

A Bayesian approach to probabilistic sensitivity analysis in structured benefit-risk assessment

Ed Waddingham^{*1}, Shahrul Mt-Isa¹, Richard Nixon² and Deborah Ashby¹

¹ Imperial Clinical Trials Unit, School of Public Health, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PG, UK

² Advanced Quantitative Sciences, Economic Modeling, Novartis Pharma AG, Postfach, CH-4002 Basel, Switzerland

on behalf of PROTECT Benefit-Risk (Billy Amzal, Deborah Ashby, Simon Ashworth, Torbjorn Callreus, Ed Chan, Christoph Dierig, Yasemin Dil, Gerry Downey, Georgy Genov, Christine Hallgreen, Ian Hirsch, Steve Hobbiger, Kimberley Hockley, Diana Hughes, Juhaeri Juhaeri, Silvia Kuhls, Alfons Lieftucht, Davide Luciani, Marilyn Metcalf, Alain Micaleff, Shahrul Mt-Isa, Jeremiah Mwangi, Richard Nixon, Rebecca Noel, John Pears, Ruth Peters, Lawrence D. Phillips, George Quartey, Susan Shepherd, Isabelle Stoeckert, Andrew Thomson, Ioanna Tzoulaki, Tjeerd van Staa, Ed Waddingham, Nan Wang, and Lesley Wise)

Received zzz, revised zzz, accepted zzz

Quantitative decision models such as multiple criteria decision analysis (MCDA) can be used in benefit-risk assessment to formalize trade-offs between benefits and risks, providing transparency to the assessment process. There is however no well-established method for propagating uncertainty of treatment effects data through such models to provide a sense of the variability of the benefit-risk balance. Here we present a Bayesian statistical method that directly models the outcomes observed in randomized placebo-controlled trials and uses this to infer indirect comparisons between competing active treatments. The resulting treatment effects estimates are suitable for use within the MCDA setting, and it is possible to derive the distribution of the overall benefit-risk balance through Markov Chain Monte Carlo simulation. The method is illustrated using a case study of natalizumab for relapsing-remitting Multiple Sclerosis.

Key words: Bayes; Benefit-Risk; Decision Making; MCDA; Statistics

Supporting Information for this article is available from the author or on the WWW under <http://dx.doi.org/10.1002/bimj.XXXXXXX>

1 Introduction

1.1 Structured benefit-risk assessment and MCDA

Benefit-risk assessment involves bringing together evidence on the favourable and unfavourable effects of a treatment in order to inform a medical decision, such as the decision by a doctor to prescribe a treatment or by a regulator to grant marketing authorisation.

Within the regulatory context, the importance of transparency in the benefit-risk decision making process is increasingly recognised. Regulators are accountable to the public and should be able to explain and defend the basis on which their decisions are made - in other words, it should be possible to demonstrate how the evidence has logically led to a certain course of action. Adopting formal structures for benefit-risk assessment can help to provide this level of transparency (Mt-Isa et al., 2011).

One method that has attracted attention is multiple criteria decision analysis (MCDA), a formal decision-making framework, various formulations of which have been developed and applied in a number of fields over the years (Keeney and Raiffa, 1976). The method involves specifying the expected consequences of each alternative course of action, and then establishing the utility (or value)

* Corresponding author: e-mail: e.waddingham@imperial.ac.uk, Phone: +44-(0)20-7594-1723, Fax: +44-(0)20-7594-0768

of these consequences. The optimal decision is that which maximizes the value to the decision-maker (Ashby and Smith, 2000) and, as the method makes value judgements explicit, it is possible for the decision-maker to adopt other stakeholders' perspectives. MCDA makes comparisons based on multiple criteria – this is achieved by establishing the consequences of each possible course of action in relation to each criterion and then calculating the overall value as the sum of the values across all criteria, weighted according to their relative importance (). For this analysis, detailed knowledge of the decision-maker's preferences is required. For each criterion, a partial value function must be established: this ascribes a numerical value to different levels of an outcome variable, indicating how importantly the decision maker values a change in that outcome. Preference weights that specify the relative importance of the criteria are also required. A key assumption here is that the partial values of different outcomes are independent, a condition known as preference independence.

Although this method is well-established in many other fields, the use of MCDA in medical settings, where the criteria are individual benefits and risks, only emerged in around 2007 (Mussen et al., 2007). In the medical context, possible decision-makers include regulators, drug developers, healthcare providers, doctors and patients themselves. Decisions are ultimately made on patients' behalf and so, where value judgements are required as part of a decision model, these are typically elicited from patients or their representatives.

1.2 Natalizumab case study

Natalizumab is a treatment for relapsing-remitting multiple sclerosis, an autoimmune disease characterised by periodic symptomatic attacks (relapses), frequently progressing to a state of sustained disability due to accumulated neurological damage. Natalizumab was licensed in the USA in 2004 but subsequently withdrawn from the market after post-marketing surveillance revealed a number of fatal cases of progressive multifocal leukoencephalopathy (PML) among patients taking the drug. PML is a frequently fatal brain infection caused by the John Cunningham (JC) virus, which is widespread but does not pose a threat to immunocompetent individuals. Natalizumab is believed to allow the JC virus to infect the nervous system, leading to PML. In spite of this risk, the FDA re-introduced the drug in 2005 due to patient demand. In the EU, the Committee for Medicinal Products for Human Use first approved the drug in 2006. In 2008 and 2010 it reviewed this decision after further cases of PML were observed, deciding to continue approval in both instances.

The natalizumab case study is part of IMI-PROTECT Work Programme 5 (IMI-PROTECT, 2012). This programme aimed to develop methods for use in benefit-risk assessment, including both the underpinning modelling and the presentation of the results. The purpose of PROTECT is not to revisit any decisions made by regulatory authorities or any other decision maker, but rather to evaluate the appropriateness of frameworks and methodologies for benefit-risk assessment.

The case study seeks to address the question of whether natalizumab should be licensed, based on publicly available treatment effects data and using preference information elicited from patients or their representatives. In the initial case study, Nixon et al. (2014) used MCDA as a quantitative tool in combination with the ProACT-URL decision making framework (Hammond et al., 1999). This paper aims to extend this work by considering the impact of data uncertainty using a probabilistic model.

1.3 Sensitivity to uncertainty of treatment effects

When applying MCDA to a benefit-risk assessment, the alternatives are treatments, and the consequences of each alternative are derived from treatment effects data, typically observed in clinical trials (or sometimes observational studies or spontaneous reports). Naturally, any such estimates are subject to sampling error. It is standard practice in medicine for this uncertainty to be reported in study results, typically through a 95% confidence interval. However, there is currently no standard probabilistic method for propagating this uncertainty through an MCDA model in order to assess its impact on the overall utility of each treatment, as the method uses only point estimates of the treatment effects. Most implementations of MCDA rely on a one-way sensitivity analysis, but this process is cumbersome and lacks statistical rigour.

A more rounded approach is to use a probabilistic model that uses assumptions about the random processes underlying the data in order to gain insight into the expected spread of values. In benefit-risk decision-making, we are not directly concerned with the variability of observed data, or with testing hypotheses, but rather with what we can infer about the range of plausible values for the underlying population parameters, as it is these that are relevant to the benefit-risk balance. This leads us to focus on Bayesian models, which assign a probability distribution to each input parameter based on the likelihood of the data observed in the source studies. The distribution of the model output (in this case, the partial value of each criterion and the overall value associated with each treatment) can then be obtained by simulating values from the input distributions and using these in the MCDA model.

There have been various attempts to use probabilistic models to allow for parameter uncertainty in benefit-risk. Sutton et al. (2005) used Bayesian modelling to derive distributions of the benefit-risk balance of warfarin, an anticoagulant. The benefit-risk balance was expressed within a framework known as Net Clinical Benefit (NCB). The NCB approach is limited, however. In a mathematical sense it can be seen as a special case of MCDA in which only certain kinds of outcomes and/or partial value functions are permitted. We seek to develop a Bayesian model that can be used in a more general MCDA setting.

Hughes et al. (2007) used probabilistic methods to allow for uncertainty of treatment effects in benefit-risk assessment using a “decision tree” framework that is very similar to MCDA. However, their method relies on evidence from studies directly comparing each treatment, which was not available for the natalizumab case study. Caster et al. (2012) extended the approach of Hughes et al. (2007) by considering the uncertainty of preference data, an extension that could also be applied here.

Stochastic multi-criteria acceptability analysis (SMAA) is a variation of MCDA, designed for situations where preference weights are unknown or only partially known (Lahdelma et al., 1998). SMAA ranks alternatives by exploring all possible combinations of weights using MCDA and calculating how often each alternative is chosen. Tervonen et al. (2011) applied this method to benefit-risk assessment, using the J-SMAA software package (Tervonen, 2012) to carry out a probabilistic benefit-risk assessment using Monte Carlo simulations to allow for uncertainty in both clinical parameters and preferences.

SMAA provides a useful method for obtaining results from a multi-criteria decision model in the absence of preference data, but it must be recognized that there will usually be a high degree of uncertainty associated with the results. Greater certainty – and therefore better decisions – will generally be achieved if preference data are elicited *a priori* from the relevant stakeholders.

Furthermore, in its current form J-SMAA only allows uncertainty of treatment effects data to be characterized by a small number of closed-form distributions. This may be adequate for simple models, such as when comparisons are made between treatment groups in a single randomized controlled trial. However, within benefit-risk assessment, it is often necessary to pull together evidence from many studies using meta-analytical methods to synthesize evidence on a comparable basis. These situations require more complex statistical models. We use a case study to apply a Bayesian approach, directly modelling the outcomes observed in placebo-controlled trials for several competing treatments for the same indication. Indirect comparisons between the active treatments are then made, and ultimately we derive the distributions of the benefit-risk score for each treatment within a pre-specified MCDA model for the case study, details of which are set out below.

1.4 Objectives

The aims of this paper are: (i) to fit a suitable Bayesian model for the treatment outcome parameters in the natalizumab case study, using the original case study data; (ii) to use Markov Chain Monte Carlo simulation to generate the distribution of the overall benefit-risk score for each treatment in the case study, using the existing MCDA model; and (iii) to consider how the information on the benefit-risk distributions, together with appropriate metrics and visualisations, could influence the benefit-risk assessment for natalizumab and benefit-risk decision making in general.

2 Methods

2.1 MCDA model and data

The treatments to be compared, source data and benefit and risk outcomes are the same as those in the case study reported by Nixon et al. (2014). There are three active treatments for relapsing-remitting multiple sclerosis: natalizumab, interferon-beta and glatiramer acetate. The data is drawn principally from three clinical trials, one for each active treatment (Polman et al., 2006), (Jacobs et al., 1996), (Johnson et al., 1995). Each trial was of 2 years' duration. Not all of the benefits and risks were reported in every trial. In cases where no data is available it is assumed either that the number of patients experiencing (and not experiencing) an event was the same as in the common placebo group, or - for rare events that were not reported - that no events occurred in the relevant trial.

For PML incidence, no clinical trial data was available. For natalizumab, post-marketing data from an observational study was used (Bloomgren et al., 2012). As of February 2012, there were 2.1 cases of PML per 1000 patients receiving natalizumab. For all treatments except natalizumab, it was assumed that PML incidence was sufficiently low that it could be ignored (i.e. a deterministic value of 0 was used).

A disability progression event is defined as a ≥ 1.0 point increase in the Expanded Disability Status Scale with a baseline score ≥ 1.0 or a ≥ 1.5 point increase in score with a baseline score of 0, confirmed after 6 months. However, the natalizumab and glatiramer acetate trials used a different definition, with disability progression needing to be confirmed after only 3 months. To allow for this, the proportion of patients undergoing disability progression was adjusted by a factor of 0.71 in the treatment group and 0.79 in the placebo group, based on the results of a trial for fingolimod (another multiple sclerosis treatment), which monitored both definitions of the outcome (Kappos et al., 2010). It is acknowledged that the sampling variability of the conversion factor has not been included in the statistical model and therefore the variability of the disability progression outcome is likely to have been understated for natalizumab and glatiramer acetate; however, we regard this as an issue of data quality rather than a fundamental problem with the model.

Data on hypersensitivity reactions among patients taking glatiramer acetate was taken from the drug's package insert, which is publicly available online (Teva Neuroscience, 2012).

There were some outcome/treatment combinations for which no data of sufficient quality could be found. In these cases it was assumed either that the proportion of events and sample size observed in the relevant study was the same as in the common placebo group, or (for rare events that were not reported) that no events occurred in the relevant trial. These are assumed observations designed to fill in gaps in the data and hence allow full testing of the model, rather than deterministic assumptions about the underlying event rates which would restrict the model's functionality.

A data summary table, showing the source of each item of data, and highlighting where additional assumptions have been made, is included in the Supporting Information for this paper.

The MCDA method requires a pre-specified partial value function for each outcome, which maps the observed measure (such as the proportion of patients experiencing an adverse event) to a subjective measure of value or utility between 0 (the worst case scenario) and 1 (the best case scenario). Similarly, weights must be elicited in order to allow calculation of the overall utility for each treatment as a weighted sum of the partial values for all outcomes. This paper uses the partial value functions and weights selected by Nixon et al. (2014) for the natalizumab case study. The case study aimed to assess the benefit-risk balance from a regulatory perspective, with the regulator incorporating the views of patients. Accordingly, linear partial value functions were used, as these implicitly value each patient equally. The best case and worst case scenarios for each outcome were selected to cover the range of plausible values: from 0 to 100% for outcomes expressed as proportions, and from 0 to 2 (exceeding the upper confidence limits in the source studies), for the 2-year relapse rate. The final outcome is the route of administration, as patients can attach different convenience values to different administration methods. This aspect of treatment is included as a categorical variable, whose outcome is known with certainty for each treatment. The value placed on each route of administration, and the preference weights for all outcomes in the value tree, were elicited in decision conferences with patient

representatives. Further information on the case study has been given by Nixon et al. (2014); a summary of the key data is provided in Table 1.

<<Table 1 about here>>

The value functions are expressed in terms of the absolute value of the clinical outcomes, as this is the only scale on which the preferences that are fed into MCDA can be unambiguously defined. However, direct comparison of the outcomes in the treatment arms of different trials will be vulnerable to confounding due to heterogeneities in the baseline rate of events in the trial populations. In order to compare treatments on an absolute scale, a reference population must be chosen. Here we select the population of the natalizumab trial, as this can be assumed to be most relevant to the population that may benefit from the drug.

The absolute outcomes in the placebo and natalizumab arms are compared directly: as the trial subjects were randomized there should be no baseline heterogeneity. For the two active comparators, direct comparisons are always avoided: instead, the relative effect of each treatment compared to placebo is extracted from its own trial, and combine this with the baseline rate in the reference placebo population. This allows for a fair comparison as the reference placebo group is the same for all treatments and the relative treatment effect is assumed not to depend on the patient population. Similar methods have been employed for network meta-analysis; see for example Woods et al. (2010). The evidence network is represented in Figure 1, which indicates where direct and indirect comparisons have been made.

<<Figure 1 about here>>

2.2 Bayesian model

For a probabilistic implementation of a quantitative benefit-risk assessment, the aim is to express how uncertainties in the input parameters affect the uncertainty of the overall result. This makes Bayesian statistics, which permits direct probability statements concerning unknown parameters, a natural fit for this purpose.

In the natalizumab case study there are many parameters. Ten uncertain clinical outcomes (denoted by $k=1$ to $k=10$) contribute to the benefit-risk analysis. Each outcome is observed in as many as six different groups of patients (the data is drawn from three clinical trials denoted by $i=1,2,3$, one for each treatment, and each trial has a placebo arm and a treatment arm denoted by $j=1,2$ respectively). It is assumed that outcomes are independent between trials. Within each trial it is necessary to model the relative treatment effect as this is required for the indirect treatment comparisons.

The majority of the outcomes to be modeled (PML, seizures, congenital anomalies, reactivation of serious viral infections, hypersensitivity reactions, flu-like reactions, elevation of transaminases, infusion/injection reactions, disability progression) are binary events at the patient level. If X_{ijk} is the number of patients experiencing event k in arm j of trial i , then $X_{ijk} \sim \text{Bin}(n_{ijk}, \pi_{ijk})$ where n_{ijk} is the total number of patients in the trial arm and π_{ijk} is the underlying risk of the event occurring. Events are assumed to occur independently; however an allowance for possible correlations is introduced later as part of the sensitivity analysis.

Let $\mu_{ik} = (\text{logit}(\pi_{i1k}) + \text{logit}(\pi_{i2k}))/2$, the mean proportion over both arms of the trial, expressed on the log-odds scale, and $\delta_{ik} = \text{logit}(\pi_{i1k}) - \text{logit}(\pi_{i2k})$, the log odds ratio – our relative treatment effect for these outcomes. Then $\text{logit}(\pi_{i1k}) = \mu_{ik} - \delta_{ik}/2$, and $\text{logit}(\pi_{i2k}) = \mu_{ik} + \delta_{ik}/2$. By use of an appropriate prior, δ_{ik} can be constrained to a range that contains all plausible values for the log odds ratio and allows for some correlation between the outcome in the control and treatment arms. This is consistent with the main assumption underlying indirect treatment comparisons, i.e. that the relative treatment effect is relatively homogeneous as the baseline event rate varies.

A strong assumption is made regarding the exchangeability of the δ_{ik} , which are all assumed to be drawn from a common prior across all treatments i and outcomes k . This prior is $\delta_{ik} \sim N(d, \tau)$, where τ is the precision or inverse variance, with hyperpriors $d \sim \text{Unif}(-1, 1)$ (noting that this is an average effect across all treatments and outcomes, some negative and some positive on the log odds ratio scale) and $\tau \sim \text{Gamma}(3, 1)$. It is also necessary to specify a prior for μ : this is $\mu_{ik} \sim N(0, 0.25)$ where 0.25 is the inverse variance.

The log-odds transformation is used in order to avoid the possibility of obtaining proportions outside the interval (0,1). This is not the only approach that could be used, however; Bayesian models have been developed that operate on other scales such as relative risk and absolute risk difference (Warn et al., 2002).

The binomial likelihood implicitly assumes that all patients have equal exposure to the risk of an event. For disability progression, however, this is not the case: if an individual undergoes a disability progression partway through a trial, they are removed from the pool of susceptible individuals, and therefore contribute less than 2 years of exposure. For the sake of simplicity, and due to lack of more precise data on the timing of the disability progression events, the model does not allow for this.

A different model is used for PML. As this outcome was not observed in any clinical trials, data from a post-marketing observational study of natalizumab patients was used along with a different statistical model. Let $X \sim \text{Bin}(n, \pi)$ where n is the number of patients in the post-marketing study, X is the number experiencing PML and π is the risk of PML. We use a simple conjugate Bayesian inference model for π , with a $\text{Beta}(1/3, 1/3)$ prior after Kerman (2011). For placebo and the two remaining active treatments it was simply assumed that PML never occurs, which seems a reasonable assumption given the nature of PML. This allows a direct comparison to be made between treatments, so it is not necessary to model a relative effect in this case.

The remaining outcome, relapses ($k=10$), is expressed in terms of a rate per person-year of exposure to treatment. If $X_{ij,10}$ is the number of relapses in arm j of trial i (where 1 is the control arm and 2 is the placebo arm), then $X_{ij,10} \sim \text{Poisson}(n_{ij,10}\pi_{ij,10})$ where $n_{ij,10}$ is the total patient-years of exposure in the trial arm and $\pi_{ij,10}$ is the underlying event rate.

Let $\mu_{i,10} = (\log(\pi_{i,1,10}) + \log(\pi_{i,2,10}))/2$, the mean log relapse rate over both arms of the trial, and $\delta_{i,10} = \log(\pi_{i,1,10}) - \log(\pi_{i,2,10})$, the log rate ratio, our relative treatment effect for this outcome.

Then $\log(\pi_{i,1,10}) = \mu_{i,10} - \delta_{i,10}/2$, and $\log(\pi_{i,2,10}) = \mu_{i,10} + \delta_{i,10}/2$.

The parameters π , δ , μ , the variable X , and the observed data n have slightly different meanings for the relapse outcome ($k=10$ in the BUGS code) than for the binary outcomes ($k=1$ to 9). Arguably, δ and μ should also be drawn from a different prior. However, the priors specified above were judged to be appropriate for both sets of variables, and so for the sake of parsimony δ and μ have been drawn from the same priors for all outcomes in the value tree.

It is worth noting that the relapse rates corresponding to the raw event and follow-up data (as used here) are slightly different from the average rates reported in the published papers (perhaps due to the use of covariates in the original analyses of the trials). This results in some minor differences in our relapse figures from those reported in Nixon et al. (2014).

Empirical estimates $\hat{\delta}$ and $\hat{\mu}$ were calculated using the formulae below. These estimates were used to produce diagnostic plots as an initial check on the suitability of the model. The formulae for binary outcomes are taken from Smith et al. (1995); the formulae for the relapse rate parameters are slightly different.

For the binary outcomes we have:

$$\hat{\delta}_{ik} = \log\left(\frac{x_{i2k} + \frac{1}{2}}{n_{i2k} - x_{i2k} + \frac{1}{2}}\right) - \log\left(\frac{x_{i1k} + \frac{1}{2}}{n_{i1k} - x_{i1k} + \frac{1}{2}}\right) \text{ and } \hat{\mu}_{ik} = \left[\log\left(\frac{x_{i2k} + \frac{1}{2}}{n_{i2k} - x_{i2k} + \frac{1}{2}}\right) + \log\left(\frac{x_{i1k} + \frac{1}{2}}{n_{i1k} - x_{i1k} + \frac{1}{2}}\right)\right] / 2$$

For the relapse rate the formulae used are:

$$\hat{\delta}_{ik} = \log\left(\frac{x_{i2k} + \frac{1}{2}}{n_{i2k} + \frac{1}{2}}\right) - \log\left(\frac{x_{i1k} + \frac{1}{2}}{n_{i1k} + \frac{1}{2}}\right) \text{ and } \hat{\mu}_{ik} = \left[\log\left(\frac{x_{i2k} + \frac{1}{2}}{n_{i2k} + \frac{1}{2}}\right) + \log\left(\frac{x_{i1k} + \frac{1}{2}}{n_{i1k} + \frac{1}{2}}\right)\right] / 2$$

For outcomes where both treatment groups in a trial are assumed to experience an event at the same rate as the reference placebo group, both the formula for $\hat{\delta}_{ik}$ and the priors used in the analysis result in an odds ratio that implies an excess of events in whichever trial arm was the larger. To avoid any resulting bias, the sample size in each treatment group is also assumed to match the reference placebo group in these instances.

The final benefit in the case study is the route of administration for each treatment, which can be included in an MCDA model as patients and other stakeholders may have preferences for particular routes of administration based on their convenience. As the route of administration for each treatment is known with certainty, and uncertainty of preference data is beyond the scope of this paper, the utility associated with this aspect of treatment is modeled deterministically.

The model for binary outcomes is an adaptation of a model presented by Smith et al. (1995) for Bayesian meta-analysis of infections acquired in intensive care units. Here, the mathematical structure of the problem is somewhat different (we have many outcomes, many treatments, and one study per treatment rather than one outcome, one treatment, and many studies) but the model is largely analogous. We have used the same priors, which we believe are justified on the grounds set out by Smith et al. (1995). The sensitivity to the selection of priors was tested by repeating the analysis using various alternative non-informative priors.

The assumption of exchangeability of δ_{ik} across treatments (i) for a given outcome (k) seems reasonable in the absence of any prior knowledge of each treatment's clinical effects. The assumed exchangeability across the different clinical outcomes k is perhaps less easy to justify – for any particular treatment, there is no reason to suppose a priori that the odds ratio of two seemingly unrelated events should be drawn from the same distribution. To explore the impact of this assumption, four alternative model structures for δ_{ik} were evaluated: (a) a separate random effects model for each outcome, exchangeable across treatments; (b) a separate random effects model for each treatment, exchangeable across outcomes; (c) no exchangeability across either outcomes or treatments, i.e. $\delta_{ik} \sim N(d_{ik}, \tau)$; and (d) retaining the assumption of exchangeability across all outcomes and treatments but with $\delta_{ik} \sim \text{Student-}t(d, \tau, 4)$, a Student- t distribution with 4 degrees of freedom that is more robust to outliers than the Normal distribution, as also suggested by Smith et al. (1995).

Although the main aim for this case study was to select priors that are not unduly informative, it is possible to convey additional information through the use of informative priors if desired.

The total benefit-risk score is a linear combination of outcome variables, all with negative sign, as can be noted by observing that it is the sum of the partial value functions in Table 1. It is therefore to be expected that positive correlations between the outcomes will increase the variance of the overall score. To explore the impact of this, each outcome's contribution to the benefit-risk score was estimated in the primary analysis; this was then used to calculate the inflated variance of the total score assuming various levels of correlation between the outcomes. Further sets of simulations were then run in order to compare the treatments' scores based on the inflated variances and thus assess the potential impact of correlations on the overall result.

2.3 Simulations

Each outcome's contribution to the benefit-risk score is calculated as the partial value function $v_k()$ of the relevant outcome measure, which (except in the case of flu-like reactions) is used as observed for placebo and natalizumab (a direct comparison) but calibrated on the reference placebo group for beta-interferon and glatiramer acetate (indirect comparisons). In other words: the partial value of outcome k for placebo and natalizumab is $v_k[\pi_{1,1,k}]$ for placebo and $v_k[\pi_{1,2,k}]$ for natalizumab. For beta-interferon ($i=2$) and glatiramer acetate ($i=3$), indirect comparisons are necessary: the partial value is $v_k[\omega^{-1}\{\omega(\pi_{1,1,k}) \times \exp(\delta_{ik})\}]$ for binary outcomes (where ω maps probabilities to odds) or $v_k[\pi_{1,1,10} \times \exp(\delta_{10,k})]$ for the relapse rate. For flu-like reactions, the paucity of data prompted a different approach: the reference placebo group is taken from the beta-interferon trial, so some direct comparisons become indirect and vice-versa; the calculations are adjusted accordingly (following the principle that outcomes observed in different trials are never compared directly). The overall benefit-risk score for each treatment is then calculated as the weighted sum of every outcome's contribution. For this illustrative case study the "true" parameter in the reference population is used but, depending on the population and the purpose of the benefit-risk assessment, other approaches such as using the observed value may be considered appropriate.

The Markov Chain Monte Carlo simulations were implemented in the WinBUGS software package (Lunn et al., 2000). The code is included in the Supporting Information for this paper.

Comment [EW1]: Move to next subsection?

In the standard deterministic approach to MCDA, it is straightforward to make decisions based on the results of the analysis: the best alternative to choose is that with the highest overall preference-weighted utility or benefit-risk score. When faced with overlapping distributions of uncertain utilities from a probabilistic model, the decision-maker's task may be less straightforward. The decision could be based on the average result, although this would be to ignore the additional information on variability provided by a probabilistic analysis. An alternative decision metric is to rank each alternative based on the proportion of simulations in which it had the highest overall utility; this is known in SMAA as the rank 1 acceptability index (Tervonen et al., 2011).

3 Results

The first 5000 simulations were discarded to allow for burn-in, and convergence was checked using the diagnostic tools in WinBUGS. 10000 further iterations were then run to obtain samples from the joint posterior distribution. The median overall benefit-risk score for each treatment was checked against the results of the corresponding deterministic MCDA model and in each case the relative error was below 0.5%.

<<Figure 2 about here>>

Figure 2 appears broadly supportive of the assumption that δ and μ are independent. The values all lie within the ranges indicated by the priors, suggesting that the choice of prior was reasonable and not unduly informative. The distribution of δ , assumed to be Normal, appears somewhat skewed by two outlying values on the right of Figure 2: these are the empirical log odds ratios for outcomes that occurred among patients taking a treatment but never occurred in the control arm of the trial (hypersensitivity reactions for natalizumab and seizures for beta-interferon). It can also be noted that a number of the δ_{ik} take the precise value zero. These are the instances where no difference was assumed to have been observed between treatment arms (flu-like reactions for natalizumab and glatiramer acetate, transaminases elevation for beta-interferon and glatiramer acetate, and hypersensitivity reactions for beta-interferon).

<<Figure 3 about here>>

Figure 3 is a density plot of the posterior distribution of the expected overall benefit-risk score for each treatment. The medians and 95% credibility intervals of the benefit-risk scores for each treatment (both in absolute terms and relative to natalizumab) are shown in Table 2, together with the rank 1 acceptability index, i.e. the probability that each treatment has the highest score. The total benefit-risk scores correspond with those given by Nixon et al (2014) (aside from the slight difference in relapse rates noted above) but some figures in that paper are expressed in units corresponding to the utility of one disability progression event. To convert from our measure to theirs (or vice versa), one must simply divide (or multiply) by the weight of one disability progression event, i.e. 5.4%.

<<Table 2 about here>>

<<Table 3 about here>>

Table 3 illustrates the differential impact of each outcome on the benefit-risk assessment by displaying the difference in benefit-risk score between each active treatment and placebo, broken down by outcome. It is a combination of strong preference weighting and differential performance between the treatments that results in a large impact for a given outcome. Most of the impact comes from the benefits in the value tree, as the risks are all either infrequently observed or given a low weight in the MCDA model. It is possible however that the risks could have more impact if different weights were used.

The sensitivity to choice of model structure and hyperpriors for the δ_{ik} are shown in Table 4, using a variety of noninformative priors. Although these variations did have some effect on the posterior distributions of the benefit-risk score, it was not enough to have an impact on the overall result (natalizumab was still ranked 1st in 100% of the simulations in each case). [Inflating the variance of the](#)

benefit-risk score to allow for correlations between outcomes had more effect, but the overall result remained robust: with a correlation coefficient of 0.9 between all pairs of outcomes (the strongest correlation setting tested) the estimated probability of natalizumab being ranked first was reduced only to 96.5%.

<<Table 4 about here>>

Comment [EW2]: Add a table of results?

4 Discussion

MCDA is a powerful decision-making tool but, in its standard deterministic implementation, no indication is given of the probability that any observed differences are simply due to sampling error. For example, in the case study, interferon beta-1a outperforms glatiramer acetate on the basis of a deterministic analysis. It is only when uncertainty is introduced that one gets a sense of just how flimsy this result is. Similarly, it is only in the light of the probabilistic model – and in particular, the rank acceptability index figures – that the robustness of natalizumab’s advantage over the other treatments becomes clear.

The drawback of this approach compared to a deterministic analysis is its relative complexity. Fitting statistical models to data on multiple benefits and risks, possibly from many different sources, takes considerable resources, and so may not be appropriate for relatively simple benefit-risk assessments. For more finely balanced problems, such as periodic reviews of high-risk products, however, probabilistic methods could play an important role. Bayesian modelling in particular is well suited to applications involving hierarchical models or complex evidence networks, or where it is necessary to update the analysis from time to time in light of new evidence. These features are all often present in benefit-risk assessment.

The average benefit-risk score and the rank acceptability index are both reductive measures that cannot give the full picture of the distribution of results or the contribution of the individual benefits and risks. There are however a range of visualisations that can be used to communicate this information, and the application of visual communication tools to complex benefit-risk problems is an active area of study (Mt-Isa et al., 2013).

Although our focus is on the methodology and we do not seek to comment on the benefit-risk balance of natalizumab or its comparators in the real world, it is reassuring to note that our results are consistent with previous benefit-risk assessments of the same treatments. Thompson et al. (2008) found natalizumab to have a better benefit-risk profile than beta-interferon, which itself was better than placebo. Qizilbash et al. (2008) found that glatiramer acetate and beta-interferon both clearly outperformed placebo, but a direct comparison between the two active treatments was more finely balanced. We have investigated the robustness of our result to alternative preference weights and found that increasing the weight for PML from 53.8% to 93.9% (with the sum of the other weights rescaled to 6.1% and the ratios between them remaining constant) reduced the probability of natalizumab achieving rank 1 from 100% to 95% – still a very strong result.

The intention for this paper was to develop a model that could be applied as standard to all of the outcomes in the value tree, but in the end issues of data quality and asymmetry meant that pragmatic alterations had to be made on an outcome-by-outcome and drug-by-drug basis. For example, the lack of consistent endpoint reporting for disability progression meant that a conversion factor had to be used, and this introduced additional uncertainty that was not captured in the model. It was not possible to use the same reference placebo group for all outcomes, again due to lack of data. The pragmatic solutions we used to deal with these data problems may not be appropriate in benefit-risk assessments for real decision-making purposes, but we consider that we have demonstrated the general utility of our approach given appropriate data. The model in this paper incorporated elicited preference information, albeit in a deterministic fashion. There is no reason why the model could not be extended to allow for uncertainty of the preference data, and there are many ways this could be achieved. One possibility is to use an SMAA-style algorithm to explore the space of all possible weights. Another is to directly elicit preference distributions. Alternatively, Caster et al. (2012) proposed a method to take account of qualitative preference judgements. Any such model for preference uncertainty could simply be “bolted

on” to the model presented here, as within the MCDA framework there is no interdependency between preferences and outcome measures.

Reproducible Research “code.zip” in the Supporting Information contains the code and data files required to generate the results in this paper together with instructions for their use. Available at: <http://dx.doi.org/10.1022/bimj.XXXXXXX>

Acknowledgements The research leading to these results was conducted as part of the PROTECT consortium (Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium, www.imi-protect.eu) which is a public-private partnership coordinated by the European Medicines Agency. The PROTECT project has received support from the Innovative Medicine Initiative Joint Undertaking (www.imi.europa.eu) under Grant Agreement n° 115004, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution.

The following persons have participated in the IMI-PROTECT Work Package 5 and have given written consent for their acknowledgement: Billy Amzal, Deborah Ashby, Simon Ashworth, Alex Asimwe, Sinan B. Sarac, Johan Bring, Torbjorn Callreus, Ed Chan, Christoph Dierig, Yasemin Dil, Gerry Downey, Susan Duke, Adam Elmachtoub, David Gelb, Georgy Genov, Alesia Goginsky, Christine Hallgreen, Richard Hermann, Ian Hirsch, Steve Hobbiger, Kimberley Hockley, Diana Hughes, Juhaeri Juhaeri, Silvia Kuhls, Alfons Liefucht, Davide Luciani, Marilyn Metcalf, Alain Micaleff, Shahrul Mt-Isa, Jeremiah Mwangi, Richard Nixon, Rebecca Noel, John Pears, Mike Pearson, Ruth Peters, Lawrence Phillips, George Quartey, Susan Shepherd, Isabelle Stoeckert, Elizabeth J. Swain, Andrew Thomson, Laurence Titeux, Thaison Tong, Ioanna Tzoulaki, Rianne van den Ham, Tjeerd van Staa, Ed Waddingham, Nan Wang and Lesley Wise.

Conflicts of Interest

Richard Nixon is employed by a pharmaceutical company which either has, or may in the future develop treatments for Multiple Sclerosis. Deborah Ashby, Shahrul Mt-Isa and Ed Waddingham have declared no conflicts.

References

- Ashby, D. and Smith, A. (2000). Evidence-based medicine as Bayesian decision-making. *Statistics in Medicine*, **19**, 3291-3305.
- Bloomgren, G., Richman, S., Hotermans, C., Subramanyam, M., Goelz, S., Natarajan, A., Lee, S., Plavina, T., Scanlon, J.V., Sandrock, A. and Bozic, C. (2012). Risk of natalizumab-associated progressive multifocal leukoencephalopathy. *New England Journal of Medicine*, **366**, 1870-80.
- Caster, O., Noren, G.N., Ekenberg, L. and Edwards, I.R. (2012). Quantitative Benefit-Risk Assessment Using Only Qualitative Information on Utilities. *Medical Decision Making*, **32**, E1-E15.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*; **1**, 515-534.
- Hammond, J., Keeney, R. and Raiffa, H. (1999). *Smart Choices: A Practical Guide to Making Better Life Decisions*. (1st edn.). Harvard Business School Press, Boston.
- Hughes, D.A., Bayoumi A.M. and Pirmohamed M. (2007). Current Assessment of Risk-Benefit by Regulators: Is It Time To Introduce Decision Analyses? *Clinical Pharmacology & Therapeutics*, **82**, 123-127.
- IMI-PROTECT (2012). About PROTECT. Available at: <http://www.imi-protect.eu/about.shtml>.
- Jacobs, L., Cookfair, D., Rudick, R., Herndon, R., Richert, J., Salazar, A., Fischer, J., Goodkin, D., Granger, C., Simon, J., Alam, J., Bartoszak, D., Bourdette, D., Braiman, J., Brownscheidle, C., Coats, M., Cohan, S., Dougherty, D., Kinkel, R., Mass, M., Munschauer, F., Priore, R., Pullicino, P., Scherokman, B. and Whitham, R., and the Multiple Sclerosis Collaborative Research Group (1996). Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. *Annals of Neurology*, **39**, 285-294.
- Johnson, K.P., Brooks, B.R., Cohen, J.A., Ford, C.C., Goldstein, J., Lisak, R.P., Myers, L.W., Panitch, H.S., Rose, J.W. Schiffer, R.B. and the Copolymer 1 Multiple Sclerosis Study Group (1995). Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis: results of a phase III multicenter, double-blind placebo-controlled trial. *Neurology*, **45**, 1268-1276.
- Kappos, L., Radue, E., O'connor, P., Polman, C., Hohlfeld, R., Calabresi, P., Selmaj, K., Agoropoulou, C., Leyk, M., Zhang-Auberson, L. and Burtin, P. (2010). A placebo-controlled trial of oral fingolimod in relapsing multiple sclerosis. *New England Journal of Medicine*, **362** 387-401.

- Keeney, R. and Raiffa, H. (1976). *Decisions With Multiple Objectives: Preferences and Value Tradeoffs*. Wiley, New York.
- Kerman, J. (2011). Neutral noninformative and informative conjugate beta and gamma prior distributions. *Electronic Journal Of Statistics*, **5**, 1450-1470.
- Lahdelma, R., Hokkanen, J. and Salminen, P. (1998). SMAA - Stochastic multiobjective acceptability analysis. *European Journal of Operational Research*, **106**, 137-143.
- Lunn, D.J., Thomas, A., Best, N. and Spiegelhalter, D., 2000. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics And Computing*, **10**, 325-337.
- Mt-Isa, S., Hallgreen, C., Asimwe, A., Downey, G., Genov, G., Hermann, R., Hughes, D., Lieftucht, A., Noel, R., Peters, R., Phillips, L., Shepherd, S., Micallef, A., Ashby, D. And Tzoulaki, I. (2013). Review of visualisation methods for the representation of benefit-risk assessment of medication. Available at: <http://www.imi-protect.eu/benefitsRep.shtml>.
- Mt-Isa, S., Tzoulaki, I., Callréus, T., Micallef, A. and Ashby, D. (2011). Weighing benefit-risk of medicines: concepts and approaches. *Drug Discovery Today: Technologies*, **8**, e29-e35.
- Mussen, F., Salek, S. and Walker, S. (2007). A quantitative approach to benefit-risk assessment of medicines - Part 1: the development of a new model using multi-criteria decision analysis. *Pharmacoepidemiology and drug safety*, **16**, S2-S15.
- Nixon, R., Dierig, C., Mt-Isa, S., Stockert, I., Tong, T., Kuhls, S., Hodgson, G., Pears, J., Waddingham, E., Hockley, K. And Thomson, A. (2014). A case study using the ProACT-URL and BRAT frameworks for structured benefit risk assessment. *Biometrical Journal*, in press.
- Polman, C., O'connor, P., Havrdova, E., Hutchinson, M., Kappos L, Miller, D.H., Phillips, J., Lublin, F., Giovannoni, G., Wajgt, A., Toal, M., Lynn F, Panzara, M.A. and Sandroek, A.W. (2006). A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. *New England Journal of Medicine*, **354**, 899-910.
- Qizilbash, N., Mendez, I. and Sanchez-De La Rosa, R. (2012). Benefit-Risk Analysis of Glatiramer Acetate for Relapsing-Remitting and Clinically Isolated Syndrome Multiple Sclerosis. *Clinical Therapeutics*, **34**, 159-176.
- Smith, T., Spiegelhalter, D. and Thomas, A. (1995). Bayesian approaches to random-effects meta-analysis: a comparative study. *Statistics in Medicine*, **14**, 2685-2699.
- Sutton, A.J., Cooper, N.J., Abrams, K.R., Lambert, P.C. and Jones, D.R. (2005). A Bayesian approach to evaluating net clinical benefit allowed for parameter uncertainty. *Journal of Clinical Epidemiology*, **58**, 26-40.
- Tervonen, T. (2012). JSMAA: open source software for SMAA computations. *International Journal of Systems Science*, **45**, 69-81.
- Tervonen, T., Van Valkenhoef, G., Buskens, E., Hillege, H.L. and Postmus, D. (2011). A stochastic multicriteria model for evidence-based decision making in drug benefit-risk analysis. *Statistics in Medicine*, **30**, 1419-1428.
- Teva Neuroscience Inc. (2012). Copaxone: Package Insert and Label Information. Available at: <http://druginserts.com/lib/rx/meds/copaxone/>
- Thompson, J.P., Noyes, K., Dorsey, E.R., Schwid, S.R. and Holloway, R.G. (1992). Quantitative risk-benefit analysis of natalizumab. *Neurology*, **71**, 357-364.
- Warn, D., Thompson, S. And Spiegelhalter, D.J. (2002). Bayesian random effects meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales. *Statistics in Medicine*, **21**, 1601-1623.
- Woods, B.S., Hawkins, N., Scott, D.N. (2010). Network meta-analysis on the log-hazard scale, combining count and hazard ratio statistics accounting for multi-arm trials: A tutorial. *BMC Medical Research Methodology*, **10**, 54.

Figure legends and tables

Figure 1 Evidence network for the case study. Solid lines represent direct comparisons in studies; dashed lines represent indirect comparisons.

Figure 2 Scatter plot of empirical estimates $\hat{\delta}$ and $\hat{\mu}$

Figure 3 Posterior distribution of expected overall benefit-risk score

Table 1 Risks, benefits, measures, partial value functions (PVFs) and weights for the natalizumab case study

Endpoint	Measure	PVF	Weight
<i>Risks:</i>			
PML	% with event in 2 years	$1 - \text{proportion}$	53.8%
Seizures	% with event in 2 years	$1 - \text{proportion}$	5.4%
Congenital abnormalities	% with event in 2 years	$1 - \text{proportion}$	5.4%
Herpes reactivation	% with event in 2 years	$1 - \text{proportion}$	6.4%
Hypersensitivity reactions	% with event in 2 years	$1 - \text{proportion}$	1.1%
Flu-like reactions	% with event in 2 years	$1 - \text{proportion}$	1.1%
Transaminases elevation	% with event in 2 years	$1 - \text{proportion}$	10.7%
Infusion/injection reactions	% with event in 2 years	$1 - \text{proportion}$	2.7%
<i>Benefits:</i>			
Disability progression	% progressing in 2 years	$1 - \text{proportion}$	5.4%
Relapses	2-year relapse rate	$1 - 0.5 \times \text{rate}$	7.5%
Convenience	Route of administration*	1 (daily oral), 0.7 (monthly infusion), 0.5 (weekly intramuscular), 0 (daily subcutaneous)	0.5%

Table 2 Results: overall benefit-risk (B-R) score. Medians and (95% credibility intervals). BI = beta-interferon, GA = glatiramer acetate.

Treatment	B-R score overall	Relative to natalizumab	Probability of rank 1, 2, 3, 4			
Placebo	0.923 (0.919, 0.928)	0.034 (0.029, 0.040)	0.00	0.00	0.12	0.88
Natalizumab	0.957 (0.954, 0.961)	0.000 (0.000, 0.000)	1.00	0.00	0.00	0.00
BI	0.931 (0.919, 0.942)	0.026 (0.015, 0.039)	0.00	0.39	0.53	0.08
GA	0.933 (0.920, 0.944)	0.024 (0.013, 0.038)	0.00	0.60	0.35	0.05

Table 3 The impact of each outcome on the benefit-risk balance of each active treatment relative to placebo.

Endpoint	Contribution to benefit-risk score $\times 10^3$: active treatment less placebo		
	Natalizumab	Beta-interferon	Glatiramer acetate
<i>Risks:</i>			
PML	-1.14	0.00	0.00
Seizures	0.15	-0.25	-0.05
Congenital abnormalities	0.16	-0.04	-0.05
Herpes reactivation	0.18	-0.08	-0.10
Hypersensitivity reactions	-0.33	-0.01	-0.02
Flu-like reactions	0.00	-1.92	-0.02
Transaminases elevation	-0.91	-0.23	-0.21
Infusion/injection reactions	-1.53	-0.13	-0.43
<i>Benefits:</i>			
Disability progression	5.59	4.21	1.60
Relapses	33.6	8.86	14.4
Convenience	-1.62	-2.70	-5.40
TOTAL	34.2	7.71	9.72

Table 4 Median (standard error $\times 10^3$) benefit-risk score for alternate structures and priors. In each row, the specified prior is changed and the others remain as per the main text. The Normal parameters are mean and precision (inverse variance); the Gamma parameters are shape and rate (inverse scale).

Prior	Natalizumab	Beta-interferon	Glatiramer acetate
Main model: δ_{ik} exchangeable across treatments and outcomes			
Priors as per main text	0.957 (1.661)	0.931 (5.904)	0.933 (5.959)
$\mu \sim N(0,1)$	0.957 (1.666)	0.931 (5.544)	0.934 (5.667)
$\mu \sim Unif(-100,100)$	0.958 (1.629)	0.933 (5.670)	0.934 (6.103)
$d \sim Unif(-0.1,0.1)$	0.957 (1.625)	0.931 (5.493)	0.933 (5.648)
$d \sim N(0,1)$	0.957 (1.637)	0.931 (5.685)	0.933 (5.966)
$\tau \sim Gamma(1,0.1)$	0.957 (1.640)	0.931 (5.718)	0.933 (5.852)
$\tau \sim Half-Cauchy(25)^*$	0.958 (1.592)	0.931 (5.733)	0.933 (6.053)
$\sigma \sim Unif(0,100)^{**}$	0.957 (1.614)	0.931 (5.645)	0.933 (6.044)
(a) δ_{ik} exchangeable across treatments only			
Priors as per main text	0.957 (1.661)	0.932 (5.904)	0.934 (5.959)
$\mu \sim N(0,1)$	0.957 (1.684)	0.932 (5.504)	0.934 (5.602)
$\mu \sim Unif(-100,100)$	0.958 (1.610)	0.932 (5.525)	0.934 (5.914)
$d \sim Unif(-0.1,0.1)$	0.957 (1.619)	0.931 (5.583)	0.933 (5.649)
$d \sim N(0,1)$	0.958 (1.626)	0.932 (5.719)	0.934 (5.875)
$\tau \sim Gamma(1,0.1)$	0.957 (1.608)	0.933 (5.422)	0.935 (5.393)
$\tau \sim Half-Cauchy(25)^*$	0.957 (1.638)	0.933 (5.431)	0.934 (5.555)
$\sigma \sim Unif(0,100)^{**}$	0.957 (1.602)	0.932 (5.553)	0.934 (5.592)
(b) δ_{ik} exchangeable across outcomes only			
Priors as per main text	0.958 (1.626)	0.930 (6.038)	0.934 (6.267)
$\mu \sim N(0,1)$	0.957 (1.658)	0.930 (5.926)	0.933 (6.023)
$\mu \sim Unif(-100,100)$	0.958 (1.641)	0.930 (6.064)	0.934 (6.706)
$d \sim Unif(-0.1,0.1)$	0.957 (1.622)	0.931 (5.695)	0.933 (5.735)
$d \sim N(0,1)$	0.958 (1.622)	0.930 (5.982)	0.933 (6.399)
$\tau \sim Gamma(1,0.1)$	0.958 (1.625)	0.930 (5.986)	0.933 (6.349)
$\tau \sim Half-Cauchy(25)^*$	0.958 (1.659)	0.930 (6.023)	0.933 (6.443)
$\sigma \sim Unif(0,100)^{**}$	0.958 (1.670)	0.930 (5.969)	0.933 (6.486)
(c) δ_{ik} not exchangeable across outcomes or treatments			
Priors as per main text	0.958 (1.632)	0.931 (5.746)	0.933 (6.279)
$\mu \sim N(0,1)$	0.957 (1.658)	0.928 (6.079)	0.930 (6.265)
$\mu \sim Unif(-100,100)$	0.958 (1.638)	0.933 (5.705)	0.935 (6.267)
$d \sim Unif(-0.1,0.1)$	0.958 (1.621)	0.931 (5.632)	0.933 (5.620)
$d \sim N(0,1)$	0.958 (1.646)	0.931 (5.850)	0.933 (6.854)
$\tau \sim Gamma(1,0.1)$	0.957 (1.620)	0.931 (5.725)	0.934 (5.988)
$\tau \sim Half-Cauchy(25)^*$	0.957 (1.634)	0.931 (5.697)	0.933 (6.061)
$\sigma \sim Unif(0,100)^{**}$	0.957 (1.662)	0.931 (5.756)	0.934 (6.046)
(d) $\delta_{ik} \sim Student-t(d, \tau, A)$ (exchangeable across treatments and outcomes)			
Prior as per main text	0.958 (1.638)	0.931 (5.736)	0.933 (6.080)
$\mu \sim N(0,1)$	0.957 (1.650)	0.931 (5.772)	0.933 (5.576)
$\mu \sim Unif(-100,100)$	0.958 (1.623)	0.930 (7.457)	0.933 (6.640)
$d \sim Unif(-0.1,0.1)$	0.958 (1.614)	0.931 (5.773)	0.933 (5.753)
$d \sim N(0,1)$	0.958 (1.631)	0.931 (5.796)	0.933 (6.005)
$\tau \sim Gamma(1,0.1)$	0.958 (1.617)	0.931 (5.787)	0.933 (5.840)
$\tau \sim Half-Cauchy(25)^*$	0.958 (1.632)	0.931 (5.725)	0.933 (5.927)
$\sigma \sim Unif(0,100)^{**}$	0.958 (1.634)	0.931 (5.813)	0.933 (5.863)

*Half-Cauchy distribution with scale parameter 25; see Gelman (2006); ** where $\sigma = \tau^{-0.5}$.