondence to: Fiona C. Warren, University of Exeter Medical School, Smeall Building, St Luke's Campus, Magdalen Road, Exeter, EX1 2LU, UK.
[†]E-mail: f.c.warren@exeter.ac.uk

Abstract
Meta-analysis for adverse events resulting from medical interventions has many challenges, in part due to small numbers of such events within primary studies. Furthermore, variability in drug dose, potential differences between drugs within the same pharmaceutical class, and multiple indications for a specific treatment, can all add to the complexity of the evidence base.
This paper explores the use of synthesis methods, incorporating mixed treatment comparisons, to estimate the risk of adverse events for a medical intervention, while acknowledging and modelling the complexity of the structure of the evidence base. The motivating example was the effect on malignancy of three anti-TNF drugs (etanercept, adalimumab and infliximab) indicated to treat rheumatoid arthritis. Using data derived from 13 primary studies, a series of meta-analysis models of increasing complexity were applied. Models ranged from a straightforward comparison of anti-TNF against non-anti-TNF controls, to more complex models in which a treatment was defined by individual drug and its dose. Hierarchical models to allow 'borrowing strength' across treatment classes and dose levels, and models involving constraints on the impact of dose level, are described.
These models provide a flexible approach to estimating sparse, often adverse, outcomes associated with interventions. Each model makes its own set of assumptions, and approaches to assessing goodness of fit of the various models will usually be extremely limited in their effectiveness, due to the sparse nature of the data. Both methodological and clinical considerations are required to fit realistically complex models in this area, and to evaluate their appropriateness.

Keywords: network meta-analysis; mixed treatment comparisons; hierarchical models; anti-TNF drugs; rheumatoid arthritis.

1. Introduction

Meta-analysis of adverse outcomes is frequently used [1-2], and is valuable because individual studies often lack power to detect adverse outcomes. Application of meta-analysis methods to adverse event data is associated with a specific set of challenges [3], including the often infrequent occurrence of such adverse events. This issue is compounded when combined with issues whereby data may be available for similar but not identical treatments (and for different treatment indications). Ignoring such subtleties in the structure of the evidence base and pooling all 'intervention' vs 'control' data into an overall estimate would maximise the power of the analysis to detect an increased risk. However, this approach would potentially lose valuable information regarding individual treatment aspects (such as dose or individual drug), or the indications for treatment. This information would be conserved using a 'splitting' approach across treatments and/or indications, but the data would inevitably be 'spread thinner' across a greater number of nodes within a meta-analysis network.

The aim of this paper is to explore, develop and apply evidence synthesis methods for the synthesis of adverse event data, drawing on recent developments in evidence synthesis methodology used in other contexts. Section 2 briefly reviews relatively recent developments in evidence synthesis (primarily under a Bayesian paradigm), highlighting approaches which are adopted in our modelling of adverse event data. Section 3 describes the motivating dataset on the use of anti-TNF drugs for rheumatoid arthritis and the risk of malignancy. Section 4 describes the models developed, with Section 5 describing the results of fitting the different models to the motivating dataset. Finally, Section 6 discusses how these models can be applied in clinical scenarios and how the appropriateness of models may be determined.

2. Recent developments in evidence synthesis with potential relevance for adverse event synthesis

Many recent developments in evidence synthesis methodology have used Bayesian approaches, as implemented using Markov Chain Monte Carlo (MCMC) techniques due to the flexibility offered by such an approach [4], and the ability to perform complex statistical modelling [5]; for these reasons, Bayesian approaches are implemented in the adverse event scenario considered here.

The Bayesian approach allows the use of ‘exact’ simulation approaches to the synthesis of sparse data, an advantage of which is that they circumvent the need for continuity corrections (adding a constant to the event numbers across all cells of the 2 x 2 table where one arm has no events [6]), unlike many classical statistical meta-analytical approaches.

Furthermore, mixed treatment comparison (MTC) methods [7–15], also referred to as network meta-analysis, offer great potential to model differences in intervention definitions, by allowing the synthesis of evidence across arbitrarily complex networks of treatment comparisons; such methods have been applied in a simple approach to adverse events analyses previously [16]. Graphical representations of evidence networks in relation to MTC analyses have been considered in detail [17], and the challenges of defining such networks [18] have been given recent consideration, acknowledging that alternative approaches may also be viable. The concept of ‘connectedness' in a network of treatments is central to the application of MTC analysis, as it ensures that the randomisation of the primary studies is maintained whilst allowing inclusion of all available comparisons between treatments [16]. Recent extensions to MTC approaches encompass the inclusion of (fixed) covariate effects [19].

The use of hierarchical models has a long history in Bayesian methods generally; such models, which facilitate the implementation of assumptions of exchangeability, are commonly used in evidence synthesis applications [20]. The exchangeability assumption acknowledges potential differences between study results despite the fact that they cannot be explained by fixed covariate effect, thus allowing a ‘borrowing of strength’ across studies [21], which then allows the synthesis of primary studies with acknowledged differences. Advanced application of Bayesian hierarchical models has also allowed constraints to be placed on the parameters to be estimated, an example of which acknowledges the differences in reliability of evidence across study types when combining RCTs and observational studies [22].

Hence, by considering these recent developments in evidence synthesis methodology, this paper brings together concepts of MTC modelling with hierarchical models and constraints, to address issues in adverse events meta-analysis that are characterised by (i) sparsity of events; and (ii) multiple treatments, which are connected in terms of the evidence network and may vary according to how the treatments are defined.

3. Motivating dataset

The motivating topic used in this paper is that of risk of malignancy in patients with rheumatoid arthritis who take anti-tumour necrosis factor (anti-TNF) drugs to alleviate their symptoms and modify the course of their disease. A possible association between anti-TNF drugs and malignancy risk has been investigated previously using pairwise meta-analysis methods [23–25]. The earliest of these meta-analyses [23] concentrated on two anti-TNFs, infliximab and adalimumab, but combined data across both drugs and did not consider dose effects. This meta-analysis purposively excluded a third anti-TNF, etanercept, which has a different molecular structure and mechanism of action compared with infliximab and adalimumab. The odds ratio (OR) for malignancy of anti-TNFs vs. placebo was 3.3, with a 95% confidence interval (CI) of 1.2; 9.1. When etanercept was considered separately in a time-to-event meta-analysis of data from trials [25], the pooled hazard ratio (HR) for malignancy of etanercept vs. controls (not receiving etanercept) was 1.84 (95% CI 0.79; 4.28). These analyses provided the motivation to carry out a synthesis of all three drugs simultaneously, with specific interest in addressing the issues surrounding the effect of dose and specific anti-TNF with regard to malignancy risk.

The dataset used in this paper was derived from previous reviews [23, 25] and in some cases, by recourse to primary sources [26–27], and is set out in Table 1. Data are included for six etanercept studies [28–33] – noting that Weisman *et al.* 2007 is referenced by an earlier citation (Baumgartner *et al.* 2004) in Bongartz *et al.* 2009 – five adalimumab studies [27, 34–37], and two infliximab studies [26, 38]. All 13 primary studies included a non-anti-TNF arm. Studies that included no malignancy events in any trial arm were excluded as these studies do not provide any information on relative treatment effects [6]. For ease of classification of different treatments, the definitions of ‘low’, ‘recommended’ and ‘high’ dose of anti-TNF are used [24]. Over the 13 studies there were 76 malignancy events in the 7233 participants. This breaks down to 14 in the control treatments out of 2275 participants (0.62%), and 62 in the anti-TNF treatments out of 4958 participants (1.25%). Non-anti-TNF therapies were assumed to have no malignancy risk, and were therefore collapsed with placebo controls. The dataset comprised six studies with two arms, five studies with three arms, and two studies with four arms.

4. Description of the models used

4.1. Overview

A series of random effects network meta-analyses, increasing in complexity, were performed. The initial random effects meta-analysis was a simple comparison between anti-TNF therapy and non-anti-TNF controls (Model A). This was followed by a network meta-analysis with each of the three anti-TNFs considered as having a distinct treatment effect, but synthesising across (i.e. ignoring) all doses for each drug (Model B). The third model was a network meta-analysis where the treatments were defined by dose of anti-TNF alone (i.e. regardless of the specific drug); Model C. Model D defined each treatment as a combination of specific anti-TNF and dose. Hierarchical models were then imposed over the network structure defined in Model D. Model E placed random effects across individual drug within each dose level (i.e. dose was considered to have a stronger influence on malignancy risk than drug), while Model F placed random effects across dose levels within each individual drug (i.e. drug was considered to have a stronger influence on malignancy risk than dose). Finally, constraints were added to Model E, with the assumption that malignancy risk increased with dose of anti-TNF (Model G). All models included a non-anti-TNF control treatment for comparison purposes. Each model is discussed individually in more detail below and the results considered in Section 5. All models are applied to the same dataset, with treatment groupings varying according to the model which meant that, in some instances, data were pooled across multiple arms of the same study (e.g. when different doses of the same drug had been included in a trial, but dose was ignored in the model; we had no interest in differences in administration regime if the overall dose was the same for each regime, and this approach simplified the analyses and reduced the likelihood of an arm with zero events). In this dataset, every study had a control arm in which no anti-TNF treatment was given, although the methodology which follows is still relevant in the broader context where there is no common arm across all studies, but the network is connected (all nodes are directly linked to at least one other node in the network, i.e. by having both nodes present in the same study, and all nodes are connected to all other nodes, either directly, or indirectly via one or more intermediate nodes).

4.2. Model A: Pairwise meta-analysis comparing anti-TNFs with non-anti-TNF control

The simplest analysis considered is a straightforward pairwise comparison, defining treatments as being either 'anti-TNF' or 'non-anti-TNF' and thus treating each study as providing the same two-group comparison. Figure 1a displays the classification of treatments graphically in the form of a network diagram; a format which will become more informative for the analysis of the more complex networks which follow. This allows a standard (pairwise) random effects meta-analysis model to be fitted to the data [39]:

$$r_{jk} \sim Binomial(p_{jk}, n_{jk}) \quad k = 1, 2 \text{ and } j = 1, 2, \ldots, N \text{ studies}$$

$$logit(p_{jk}) = \begin{cases} \mu_j, & if\ k = 1 \\ \mu_j + \delta_j, & if\ k = 2 \end{cases} \qquad [1]$$

$$\delta_j \sim Normal(d, \sigma^2),$$

where r_{jk} is the number of events that have occurred out of a total of n_{jk} patients randomised to treatment-arm *k* of trial *j*. These are assumed to be sampled from a binomial distribution with underlying true event probability p_{jk}. Thus, μ_j is the true log-odds of an event for treatment *k* = 1 (in this case the non-anti-TNF control), δ_j is the true trial-specific log-odds ratio of the active treatment (*k* = 2) relative to treatment *k* =1 in the *j*th trial. The trial-specific log-odds are assumed to be generated from a normal distribution with mean *d* and variance σ^2, which is the standard random effects assumption. Vague prior distributions for the unknown parameters are specified as:

$$d \sim Normal(0, 1000); \quad \mu_j \sim Normal(0, 10000); \quad \sigma \sim Uniform(0, 2).$$

This model can be thought of as a special two-treatment case of the more general network meta-analysis model presented in the next section. For this model, and all subsequent models, in order to produce an absolute goodness of fit statistic for the model, the deviance for each datapoint is calculated. This deviance can then be summed and compared to the total number of datapoints. The fitted value for each of *i* datapoints is given by:

$$\hat{r}_i = \hat{p}_i \times n_i. \qquad [2]$$

These are then used to calculate the deviance for each datapoint via [40]:

$$deviance_i = 2 \times \left(r_i \times \left(log(r_i) - log(\hat{r}_i) \right) + (n_i - r_i) \times \left(log(n_i - r_i) - log(n_i - \hat{r}_i) \right) \right).$$

These deviances are summed to give the overall sum of deviance.To provide a contrast with the Bayesian pairwise analysis, an equivalent frequentist random effects meta-analysis [41] was also performed, using Stata v.12.

4.3. Model B: Network meta-analysis comparing individual anti-TNFs with non-anti-TNF control

A network meta-analysis model is applied to the dataset, in which each drug (i.e. etanercept, adalimumab and infliximab) is treated distinctly, but different dose levels are ignored. A network diagram for this four node model is shown in Figure 1b. This model has been described in depth [10, 42] and is outlined below. A standard mixed treatment random effects model with a binary outcome can be specified. Suppose that *K* treatments (where $k = 1,2,3,$ etc.) are being compared in a meta-analysis of *j* = 1, 2, …., *N* trials, and treatment 1 (non-anti-TNF in this case) is taken to be the reference or baseline treatment in the analysis. Let the number of events, r_{jk}, that have occurred out of a total of n_{jk} patients randomised to treatment-arm *k* of trial *j*, be sampled from a binomial distribution with underlying true event probability p_{jk}. Thus, the likelihood of an event and the logistic regression model are the same as those in Model A, but here it involves *K* rather than 2 treatments:

arm-level likelihood: $r_{jk} \sim Binomial(p_{jk}, n_{jk})$

$$logit(p_{jk}) = \begin{cases} \mu_{jb}; & k = b; \quad b = 1,2,3,\ldots. \\ \mu_{jb} + \delta_{jbk}; & k\ numerically\ after\ b; \quad b = 1,2,3\ ,\ldots. \end{cases} \qquad (2)$$

$$\delta_{jbk} \sim Normal(d_{bk}, \sigma^2),$$

where μ_{jb} is the true log-odds of an event in baseline treatment *b* in trial *j*, δ_{jbk} is the true trial-specific log-odds ratio of treatment *k* relative to treatment *b*, $d_{bk} = d_{1k} - d_{1b}$, and $d_{11} = 0$.

The trial-specific log-odds are assumed to be normally distributed with mean d_{bk} and variance σ^2 under a homogenous variance assumption [7, 19]. As before, vague prior distributions are specified for all unknown parameters:

$\mu_{jb} \sim Normal(0, 10000); \quad d_{bk} \sim Normal(0, 1000); \quad \sigma \sim Uniform(0, 2).$

4.4. Model C: Network meta-analysis comparing dose of anti-TNF with non-anti-TNF control

Model C changes the intervention definitions used compared to Model B by not distinguishing between the different anti-TNF drugs, but categorising by dose level, defined as recommended, low and high doses. The resulting evidence network is presented in Figure 1c. As some of the primary studies trialled multiple doses, three or more distinct treatment regimes exist for certain studies, whereas in models A and B, multiple active treatment arms could always be pooled, as no trial compared two or more anti-TNFs. This means multiple, correlated, comparisons were available for some studies, hence the network meta-analysis model required adjustment to account for this correlation [21]. Multi-arm trials introduce a correlation between each pair $(\delta_{jk}, \delta_{jh})$, $k \neq h$, of ½ and a covariance of $\frac{\sigma^2}{2}$ for homogeneous variance models [10, 42]. Thus, the correlated treatment effects in multi-arm trials, with *p* distinct arms, are assumed to be sampled from a multivariate (MVN) normal distribution given by Equation (3). The model is fitted in WinBUGS by decomposing the MVN as a series of conditional univariate distributions [7, 42] as in Equation (4).

$$\begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2/2 & \dots & \sigma^2/2 \\ \sigma^2/2 & \sigma^2 & \cdots & \sigma^2/2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2/2 & \sigma^2/2 & \cdots & \sigma^2 \end{pmatrix} \right), \tag{3}$$

$$x_i \left| \begin{pmatrix} x_1 \\ \vdots \\ x_{i-1} \end{pmatrix} \right. \sim N(\mu_i + \frac{1}{i}\sum_{j=1}^{i-1}(x_j - \mu_j), \frac{(i+1)}{2i}\sigma^2)\,. \tag{4}$$

The MTC model specified above assumes prior independence between μ_{jA} and the δ_{jAk} parameters [7], with the μ_{jA} parameters treated as fixed nuisance parameters.

This model is first in the series to introduce loops into the model; in this model, there are seven loops in total, four with three nodes and three with four nodes. Hence, the possibility of inconsistency [12] within the model is encountered for the first time.

4.5. Model D: Network meta-analysis comparing active treatments defined by dose-drug combination with non-anti-TNF control

Model D distinguishes distinct intervention nodes by the combination of specific anti-TNF used and its dose (e.g. infliximab at high dose); all active interventions were compared against a non-anti-TNF control, as in previous models. The network used for this analysis is presented in Figure 1d. As some trials contributed more than two 'arms' to the analysis, the network model that accounts for multi-arm trials was utilised, as described for Model C above. As with Model C, Model D includes treatment loops, hence there is a possibility for inconsistency within the model (also applying to Models E, F and G, which are based on the same treatment network). In Model D, there are nine loops in total, six with three nodes and three with four nodes.

4.6. Model E: Network meta-analysis, assuming exchangeability of individual anti-TNF drug across dose level

The use of random effects models to assume exchangeability between treatment effects from multiple studies has a long history in evidence synthesis [4]. Very recently, exchangeability has also been assumed across certain nodes in a network meta-analysis [43], which we also consider here.

In Model E, we define treatment nodes by both drug and dose, in exactly the same way as for Model D. However, Model E makes the assumption that the specific dose has a stronger influence on malignancy risk than the specific drug that defines an individual 'treatment', whereas in Model D, drug and dose combined to create an individual treatment, with no assumptions regarding the relative strength of influence of either drug or dose on malignancy risk. Hence, using Model E, information within a dose level is 'exchangeable' across drugs, thus allowing 'borrowing of strength' across a dose level.

The additions to Model D that are required to implement the exchangeability assumptions are formally outlined below:

$$d_{Ak} \sim \begin{cases} Normal(D_{Low}, \sigma^2_{Dose}) & \text{for each treatment } k \text{ that represents a low dose} \\ Normal(D_{Recommended}, \sigma^2_{Dose}) & \text{for each treatment } k \text{ that represents a recommended dose} \\ Normal(D_{High}, \sigma^2_{Dose}) & \text{for each treatment } k \text{ that represents a high dose.} \end{cases} \quad (5)$$

where D_{Low}, $D_{Recommended}$ and D_{High} are the population log-odds ratios for low, recommended and high doses of anti-TNF respectively, and σ^2_{Dose} is the between-estimate variance of study results within each of the three doses. The degree of

heterogeneity within the dose levels is assumed to be equal for each of the three levels, due to the sparsity of the data. Vague prior distributions were placed on these new parameters:

$$D_{Low}, D_{Recommended}, D_{High} \sim Normal(0, 1000)$$

$$\sigma^2_{Dose} \sim Uniform(0, 2).$$

4.7. Model F: Network meta-analysis, assuming exchangeability of different dose levels within individual anti-TNF drug

Model F is of the same format as Model E, but here it is assumed that the individual anti-TNF drugs have a stronger influence on malignancy risk compared to dose. Hence, this model incorporates exchangeability of information across dose level within anti-TNF drug, allowing 'borrowing of strength' across an individual anti-TNF drug. Formally:

$$d_{Ak} \sim \begin{cases} Normal(D_{Etanercept}, \sigma^2_{Drug}) & \text{for each treatment } k \text{ that represents a dose of etanercept} \\ Normal(D_{Adalimumab}, \sigma^2_{Drug}) & \text{for each treatment } k \text{ that represents a dose of adalimumab} \\ Normal(D_{Infliximab}, \sigma^2_{Drug}) & \text{for each treatment } k \text{ that represents a dose of infliximab,} \end{cases} \quad (6)$$

where $D_{Etanercept}$, $D_{Adalimumab}$ and $D_{Infliximab}$ are the population log-odds ratios for etanercept, adalimumab and infliximab, and σ^2_{Drug} is the between-estimate variance of study results within each of the three drugs. As for Model E, heterogeneity within the (drug) levels is assumed to be equal for each of the three levels, due to the sparsity of the data. The following vague priors are specified:

$$D_{Etanercept}, D_{Adalimumab}, D_{Infliximab} \sim Normal(0, 1000)$$

$$\sigma^2_{Drug} \sim Uniform(0, 2).$$

4.8. Model G: Applying constraints on the effect of dose to Model E

The hierarchical network meta-analysis model was extended further by placing constraints on parameters in the model. The network meta-analysis model in which dose is considered to have a stronger association with malignancy risk (Model E), leading to the assumption that information on individual anti-TNF drugs is exchangeable within dose, is developed further, by specifying constraints that enforce the assumption that the effects of lower anti-TNF doses on risk of malignancy cannot be greater than those of higher doses:

$$D_{Low} \leq D_{Recommended} \leq D_{High}. \quad (7)$$

To achieve this, the prior distributions placed on each dose level were truncated to ensure that only higher values for the log-odds ratio of malignancy (compared with the non-anti-TNF control group) could be sampled for each dose level compared with the dose level immediately below it. Using recommended dose as the baseline, as this dose was present across all three anti-TNFs, the log-odds ratio for malignancy in the low and high doses were set to be related to the log-odds ratio for the recommended dose, by addition of a difference factor for each dose. The difference factor for the log-odds ratio comparing the low with recommended dose was then set to be negative, based on a half-normal distribution truncated to be below zero; similarly, the difference factor for the log-odds ratio for the high dose compared with the recommended dose was set to be positive (truncated above zero). The changes required to Model E to implement these constraints are presented below:

$$\begin{aligned} D_{Low} &= D_{Recommended} + \varphi_1 \\ D_{High} &= D_{Recommended} + \varphi_2 \end{aligned} \qquad (8)$$

$$\begin{aligned} \varphi_1 &\sim Normal(0,10000)I(,0) \\ \varphi_2 &\sim Normal(0,10000)I(0,), \end{aligned}$$

where φ_1 and φ_2 are the differences in the log-odds ratios between low and recommended doses, and high and recommended doses, respectively; I(,0) indicates that the normal distribution is truncated above zero (can take only negative values); and I(0,) indicates that the normal distribution is truncated below zero (can take only positive values).

4.9. Implementation in WinBUGS

All models were implemented using WinBUGS v.1.4.3. Three chains with different initial values were used for each model, with assessment for convergence using the Brooks–Gelman–Rubin method [44], as well as visual inspection of the trace. Convergence was confirmed prior to the selection of an adequate burn-in period which was always at least 10,000 iterations. Following burn-in, at least 50,000 iterations were performed to provide the results for each model. The WinBUGS code used for the most complex model, Model G, is provided in Appendix A. Highest posterior density (HPD) credible intervals (CrIs) were derived using the boa package [45] for R.

4.10. Sensitivity analysis to choice of prior distributions

In all the analyses presented above, with the exception of the constraints specified in Model G, all prior distributions were intended to be vague. Our desire is for all such prior distributions to have minimal impact on the modelling, indeed a Bayesian analysis in WinBUGS was utilised because it provides the flexibility to fit the desired models, rather than due to the ability of such analyses to include external information (although in some contexts this may be advantageous). While very vague priors can be placed on location parameters, such as those specified on the log-odds ratios, previous work [46] has highlighted the difficulties in achieving this for scale parameters, such as variance components for random effects where data are limited, i.e. where there are few estimates (studies) contributing to the estimation of random effects. Due to this, we believe it is necessary to check robustness of estimation to the specification of such priors. To this end, we conducted a sensitivity analysis changing all priors on random effects standard deviations from Uniform(0, 2) to half-Normal(0, 1), i.e. a normal distribution truncated above 0; also, we placed Inverse-Gamma(0.001, 0.001) distributions on the variance for models B, E, F and G (where multiple random effects were specified, some of which were informed by small numbers of studies).

5. Results

Table 2 presents all the relative treatments effects, compared with control, for each of the anti-TNF treatment categories as defined by each of the models. The overall mean sum of deviance(s), in conjunction with the number of datapoints, are set out in Table 3. The assessment of model fit, based on comparing the sum of deviances with the number of datapoints (which varies across models due to the merging of arm data for some trials in some models) suggested all models would appear to be a good fit to the data, and thus has little discriminatory ability. This is probably due to the fact that there is relatively little information in the data on which the evaluation is based, due to the sparseness of events, and thus highlights a universal problem when analysing sparse data. Furthermore, in a random effects analysis, the heterogeneity parameter can increase to accommodate wide variation in effects across studies.

Model A provided evidence that anti-TNFs are associated with higher risk of malignancy than non-anti-TNF controls (OR 2.48, 95% credible interval (CrI) 1.19;

7.35), but, due to the model specification, was unable to provide insight as to whether the risk was different across anti-TNFs and/or dose levels. The equivalent random effects frequentist meta-analysis yielded contrasting results: OR 1.52, 95% CI 0.86; 2.67. These dissimilar results are due to (i) the inclusion of a continuity correction of 0.5 for all studies with zero events in one arm (a smaller continuity correction of 0.05 yielded similar results to those derived by use of 0.5); and (ii) the application of a prior distribution to the between-studies heterogeneity in the Bayesian model – although non-informative, such a distribution may impact on the posterior distribution of the between-studies standard deviation. Model B, which estimated the risk for each anti-TNF separately, indicated that all three drugs exhibit a higher risk of malignancy than the non-anti-TNF control groups, although there was considerable uncertainty around all three estimates with the CrIs including an OR of 1 in each instance.

The third model, Model C, provided estimates for each dose level, ignoring specific anti-TNF effects. This model was illuminating because the OR for the high dose group was elevated considerably (OR 7.36, 95% CrI 1.94; 39.95), indicating at least a doubling of risk of malignancy for the high dose anti-TNF compared with controls. This was in contrast to the recommended and low dose treatments, for both of which an OR of approximately 2 (compared with non-anti-TNF controls) was estimated, with a CrI including 1 in both cases.

Model D defined the treatments by both dose and specific anti-TNF, resulting in a potential of nine different treatment combinations (three drugs and three dose levels), of which seven had observed data from the available trials. However, this degree of resolution for the treatment definitions, in conjunction with the sparsity of data, resulted in very wide credible intervals, making it difficult to discern whether any specific anti-TNF/dose combination was associated with greater malignancy risk, with all CrIs including an OR of 1. Notably, the ordering of the point estimates for doses within drugs was consistent with that expected for a dose–response relationship.

The motivation for Models E and F, which make the assumption of exchangeability across certain treatment categorisation factors, was to try and reduce uncertainty in treatment effects, through borrowing strength across units in the hierarchy, while still being able to estimate distinct drug/dose combinations. Thus, these models can be seen as a ‘half-way house’ between the assumptions made in models B and C,

which pool across doses and drugs respectively, and model D, which allows estimation of each drug/dose combination independently.

The impact on the estimation of the treatment effects is somewhat predictable for both of these models. In each case there is a degree of shrinkage across the treatment groups assumed to be exchangeable. For example, with Model E, within each dose, the treatment effects for each individual drug became more similar, compared with the equivalent treatment effects for Model D. For completeness, the parameter estimates for the random effects relating to dose (Models E and G) and drug (Model F) are given in the Appendix in Table A1. The standard deviation across doses (Model E; 0.361, 95% highest posterior density (HPD) CrI 0.0008; 1.152) was very similar to that across drugs (Model F; 0.352, 95% HPD CrI 5.0 e-5; 1.187). For Model F, the three levels of dosage were shrunk towards each other within each drug (in comparison with Model D). In addition, for both Models E and F, the uncertainty, and thus the width of the CrIs, was also reduced throughout. While the estimated median treatment effect for each drug/dose combination remained elevated in both models E and F, the CrI for high dose infliximab no longer included an OR of 1, despite the point estimate being considerably reduced (the OR of 10.00 from Model D was reduced to 3.75 and 3.17 in Models E and F respectively). Based on the sum of deviances (Table 3), Model E appeared to be a fractionally improved fit to the data over Models D and F, perhaps supporting (although very weakly) the notion that dose has a stronger influence on malignancy risk than individual drug.

Model G modified Model E by placing constraints on the effect of drug dose, enforcing the assumption that lower doses cannot have higher risks of malignancy (in terms of the log-odds ratio compared with non-anti-TNF controls) associated with them. These constraints made the differences between dose levels more pronounced; although conclusions remain similar to previous models, with all treatment/dose combinations having inflated ORs, using this model, the CrIs for high doses of both infliximab and adalimumab did not include an OR of 1.

Table A2 in the appendix presents the results of the sensitivity analyses where the prior distributions were changed on all variance components in models B, E, F and G, to ascertain the influence of such prior specifications on parameter estimation. These analyses demonstrated that the prior distributions used do indeed have some influence on parameter estimation; we would suggest the differences are not of a magnitude that would influence the overall conclusions of any of the analyses.

6. Discussion

In this paper we have outlined a number of network meta-analysis models for estimating the risk of adverse events. Sparse data has been shown to be challenging for standard pairwise meta-analysis in the past, especially if there were no events at all in some studies or no events in at least one treatment arm within a study [6]. The analyses presented here demonstrate the range of potential models available; at one extreme are models that are unrestricted in terms of making assumptions regarding treatment effects according to different treatment parameters (such as drug and dose in Model D), for which the parameter estimates are very uncertain; at the other extreme are models which explicitly make stronger assumptions through expressions of exchangeability (Models E and F) and the addition of constraints on parameter estimation (Model G). Although there may be concerns that these approaches are overly complex, and make too many strong assumptions, it is important not to forget that a simple pairwise meta-analysis of this data (Model A), which *implicitly* assumes all treatments (in this example all combinations of individual drug and dose) to have *exactly the same effect*, does in fact make the *strongest* assumptions of all the models presented here.

It is very difficult to answer the question of how strong the assumptions should be (and therefore which approach to modelling should be used) in any particular clinical context. We have shown here and elsewhere that usual approaches to establishing best fitting models are problematic in a sparse data context [6]. In this paper, global measures of fit had very little discriminatory value, and relying on statistical significance of model parameters will often be inappropriate due to the often low power of the analysis. Previous simulation work showed that, in a frequentist paradigm, between-study variance parameters will be estimated as zero even when the data have been simulated from a scenario with considerable heterogeneity [6]; this suggests that informative prior distributions for variance components in these models would be advantageous [21] and circumvent problems that the (unintended) influence of intentionally vague prior distributions can have [46]. The use of multiple models making different assumptions, as done here in the form of a sensitivity analysis, is possibly the most sensible approach to address these issues. Alternatively, prior distributions that are empirically based and informative have been proposed for the heterogeneity parameter [47], although not specifically with regard to adverse events or sparse data. Heterogeneity was seen to be considerably higher

for Models B and D. This reflects the fact that in these models the nodes were defined by the individual anti-TNF only, without reference to dose, the more dominant factor in influencing malignancy risk (Model B), or by anti-TNF and dose (Model D), but without any specific modelling to impose any form of exchangeability across anti-TNFs/doses (as in Models E, F and G).

A simple solution to these problems would be simply to collect more data so that sparseness (in terms of number of events and numbers of trials) becomes less of an issue. However, this will often be time-consuming, costly, and potentially unethical in some contexts. As an alternative to further research, it may be possible to extend the potential number of trials that provide data of relevance to the clinical issue, for example by including data from trials where a specific intervention is used for indications other than those directly of interest. In the motivating example considered here, data were derived from studies using anti-TNF drugs to treat rheumatoid arthritis, but these drugs are licensed for other conditions, so data may be available from trials of anti-TNFs used for other indications. Data derived from trials for different indications would not be uncritically pooled; but it would be possible to extend the hierarchical network models presented here to allow data from trials for differing interventions to be incorporated, using the methods presented here, for example, to allow exchangeability within indications. A further, and less far-reaching, extension to the modelling presented here would be to specify simultaneous random effects for drug and dose levels, and in a sense combine the modelling extensions presented in Models E and F. Such modelling would have similarities with synthesis approaches pioneered in the context of extrapolating exposure risks across both species and dose levels simultaneously [48]. However, this modelling approach was not pursued here due to concerns with over-parameterising a dataset of modest size.

Within a network meta-analysis including loops, there is a possibility of inconsistency of evidence. In this example, Model C, and Models D, E, F and G, all have the potential for inconsistency, as they include multiple loops and trials with up to four arms. Additional investigation into inconsistencies within the models is warranted, for example using techniques of node-splitting [12], but such analyses are, due to the mixture of multi-arm trials (i.e. trials with two, three or four arms), inherently technically difficult and thus beyond the scope of this paper.

In summary, synthesis of sparse event data presents unique challenges. Novel modelling approaches have been developed and applied to a dataset that incorporates sparsity of events. Although these approaches may be valuable in this context, none provide a generic solution to the issue of data sparsity. Hence, consideration should be given to both methodological and clinical issues in order to fit realistically complex and contextually appropriate models in this area.

Acknowledgements

FCW was supported by a Department of Health Sciences award whilst conducting this work. KRA is partially supported by a National Institute for Health Research (NIHR) Senior Investigator Award (NI-SI-0508-10061). There were no external sources of funding for this work. We thank the reviewers, and the editors at *Statistics in Medicine*, for their valuable comments, which enabled us to improve the paper.

Table 1. Anti-TNF rheumatoid arthritis trial data used in synthesis models.

First author (Year)	Treatment	Anti-TNF regime(s)*	Anti-TNF dose	Number of patients	Number of malignancies
		Etanercept			
Ericson (1999)	Control	NA	NA	105	0
”	Etanercept	25mg biw	Rec	111	0
”	Etanercept	10mg qw or 25mg qw or 10mg biw	Low	343	2
Moreland (1999)	Control	NA	NA	80	0
”	Etanercept	25mg biw	Rec	78	1
”	Etanercept	10mg biw	Low	76	0
Genovese (2002)	Control	NA	NA	217	4
”	Etanercept	25mg biw	Rec	207	5
”	Etanercept	10mg biw	Low	208	5
Combe (2006)	Control	NA	NA	50	0
”	Etanercept	25mg biw	Rec	204	1
Van der Heijde (2006)	Control	NA	NA	228	1
”	Etanercept	25mg biw	Rec	454	10
Weisman(2007)/ Baumgartner (2004)	Control	NA	NA	269	2
”	Etanercept	25mg biw	Rec	266	2
		Adalimumab			
Furst (2003)	Control	NA	NA	318	0
”	Adalimumab	40mg eow	Rec	318	4
Weinblatt (2003)	Control	NA	NA	62	0
”	Adalimumab	40mg eow	Rec	67	0
”	Adalimumab	20mg eow	Low	69	0
”	Adalimumab	80mg eow	High	73	1
Keystone (2004)	Control	NA	NA	200	1
”	Adalimumab	20mg qw or 40mg eow	Rec	419	8
Van de Putte (2004)	Control	NA	NA	110	1
”	Adalimumab	20mg qw or 40mg eow	Rec	225	2
”	Adalimumab	20mg eow	Low	106	1
”	Adalimumab	40mg qw	High	103	1
Breedveld (2006)	Control	NA	NA	257	4
”	Adalimumab	40mg eow	Rec	542	6
		Infliximab			
Maini (2004)	Control	NA	NA	88	1
”	Infliximab	3mg/kg q8w	Rec	86	1
”	Infliximab	3mg/kg q4w or 10mg/kg q8w or 10mg/kg q4w	High	254	8
St Clair (2004)	Control	NA	NA	291	0
”	Infliximab	3mg/kg q8w	Rec	372	0
”	Infliximab	6mg/q8w	High	377	4

* Key: biw: twice weekly; eow: every other week; NA: not applicable; Rec: recommended; q4w: 4-weekly; q8w: 8-weekly; qw: weekly

Table 2. Odds ratios obtained from fitting models A to G (anti-TNF compared against non-anti-TNF control).

Treatment	Odds Ratio*	95% Credible Interval (equal-tailed)
Model A : Pairwise meta-analysis, all anti-TNFs combined		
Anti-TNF	2.480	1.192; 7.354
Model B : Network meta-analysis, comparing individual anti-TNFs		
Etanercept	2.485	0.730; 13.44
Adalimumab	2.399	0.685; 13.45
Infliximab	6.878	0.664; 285.8
Model C : Network meta-analysis, distinguishing only by dose of anti-TNF		
Low dose (Anti-TNF)	2.012	0.531; 8.627
Recommended dose (Anti- TNF)	2.099	0.961; 5.455
High dose (Anti-TNF)	7.362	1.942; 39.95
Model D : Network meta-analysis: comparing active treatments defined by dose–drug combination		
Etanercept – low dose	2.279	0.378; 19.42
Etanercept – recommended dose	2.423	0.627; 12.64
Adalimumab – low dose	1.336	0.035; 30.91
Adalimumab – recommended dose	2.253	0.595; 12.95
Adalimumab – high dose	4.110	0.275; 85.83
Infliximab – recommended dose	0.881	0.016; 47.71
Infliximab - high dose	10.00	0.948; 393.3
Model E : Network meta-analysis, assuming exchangeability of anti-TNF drugs within dose levels		
Etanercept – low dose	2.311	0.761; 8.145
Etanercept – recommended dose	2.288	0.936; 6.620
Adalimumab – low dose	2.290	0.484; 9.398
Adalimumab – recommended dose	2.255	0.938; 6.665
Adalimumab – high dose	3.170	0.927; 16.55
Infliximab – recommended dose	1.965	0.453; 6.973
Infliximab - high dose	3.753	1.253; 23.02
Model F : Network meta-analysis, assuming exchangeability of dose levels within anti-TNF drugs		
Etanercept – low dose	2.323	0.795; 8.198
Etanercept – recommended dose	2.369	0.954; 7.253
Adalimumab – low dose	2.280	0.501; 9.194
Adalimumab – recommended dose	2.299	0.914 ; 7.315
Adalimumab – high dose	2.538	0.776; 12.20
Infliximab – recommended dose	2.109	0.398; 9.309
Infliximab - high dose	3.167	1.061; 23.08
Model G : Applying constraints on the effect of dose to Model E		
Etanercept – low dose	1.440	0.401; 5.027
Etanercept – recommended dose	2.170	0.865; 6.650
Adalimumab – low dose	1.313	0.212; 5.933

Adalimumab – recommended dose	2.275	0.915; 7.323
Adalimumab – high dose	6.711	1.449; 48.58
Infliximab – recommended dose	2.048	0.426; 8.716
Infliximab - high dose	8.311	2.060; 67.40

* Median of samples from posterior distribution

Table 3. Goodness of model fit statistics for models A to G.

Model	Between study standard deviation – i.e. heterogeneity parameter: median (95% highest posterior density credible interval)	Sum of deviance: mean (number of data points)
A	0.583 (0.0005; 1.558)	24.14 (26)
B	0.802 (4.0 e-5; 1.775)	25.06 (26)
C	0.548 (0.0003; 1.485)	32.41 (35)
D	0.862 (0.007; 1.811)	34.28 (35)
E	0.550 (0.0002; 1.460)	33.13 (35)
F	0.550 (0.0003; 1.489)	33.61 (35)
G	0.590 (0.001; 1.541)	32.67 (35)

Appendix A. WinBUGS code for Model G.

```
# ================================
# Notes
# 13 trials (6 with 2 arms,  5 with 3 arms, 2 with 4 arms, 35 arms total),
# 35 data points,
# 8 treatments:

# i indexes datapoints
# j indexes trials
# k indexes treatments

#Treatments coding:
# 1: Placebo and/or DMARD
# 2: Etanercept Recommended
# 3: Etanercept Low
# 4. Adalimumab Recommended
# 5. Adalimumab Low
# 6. Adalimumab High
# 7. Infliximab Recommended
# 8. Infliximab High

# ================================

model{
for(i in 1:NS){
        w[i,1] <-0
        delta[i,t[i,1]]<-0
        mu[i] ~ dnorm(0,.0001)          # vague priors for trial baselines
        for (k in 1:na[i])  {
                r[i,k] ~ dbin(p[i,t[i,k]],n[i,k])               # binomial likelihood
                logit(p[i,t[i,k]])<-mu[i] + delta[i,t[i,k]] }

        # model
                for (k in 2:na[i]) {
                        delta[i,t[i,k]] ~ dnorm(md[i,t[i,k]],taud[i,t[i,k]])
                        md[i,t[i,k]] <-  d[t[i,k]] - d[t[i,1]]  + sw[i,k]
                        taud[i,t[i,k]] <- tau *2*(k-1)/k
                        w[i,k] <- (delta[i,t[i,k]]  - d[t[i,k]] + d[t[i,1]]) #adjustment, multi-arm
RCTs
                        sw[i,k] <-sum(w[i,1:k-1])/(k-1) }  }
        for(i in 1:NS){
                for (k in 1:na[i]){
         #Deviance residuals for data i
                rhat[i,k] <- (p[i,t[i,k]] * n[i,k])
                dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))  +  (n[i,k]-r[i,k]) * (log(n[i,k]-
r[i,k]) - log(n[i,k]-rhat[i,k])))  }
                sumdev[i] <- sum(dev[i,1:na[i]])   }
                ssumdev <- sum(sumdev[])
        d[1]<-0
```

#D.d[1] refers to recommended dose, D.d[2] refers to low dose, D.d[3] refers to high dose

```
d[2]~dnorm(D.d[1], prec.d)
d[3]~dnorm(D.d[2], prec.d)
d[4]~dnorm(D.d[1], prec.d)
d[5]~dnorm(D.d[2], prec.d)
d[6]~dnorm(D.d[3], prec.d)
d[7]~dnorm(D.d[1], prec.d)
d[8]~dnorm(D.d[3], prec.d)

#  vague priors for basic parameters

sd~dunif(0,2)
tau<-1/pow(sd,2)

prec.d<-1/(sd.d*sd.d)
sd.d~dunif(0,2)

# Constraints model for D.d

D.d[2] <- D.d[1] + diff1
D.d[3] <- D.d[1] + diff2

diff1 ~ dnorm(0,0.0001)I(,0)
diff2 ~ dnorm(0,0.0001)I(0,)

D.d[1] ~ dnorm(0,0.0001)

pdiff1 <- 1-step(diff1)
pdiff2 <- step(diff2)

}
```

DATA

NS=no. studies;
NB : set up M vectors each r[,]. n[,] and t[,], where M is the maximum number of
treatments per trial in the dataset. In this dataset M is 4.

list(NS=13)

r[,1]	n[,1]	r[,2]	n[,2]	r[,3]	n[,3]	r[,4]	n[,4]	t[,1]	t[,2]	t[,3]	t[,4]	na[]
1	228	10	454	NA	1	NA	1	1	2	NA	NA	2
0	50	1	204	NA	1	NA	1	1	2	NA	NA	2
0	80	1	78	0	76	NA	1	1	2	3	NA	3
4	217	5	207	5	208	NA	1	1	2	3	NA	3
2	269	2	266	NA	1	NA	1	1	2	NA	NA	2
0	106	0	111	2	343	NA	1	1	2	3	NA	3
0	318	4	318	NA	1	NA	1	1	4	NA	NA	2

```
0	62	0	67	0	69	1	73	1	4	5	6	4
1	200	8	419	NA	1	NA	1	1	4	NA	NA	2
1	110	2	225	1	106	1	103	1	4	5	6	4
4	257	6	542	NA	1	NA	1	1	4	NA	NA	2
1	88	1	86	8	254	NA	NA	1	7	8	NA	3
0	291	0	372	4	377	NA	1	1	7	8	NA	3
END


# INITIAL VALUES

# d refers to number of treatments, mu refers to number of studies,
# delta refers to number of datapoints

list(d=c(NA,0.75,0.70,0.77,0.65,1.87,0.54,2.1),  mu=c(-5,-7,-7,-4,-5,-7,-6,-7,-5,-6,-4,-5,-
7),sd=0.7,sd.d=0.6, D.d=c(0,NA,NA),diff1=-1,diff2=1,
delta = structure(.Data=c(
        NA,1, NA, NA, NA,
        NA, NA, NA, NA, 0.8,
        NA, NA, NA, NA, NA,
        NA, NA,1.0, 0.6, NA,
        NA, NA, NA, NA, NA,
         0.5, 0.5, NA, NA, NA,
        NA, NA, NA, 0.7, NA,
        NA, NA, NA, NA, NA,
        NA, 0.7, 0.9, NA, NA,
        NA, NA, NA, NA, NA,
        NA, 1.2, NA, NA, NA,
        NA, NA, NA, NA, 0.7,
         0.6, 2.0, NA, NA, NA,
        NA, NA, 0.9, NA, NA,
        NA, NA, NA, NA, NA,
         0.7, 0.5, 1.5, NA, NA,
        NA, NA, NA, 0.3, NA,
        NA, NA, NA, NA, NA,
        NA, NA, NA, NA, 0.4,
         1.86, NA, NA, NA, NA,
        NA, NA,0.5,2.4),
.Dim = c(13,8)))
```

Table A1. Dose and drug random effect parameter estimates from Models E, F and G.

Model	Parameter	Median (95% Credible Interval)[1]
E	$\exp(D_{Low})$	2.389 (0.691 to 8.950)
	$\exp(D_{Recommended})$	2.259 (0.813; 6.931)
	$\exp(D_{High})$	3.135 (1.046; 15.25)
	σ_{Dose}	0.361 (0.0008; 1.152)*
F	$\exp(D_{Etanercept})$	2.380 (0.808; 8.277)
	$\exp(D_{Adalimumab})$	2.391 (0.796; 8.539)
	$\exp(D_{Infliximab})$	2.515 (0.749; 11.84)
	σ_{Drug}	0.352 (5.0 e-5; 1.187)*
G	$\exp(D_{Low})$	1.228 (0.232; 3.948)
	$\exp(D_{Recommended})$	2.274 (0.802; 7.989)
	$\exp(D_{High})$	7.758 (1.984; 64.54)
	σ_{Dose}	0.394 (0.0008; 1.409)*

[1]Credible interval is equal-tailed unless indicated by * (highest posterior density interval).

Table A2. Results of sensitivity analysis changing the prior distributions placed on variance parameters.

Treatment effect/parameters	Prior Distributions on Variance Components: Standard Deviations ~ Uniform(0,2)		Standard Deviations ~ Half-Normal(0,1)		Variances ~ Inverse-Gamma(0.001,0.001)	
Model B : Network meta-analysis, comparing individual anti-TNFs						
	Odds Ratio[1]	**95% CrI[2]**	**Odds Ratio[1]**	**95% CrI[2]**	**Odds Ratio[1]**	**95% CrI[2]**
Etanercept	2.485	0.730; 13.44	2.353	0.797; 10.22	2.302	0.828; 10.87
Adalimumab	2.399	0.685; 13.45	2.264	0.735; 10.39	2.134	0.758; 10.82
Infliximab	6.878	0.664; 285.8	6.054	0.741; 174.4	5.998	0.785; 273.3
	Median	**95% CrI[3]**	**Median**	**95% CrI[3]**	**Median**	**95% CrI[3]**
σ	0.802	4.0 e-5; 1.775	0.590	0.0003; 1.513	0.313	0.013 ; 1.690
Model E : Network meta-analysis, assuming exchangeability of different anti-TNF drugs within dose levels						
	Odds Ratio[1]	**95% CrI[2]**	**Odds Ratio[1]**	**95% CrI[2]**	**Odds Ratio[1]**	**95% CrI[2]**
Etanercept – low dose	2.311	0.761; 8.145	2.320	0.830; 7.123	2.178	0.904; 5.432
Etanercept – recommended dose	2.288	0.936; 6.620	2.284	1.002; 6.005	2.143	1.068; 4.902
Adalimumab – low dose	2.290	0.484; 9.398	2.349	0.644; 8.269	2.169	0.772 ; 6.061
Adalimumab – recommended dose	2.255	0.938; 6.665	2.231	0.974; 6.036	2.099	0.997; 4.791
Adalimumab – high dose	3.170	0.927; 16.55	3.053	1.055; 13.67	2.538	1.059; 8.692
Infliximab – recommended dose	1.965	0.453; 6.973	1.977	0.519; 6.090	1.979	0.664; 4.870
Infliximab - high dose	3.753	1.253; 23.02	3.472	1.292; 17.57	2.774	1.251; 11.21
$\exp(D_{Low})$	2.389	0.691; 8.950	2.397	0.814;7.611	2.205	0.919; 5.759
$\exp(D_{Recommended})$	2.259	0.813; 6.931	2.236	0.898; 6.126	2.114	0.996; 4.890
$\exp(D_{High})$	3.135	1.046; 15.25	2.991	1.126; 12.11	2.524	1.153; 8.128
	Median	**95% CrI[3]**	**Median**	**95% CrI[3]**	**Median**	**95% CrI[3]**
σ_{Dose}	0.361	0.0008; 1.152	0.313	0.0002; 0.923	0.170	0.014; 0.679
σ	0.550	0.0002; 1.460	0.440	0.002; 1.182	0.192	0.014; 0.957
Model F : Network meta-analysis, assuming exchangeability of different dose levels within each anti-TNF drug						

	Odds Ratio[1]	95% CrI[2]	Odds Ratio[1]	95% CrI[2]	Odds Ratio[1]	95% CrI[2]
Etanercept – low dose	2.323	0.795; 8.198	2.258	0.840; 6.591	2.174	1.016; 5.532
Etanercept – recommended dose	2.369	0.954; 7.253	2.293	0.985; 5.989	2.201	1.085; 5.377
Adalimumab – low dose	2.280	0.501; 9.194	2.194	0.628; 7.771	2.154	0.852; 5.889
Adalimumab – recommended dose	2.299	0.914; 7.315	2.219	0.969; 6.002	2.157	1.061; 5.252
Adalimumab – high dose	2.538	0.776; 12.20	2.390	0.796; 9.673	2.259	0.977; 6.852
Infliximab – recommended dose	2.109	0.398; 9.309	2.061	0.462; 7.488	2.099	0.729; 6.188
Infliximab - high dose	3.167	1.061; 23.08	2.907	1.097; 18.03	2.533	1.137; 10.39
$\exp(D_{Etanercept})$	2.380	0.808; 8.277	2.292	0.884; 6.655	2.199	1.056; 5.577
$\exp(D_{Adalimumab})$	2.391	0.796; 8.539	2.277	0.881; 7.037	2.195	1.039; 5.641
$\exp(D_{Infliximab})$	2.515	0.749; 11.84	2.401	0.831; 9.386	2.276	1.015; 6.911
	Median	**95% CrI[3]**	**Median**	**95% CrI[3]**	**Median**	**95% CrI[3]**
σ_{Drug}	0.352	5.0 e-5; 1.187	0.295	3.0 e-5; 0.959	0.146	0.013; 0.666
σ	0.550	0.0003; 1.489	0.464	0.0002; 1.194	0.211	0.014; 0.992

Model G : Applying constraints on the effect of dose to Model E

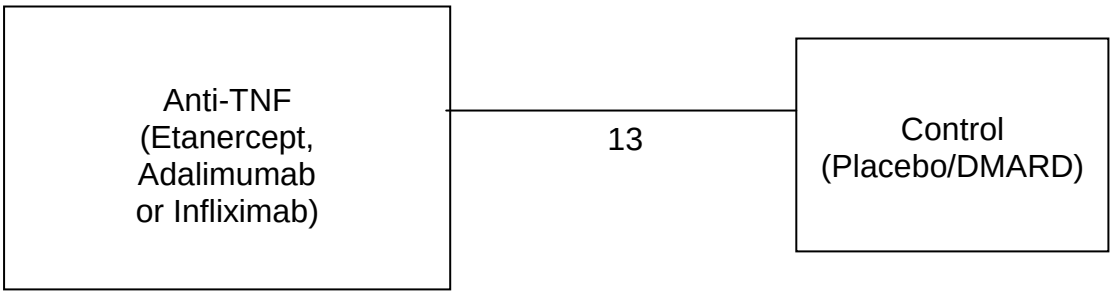
	Odds Ratio[1]	95% CrI[2]	Odds Ratio[1]	95% CrI[2]	Odds Ratio[1]	95% CrI[2]
Etanercept – low dose	1.440	0.401; 5.027	1.466	0.448; 4.505	1.388	0.498; 3.632
Etanercept – recommended dose	2.170	0.865; 6.650	2.151	0.924; 5.668	2.051	0.988; 4.681
Adalimumab – low dose	1.313	0.212; 5.933	1.342	0.288; 5.030	1.345	0.391; 3.953
Adalimumab – recommended dose	2.275	0.915; 7.323	2.241	0.975; 6.190	2.089	1.038; 4.851
Adalimumab – high dose	6.711	1.449; 48.58	6.727	1.610; 36.15	6.493	1.918; 27.07
Infliximab – recommended dose	2.048	0.426; 8.716	2.055	0.525; 7.144	2.012	0.770; 5.308
Infliximab – high dose	8.311	2.060; 67.40	7.919	2.208; 46.23	7.024	2.238; 32.54
$\exp(D_{Low})$	1.228	0.232; 3.948	1.281	0.305; 3.701	1.313	0.409; 3.307
$\exp(D_{Recommended})$	2.274	0.802;	2.239	0.903;	2.088	1.001;

$\exp(D_{High})$	7.758	7.989 1.984; 64.54	7.475	6.495 2.091; 43.73	6.824	5.046 2.188; 30.46
	Median	**95% CrI[3]**	**Median**	**95% CrI[3]**	**Median**	**95% CrI[3]**
σ_{Dose}	0.394	0.0008; 1.409	0.465	0.0005; 1.111	0.145	0.015; 0.767
σ	0.590	0.001; 1.541	0.326	7.0 e-5; 1.237	0.209	0.015; 1.035

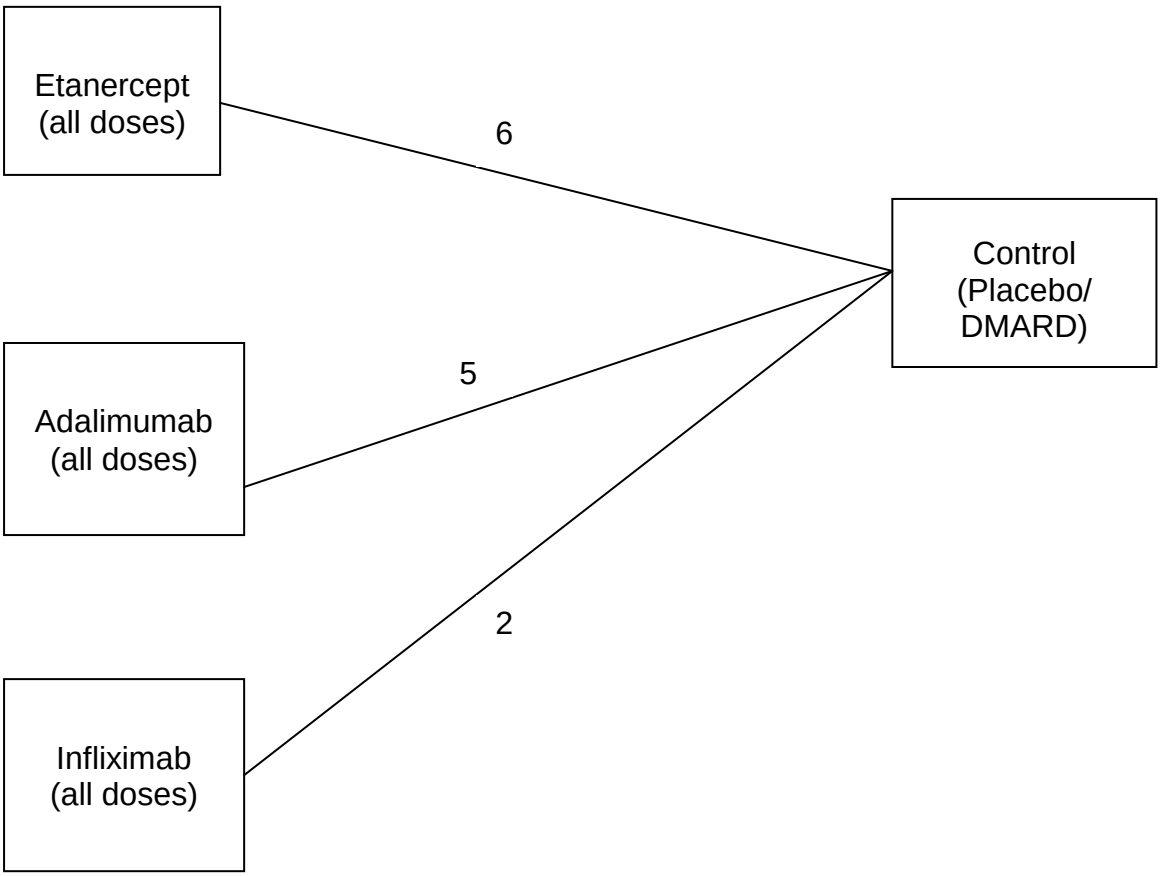
[1] Median of samples from posterior distribution. [2] Equal-tailed credible interval. [3] Highest posterior density credible interval.

Figure 1 Network diagrams showing categorisation of treatment definitions used in the various models considered (numbers on lines represent the total number of studies which make each comparison; where there is a second number, this indicates the number of studies within the total for which there are zero events in both groups making the comparison).

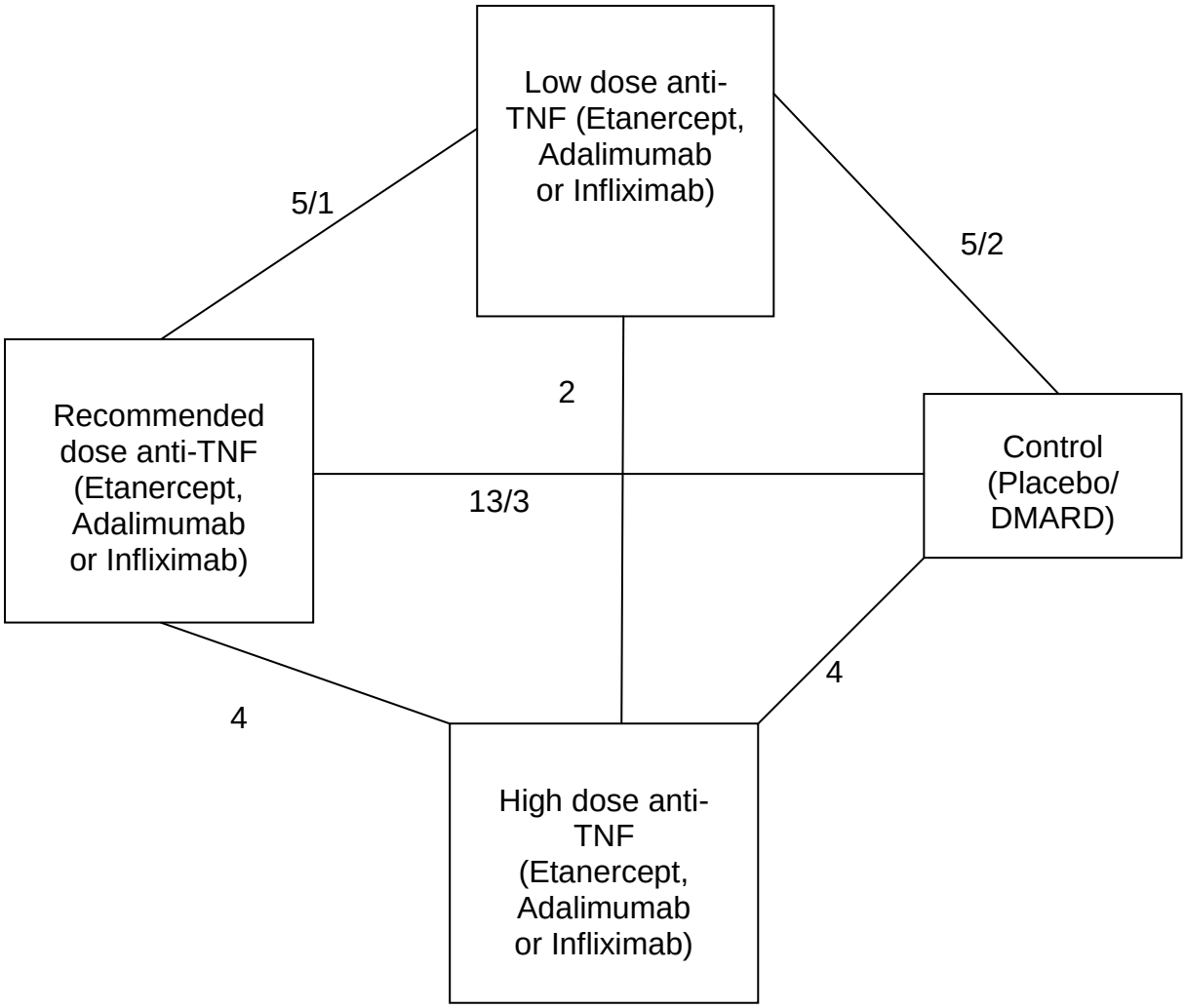
a) Pairwise meta-analysis comparing all anti-TNF data to non-anti-TNF control; used in Model A.



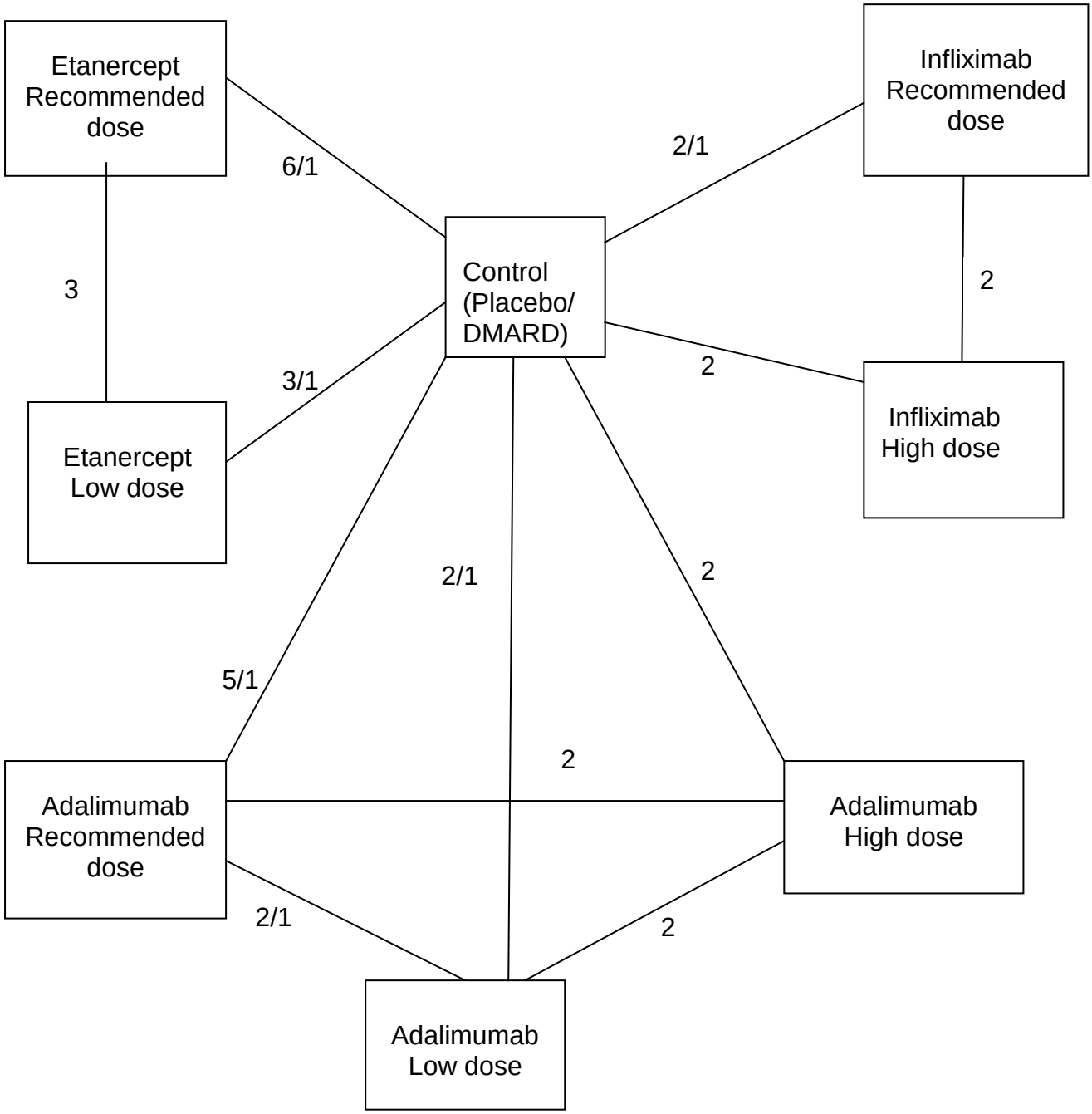
b) Network meta-analysis comparing individual anti-TNFs (combining all doses) with non-anti-TNF control; used in Model B.



c) Network meta-analysis comparing distinct dose level of anti-TNFs (combining individual drugs at the same dose) with non-anti-TNF control; used in Model C.



d) Network meta-analysis comparing individual anti-TNFs/dose combinations with non-anti-TNF control; used in Models D, E, F and G.



References

1. Cornelius VR, Perrio MJ, Shakir SAW, Smith LA. Systematic reviews of adverse effects of drug interventions: a survey of their conduct and reporting quality. *Pharmacoepidemiology and Drug Safety* 2009; **18**: 1223–1231.
2. Warren FC, Abrams KR, Golder S, Sutton AJ. Systematic review of methods used in meta-analyses where a primary outcome is an adverse or unintended event. *BMC Medical Research Methodology* 2012; **12**.
3. Sutton AJ, Cooper NJ, Lambert PC, Jones DR, Abrams KR, Sweeting MJ. Meta-analysis of rare and adverse event data. *Expert Review of Pharmacoeconomics & Outcomes Research* 2002; **2**: 367–379.
4. Sutton AJ, Higgins JPT. Recent developments in meta-analysis. *Statistics in Medicine* 2008; **27**: 625–650.
5. Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research* 2001; **10**: 277–303.
6. Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in Medicine* 2004; **23**: 1351–1375.
7. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine* 2004; **23**: 3105–3124.
8. Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005; **331**: 897–900.
9. Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 1: Introduction. *Medical Decision Making* 2013; **33**: 597-606.
10. Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: A generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Medical Decision Making* 2013; **33**: 607-617.
11. Dias S, Sutton AJ, Welton NJ, Ades AE. Evidence synthesis for decision making 3: Heterogeneity—subgroups, meta-regression, bias and bias-adjustment. *Medical Decision Making* 2013; **33**: 618-640.
12. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. Evidence sythesis for decision making 4: Inconsistency in networks of evidence based on randomized controlled trials. *Medical Decision Making* 2013; **33**: 641-656.
13. Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 5: The baseline natural history model. *Medical Decision Making* 2013; **33**: 657-670.
14. Dias S, Sutton AJ, Welton NJ, Ades AE. Evidence synthesis for decision making 6: Embedding evidence synthesis in probabilistic cost-effectiveness analysis. *Medical Decision Making* 2013; **33**: 671-678.
15. Ades AE, Caldwell DM, Reken S, Welton NJ, Sutton AJ, Dias S. Evidence synthesis for decision making 7: A reviewer's checklist. *Medical Decision Making* 2013; **33**: 679-691.
16. Cooper NJ, Sutton AJ, Lu G, Khunti K. Mixed comparison of stroke prevention treatments in individuals with nonrheumatic atrial fibrillation. *Archives of Internal Medicine* 2006; **166**: 1269–1275.
17. Salanti G, Kavvoura FK, Ioannidis JPA. Exploring the geometry of treatment networks. *Annals of Internal Medicine* 2008; **148**: 544–553.

18. Del Giovane C, Vacchi L, Mavridis D, Filippini G, Salanti G. Network meta-analysis models to account for variability in treatment definitions: application to dose effects. *Statistics in Medicine* 2013; **32**: 25-39.
19. Cooper NJ, Sutton AJ, Morris D, Ades AE, Welton NJ. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Statistics in Medicine* 2009; **28**: 1861–1881.
20. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society Series A – Statistics in Society* 2009; **172**: 137–159.
21. Higgins J, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine* 1996; **15**: 2733–2749.
22. Prevost TC, Abrams KR, Jones DR. Hierarchical models in generalized synthesis of evidence: an example based on studies of breast cancer screening. *Statistics in Medicine* 2000; **19**: 3359–3376.
23. Bongartz T, Sutton AJ, Sweeting MJ, Buchan I, Matteson EL, Montori V. Anti-TNF antibody therapy in rheumatoid arthritis and the risk of serious infections and malignancies: systematic review and meta-analysis of rare harmful effects in randomized controlled trials. *Journal of the American Medical Association* 2006; **295**: 2275–2285.
24. Leombruno JP, Einarson TR, Keystone EC. The safety of anti-Tumor Necrosis Factor treatments in rheumatoid arthritis: meta and exposure adjusted pooled analyses of serious adverse events. *Annals of the Rheumatic Diseases* 2008; **68**: 1136–1145.
25. Bongartz T, Warren FC, Mines D, Matteson EL, Abrams KR, Sutton AJ. Etanercept therapy in rheumatoid arthritis and the risk of malignancies: a systematic review and individual patient data meta-analysis of randomized controlled trials. *Annals of the Rheumatic Diseases* 2009; **68**: 1177–1183.
26. Maini RN, Breedveld FC, Kalden JR, Smolen JS, Furst D, Weisman MH, St. Clair EW, Keenan GF, van der Heijde D, Marsters PA, Lipsky PE. Sustained improvement over two years in physical function, structural damage, and signs and symptoms among patients with rheumatoid arthritis treated with infliximab and methotrexate. *Arthritis & Rheumatism* 2004; **50**: 1051–1065.
27. Breedveld FC, Weisman MH, Kavanaugh AF, Cohen SB, Pavelka K, van Vollenhoven R, Sharp J, Perez JL, Spencer-Green GT. The PREMIER study: A multicenter, randomized, double-blind clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone in patients with early, aggressive rheumatoid arthritis who had not had previous methotrexate treatment. *Arthritis & Rheumatism* 2006; **54**: 26–37.
28. Ericson ML, Wajdula J, European Etanercept Investigators. A double-blind placebo-controlled study of the efficacy and safety of four different doses of etanercept in patients with rheumatoid arthritis. *Arthritis & Rheumatism* 1999; **42**: S82.
29. Moreland LW, Schiff MH, Baumgartner SW, Tindall EA, Fleischmann RM, Bulpitt KJ, Weaver AL, Keystone EC, Furst DE, Mease PJ, Ruderman EM, Horwitz DA, Arkfeld DG, Garrison L, Burge DJ, Blosch CM, Lange MLM, McDonnell ND, Weinblatt

ME. Etanercept therapy in rheumatoid arthritis: a randomized, controlled trial. *Annals of Internal Medicine* 1999; **130**: 478–486.
30. Genovese MC, Bathon JM, Martin RW, Fleischmann RM, Tesser JR, Shiff MH, Keystone EC, Wasko MC, Moreland LW, Weaver AL, Markenson J, Cannon GW, Spencer-Green G, Finck BK. Etanercept versus methotrexate in patients with early rheumatoid arthritis: two-year radiographic and clinical outcomes. *Arthritis & Rheumatism* 2002; **46**: 1443–1450.
31. Combe B, Codreanu C, Fiocco U, Gaubitz M, Geusens PP, Kvien TK, Pavelka K, Sambrook PN, Smolen JS, Wajdula J, Fatenejad S. Etanercept and sulfasalazine, alone and combined, in patients with active rheumatoid arthritis despite receiving sulfasalazine: a double-blind comparison. *Annals of the Rheumatic Diseases* 2006; **65**: 1357–1362.
32. van der Heijde D, Klareskog L, Rodriguez-Valverde V, Codreanu C, Bolosiu H, Melo-Gomes J, Tornero-Molina J, Wajdula J, Pedersen R, Fatenejad S. Comparison of etanercept and methotrexate, alone and combined, in the treatment of rheumatoid arthritis: two-year clinical and radiographic results from the TEMPO study, a double-blind randomized trial. *Arthritis & Rheumatism* 2006; **54**: 1063–1074.
33. Weisman MH, Paulus HE, Burch FX, Kivitz AJ, Flerer J, Dunn M, Kerr DR, Tsuji W, Baumgartner SW. A placebo-controlled, randomized, double-blinded study evaluating the safety of etanercept in patients with rheumatoid arthritis and concomitant comorbid diseases. *Rheumatology* 2007; **46**: 1122–1125.
34. Furst DE, Schiff MH, Fleischmann RM, Strand V, Birbara CA, Compagnone D, Fischkoff SA, Chartash EK. Adalimumab, a full human anti tumor necrosis factor-alpha monoclonal antibody, and concomitant standard antirheumatic therapy for the treatment of rheumatoid arthritis: results of STAR (Safety Trial of Adalimumab in Rheumatoid Arthritis). *The Journal of Rheumatology* 2003; **30**: 2563–2571.
35. 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 \alpha monoclonal antibody, for the treatment of rheumatoid arthritis in patients taking concomitant methotrexate: The Armada trial. *Arthritis & Rheumatism* 2003; **48**: 35–45.
36. 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 & Rheumatism* 2004; **50**: 1400–1411.
37. Van de Putte LBA, Atkins C, Malaise M, Sany J, Russell AS, van Riel PLCM, Settas L, Bijlsma JW, Todesco S, Dougados M, Nash P, Emery P, Walter N, Kaul M, Fischkoff S, Kupper H. Efficacy and safety of adalimumab as monotherapy in patients with rheumatoid arthritis for whom previous disease modifying antirheumatic drug treatment has failed. *Annals of the Rheumatic Diseases* 2004; **63**: 508–516.
38. St. Clair EW, van der Heijde DMFM, Smolen JS, Maini RN, Bathon JM, Emery P, Keystone EC, Schiff M, Kaiden JR, Wang B, DeWoody K, Weiss R, Baker D, Active-controlled Study of Patients Receiving Infliximab for the Treatment of Rheumatoid Arthritis of Early Onset Study Group. Combination of infliximab and methotrexate

therapy for early rheumatoid arthritis: A randomized, controlled trial. *Arthritis & Rheumatism* 2004; **50**: 3432–3443.
39. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. *Statistics Medicine* 1995; **14**: 2685–2699.
40. McCullagh P, Nelder JA. *Generalized linear models* (2nd edition). Chapman & Hall/CRC: Boca Raton, FL, USA, 1989.
41. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled clinical trials* 1986; **7**: 177–188.
42. Ades AE, Welton NJ, Lu G. Introduction to mixed treatment comparison. Available at: http://www.bristol.ac.uk/social-community-medicine/media/mpes/intro-to-mtc.pdf [02 April 2013].
43. Dakin HA, Welton NJ, Ades AE, Collins S, Orme M, Kelly S. Mixed treatment comparison or repeated measurements of a continuous endpoint: an example using topical treatments for primary open-angle glaucoma and ocular hypertension. *Statistics in Medicine* 2011; **30**: 2511–2535.
44. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 1998; **7**: 434–455.
45. Smith BJ. boa: An R package for MCMC output convergence assessment and posterior inference. *Journal of Statistical Software* 2007; **21**(11).
46. Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine* 2005; **24**: 2401–2428.
47. Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JPT. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *International Journal of Epidemiology* 2012; 41: 818–827.
48. Dumouchel WH, Harris JE. Bayes methods for combining the results of cancer studies in humans and other species. *Journal of the American Statistical Association* 1983; **78**: 293-308.