STATISTICS IN MEDICINE
Statist. Med. 2008; **27**:418–434
Published online 4 May 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.2897

Flexible parametric models for random-effects distributions

Katherine J. Lee*, † and Simon G. Thompson

MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 0SR, U.K.

SUMMARY

It is commonly assumed that random effects in hierarchical models follow a normal distribution. This can be extremely restrictive in practice. We explore the use of more flexible alternatives for this assumption, namely the *t* distribution, and skew extensions to the normal and *t* distributions, implemented using Markov Chain Monte Carlo methods. Models are compared in terms of parameter estimates, deviance information criteria, and predictive distributions. These methods are applied to examples in meta-analysis and health-professional variation, where the distribution of the random effects is of direct interest. The results highlight the importance of allowing for potential skewing and heavy tails in random-effects distributions, especially when estimating a predictive distribution. We describe the extension of these random-effects models to the bivariate case, with application to a meta-analysis examining the relationship between treatment effect and baseline response. We conclude that inferences regarding the random effects can crucially depend on the assumptions made and recommend using a distribution, such as those suggested here, which is more flexible than the normal. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: random effects; flexible modelling; skewing; predictive distribution; clustering; meta-analysis

1. INTRODUCTION

Hierarchical models are used to analyse data from a wide range of medical settings, including cluster randomized trials, repeated measures designs, meta-analysis and to allow for health-professional variation. In each of these situations there are two sources of variation, one at the observation level and another at a higher level (between clusters, individuals, studies or health-professionals in the examples above). A standard method to allow for this second level of variability is to use random effects. For mathematical convenience these are traditionally assumed to be normally distributed [1], although this is not necessarily appropriate in practice. This assumption for random effects has been criticized in case studies for providing restricted and potentially inappropriate inferences [2], for causing 'over shrinkage' of the random-effects distribution [3], and for being

*Correspondence to: Katherine J. Lee, MRC Clinical Trials Unit, 222 Euston Road, London NW1 2DA, U.K.
†E-mail: kjl@ctu.mrc.ac.uk

Received 11 May 2006
Accepted 7 March 2007

less efficient under systematic departures from normality [4]. In addition, it is not easy to assess the validity of this normality assumption, particularly when there are only a small number of higher-level units [5]; normal probability plots for unobserved latent random effects need to be interpreted with caution, and there is no straightforward diagnostic to test for non-normality [6].

Previous studies suggest that fixed covariate effects are fairly robust to misspecification of the random-effects distribution [7], although such misspecification can more seriously affect the estimates of the variance of the random-effects distribution [5]. Previous results have found inflated variances for fixed and random effects when normality is assumed for the random effects in heavy-tailed data [8] or if there is skewing present [9]. A number of more flexible distributions have been proposed as alternatives to the normal distribution for modelling random effects, including non-parametric distributions [10, 11], the t distribution [3, 12], the gamma distribution [5], and a mixture of normals [3, 13]. Skewed versions of the normal and t distribution [9] and stable distributions [14] have been applied to non-normal data but not to random effects. The aim of this paper is to investigate the use of some of these flexible alternatives for random effects and to assess their impact on the results and conclusions in relevant examples. We concentrate on the t distribution and novel use of skew extensions to the normal and t distributions, but return to other potential alternatives in the discussion.

Section 2 introduces the standard normal random-effects model and describes the t, skewed normal and skewed t distributions as alternatives, along with details on fitting these distributions using Markov chain Monte Carlo (MCMC) methods. Section 3 investigates the use of these flexible distributions, applying them to simulated data and examples in meta-analysis and health-professional variation. Section 4 considers a bivariate extension of these flexible models using a regression construction approach, which is illustrated in a bivariate meta-analysis to assess the relationship between treatment effect and baseline response. Section 5 provides a discussion of the issues raised.

2. FLEXIBLE DISTRIBUTIONS FOR RANDOM EFFECTS

We first consider the standard random-effects model and refer, for generality, to observations grouped into 'clusters' in the context of a clinical trial. For example, a simple random-effects model for estimating a treatment effect in a continuous outcome y_{ij}, the ith observation in the jth cluster, is

$$\begin{aligned} y_{ij} &\sim \text{Normal}(\mu_{ij}, \sigma_e^2) \\ \mu_{ij} &= \beta t_{ij} + u_j \\ u_j &\sim \text{Normal}(\mu, \sigma^2) \end{aligned} \tag{1}$$

where $t_{ij} = 0$ or 1 is the treatment group indicator, β the treatment effect, and u_j the random effect of cluster j. In a standard hierarchical model these random effects are assumed to have a normal distribution; the mean μ represents the mean outcome in the control group and σ^2 the between cluster variance. It is this normality assumption that is of interest here.

For binary outcomes $y_{ij} = 0$ or 1, we can replace the first two lines in (1) by $y_{ij} \sim \text{Binomial}(1, \pi_{ij})$ and $\text{logit}(\pi_{ij}) = \beta t_{ij} + u_j$, leaving the third line unchanged [15]. Similar extensions are available for ordinal outcomes [16] and survival outcomes [17].

DOI: 10.1002/sim

2.1. The t distribution

One possible alternative to the normal assumption for the random effects u_j is to use a t distribution, $t(\mu, v, k)$, which has probability density function (pdf)

$$p(u_j|\mu, v, k) = \frac{\Gamma((k+1)/2)}{\Gamma(k/2)\sqrt{\pi k}} \left(1 + \frac{(u_j - \mu)^2}{kv}\right)^{-(k+1)/2} \tag{2}$$

where μ is the mean, v the scale parameter, and k the degrees of freedom determining the weight of the tails giving variance $\sigma^2 = (k/(k-2))v$. The t distribution has more weight in the tails than the normal distribution, so that outlying random effects are less influential and undergo less shrinkage [3], with $t(\mu, v, k) \to \mathrm{N}(\mu, \sigma^2)$ as $k \to \infty$ so that the normal distribution is a special case. However, this distribution still has limited flexibility because it is symmetric.

2.2. Skewed extensions

A more flexible distribution would allow for potential skewing. Fernandez and Steel [9] suggest introducing skewing into any unimodal distribution symmetric about 0, with pdf $f(.)$, using a scale factor on each side of 0. For random variable $-\infty < x < \infty$ this gives

$$p(x|\gamma) = \frac{2}{\gamma + 1/\gamma} \left[f\left(\frac{x}{\gamma}\right) I_{[0,\infty)}(x) + f(\gamma x) I_{(-\infty,0)}(x) \right] \tag{3}$$

where $I_{[a,b]}(x)$ is the indicator function for x being between a and b, and γ controls the mass on each side of 0, and hence represents the skewness. When $\gamma = 1$ the pdf in (3) reduces to the original symmetric distribution $f(x)$, with $0 < \gamma < 1$ and $\gamma > 1$ corresponding to negative and positive skewing, respectively (Figure 1). This skewed distribution retains its mode at 0 but the mean is shifted away from 0 due to the asymmetry. From (3) the moments of the distribution can be expressed as

$$E[x^r|\gamma] = M_r \frac{\gamma^{r+1} + ((-1)^r/\gamma^{r+1})}{\gamma + 1/\gamma} \tag{4}$$

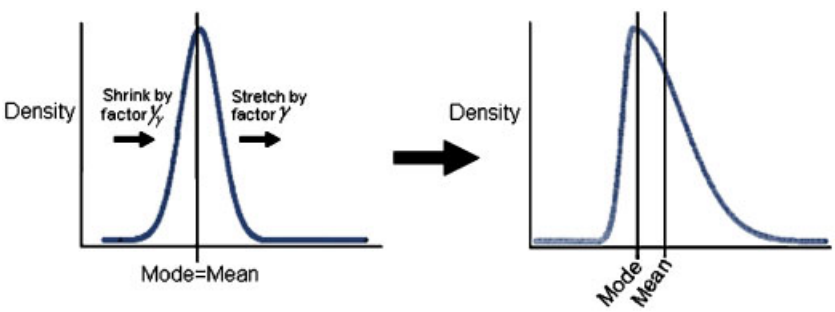

Figure 1. Introducing skewing into a symmetric density as suggested by Fernandez and Steel [9].

DOI: 10.1002/sim

where

$$M_r = 2\int_0^\infty s^r f(s)\,\mathrm{d}s$$

The distribution can be extended to include location and scale parameters which can then be applied to random effects using the transformation $u = \mu_0 + \sigma_0 x$ prior to skewing the distribution (now skewed around μ_0 rather than 0), so that μ_0 represents the mean and σ_0 the standard deviation of the underlying symmetric distribution.

Taking f to be the normal distribution gives the three-parameter skewed normal distribution, $\mathrm{SN}(\mu_0, \sigma_0^2, \gamma)$, with mean and variance

$$\begin{aligned}\mu &= \mu_0 + \sigma_0\sqrt{\frac{2}{\pi}}(\gamma - 1/\gamma)\\ \sigma^2 &= \sigma_0^2((1-2/\pi)(\gamma^2 + 1/\gamma^2) + (4/\pi - 1))\end{aligned} \tag{5}$$

Alternatively we can take f to be the t distribution to give the four-parameter skewed t distribution, $\mathrm{ST}(\mu_0, v, k, \gamma)$, with mean and variance

$$\begin{aligned}\mu &= \mu_0 + \frac{2(\gamma - 1/\gamma)\sqrt{vk}\Gamma((k+1)/2)}{\sqrt{\pi}(k-1)\Gamma(k/2)}\\ \sigma^2 &= \frac{vk\left((1-\gamma^2+\gamma^4)\Gamma(k/2-1)\Gamma(k/2) - \left(\dfrac{8(\gamma^2-1)^2\Gamma((k+1)/2)^2}{\pi(k-1)^2}\right)\right)}{2\gamma^2\Gamma(k/2)^2}\end{aligned} \tag{6}$$

These distributions provide useful alternatives to the normal assumption that allow much more flexibility in the distribution of random effects, with the normal distribution being a special case of each. The parameters of both the skewed normal and skewed t distributions have intuitive interpretations, with k and γ controlling the kurtosis and skewness almost exclusively [9].

2.3. Implementation

All of these distributions can be fitted using a Bayesian MCMC approach in WinBUGS. The normal and t distributions are incorporated in WinBUGS and hence can be fitted directly. In each case we use vague priors for μ and σ; in our examples these are Normal(0, 100) for μ and Uniform(0, 100) for σ [18]. An Exponential(0.1) prior is used for k in the t distribution as suggested by Fernandez and Steel [9], but restricted to $k>2.5$ to prevent problems with an undefined variance when $k\leqslant 2$. This provides reasonable support for values of k between 2.5 and 50 representing heavy tails as well as almost normal data. The skewed normal and skewed t distributions are more difficult to implement because they are not standard distributions and cannot be fitted directly. Instead they may be implemented in WinBUGS using either the 'ones trick' [19] or the WinBUGS Development (WBDev) Interface [20]. Both of these methods enable any distribution to be fitted in WinBUGS providing that the likelihood can be expressed and calculated in closed form. The advantage of WBDev is that it is computationally much more efficient, and is easier to use in practice since the coding is hidden within the program. This approach has therefore been adopted here.

Ideally, the priors for these skewed distributions would be on the overall mean and variance of the distribution. Although this is theoretically possible, the relationship between these overall

DOI: 10.1002/sim

parameters and those defining the distribution is complex causing the MCMC series to trap. Instead we use vague priors for the parameters of the underlying symmetric model, chosen in our examples to be a Normal(0, 100) prior for μ_0, a Uniform(0, 100) for σ_0, calculated as $\sqrt{v(k/(k-2))}$ in the skewed t case, and an Exponential(0.1) prior for k restricted to $k>2.5$. For the skew parameter γ, we use a Gamma(0.5, 0.318) prior for $\varphi=\gamma^2$ as suggested by Fernandez and Steel [9]. This prior has an expectation of 1 for γ, representing a symmetric distribution, and a variance of 0.57, providing reasonable support for values of γ from near 0 up to 4 corresponding to extreme negative and positive skewing.

In each case the models are fitted using five chains of 10 000 iterations, with a burn in of 5000 iterations [19]. Convergence was monitored using standard divergence diagnostics with the consistency of the chains assessed using the Gelman–Rubin diagnostic [19]. The results are compared in terms of the parameter estimates and standard errors (formally marginal posterior medians and standard deviations) and the predictive distribution. The latter is derived as the distribution for a new random effect drawn from the modelled distribution [19]. Model comparisons are made using the deviance information criterion (DIC), a measure of the adequacy of each model as a trade-off between its complexity and fit. The DIC is designed to have similar prediction properties to the Akaike information criterion, to optimize short-term predictions made on data sets of similar size and structure as that being analysed currently. It is increasingly used in a wide variety of contexts and appears to have good practical properties. A rule of thumb is that a difference of less than 2 in DIC represents similarly supported models, with a lower DIC by 3 or more indicating a more appropriate model [19]. Example, WinBUGS code is available from www.bsu-cam.ac.uk.

3. EXAMPLES OF APPLYING THE FLEXIBLE DISTRIBUTIONS IN PRACTICE

We now explore the basic performance of these models in practice. We begin by applying these models to two simulated data sets before investigating the use of such models in two practical situations where we wish to model a single random effect.

3.1. Simulated data

Initially we employ simulated normal and non-normal random-effects data to illustrate the use of such models when the structure of the data is known. Both data sets consist of 100 observations $y_i \sim \text{Normal}(u_i, 1)$, with random effects $u_i \sim \text{Normal}(5, 10)$ in Example 1 and $u_i \sim \text{ST}(3, 1, 5, 2)$ in Example 2. This structure corresponds to the practical setting in Sections 3.2 and 3.3. The parameters were chosen to ensure that the variability in the random effects was considerably larger than the variability at the observation level, with $k=5$ and $\gamma=2$ in the skewed t example to provide considerable skewing and heavy tails. The models fitted were $y_i \sim \text{Normal}(u_i, 1)$, with u_i following normal, t, skewed normal and skewed t distributions with mean μ and variance σ^2 (and additional parameters k and γ as appropriate).

Fitting a random-effects model to the simulated normal data set shows little difference in the parameter estimates irrespective of the distribution used to model the random effects (Table I). There is no evidence of heavy tails or skewing with $k>10$ and γ close to 1 in each model. There is a slight increase in σ^2 and its uncertainty in the t models, possibly due to the difficulty in

DOI: 10.1002/sim

Table I. Parameter estimates (and standard errors) from fitting random-effects models with various distributional assumptions for the random effects to simulated normal and skewed t random effects data, along with the empirical estimates from the data.

	μ	σ^2	k	γ	DIC
Simulated normal random effects					
Empirical	5.440	13.02	—	—	—
Normal	5.439 (0.366)	12.23 (1.954)	—	—	376.5
t	5.486 (0.363)	12.73 (2.739)	12.19 (9.916)	—	376.3
Skewed normal	5.425 (0.366)	12.42 (2.012)		0.898 (0.137)	376.8
Skewed t	5.392 (0.382)	13.10 (3.392)	10.61 (9.549)	0.857 (0.143)	376.5
Simulated skewed t random effects					
Empirical	4.421	3.830	—	—	—
Normal	4.420 (0.198)	2.877 (0.570)	—	—	359.1
t	4.353 (0.198)	2.977 (0.821)	9.724 (9.391)	—	358.5
Skewed normal	4.436 (0.199)	2.854 (0.595)	—	1.858 (0.571)	354.0
Skewed t	4.425 (0.202)	3.016 (0.964)	10.98 (9.890)	1.779 (0.539)	354.0

estimating k when the data are in fact normal. The DIC is similar for all of the models; in this example the DIC does not clearly favour the normal model as one might expect.

The same models were fitted to the random effects simulated from a skewed t distribution (Table I). Fitting skewed distributions detects the skewing in the random effects (γ close to 2), although there is little evidence of heavy tails from the t models indicating that the random effects were not as heavy tailed as intended. The DIC is reduced in the skewed models highlighting the importance of the skewing in this example. We note these examples demonstrate only the use of such models in two particular cases. Ideally a more formal simulation exercise would be carried out, but this would be computationally intensive.

3.2. *Meta-analysis*

We now apply these flexible distributions to real data sets and one context where these models can be used is in random-effects meta-analysis [21]. Here the distribution of random treatment effects is of direct interest, as is the predictive distribution for the treatment effect in a new study [22]. The standard random-effects meta-analysis model is

$$\begin{aligned} d_i &\sim \text{Normal}(\delta_i, \sigma_i^2) \\ \delta_i &\sim \text{Normal}(\mu, \sigma^2) \end{aligned} \tag{7}$$

where d_i is the estimate of the treatment effect in study i, with variance σ_i^2 estimated from the data and assumed to be known. In the standard analysis δ_i, the true treatment effect in study i, is assumed to follow a normal distribution with mean μ, representing the overall treatment effect, and between study variance σ^2. It is this normality assumption that is of interest here.

We investigate the importance of the normality assumption in a meta-analysis of 70 trials assessing the effectiveness of fluoride toothpaste compared to a placebo conducted between 1954 and 1994 [23]. The treatment effect in this meta-analysis is the 'prevented fraction': the mean

DOI: 10.1002/sim

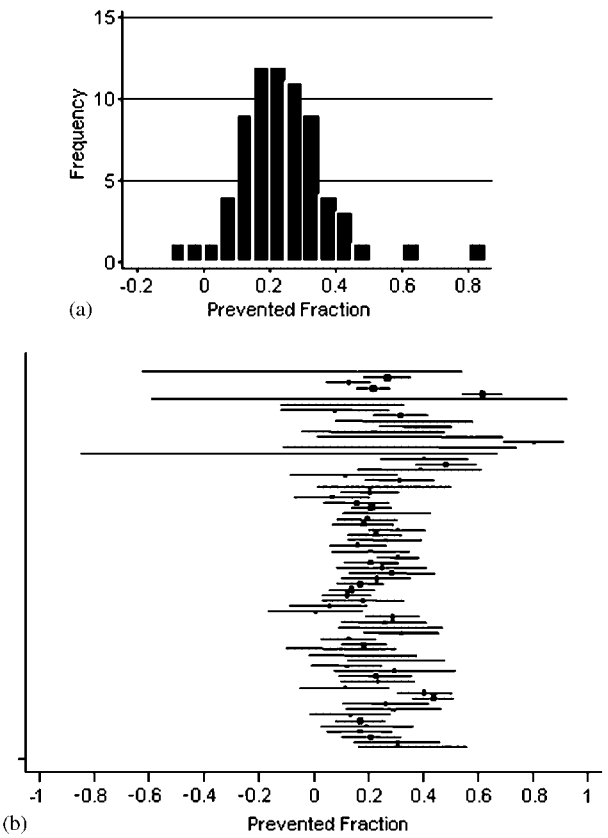

Figure 2. Exploratory view of the prevented fraction outcome for 70 trials assessing the effectiveness of fluoride toothpaste: (a) histogram of the prevented fraction estimates and (b) plot of the prevented fraction estimates and 95 per cent confidence intervals.

increment (change from baseline in the decayed and filled dental surfaces) in the controls minus the mean increment in the treated group, divided by the mean increment in the controls [23]. An exploratory view of these data shows a couple of positive outliers, with a clear evidence of heterogeneity across trials (Figure 2).

The standard analysis assuming normality gives a pooled estimate of 0.24 (95 per cent CI 0.21–0.28) for the treatment effect (Table II). Relaxing the normality assumption shows clear evidence of both skewing and heavy tails, with k around 4 and γ around 2. This is confirmed by a drop of nearly 10 in the DIC for the skewed t model compared to the normal model indicating that the skewed t is the preferred distribution in this case. There are slight differences in the estimates of μ and σ^2 across the various distributional assumptions, but the main effect of relaxing the normal assumption can be seen in the predictive distribution (Figure 3). For example, the probability of a negative treatment effect in a new trial is reduced from 3 per cent under the normal assumption to 0.3 per cent under the skewed t assumption.

DOI: 10.1002/sim

Table II. Parameter estimates (and standard errors) from a meta-analysis of 70 trials of the effectiveness of fluoride toothpaste using various distributional assumptions for the random effects.

	μ	σ^2	k	γ	DIC
Normal	0.242 (0.017)	0.015 (0.003)	—	—	−139.2
t	0.224 (0.014)	0.012 (0.005)	3.332 (1.885)	—	−146.2
Skewed normal	0.250 (0.015)	0.012 (0.003)	—	2.219 (0.593)	−143.5
Skewed t	0.238 (0.015)	0.012 (0.006)	4.005 (4.028)	1.951 (0.567)	−148.7

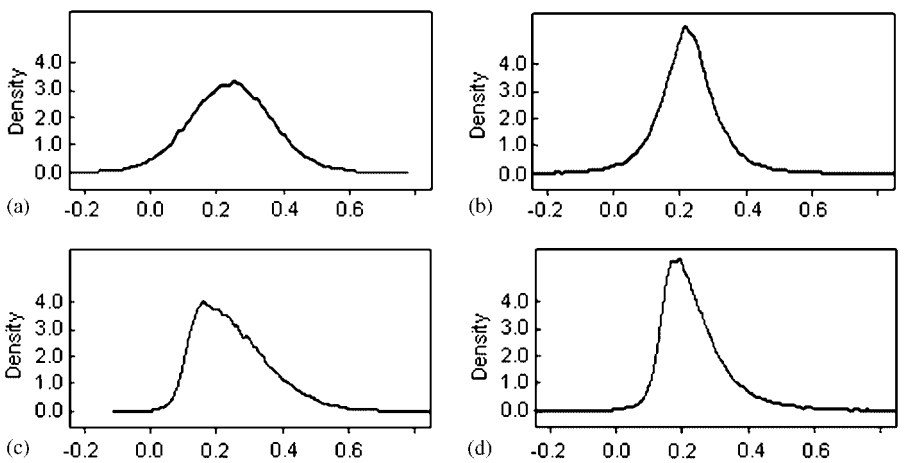

Figure 3. Predictive distributions for the treatment effect in a new trial in the effectiveness of fluoride toothpaste using various distributional assumptions for the random effects: (a) normal; (b) t-distribution; (c) skewed normal; and (d) skewed t.

3.3. Health-professional variation

Another context where these flexible models can be used is to allow for the variation in a treatment effect across the health-professionals who deliver the intervention in a trial [24]. We take the example of a major UK trial comparing the effectiveness of teleconsultations (video-conferences to allow virtual meetings between the patient, general practitioner and hospital consultant) with standard outpatient hospital appointments (between the patient and consultant only) [25]. The trial involved 2094 patients, referred to a consultant from their local primary care practice, who were individually randomized to the teleconsultation or control group. The primary outcome was the offer of a follow-up hospital appointment by the consultant. Here we concentrate on the variation between the 20 consultants participating in the trial.

Ideally these data would be analysed using the individual patient data allowing for clustering of the outcomes and the intervention effect by consultant, but this requires fitting two correlated random effects (see Section 4). Here we consider a summary measures analysis by taking an estimate of the treatment effect and its standard error for each consultant, and combine them using meta-analysis as in equation (7). Here d_i in (7) is the log odds ratio of being offered a follow-up hospital appointment, comparing teleconsultation to control. An initial look at the data highlights heterogeneity in treatment effects across consultants, with some doubt over their

DOI: 10.1002/sim

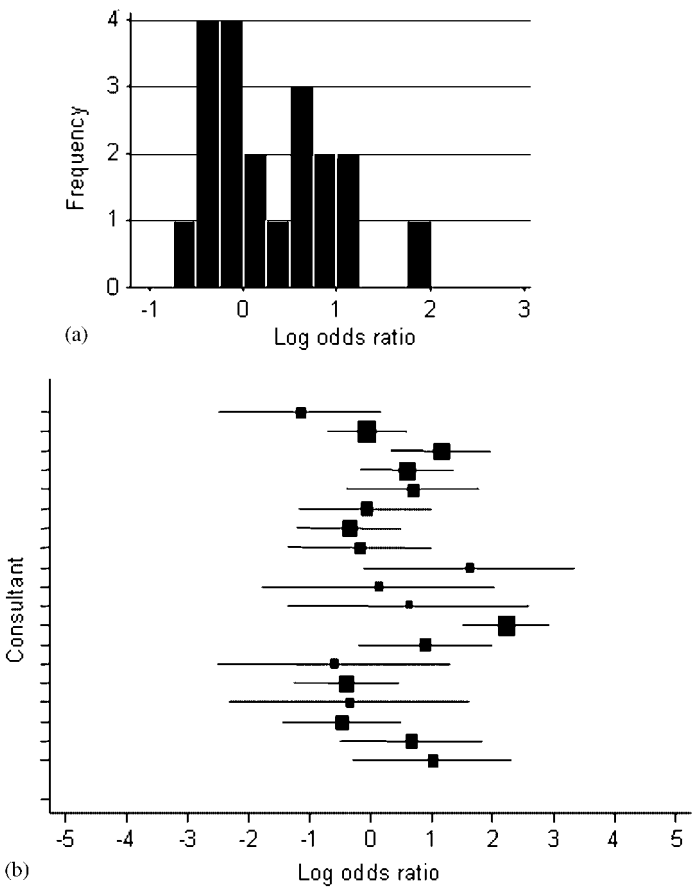

Figure 4. The results from the teleconsultation trial stratified by consultant: (a) histogram of the estimated treatment effects (log-odds ratio of a follow-up appointment) and (b) plot of the estimated treatment effects and 95 per cent confidence intervals.

Table III. Parameter estimates (and standard errors) for the random-effect summary measures analysis by consultant in the teleconsultation trial using various distributional assumptions for the random effects.

	μ	σ^2	k	γ	DIC
Normal	0.308 (0.222)	0.609 (0.340)	—	—	42.80
t	0.269 (0.215)	0.650 (0.517)	8.588 (9.496)	—	42.79
Skewed normal	0.336 (0.223)	0.586 (0.361)	—	1.435 (0.610)	42.36
Skewed t	0.309 (0.225)	0.643 (0.573)	8.882 (9.693)	1.481 (0.617)	42.35

normality (Figure 4). Assuming a flexible distribution for the random effects (Table III) indicates positive skewing with γ around 1.5, but little evidence of heavy tails with k around 8 primarily reflecting the prior. Again the main differences between the model assumptions can be seen in the predictive distributions for the treatment effect, being slightly skewed under the skewed assumptions

DOI: 10.1002/sim

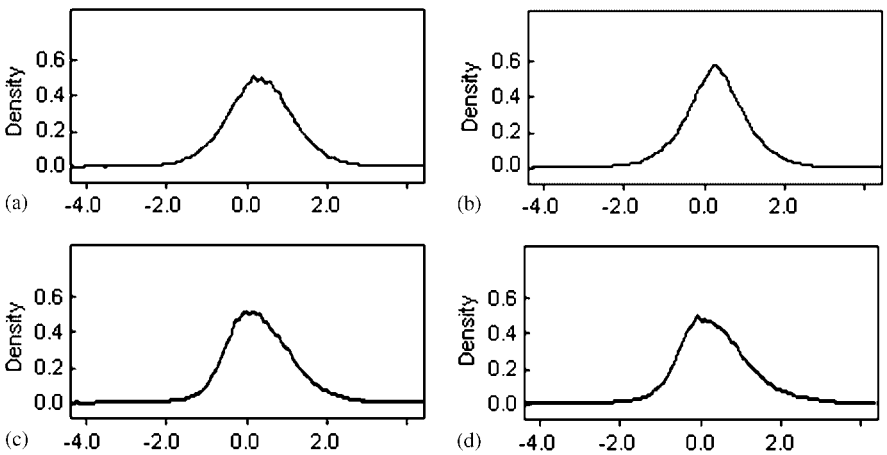

Figure 5. Predictive distributions for the treatment effect (log-odds ratio of a follow-up appointment in the intervention group compared to the control group) for a new consultant, using various distributional assumptions for the random effects: (a) normal; (b) t-distribution; (c) skewed normal; and (d) skewed t.

and more kurtosed in the t model (Figure 5). In this example the DIC is similar in the more complex models compared to the normal model due to the inclusion of extra parameters. Although these flexible models suggest some skewing and heavy tails in the random-effect distribution they are less appropriate for inference in this example due to their additional complexity.

4. BIVARIATE EXTENSION

So far we have considered situations where there is a single random effect. Flexible random-effects models can be similarly used when there are two independent random effects, for example, in health-professional variation where different professionals deliver the intervention in the two treatment arms [26]. In other situations the two random effects may be correlated. This occurs in health-professional variation when the same professional delivers both interventions [26], in meta-analysis when assessing the relationship between the treatment effect and baseline risk [27] or when there are more than two intervention arms [28], and in other cases when the outcome of interest is bivariate [29]. All of these situations require the use of a bivariate distribution for the random effects.

An example of such a model used for modelling health-professional variation where the same group of professionals deliver both interventions, as in the teleconsultation trial above, is [26]

$$\begin{aligned} y_{ij} &\sim \text{Normal}(\mu_{ij}, \sigma_e^2) \\ \mu_{ij} &= \alpha + \beta t_{ij} + u_{1j} + u_{2j} t_{ij} \\ \begin{pmatrix} u_{1j} \\ u_{2j} \end{pmatrix} &\sim \text{BVN}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \right) \end{aligned} \tag{8}$$

DOI: 10.1002/sim

where u_{1j} is the random effect of cluster j across all patients with between-cluster variance σ_1^2 representing the variability of the outcomes, u_{2j} is the random intervention effect with between-cluster variance σ_2^2 representing the variability of the intervention effects, and σ_{12} is the covariance between the random effects. This is an extension of model (1) to include two random effects.

It is possible to extend our models for random effects to the bivariate case, and to replace the bivariate normal (BVN) assumption above with a more flexible one. Implementing BVN and t distributions (with the same degrees of freedom parameter for both random effects) is relatively straightforward, since these distributions are fairly standard and for example are part of the WinBUGS program. Bivariate extensions to the skewed distributions are more difficult, and for these we adopt a regression construction approach [29].

4.1. Regression construction

One method of constructing a BVN distribution is to use a normal distribution for one random effect and then assume that the second random effect is normal conditional on the first [29]. This can be achieved by setting $u_{1j} \sim \text{Normal}(0, \sigma_1^2)$, and $v_j \sim \text{Normal}(0, \sigma_v^2)$ independent of u_{1j}. Taking

$$u_{2j} = \lambda u_{1j} + v_j \tag{9}$$

gives a second normal random effect u_{2j} with mean 0 and variance $\sigma_2^2 = \lambda^2\sigma_1^2 + \sigma_v^2$. The two random effects (u_{1j}, u_{2j}) then have a BVN distribution with covariance $\sigma_{12} = \lambda\sigma_1^2$.

Using this regression approach, we can obtain bivariate extensions to any distribution that can be written down in closed form. For example, in the skewed t case, consider the independent distributions $\text{ST}(\mu_{01}, v_1, k_1, \gamma_1)$ with mean μ_1 and variance σ_1^2, and $\text{ST}(\mu_{0v}, v_v, k_v, \gamma_v)$ with mean μ_v and variance σ_v^2. Taking

$$\begin{aligned} u_{1j} &\sim \text{ST}(\mu_{01}, v_1, k_1, \gamma_1) - \mu_1 \\ v_j &\sim \text{ST}(\mu_{0v}, v_v, k_v, \gamma_v) \\ u_{2j} &= \lambda u_{1j} + v_j - \mu_v \end{aligned} \tag{10}$$

including μ_1 to ensure that $E[u_{1j}] = 0$, and including $\mu_v = E[\lambda u_{1j} + v_j]$ to ensure that $E[u_{2j}] = 0$ giving a bivariate version of the skewed t distribution for u_{1j} and u_{2j}, with covariance $\sigma_{12} = \lambda\sigma_1^2$. This bivariate model has seven parameters, three for each skewed t distribution (since the means are set to 0) and the covariance parameter, compared to three parameters for the BVN model. Similarly this method can be used to obtain the five-parameter bivariate t (with two degrees of freedom parameters unlike the standard bivariate t distribution where there is only one) and bivariate skewed normal distributions. We note that in general u_{2j} does not have the same type of marginal distribution as u_{1j}, although it does in the BVN case.

These regression construction bivariate models can be fitted in WinBUGS using either the ones trick or WBDev; again we present the results using WBDev. Ideally, priors would be placed on the between-cluster variance terms, σ_1^2 and σ_2^2, and the correlation between them, $\rho = \sigma_{12}/\sqrt{\sigma_1^2 + \sigma_2^2}$. This is possible for the normal and t models, as in the univariate case, using a $\text{Uniform}(-1, 1)$ prior for ρ. In order to fit the bivariate skewed models, the priors must in practice be placed on the parameters of the symmetric distribution as in Section 2, and on λ rather than the correlation. λ is given a vague $\text{Normal}(0, 100)$ prior in our example.

DOI: 10.1002/sim

4.2. *Application*: *Bivariate meta-analysis*

We illustrate the use of these models in a bivariate meta-analysis [30]. One important question in meta-analysis is whether the treatment effect varies across trials according to the control group response. The conventional method of analysis, by regression of the treatment effect on the observed control group response, or the average response across all patients, is flawed due to regression to the mean [31]. An alternative approach is to assume that the true treatment effects are linearly related to the true baseline responses, both of which vary across studies. We write $y_{\mathrm{T}i}$ as the observed outcome in the treatment group of trial i, and $y_{\mathrm{C}i}$ as the observed outcome in the corresponding control group, so that the estimated treatment effect is $d_i = y_{\mathrm{T}i} - y_{\mathrm{C}i}$. We consider the model [32]

$$\begin{aligned} y_{\mathrm{C}i} &\sim \mathrm{Normal}(\mu_{\mathrm{C}i}, \sigma^2_{\mathrm{C}i}) \\ d_i &\sim \mathrm{Normal}(\delta_i, \sigma_i^2) \\ \delta_i &= \mu + \beta(\mu_{\mathrm{C}i} - \mu_{\mathrm{C}}) + \varepsilon_i \end{aligned} \tag{11}$$

with independent distributional assumptions

$$\begin{aligned} \mu_{\mathrm{C}i} &\sim \mathrm{Normal}(\mu_{\mathrm{C}}, \sigma^2_{\mathrm{C}}) \\ \varepsilon_i &\sim \mathrm{Normal}(0, \sigma^2_{\delta}) \end{aligned} \tag{12}$$

In this model the observed treatment effect d_i is normally distributed about the true treatment effect δ_i with variance σ_i^2 assumed to be known. Similarly, the observed baseline response $y_{\mathrm{C}i}$ is normally distributed about the true baseline $\mu_{\mathrm{C}i}$ with variance $\sigma^2_{\mathrm{C}i}$ assumed to be known. The true treatment effects are linearly related to the true baselines, where β represents the dependence of the treatment effect on the control group response, and μ is the overall treatment effect adjusted for baseline. The baseline values are centred to ensure that μ is the treatment effect at the mean baseline. σ^2_{C} and σ^2_{δ} represent the heterogeneity across trials in the baseline responses and the treatment effects, respectively.

The above model assumes a BVN distribution of baseline and treatment effects, in which the correlation is explicitly set up by regression construction in (11). This assumption of bivariate normality is the one we investigate here.

4.3. *Example*: *Fluoride meta-analysis*

To illustrate the use of flexible bivariate distributions we return to the meta-analysis of 70 trials of fluoride toothpaste from Section 3.2. In this meta-analysis the baseline response is the mean increment in the control group, representing the amount of dental decay during the trial with no treatment. In this example, a relation between treatment effect and baseline response could reflect differences in duration across the trials. Our analyses allow the mean increment in the control group to be negative, which is unrealistic in practice as it suggests that the amount of decay has decreased; truncation of the random-effects distributions at zero would be necessary to avoid this.

A scatter plot of the treatment effect against the control group response shows some skewing along both axes (Figure 6), suggesting that there is some doubt about the normality of the random-effects distributions. Fitting various bivariate distributions to the random effects in this data set confirms that there is skewing and heavy tails in both distributions, with γ around 2 and k around 5 in each case (Table IV). β is reduced from 0.012 to 0.008 when the normality assumption

DOI: 10.1002/sim

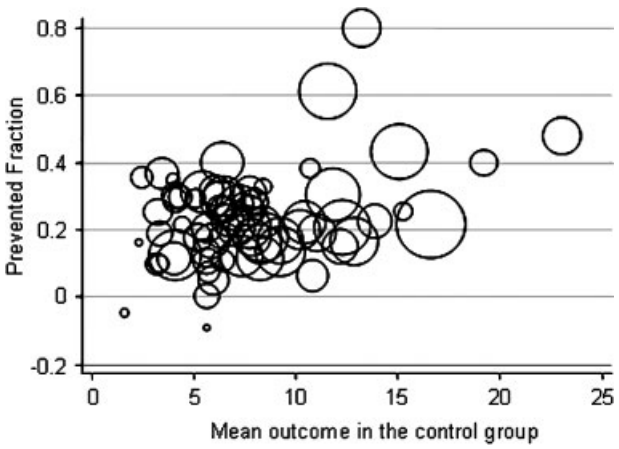

Figure 6. Scatter plot of the treatment effect against the mean outcome in the control group for the 70 trials of the effectiveness of fluoride toothpaste; the size of the circle is inversely related to the variance of the treatment-effect estimate.

Table IV. Parameter estimates (and standard errors) for the bivariate random-effects models applied to the fluoride toothpaste meta-analysis using various distributional assumptions for the random effects.

	Normal	t	Skewed normal	Skewed t
μ	0.235 (0.016)	0.222 (0.014)	0.241 (0.015)	0.234 (0.015)
β	0.012 (0.004)	0.008 (0.004)	0.010 (0.004)	0.008 (0.003)
σ^2_δ	0.013 (0.003)	0.011 (0.005)	0.011 (0.003)	0.011 (0.006)
γ_δ	—	—	2.184 (0.587)	2.014 (0.571)
k_δ	—	3.948 (3.415)	—	5.012 (5.664)
μ_C	7.708 (0.484)	7.171 (0.468)	7.871 (0.467)	7.732 (0.494)
σ^2_C	15.89 (2.961)	16.59 (6.043)	14.93 (2.818)	16.33 (6.587)
γ_C	—	—	2.042 (0.448)	1.900 (0.453)
k_C	—	5.142 (5.689)	—	8.171 (8.473)
DIC	−23.824	−27.895	−28.884	−30.461

is relaxed, indicating less of a dependence between treatment effect and baseline response. The difference can be also seen in the predictive distributions (Figure 7), where there is skewing in both marginal distributions leading to a non-elliptical bivariate posterior. The importance of modelling the skewing and heavy tails in this data set can be seen by the reduction of nearly 7 in the DIC for the skewed t compared to the normal model, indicating that this is the preferred model here. We note the predictive distribution for the mean outcome in the control group is not smooth in the skewed t model (Figure 7(b)ii) although model diagnostics suggest that the convergence is good. This may reflect the fact that the predictive distribution is a mixture of skewed t distributions averaged over the uncertainty in its defining parameters.

DOI: 10.1002/sim

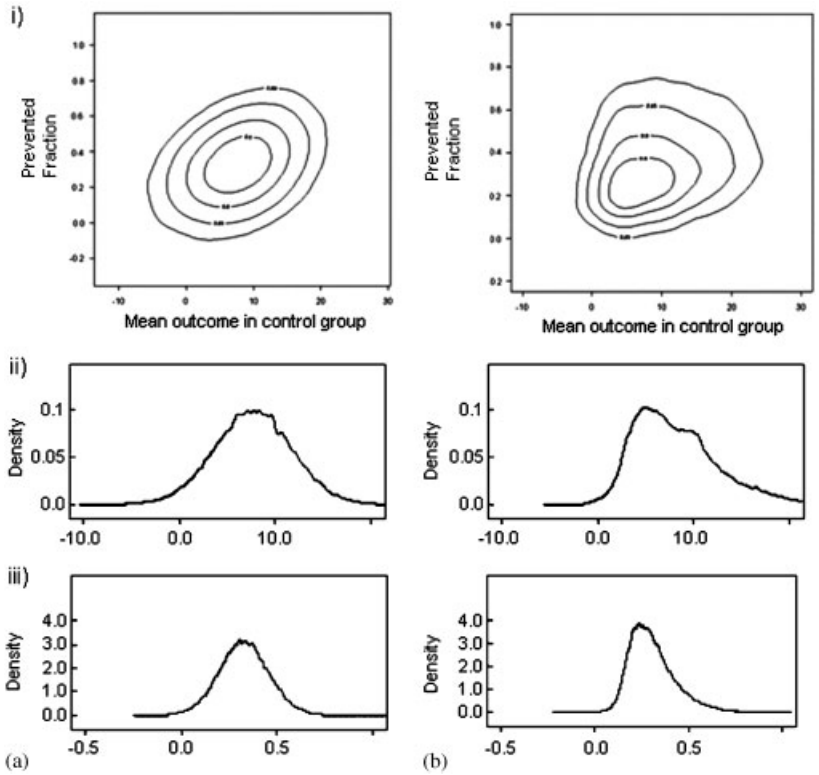

Figure 7. Predictive distributions for a new trial of the effectiveness of fluoride toothpaste under different distributional assumptions for the random effects: (a) assuming normality and (b) assuming a skewed t distribution. (i) Bivariate predictive distribution; (ii) marginal predictive distribution for the mean outcome in the control group; and (iii) marginal predictive distribution for the treatment effect.

5. DISCUSSION

In this paper, we have demonstrated the implications of the restrictive nature of the normal assumption for random effects in hierarchical models. Although fixed-effect estimates seem to remain fairly stable irrespective of the distributional assumption for the random effects, we have shown the substantial impact this assumption can have on the predictive distribution. This highlights the importance of exploring more flexible distributions for random effects when there is some doubt about the validity of the normal assumption. This is particularly important when the distribution of the random effects is of interest, as in the context of meta-analysis or health-professional variation. It is less important when the primary interest is in the fixed effects since the methods discussed here generally give similar results to the more accessible normal assumption.

Determining which model to use in practice can be difficult. If there is clear evidence of non-normality and large heterogeneity, assuming a skewed t distribution for the random effects can be the preferred model, as in the meta-analysis example from Section 3.2. In other situations, although there may be some suggestion of non-normality of the random effect, assuming a skewed t distribution may lead to a less appropriate model due to its complexity, as in the health-professional

DOI: 10.1002/sim

variation example in Section 3.3. In this case it is important not to overfit the model to the data. This becomes even more of an issue in the bivariate case where the flexible models have more parameters. However, it can be important to explore flexible models for the random effects in this context as seen in the bivariate meta-analysis example presented here, with the DIC providing a useful practical guide to model choice. Even if the random effects are truly normal, these flexible models do not misrepresent the data as the normal distribution is a special case of each. One disadvantage of the flexible bivariate models is that the estimates are affected by the ordering of the random effects, as one has a marginal distribution of the required form but the second does not [29]. This is not an issue in the bivariate meta-analysis example since the ordering is implicit in the methods (equation (11)), but can affect the estimates when the ordering is not pre-determined, for example, in health-professional variation. An initial exploration of this in simulated examples suggests, however, that these differences are minimal. In general, we do not advocate the use of the skewed t model in all situations but recommend using the models discussed here to assess the robustness of the conclusions, replacing the normal assumption with a more flexible model if this provides a more appropriate analysis.

The type of models discussed in this paper would be difficult to fit using a classical approach. Instead we have adopted a Bayesian approach with models fitted using WBDev within WinBUGS. This interface can be used with no extra knowledge of WinBUGS once the distribution has been set up, and so this method is accessible to others. Using the 'ones trick' is a viable alternative, but the models take much longer to run and require more complex programming within WinBUGS. When using either of these MCMC approaches it is important to ensure that the chains are converging, particularly in more complex models, to obtain reliable results. Convergence was not problematic in the examples in this paper, but can be when the random-effects variance is small, since there is then generally little information in the data from which to estimate the model parameters. Convergence can also be more difficult to attain for the bivariate random-effects models that we have developed, when compared to the univariate models, because of the larger number of parameters to be estimated.

The models in this paper have been fitted with priors intended to be vague, although in this context priors can be unintentionally informative and it is important to consider the sensitivity of the results to the priors used. An alternative prior for the skew parameter γ could be to use a Normal$(0, \frac{1}{2})$ prior for $\log(\gamma)$. Applying this prior to the example in Section 3.2 made almost no difference to the results. Other alternatives for the degrees of freedom k include a categorical prior for k [5] or a uniform prior for $\log\left(\frac{2}{k}\right)$. Applying these priors in Section 3.2 found the exponential prior to be the least influential in fitting a t distribution, and the only prior that would allow the model to run reliably for the skewed t distribution.

Skewness and heavy tails of random-effects distributions might be anticipated to be the most common forms of non-normality that would be encountered in practice. The skewed t distribution defined by Fernandez and Steel [9], as used in this paper, is an effective model for encompassing these aspects, with separate parameters controlling the skewness and kurtosis. One criticism of the Fernandez and Steel approach is that the even derivatives are discontinuous at μ_0, since it comprises two half-distributions. This makes standard likelihood theory non-applicable [33]. Another disadvantage is that the model is parameterized in terms of the mode, so that the mean is a rather complex function of the other parameters. Alternative approaches of introducing skewing into a t distribution have been suggested by Azzalini and Capitanio [34] and Jones and Faddy [33], although we find these approaches to be less intuitive and more difficult to fit in practice. In addition,

DOI: 10.1002/sim

there is some dependence between the model parameters in the Jones and Faddy approach, with no clear distinction between the skewness and kurtosis.

Although we have concentrated on the t distribution and skew extensions to the normal and t distribution, there is a range of alternative distributions that could be used to allow extra flexibility. One possibility is to use a non-parametric approach, which makes no assumptions about the shape of the random-effects distribution [35]. However, in a maximum-likelihood setting, the estimated random-effects distribution is discrete. A Bayesian non-parametric approach is preferable in providing more realistic inferences about random-effects distributions [11, 36, 37], although these models are computationally complex; such models are useful when the aim is to detect groups of clusters with unusual results. A fully parametric alternative is to use a mixture of normal distributions [3, 13]. However, such models can fail in situations where there are a few outliers, since one normal is often fitted to these outliers and there can be over-fitting of the distribution [13]. The gamma distribution is another alternative to introduce skewness, but is insufficiently flexible [5], and there are large dependencies between the model parameters which cause convergence problems in practice.

We conclude that inferences regarding random effects can critically depend on the assumptions made, and recommend investigating distributions more flexible than the normal, such as those developed in this paper.

ACKNOWLEDGEMENTS

We are grateful to David Spiegelhalter for his advice, and to two referees for their helpful comments.

REFERENCES

1. Goldstein H. *Multilevel Statistical Models*. Arnold: London, 2003.
2. Matsuyama Y, Sakamoto J, Ohashi Y. A Bayesian hierarchical survival model for the institutional effects in a multi-centre cancer clinical trial. *Statistics in Medicine* 1998; **17**:1893–1908.
3. Marshall EC, Spiegelhalter DJ. Comparing institutional performance using Markov chain Monte Carlo methods. In *Statistical Analysis of Medical Data*: *New Developments*, Everitt BS, Dunn G (eds). Arnold: London, 1998; 229–249.
4. Mehrotra DV. Non-iterative robust estimators of variance components in with-in subject designs. *Statistics in Medicine* 1997; **16**:1465–1479.
5. Turner RM, Omar RZ, Thompson SG. Bayesian methods of analysis for cluster randomized trials with binary outcome data. *Statistics in Medicine* 2001; **20**:453–472.
6. Hartford A, Davidian M. Consequences of misspecifying assumptions in non-linear mixed effects models. *Computational Statistics and Data Analysis* 2000; **34**:139–164.
7. Neuhaus JM, Hauck WW. The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika* 1992; **79**:755–762.
8. Pinheiro JC, Lui C, Wu Y. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate distribution. (http://cm.bell-labs.com/cm/ms/departments/sia/project/mixedEffects/index.html), 2006.
9. Fernandez C, Steel MFJ. On Bayesian modelling of fat tails and skewness. *Journal of American Statistical Association* 1998; **93**:359–371.
10. Aitkin M. Meta-analysis by random effect modelling in generalised linear models. *Statistics in Medicine* 1999; **18**:2343–2351.
11. Burr D, Doss H, Cooke GE, Goldschmidt-Clermont PJ. A meta-analysis of studies on the association of the platelet PlA polymorphism of glycoprotein IIIa and risk of coronary heart disease. *Statistics in Medicine* 2003; **22**:1741–1760.
12. Congdon P. *Bayesian Statistical Modelling*. Wiley: Chichester, 2001.

DOI: 10.1002/sim

13. Carroll RJ, Roeder K, Wasserman L. Flexible parametric measurement error models. *Biometrics* 1999; **55**:44–54.
14. Buckle DJ. Bayesian inference for stable distributions. *Journal of American Statistical Association* 1995; **90**: 605–613.
15. Turner RM, Omar RZ, Yang M, Goldstein H, Thompson SG. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine* 2000; **19**:3417–3432.
16. Whitehead A, Omar RZ, Higgins JPT, Savaluny E, Turner RM, Thompson SG. Meta-analysis of ordinal outcomes using individual patient data. *Statistics in Medicine* 2001; **20**:2243–2260.
17. Vaida F, Xu R. Proportional hazards model with random effects. *Statistics in Medicine* 2000; **19**:3309–3324.
18. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 2006; **1**(3): 515–533.
19. Spiegelhalter DJ, Thomas A, Best NG, Lunn D. *WinBUGS Version 1.3 User Manual*. (Available from http://www.mrc-bsu.cam.ac.uk/bugs/), 2006.
20. Lunn D. WinBUGS Development Interface (WBDev). *ISBA Bulletin* 2003; 10.
21. DerSimonion R, Laird N. Meta-analysis of clinical trials. *Controlled Clinical Trials* 1986; **7**:177–188.
22. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random effects meta-analysis: a comparative study. *Statistics in Medicine* 1995; **14**:2685–2699.
23. Marinho VCC, Higgins JPT, Logan S, Sheiham A. Fluoride toothpastes for preventing dental caries in children and adolescents (Cochrane Review). *The Cochrane Library* (Issue 4 edn). Wiley: Chichester, 2003.
24. Lee KJ, Thompson SG. Clustering by health-professional in individually randomized trials. *British Medical Journal* 2005; **330**:142–144.
25. Wallace P, Haines A, Harrison R, Barber J, Thompson SG, Jacklin P, Roberts J, Lewis L, Wainwright P. Joint teleconsultations (virtual outreach) versus standard outpatient appointments for patients referred by their general practitioner for a specialist opinion: a randomized trial. *The Lancet* 2002; **359**:1961–1968.
26. Lee KJ, Thompson SG. The use of random effects models to allow for clustering in individually randomized trials. *Clinical Trials* 2005; **2**:174–181.
27. Arends LR, Hoes AW, Lubsen J, Deiderik EG, Stijnen T. Baseline risk as a predictor of treatment benefit: three clinical meta-reanalysis. *Statistics in Medicine* 2000; **19**:3497–3518.
28. Berkey CS, Anderson JJ, Hoaglin DC. Multiple-outcome meta-analysis of clinical trials. *Statistics in Medicine* 1996; **15**:537–557.
29. Turner RM, Omar RZ, Thompson SG. Modeling multivariate outcomes in hierarchical data, with application to cluster randomised trials. *Biometrical Journal* 2006; **48**:333–345.
30. van Houwelingen HC, Zwinderman KH, Stijnen T. A bivariate approach to meta-analysis. *Statistics in Medicine* 1993; **12**:2273–2284.
31. Sharp SJ, Thompson SG, Altman DG. The relation between treatment benefit and underlying risk in meta-analysis. *British Medical Journal* 2000; **313**:735–738.
32. Sharp SJ, Thompson SG. Analysing the relationship between treatment effect and underlying risk in meta-analysis: comparison and development of approaches. *Statistics in Medicine* 2000; **19**:3251–3274.
33. Jones MC, Faddy MJ. A skew extension to the *t* distribution, with applications. *Journal of the Royal Statistical Society*, *Series B* 2003; **65**:159–174.
34. Azzalini A, Capitanio A. Distributions generated by perturbation of symmetry with emphasis on a multi-variate skew *t*-distribution. *Journal of the Royal Statistical Society*, *Series B* 2003; **61**:579–602.
35. Everitt BS, Dunn G. *Statistical Analysis of Medical Data*: *New Developments*. Arnold: London, 1998.
36. Burr D, Doss H. A Bayesian semi-parametric model for random effects meta-analysis. *Journal of American Statistical Association* 2005; **100**:242–251.
37. Ohlssen DI, Sharples LD, Spiegelhalter DJ. Flexible random-effects models using Bayesian semi-parametric models: applications in institutional comparisons. *Statistics in Medicine* 2006; **26**(9):2088–2112.

DOI: 10.1002/sim