

Network Meta-Analysis with Competing Risk Outcomes

A. E. Ades, PhD,¹ Ifigeneia Mavranzouli, MD, MSc,² Sofia Dias, PhD,¹ Nicky J. Welton, PhD,¹
Craig Whittington, PhD,² Tim Kendall, FRCPSych³

¹Department of Community Based Medicine, University of Bristol, Cotham Hill, Bristol, UK; ²National Collaborating Centre for Mental Health, Centre for Outcomes Research and Effectiveness, Research Department of Clinical, Educational and Health Psychology, University College London, London, UK; ³National Collaborating Centre for Mental Health, Royal College of Psychiatrists Research and Training Unit, London, UK

ABSTRACT

Background: Cost-effectiveness analysis often requires information on the effectiveness of interventions on multiple outcomes, and commonly these take the form of competing risks. Nevertheless, methods for synthesis of randomized controlled trials with competing risk outcomes are limited.

Objective: The aim of this study was to develop and illustrate flexible evidence synthesis methods for trials reporting competing risk results, which allow for studies with different follow-up times, and that take account of the statistical dependencies between outcomes, regardless of the number of outcomes and treatments.

Methods: We propose a competing risk meta-analysis based on hazards, rather than probabilities, estimated in a Bayesian Markov chain Monte Carlo (MCMC) framework using WinBUGS software. Our approach builds on existing work on mixed treatment comparison (network) meta-analysis, which can be applied to any number of treatments, and any

number of competing outcomes, and to data sets with varying follow-up times. We show how a fixed effect model can be estimated, and two random treatment effect models with alternative structures for between-trial variation. We suggest methods for choosing between these alternative models.

Results: We illustrate the methods by applying them to a data set involving 17 trials comparing nine antipsychotic treatments for schizophrenia including placebo, on three competing outcomes: relapse, discontinuation because of intolerable side effects, and discontinuation for other reasons.

Conclusions: Bayesian MCMC provides a flexible framework for synthesis of competing risk outcomes with multiple treatments, particularly suitable for embedding within probabilistic cost-effectiveness analysis.

Keywords: antipsychotic medication, Markov model, meta-analysis, mixed treatment comparisons, schizophrenia.

Introduction

It is common for randomized controlled trials (RCTs) to report more than one outcome. For purposes of designing a trial, it is generally felt that a single outcome should be prespecified to be the “primary” outcome. But, it is also recognized that, when pooling results from trials in a meta-analysis, there are several reasons why it may be appropriate to combine information on different outcomes. First, one might wish to “gather strength” by combining several similar outcomes, or to be able to combine results from trials that report different, but similar, outcomes [1,2]. Second, in a decision-making context, the different outcomes recorded may each have separate implications for estimating quality of life or economic consequences of each treatment.

A key requirement in the synthesis of multiple outcomes is that the correlation structures are appropriately represented [1–4]. In a meta-analysis, the correlations may occur at either or both of two levels. At the between-patient within-trial level, a patient’s outcome on one measure may be positively or negatively correlated with their outcome on another. At the between-trial level, trials in which there is a larger treatment effect on one measure may tend to be the trials on which there is a larger treatment effect on another (a positive correlation), or possibly a smaller one (a negative correlation).

Competing risk outcomes represent a special type of multiple outcome structure in which there are several different failure time outcomes that are considered mutually exclusive. Once a patient has reached any one of these end points, they are considered to be

out of the risk set. Censoring may also be occurring. When results from these trials are pooled in a meta-analysis, the competing risk structure should be taken into account so that the statistical dependencies between outcomes are correctly reflected in the analysis. These dependencies are essentially within-trial, negative correlations between outcomes, applying in each arm of each trial. They arise because the occurrence of outcome events is a stochastic process, and if more patients should by chance reach one outcome, then fewer must reach the others. The importance of these correlations in the context of meta-analysis of competing risk outcomes has been recognized by Trikalinos and Olkin [5], who suggest an approach based on normal approximation of the variances and covariances arising from multinomial data, and illustrate it with an application to a two-treatment meta-analysis with two competing outcomes.

In this article, we present an alternative approach based on hazards rather than probabilities, to more appropriately take account of time at risk. We use Bayesian Markov Chain Monte Carlo (MCMC) estimation [6], which we believe is more flexible in a situation with large numbers of treatments and outcomes. We begin by describing the data set, and we then explain the “proportional competing risks” assumption that underlies our approach. We next propose three alternative models: one fixed effects, and two random effects analyses. We also suggest some methods for model selection. In our discussion of the results, we compare our proposed method to previously described approaches, and consider some possible extensions.

Methods

Illustrative Data set

Figure 1 shows the network of comparisons of trials of antipsychotic medication for the prevention of relapse in people with

Address correspondence to: A.E. Ades, Department of Community Based Medicine, University of Bristol, Cotham House, Cotham Hill, Bristol BS6 6RT, UK. E-mail: t.ades@bristol.ac.uk
10.1111/j.1524-4733.2010.00784.x

cause-specific hazard at time t for outcome m , then the conditional probability that failure at time t is of type m , given there is a failure at time t is

$$\pi_m(t) = \lambda_m(t) / \sum_u \lambda_u(t) \tag{1}$$

The probability that failure occurs before time D and is of type m is: (the probability of surviving to t , times the probability of failure at t , times the conditional probability that failure is of type m), integrated over all times t between zero and D . This is:

$$\int_0^D \exp\left[-\int_0^t \sum_u \lambda_u(z) dz\right] \left[\sum_u \lambda_u(t)\right] \pi_m(t) dt = \int_0^D \lambda_m(t) \exp\left[-\int_0^t \sum_u \lambda_u(z) dz\right] dt \tag{2}$$

A form of “proportional hazards” assumption can now be made, which might be better termed “proportional competing risks,” in which the ratio (Equation 1) is constant over all t (i.e., $\pi_m(t) = \pi_m$). Under this restriction, Equation 2 becomes:

$$\pi_m \left[-\exp\left(-\int_0^D \sum_u \lambda_u(z) dz\right) \right] = \pi_m \left[1 - \exp\left(-\int_0^D \sum_u \lambda_u(z) dz\right) \right] \tag{3}$$

which is the probability of failure before time D times the probability that failure was of type m . In what follows, we assume constant underlying hazards, but Equation 3 shows that with proportional competing risks we are free to fit more complex survival distributions. This suggests some useful extensions to which we return in the discussion.

We now number the treatments from 1 to 9 (as shown in Table 1 and Fig. 1). Placebo is selected as the reference treatment 1. This is an arbitrary choice, but made to ease interpretation. We define the three outcomes as: $m = 1$ relapse, $2 =$ discontinuation caused by side effects, and $3 =$ discontinuation for other reasons. Then, each outcome is modeled separately on the log hazard rate scale. Considering a trial j that compares treatments k and b , the cause-specific log hazard for outcome m for treatment T is:

$$\log(\lambda_{j,T,m}) = \theta_{j,T,m} = \begin{cases} \mu_{j,m} & \text{if } (T = b), \quad b = 1, 2 \dots 8 \\ \mu_{j,m} + \delta_{j,b,k,m} & \text{if } (T = k), \quad k > b \end{cases} \quad m = 1, 2, 3 \tag{4}$$

where $\delta_{j,b,k,m}$ is the trial-specific log hazard ratio of treatment k relative to treatment b for outcome m . This can be interpreted as meaning that the b arm of the trial estimates the baseline log hazard $\mu_{j,m}$, while the k arm estimates the sum of the baseline hazard and the log hazard ratio. Note that b is not necessarily treatment 1, nor is it the same treatment in every trial; instead, it is simply the treatment with the lowest index in that trial. Thus, in a trial comparing treatments 2 and 3, $b = 2$. The trial-specific log hazard ratios are assumed to come from a common normal distribution, following the standard “random effects” approach:

$$\delta_{j,b,k,m} \sim \text{Normal}(d_{k,m} - d_{b,m}, \sigma_m^2), \tag{5}$$

The mean of this distribution is a difference between mean relative effects $d_{k,m}$ and $d_{b,m}$, which are the mean effects of treatments k and b , respectively, relative to (placebo) treatment 1, for outcome m , and we define $d_{1,m} = 0$. This formulation of the problem expresses the consistency equations [9], by which the dimensionality of the 11 treatment contrasts on which there are direct data (Table 1 and Fig. 1), are reduced to functions of the eight contrasts between the active treatments and placebo. The

between-trial variance of the random effect distribution, σ_m^2 , is specific to each outcome m . Three models for the variance are considered below.

We may write the model as Equation 4 because all the trials in this example are two-arm trials. An advantage of our approach, however, is that it can be readily extended to multi-arm trials, and Equation 5 should in fact be interpreted as a “fragment” of a multivariate normal distribution:

$$\begin{pmatrix} \delta_{j,1,2,m} \\ \cdot \\ \delta_{j,1,2,m} \\ \cdot \\ \delta_{j,1,9,m} \end{pmatrix} \sim \text{MVN} \left(\begin{pmatrix} d_{1,2,m} \\ \cdot \\ d_{1,2,m} \\ \cdot \\ d_{1,9,m} \end{pmatrix}, \begin{pmatrix} \sigma_m^2 & \sigma_m^2/2 & \sigma_m^2/2 \\ \cdot & \sigma_m^2 & \sigma_m^2/2 \\ \sigma_m^2/2 & \sigma_m^2 & \sigma_m^2/2 \\ \cdot & \cdot & \cdot \\ \sigma_m^2/2 & \sigma_m^2/2 & \sigma_m^2 \end{pmatrix} \right) \tag{6}$$

The underlying assumption in Equations 5 and 6 is therefore that every trial may be considered as if it were a multiarm trial on all nine treatments, that trial-specific relative treatment effects are sampled from the multivariate normal in Equation 6, and that treatments are missing at random. (Note that missing at random means missing without regard to treatment efficacy; it does not mean that treatment arms are equally likely to be included in a trial).

The linking function that relates the arm-specific log hazards $\theta_{j,k,m}$ to the likelihood is developed as follows. Figure 2 shows a Markov transition model with a starting state (remission) and three absorbing states (relapse, discontinuation caused by side effects, and discontinuation caused by other reasons). Based on Equation 3, if we assume constant hazards $\lambda_{j,k,m}$ acting over the period of observation D_j in years, the probability that outcome m had occurred by the end of the observation period for treatment T in trial j is:

$$p_{j,T,m}(D_j) = \frac{\lambda_{j,T,m}}{\sum_{u=1}^3 \lambda_{j,T,u}} \left[1 - \exp\left(-D_j \sum_{u=1}^3 \lambda_{j,T,u}\right) \right], \quad m = 1, 2, 3 \tag{7}$$

The probability of staying in the remission state ($m = 4$) is now simply 1 minus the sum of the probabilities of arriving at the three absorbing states, that is,

$$p_{j,T,m=4}(D_j) = 1 - \sum_{u=1}^3 p_{j,T,u}(D_j) \tag{8}$$

The data for each trial j and treatment T constitute a multinomial likelihood with four outcomes: moving to one of the three absorbing states, or remaining in the initial remission state. If $r_{j,T,m}$ is the number of patients on treatment T observed to reach end point m , and $n_{j,T}$ is the total number at risk on treatment T in trial j , then:

$$r_{j,T,m=1,2,3,4} \sim \text{Multinomial}(p_{j,T,m=1,2,3,4}, n_{j,T}), \text{ where } \sum_{m=1}^4 p_{j,T,m} = 1 \tag{9}$$

Three different models were fitted, differing solely in the specification of the between-trial variation in relative treatment effects, σ_m^2 . In the fixed effects model, $\sigma_m^2 = 0$, and the model collapses to: $\theta_{j,T,m} = \mu_{j,m} + d_{k,m} - d_{b,m}$, $b = 1, 2 \dots 8$, $k \geq b$.

In the random effect single-variance model, the between-trial variance $\sigma_m^2 = \sigma^2$, reflecting the assumption that the between-trial variation is the same for each outcome: a vague inverse gamma prior was put on the variance, $1/\sigma^2 \sim \text{gamma}(0.001, 0.001)$. In the random effect different variances model, each outcome has a different between-trial variation, and the vague uniform prior is put on each: $1/\sigma_m^2 \sim \text{gamma}(0.001, 0.001)$. A sensitivity analysis based on uniform priors was also examined: $\sigma \sim \text{uniform}(0, 5)$. This gave virtually identical posteriors for the

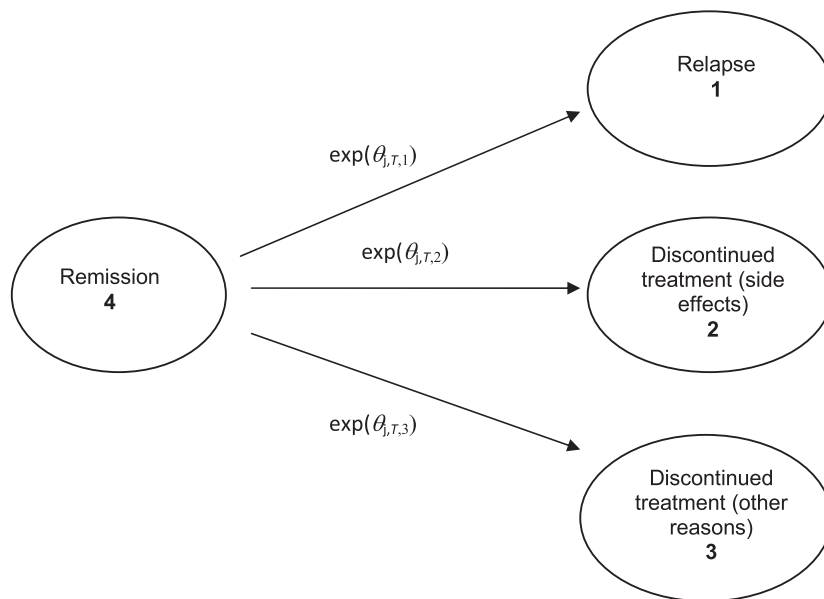


Figure 2 Markov model showing competing risk structure. Patients in trial j , on treatment T , move from remission to relapse, discontinued treatment (side effects), and discontinued treatment (other reasons), at rates $\exp(\theta_{i,T,1})$, $\exp(\theta_{i,T,2})$, and $\exp(\theta_{i,T,3})$, respectively.

treatment effects, but resulted in posterior distributions with “spikes” at σ_m values at or close to zero and spikes in the posterior mean treatment effects. Gamma priors, which give zero weight to infinite precision and hence to zero SD, were therefore used in the primary analyses reported below.

Finally, in each of the three models, vague Normal $(0, 100^2)$ priors were put on all the trial baselines $\mu_{j,m}$ and mean treatment effects $d_{k,m}$. The model for treatment effects (Equations 4 and 5) is therefore identical to that previously proposed for mixed treatment comparisons (MTCs) except that the multinomial likelihood (Equation 9) and linking function (Equation 7) are used, as is appropriate for the data at hand, in place of the binomial likelihood and logit link function proposed in most of previous work on these kinds of evidence structures [9–13].

Model Selection

Choice of models was based on the deviance information criterion [14]. This is a deviance measure of goodness of fit, \bar{D} , equal to the posterior mean of minus twice the log likelihood, penalized by an estimate of the effective number of parameters in the model, p_D . The DIC can be seen as a Bayesian measure analogous to the Akaike information criterion used in classical analysis, but which can also be applied to hierarchical models. Here, we adjust the standard deviance formula by subtracting the deviance of the saturated model (a constant). The contribution of each multinomial observation (trial j treatment T) to the deviance is:

$$\text{Dev}_{j,T} = -2 \sum_{m=1}^{m=4} r_{j,T,m} \log \left(\frac{\hat{r}_{j,T,m}}{r_{j,T,m}} \right), \quad \hat{r}_{j,T,m} = \hat{p}_{j,T,m} n_{j,T} \quad (10)$$

where $\hat{p}_{j,T,m}$ is an estimate of the probability of outcome m , for example, the estimate generated on some MCMC cycle, and \bar{D} is the posterior mean of the sum of the deviance contributions over all data points, $\sum_{j,T} \text{Dev}_{j,T}$. In a model that fits well, \bar{D} is expected to roughly approximate the number of data points. In this data set, with 15 two-arm trials each reporting three outcomes (the fourth, “censored” outcome is predicted from the number at risk less the other outcomes), the number of indepen-

dent data points is 90; p_D is equal to $\bar{D} - \hat{D}$, where \hat{D} is the sum of the deviance contributions, evaluated at the posterior mean of the fitted values [15].

Computation

Models were estimated using the freely available Bayesian MCMC software WinBUGS 1.4.3 [16]. Convergence for all models occurred within 10,000 to 25,000 iterations as assessed by the Brooks–Gelman–Rubin criteria [17]. Results are based on 150,000 samples, from three separate chains with disparate starting values for fixed effect, and five chains for the random effect models, taken after the first 60,000 were discarded. We also established that, in each model, all the chains converged to the same posterior. The code for each model is available on the ISPOR Web site as Supporting Information, in Appendix A at: http://www.ispor.org/Publications/value/ViHsupplementary/ViH13i8_Ades.asp.

Results

The global goodness-of-fit statistics rule out the fixed effects model (Table 2), which fits relatively poorly: the mean posterior deviance is 119.8 compared to the number of data points, 90. This model also has the highest DIC because the relatively poor fit is not sufficiently compensated by the fact that it has fewer effective parameters. The DIC results slightly favor the three-variance model over the single-variance model.

Posterior summaries of the log relative hazard rates $d_{k,m}$ for each treatment relative to placebo are shown in Table 2. For each of the three outcomes, the ranking of the treatments shows a high degree of consistency regardless of the choice of model. Zotepine is the most effective treatment in preventing relapse, followed by olanzapine; amisulpride causes the least discontinuation caused by intolerable side effects followed by risperidone and olanzapine; risperidone appears to cause the fewest discontinuations for other reasons, followed by amisulpride and then olanzapine. Nevertheless, the rankings of the posterior mean treatment effects do not take uncertainty into account.

Table 2 Goodness of fit, relative effects (posterior mean log hazard rates $d_{k,m}$ relative to placebo), and between-trial heterogeneity (posterior median and credible intervals for between-trial SD), for fixed and random effect models

	Fixed effect			Random effect single variance			Random effect three variances		
Goodness of fit statistics									
\bar{D}	119.8			95.5			92.6		
p_D	68.3			78.2			78.5		
DIC	188.1			173.7			171.0		
Relapse	LHR	rank	SD	LHR	rank	SD	LHR	rank	SD
1. Placebo (ref)	0.000	9		0.000	9		0.000	9	
2. Olanzapine	-1.577	2	0.230	-1.500	2	0.337	-1.474	2	0.421
3. Amisulpride	-1.151	4	0.479	-1.029	5	0.614	-0.980	6	0.727
4. Zotepine	-2.098	1	0.582	-2.101	1	0.739	-2.102	1	0.886
5. Aripiprazole	-0.726	8	0.183	-0.724	8	0.479	-0.729	7	0.689
6. Ziprasidone	-1.118	5	0.194	-1.199	4	0.407	-1.227	3	0.558
7. Paliperidone	-1.020	6	0.257	-1.020	6	0.512	-1.019	5	0.709
8. Haloperidol	-0.910	7	0.266	-0.771	7	0.436	-0.716	8	0.560
9. Risperidone	-1.284	3	0.288	-1.163	3	0.489	-1.130	4	0.640
Discontinuation caused by intolerable side effects									
1. Placebo (ref)	0.000	6		0.000	6		0.000	6	
2. Olanzapine	-1.345	3	0.333	-1.257	2	0.430	-1.189	2	0.605
3. Amisulpride	-1.688	1	0.680	-1.676	1	0.778	-1.701	1	1.001
4. Zotepine	1.144	8	0.600	1.141	8	0.750	1.162	8	1.068
5. Aripiprazole	0.026	7	0.382	0.013	7	0.583	0.021	7	0.941
6. Ziprasidone	-1.057	4	0.343	-1.046	4	0.494	-1.015	4	0.750
7. Paliperidone	1.295	9	1.397	1.320	9	1.454	1.244	9	1.661
8. Haloperidol	-0.922	5	0.380	-0.865	5	0.524	-0.848	5	0.759
9. Risperidone	-1.358	2	0.410	-1.164	3	0.592	-1.026	3	0.930
Discontinuation caused by other reasons									
1. Placebo (ref)	0.000	7		0.000	7		0.000	7	
2. Olanzapine	-0.526	3	0.227	-0.484	3	0.332	-0.494	3	0.242
3. Amisulpride	-0.603	2	0.254	-0.582	2	0.474	-0.596	2	0.299
4. Zotepine	-0.473	4	0.318	-0.476	4	0.544	-0.482	4	0.352
5. Aripiprazole	0.235	9	0.385	0.223	8	0.587	0.219	8	0.410
6. Ziprasidone	-0.450	5	0.272	-0.368	6	0.440	-0.421	5	0.301
7. Paliperidone	0.743	9	0.462	0.733	9	0.642	0.736	9	0.485
8. Haloperidol	-0.444	6	0.258	-0.426	5	0.431	-0.411	6	0.288
9. Risperidone	-1.084	1	0.285	-1.031	1	0.495	-1.072	1	0.330
Between-trial SD									
σ				Median (95% CI)			Median (95% CI)		
σ_1				0.404 (0.17–0.75)			0.561 (0.25–1.24)		
σ_2							0.484 (0.04–2.13)		
σ_3							0.094 (0.03–0.41)		

Figure 3 shows the probabilities that each treatment is ranked j th ($j = 1, 2 \dots 9$) for each of the three outcomes, based on the different variances model. As with Table 2, a rank of 1 indicates that the treatment is “best” at avoiding that (unwanted) outcome. These rankograms, introduced by Cipriani et al. [18], provide an “at a glance” summary that is hard to achieve from numbers alone, as they simultaneously demonstrate not only the relative rankings of treatments on each outcome, but also the very considerable uncertainty in inferences about relative efficacy. For example, although Table 2 shows that zotepine is ranked first in reducing relapse, Figure 3 reveals that the probability that zotepine is best in this outcome is only about 0.6. Olanzapine has a relatively high probability of being ranked among the highest three places for all three outcomes.

Although the mean effects $d_{k,m}$ are similar across models, their posterior uncertainty as assessed by the posterior SD depends strongly on the model used. As one would expect, the random effect summaries are much more uncertain than the fixed effect summaries. Nevertheless, in the random effect models, posterior uncertainty is greater in the three-variance model than in the single-variance version, but only for the relapse and discontinuation caused by intolerable side effect outcomes. The reason for this can be seen in the summaries of the between-trial SDs

(Table 2). Although the single-variance model produces an average between-trial SD with relatively narrow credible limits, the three-variance model has a particularly low SD for the relative effects on discontinuation for other reasons, and higher SDs with wide credible intervals for the other two outcomes. We have no ready explanation for this apparent difference in between-trial variability between outcomes, which impacts on the uncertainty in the mean treatment effect measures. Alternative uniform (0, 5) priors for the between-trial SD parameters were used in sensitivity analyses (not shown). These result in posterior estimates of σ and σ_m that were higher, but this had only minor effects on posterior distributions of $d_{k,m}$.

Discussion

We have described and illustrated a simple approach to meta-analysis of trials with multiple, mutually exclusive event outcomes. The special feature of such “competing risk” data is the negative correlations between outcomes within trial arms. The Bayesian MCMC framework, particularly with WinBUGS, allows the user to specify a multinomial likelihood along with conventional fixed or random effect models for relative treatment differences.

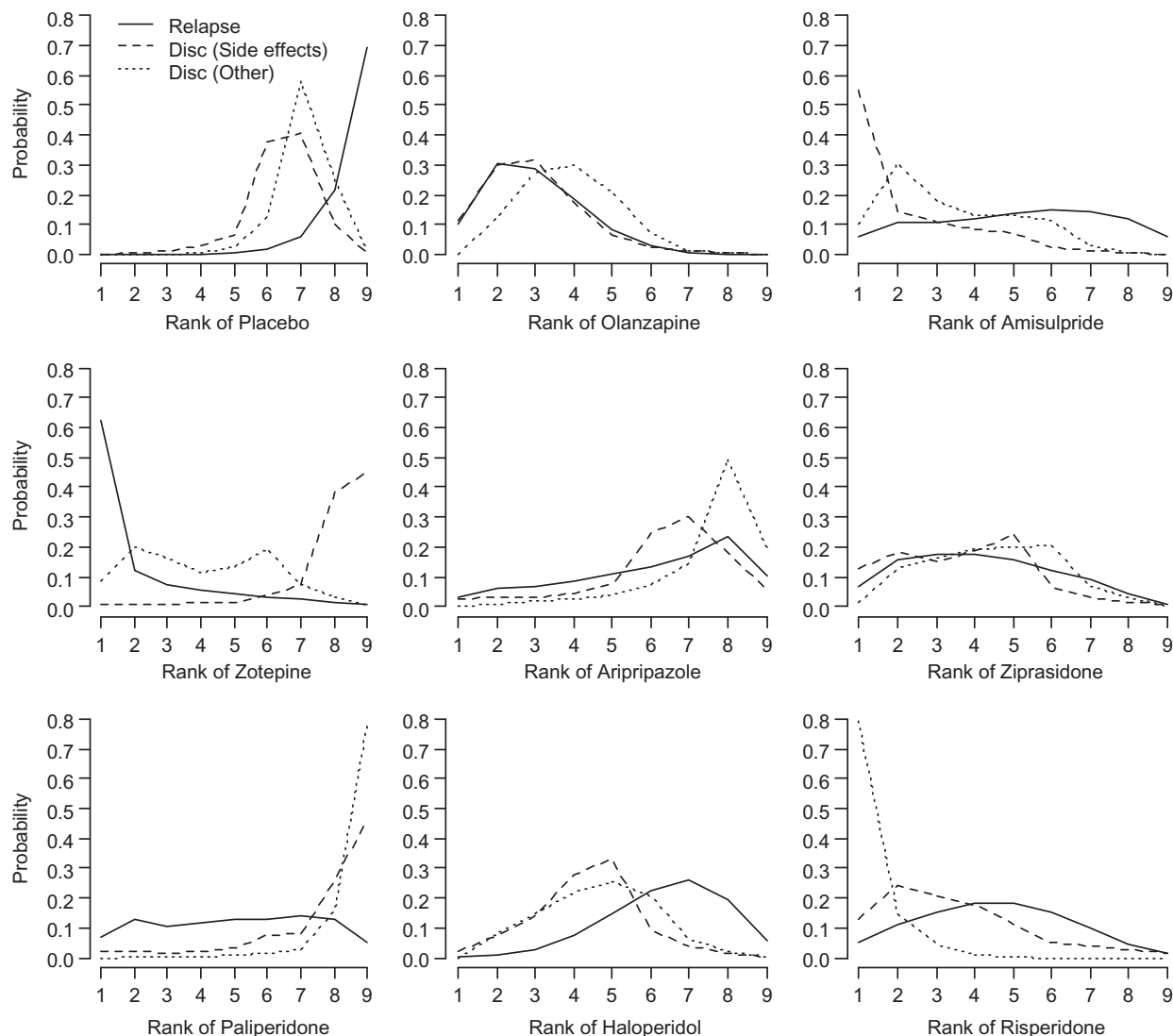


Figure 3 Treatment rankings: probability that each treatment is the j th best ($j = 1, 2, \dots, 9$), on avoiding relapse, discontinuation for side effects, and discontinuation for other reasons. Different variances model.

Almost invariably, when meta-analyses have been applied to competing risk outcomes, they have looked at outcomes one at a time, and have not attempted an analysis that looks at the multiple outcomes simultaneously within a single coherent framework. Methods for competing risk meta-analysis appear not to have been previously described, although recently Trikalinos and Olkin have presented methodology for what they describe as “mutually exclusive binary outcomes” [5]. They propose synthesis of log odds ratios, log relative risk, or risk difference outcomes from multinomial data arising from what is in fact a competing risk situation. Their approach is to take the estimated log odds ratios (or risk differences, or log relative risks) as data, and to develop expressions for the variances and covariances across outcomes based on normal theory. We believe that our approach has several advantages over this scheme. First, the use of the multinomial likelihood avoids the approximations required when the normal likelihood is used, and also avoids the need to add a constant to zero cells, a maneuver that generates bias [19,20]. Second, although calculation of the variances and cova-

riances is relatively easy when there are two competing outcomes, it has the potential to be error prone if there are more than four or five. Similarly, the covariances would become still more complex to specify in multiarm trials, whereas our Bayesian approach can be readily extended [9,11], as shown in Equation 6. A third advantage of the Bayesian simulation framework is that it is suitable both for inference and for use in probabilistic decision models, or within cost-effectiveness analyses. Frequentist methods would need to be supplemented by, for example, bootstrap sampling, in order to be used in such environments.

It is important for users to be aware of the underlying assumptions being made in applications of MTC synthesis. The key assumption is that, if all trials had included all nine treatments, the relative effects $\delta_{i,b,k,m}$ would, for each outcome m , be exchangeable across trials. This is equivalent to assuming that treatments are missing at random, which means that the absence of data on a treatment is independent of the treatment effect. Such assumptions, commonly made in evidence synthesis, are very hard to verify, and one must generally rely on experts with

knowledge of the trials and the subject area to confirm its plausibility. This ensemble of trials was considered sufficiently homogeneous to form the basis for a comparative assessment of the eight active treatments in the context of clinical guideline development [7].

But, the main advantage of the methods proposed here is that they are based on hazard rates and their ratios, and as a result they can correctly accommodate sets of trials that have been run to different follow-up times, as is the case in the present data set. Methods based on odds ratios, risk ratios, and risk differences, by contrast, cannot give consistent results for trials with different follow-up times, and it would be mathematically impossible to construct survival time distributions that would be compatible with constant measures at different durations. Use of these measures in trials with different follow-up times not only introduces unwanted heterogeneity to the estimates, but will also result in estimates that cannot, strictly speaking, be applied to other follow-up times. The Trikalinos and Olkin proposals could, however, be extended to hazard ratios, by introducing expressions for variances and covariances of log hazard ratios. Although this would yield coherent rate models with comparable estimates to those derived from our approach, it would still suffer from the disadvantages of normal approximations and an explosion of complexity with multiple competing outcomes. These factors, along with the intrinsic compatibility of MCMC posterior sampling with probabilistic decision analysis, make the modeling approach proposed here the obvious choice in cost-effectiveness analysis.

The analyses proposed here can be extended in a number of ways. The purpose of our models is to take into account the negative correlations between outcomes at the patient level. Nevertheless, it may also be worth considering the additional possibility of correlations between outcomes at the trial level. For example, it may be plausible to assume that trials in which treatments which are more effective in preventing relapse may be those in which treatments are more likely to lead to intolerable side effects leading to treatment discontinuation. This suggests a class of extensions that might focus attention on the possibility of correlations between the trial-specific treatment efficacy and discontinuation rates, in which the treatment effects are drawn from an extended multivariate normal distribution. If we write Equation 5 as $\delta_{i,m} \sim MVN(d_m, \Sigma_{mm})$, we might consider, for example:

$$\begin{pmatrix} \delta_{i,1} \\ \delta_{i,2} \\ \delta_{i,3} \end{pmatrix} \sim MVN \left(\begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{21} & \Sigma_{31} \\ \Sigma_{12} & \Sigma_{22} & \Sigma_{32} \\ \Sigma_{13} & \Sigma_{23} & \Sigma_{33} \end{pmatrix} \right) \quad (11)$$

where the off-diagonal covariance matrices carry terms for correlations between treatment effects on different outcomes. The present model would then be a special case where these correlations were assumed to be zero. Experience with heterogeneous variance models [21] in MTC suggests that very large amounts of data, especially from multiarm trials, would be required to identify the off-diagonal covariances, and that the impact on the posterior distributions of mean treatment effects may be limited. Informative priors could of course be placed on the correlations, although it is no trivial exercise to ensure the matrix is positive definite. A useful parameterization has been suggested by Lu and Ades [21].

Because of the underlying rate parameterization, it would be possible to incorporate additional data that are reported in the form of number of events and time at risk [22]. It is also relatively easy to incorporate data where outcomes are reported at more than one time point. This is best achieved by conditioning the data for subsequent intervals on survival to the end of the pre-

vious interval, in order to achieve independent Bernoulli samples, as illustrated by Lu et al. [23]. The WinBUGS code provided in the Appendix (http://www.ispor.org/Publications/value/ViHsupplementary/ViH13i8_Ades.asp) would not need to be altered to accommodate additional data structured in this way.

The existence of data at multiple time points would facilitate further extensions of the competing risk analyses to more complex underlying survival distributions, such as the Weibull or other distributions. This represents a very substantial liberalization of the modeling assumptions. Depending on how much data are available, Weibull shape parameters could be held constant or allowed to vary across trials [24]. Alternatively, a piece-wise constant underlying hazard model offers considerable flexibility [23].

These extensions, of course, all require not only a proportional hazard assumption for relative treatment effects, but also the proportional competing risk assumption described in the introduction. The latter is a strong assumption, but it can be relaxed, for example, in a piece-wise constant hazard framework. This could accommodate the possibility that, for example, the proportional risk of discontinuation caused by side effects could be higher in an initial period. It is also possible to incorporate data from trials that fail to report a subset of end points separately, for example, by aggregating discontinuation caused by side effects with other reasons for discontinuing treatment.

Source of financial support: No specific funding was received for this work. A.E.A., S.D., and N.J.W. were supported by Medical Research Council funding to the Health Services Research Collaboration, transferred to University of Bristol.

References

- Nam I-S, Mengerson K, Garthwaite P. Multivariate meta-analysis. *Stat Med* 2003;22:2309–33.
- Riley RD, Abrams KR, Lambert PC, et al. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Stat Med* 2007;26:78–97.
- Berkey CS, Anderson JJ, Hoaglin DC. Multiple-outcome meta-analysis of clinical trials. *Stat Med* 1996;15:537–57.
- Berkey CS, Hoaglin DC, Antczak-Bouckoms A, et al. Meta-analysis of multiple outcomes by regression with random effects. *Stat Med* 1998;17:2537–50.
- Trikalinos T, Olkin I. A method for the meta-analysis of mutually exclusive binary outcomes. *Stat Med* 2008;27:4279–300.
- Gilks WR, Richardson S, Spiegelhalter DJ. *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall/CRC, 1996.
- National Collaborating Centre for Mental Health. *Schizophrenia: The NICE Guidelines on Core Interventions in the Treatment and Management of Schizophrenia in Primary and Secondary Care*. Updated edition. Leicester & London: British Psychological Society and Royal College of Psychiatrists, 2010.
- Cox DR, Oakes D. *Analysis of Survival Data*. London: Chapman & Hall Ltd, 1984.
- Lu G, Ades AE. Assessing evidence consistency in mixed treatment comparisons. *J Am Stat Assoc* 2006;101:447–59.
- Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med* 2002;21:2313–24.
- Higgins JPT, Whitehead J. Borrowing strength from external trials in a meta-analysis. *Stat Med* 1996;15:2733–49.
- Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005;331:897–900.
- Spiegelhalter DJ, Best N, Carlin BP. Bayesian deviance: the effective number of parameters, and the comparison of arbitrarily complex models. *Division of Biostatistics: University of Minnesota*, 1998.
- Spiegelhalter DJ, Best NG, Carlin BP, et al. Bayesian measures of model complexity and fit. *J R Stat Soc Series B Stat Methodol* 2002;64:583–616.

- 15 Welton NJ, Ades AE. A model of toxoplasmosis incidence in the UK: evidence synthesis and consistency of evidence. *Appl Stat* 2005;54:385–404.
- 16 Spiegelhalter D, Thomas A, Best N, Lunn D. WinBUGS user manual version 1.4 January 2003. Upgraded to Version 1.4.3. Available from: <http://www.mrc-bsu.cam.ac.uk/bugs> [Accessed July 20, 2010].
- 17 Brooks SP, Gelman A. Alternative methods for monitoring convergence of iterative simulations. *J Comput Graph Stat* 1998;7:434–55.
- 18 Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 12 new generation antidepressants: a multiple-treatments meta-analysis. *Lancet* 2009;373:746–58.
- 19 Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Stat Med* 2004;23:1351–75.
- 20 Bradburn MJ, Deeks JJ, Berlin JA, Localio AR. Much ado about nothing: a comparison of the performance of meta-analysis methods with rare events. *Stat Med* 2006;26:53–77.
- 21 Lu G, Ades A. Modelling between-trial variance structure in mixed treatment comparisons. *Biostatistics* 2009;10:792–805.
- 22 Welton NJ, Ades AE. Estimation of Markov chain transition probabilities and rates from fully and partially observed data: uncertainty propagation, evidence synthesis and model calibration. *Med Decis Making* 2005;25:633–45.
- 23 Lu G, Ades AE, Sutton AJ, et al. Meta-analysis of mixed treatment comparisons at multiple follow-up times. *Stat Med* 2007;26:3681–99.
- 24 Welton NJ, Cooper NJ, Ades AE, et al. Mixed treatment comparison with multiple outcomes reported inconsistently across trials: evaluation of antivirals for treatment of influenza A and B. *Stat Med* 2008;27:5620–39.