New models for describing outliers in meta-analysis[†]

Rose Baker* and Dan Jackson

An unobserved random effect is often used to describe the between-study variation that is apparent in meta-analysis datasets. A normally distributed random effect is conventionally used for this purpose. When outliers or other unusual estimates are included in the analysis, the use of alternative random effect distributions has previously been proposed. Instead of adopting the usual hierarchical approach to modelling between-study variation, and so directly modelling the study specific true underling effects, we propose two new marginal distributions for modelling heterogeneous datasets. These two distributions are suggested because numerical integration is not needed to evaluate the likelihood. This makes the computation required when fitting our models much more robust. The properties of the new distributions are described, and the methodology is exemplified by fitting models to four datasets. © 2015 The Authors. _Research Synthesis Methods_ published by John Wiley & Sons, Ltd.

Keywords: meta-analysis; outlier; skew distribution; lagged-normal distribution; mixed distribution

1. Introduction

In many meta-analyses, the results of different studies disagree among themselves by more than their within-study statistical errors would imply. For example, in their large-scale empirical investigation of 14 886 meta-analyses, Turner _et al._ (2012) found in 43% of meta-analyses that the DerSimonian and Laird (1986) estimator of the between-study variance was positive. The between-study variance was greater in subjective meta-analyses, and least in all-cause mortality, which gives confidence that the unexplained additional error is not entirely a statistical artifact. For small sample sizes, a modest unexplained error may well be estimated as zero, so between-study variance is probably even more ubiquitous than we think.

The most commonly used solution to this difficulty is the introduction of an unobserved normally distributed random effect. This random effect has variance τ^2, where this parameter is usually referred to as the between-study variance. In particle physics on the other hand, the within-study standard errors are instead scaled up when the outcome data appear to be heterogeneous; refer to Baker and Jackson (2013) for an account of this alternative idea, and also a small-scale empirical demonstration that the conventional random-effects model usually describes the data better.

Perhaps the main motivation for adopting a normally distributed random-effect is that it is mathematically very convenient to make this assumption. The resulting marginal model for the effect y_i is then obtained from a hierarchical modelling framework, where we assume that $Y_i | \mu_i \sim N(\mu_i, \sigma_i^2)$ and $\mu_i \sim N(\mu, \tau^2)$, where μ is the mean effect and σ_i^2 the quoted within-study variance. This means that the marginal distribution, $Y_i \sim N(\mu, \sigma_i^2 + \tau^2)$, is obtained after integrating μ_i out of the joint model for (Y_i, μ_i). Hence, the likelihood function for the random-effects model differs trivially from the likelihood for the fixed-effect model, because σ_i^2 in the fixed effect model is simply replaced by $\sigma_i^2 + \tau^2$ in the random effects model. The within-study variances σ_i^2 are treated as fixed and known in analysis, so that the random effects model involves two unknown parameters: the overall or average treatment effect μ and the between-study variance τ^2. However, as Hoaglin (2015) explains, taking the within-study variances as known is a convenient approximation that can perform poorly in small studies. We explore this issue in our second simulation study, where we simulate binary outcome data for which we take the within-study variances as fixed and known in analysis.

School of Business, University of Salford, City of Salford, UK
*Correspondence to: Rose Baker, School of Business, University of Salford.
E-mail: r.d.baker@salford.ac.uk

This assumption of between-study normality can be motivated by appealing to the notion that μ_i might be thought to depend on the sum of several unknown variables, such as patient mix and operational procedures. Then, by the central limit theorem, this random effect should be approximately normally distributed. However, sometimes it is apparent that a long-tailed or even skew distribution may needed to describe the data well. One way to motivate a skew distribution of effect would be in the situation, where the treatment is reasonably effective but also where the effect is limited by the original severity of disease. This type of 'ceiling effect' is quite likely to result in a skew distribution in the true study effects μ_i.

There has been little work on fitting non-normal distributions for the random effect in meta-analysis. Baker and Jackson (2008) explored the use of several non-normal random-effects distributions, of which the *t*-distribution performed best; however, they did not examine skew distributions. A difficulty with their methodology is that by using maximum likelihood estimation in conjunction with non-tractable random effects distributions, the classical methods proposed by Baker and Jackson (2008) required both numerical integration and numerical maximisation. This means that that their numerical methods are quite liable to be fragile in repeated routine application. Lee and Thompson (2008) explored the use of the skewed *t*-distribution in a Bayesian context. This distribution comprises two halves of a *t*-distribution with differing scale factors, 'glued together' at the mode so that the probability density function and its first derivative are continuous (Fernandez and Steel, 1998). Both Baker and Jackson (2008) and Lee and Thompson (2008) used the fluoride toothpaste dataset that is re-analysed here; Lee and Thompson used Markov Chain Monte Carlo to overcome the difficulties associated with integrating out the random effect.

One may wonder if there is any point fitting complex models, with the small sample sizes found in meta-analysis. We shall show by Monte Carlo simulations that when outliers are present, simply fitting the normally distributed random effects model can be misleading.

In this paper we therefore present some new marginal models for meta-analysis, so that heavy tailed and skew distributions can be used where necessary. In order to ensure that our models are analytically tractable, we avoid the hierarchical modelling approaches of Baker and Jackson (2008) and Lee and Thompson (2008), and we instead model only the marginal distribution of Y_i. We ensure that our models reduce to the conventional random effects marginal model in special cases, so that the usual model is recovered in situations where it is adequate.

To our knowledge, only two proposals for tractable models of this kind have been previously proposed. Firstly, Beath (2014) uses a mixture distribution. This has advantages, because it is computationally simple and the probability that a study is an outlier can be readily found. However, a problem with this approach is that a mixture of two normal distributions requires two extra parameters, the mixing probability and the variance scaling factor, so this is an expensive methodology to use when there is little data, as is commonly the case. Secondly, Gumedze and Jackson (2011) have a compromise solution to this type of identifiability problem; they use normal distributions with a novel variance structure that facilitates sensitivity analyses and the identification and down-weighting of outliers. Viechtbauer and Cheung (2010) and Julious and Whitehead (2011) also present methodologies for identifying outliers.

We illustrate the use of our methodology using meta-analyses of the fluoride toothpaste and cytidine diphosphate (CDP)-choline datasets described in Baker and Jackson (2008), the pravastatin dataset described by Gehr, Weiss and Porzsolt (2006) and the paroxetine dataset described by Julious and Whitehead (2011), and conclude with a discussion.

2. New marginal models for meta-analysis

In the conventional random effects model we assume the marginal model $Y_i \sim N(\mu, \sigma_i^2 + \tau^2)$. The aim here is to develop some alternative marginal models with additional parameters that allow skew and heavy tailed distributions but also reduce to the conventional random effects model in special cases, so that this model can be recovered when it is appropriate. We will define $u_i^2 = \sigma_i^2 + \tau^2$ and omit the suffix i when presenting our new models, so that the conventional random-effects model can be written simply as as $Y \sim N(\mu, u^2)$.

2.1. Symmetric distributions

Symmetric but heavy tailed distributions are useful in order to model and downweight outliers (Baker and Jackson, 2008; Gumedze and Jackson 2011). In this section we develop such distributions for this purpose.

2.1.1. The four parameter symmetric marginal model. A suitable long-tailed symmetric distribution is the mixed normal, with probability density functions (pdf)

$$f(y) = \frac{(1-p)\exp\left(-\frac{(y-\mu)^2}{2u^2}\right)}{\sqrt{2\pi}u} + \frac{p\ \exp\left(-\frac{(y-\mu)^2}{2(u^2+v^2)}\right)}{\sqrt{2\pi(u^2+v^2)}}, \tag{1}$$

where v^2 is an additional variance parameter. This is the solution proposed by Beath (2014), because model (1) is a re-parameterisation of his robust model for meta-analysis described in his section 2.2. On setting $v=0$, or $p=0$, we regain the conventional random effects marginal model. The parameter p describes the probability of outlying

studies, and the parameter v^2 describes the additional variation associated with the outliers. Parameters μ and τ^2 retain their interpretation in the conventional random effects model.

The log-likelihood $\ell = \sum_{i=1}^{n} \ln f(y_i)$ involves the logarithm of a sum of negative exponentials. Throughout we will compute the studies' contributions to the likelihood in ways that avoid any possible numerical difficulties. For this model, it is computationally best to take the exponential out of the logarithm to avoid numerical problems associated with taking the logarithm of a very small number. The log-likelihood can be computed as

$$\ell = \sum_{i=1}^{n} \left\{ -\frac{(y_i - \mu)^2}{2(u_i^2 + v^2)} + \ln \left(\frac{(1-p)\exp\left(-\frac{(y_i-\mu)^2 v^2}{2u_i^2(u_i^2+v^2)}\right)}{u_i} + \frac{p}{\sqrt{u_i^2 + v^2}} \right) \right\} + c \tag{2}$$

where $c = -(n/2)\ln(2\pi)$. Here the logarithm contains only a constant and a negative exponential, which provides a small correction to the first term, and there is no positive exponential that could cause numerical difficulties.

Although not needed when implementing the proposed methodology, we describe further properties of the four parameter exponential distribution in Appendix A. These properties are primarily presented in order to assist with future methodological work involving this distribution.

2.1.2. The three parameter symmetric marginal model. A problem with the model proposed by Beath (2014) is that it requires two more parameters than the conventional random effects model: the mixing probability p and the additional variance v^2. These can be hard to identify together in typical small meta-analysis datasets, because the presence of a handful of unusual study results could be explained by a small p and large v^2, or a larger p and smaller v^2. Hence, the effects of these two additional parameters are hard to disentangle in practice; from the examples, there was a negative correlation between $\hat{p}$ and $\hat{v}^2$.

This difficulty in identifying the full model can be remedied by making p a function of v. Here there are several possibilities.

One is to impose the constraint $p = u^2/(u^2 + v^2)$ in order to obtain a three parameter symmetric marginal model. The parameter p is now strictly decreasing in v^2, so that as v^2 increases the probability of outliers decreases. This means that the single parameter v^2 can be used to identify the situations where outliers are because of a small proportion of studies with very considerable additional variance, or where the outliers are because of a large proportion of studies with somewhat less additional variance.

An alternative approach is to write $p = (s_0^2 + \tau^2)/(s_0^2 + \tau^2 + v^2)$, where s_0^2 is the reciprocal of the average information in the sample (the harmonic mean of the variances)

$$s_0^2 = n/\sum_{i=1}^{n} 1/\sigma_i^2. \tag{3}$$

This alternative model means that p is now common for all studies. A further alternative is to use the typical within-study variance suggested by Higgins and Thompson (2002) in (3) instead of s_0.

Using the first method, because the within-study variances differ from one study to the next, so does the value of p. The probability p now depends on i, where p_i is increasing in σ_i^2. This means that smaller studies (larger σ_i^2) are more likely to be outliers when imposing this constraint; this is appropriate because it is generally thought that small studies are more prone to producing unusual results. This is because small study effects are not uncommon in meta-analysis (Sterne 2001). This constraint also means that the probability p has a similar dependence on τ^2, which means that as τ^2 increases so does the value of p. This also seems reasonable because outliers may be thought to be be likely in very heterogeneous datasets. We shall see that this model generally fits better than the alternative model that uses $p = (s_0^2 + \tau^2)/(s_0^2 + \tau^2 + v^2)$, which we will refer to as 's_0' model.

2.2. Skew distributions

Skewness of the distributions of the study estimates is another possibility. Lee and Thompson (2008) used the skewed t-distribution for the random effect in order to allow for this possibility. In this section we propose an alternative skew and marginal model that is more easily used in applied work.

2.2.1. The five parameter skew marginal model. Our skew marginal distribution can be derived as a mixture distribution. Here we again take a mixture of two random variables, that we denote as R_1 and R_2, where $R_1 \sim N(\mu, u^2)$. However, here we take $R_2 = \varepsilon_1 - \varepsilon_2 + \varepsilon_3 + \eta$, $\varepsilon_1 \sim \text{Exp}(a)$, $\varepsilon_2 \sim \text{Exp}(b)$ and $\varepsilon_3 \sim N(0, u^2)$, where $\text{Exp}(\theta)$ denotes an exponential distribution with mean $1/\theta$ and $\eta = \mu - 1/a + 1/b$. The distribution of the sum or difference of one or more exponentials and a Gaussian random variable follows the lagged-normal distribution (Bassingthwaite *et al.*, 1966; Johnson, Kotz and Balakrishnan 1995). It is also sometimes called the exponentially modified Gaussian distribution. Any number of exponentials can be added or subtracted, but here we just add

one and subtract one. Defining η as we have ensures that the mean of both R_1 and R_2, and so the mean of the marginal model, is μ.

The probability density function of the mixture distribution is

$$f(x) = (1 - p)\phi((x - \mu)/u) + pf_l(x - \eta), \qquad (4)$$

where $\phi(\cdot)$ is the standard normal density function and

$$f_l(y) = \frac{ab}{a + b}\left\{\exp(a^2u^2/2 - ay)\Phi(y/u - ua) + \exp(b^2u^2/2 + by)\Phi(-y/u - ub)\right\}, \qquad (5)$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. When deriving the pdf of the lagged-normal density in (5) it is algebraically simpler to convolve the two exponentials first; the details are given in Appendix B. It can be seen there that the distribution of $\varepsilon_1 - \varepsilon_2$ can be written as a mixture of exponential distributions, which gives the algebraic simplification.

The model parameters are $\mu, \tau, 1/a, 1/b$ and p. The model can be reparameterised using $a = \alpha/r$, $b = \alpha r$. The five parameters are then μ, τ, α, r and p, with zero skewness if $r = 1$ or $\alpha = 0$ (refer to Appendix C for formulae for this and other moments). A value of $r > 1$ denotes skewness to the right, and $r < 1$ skewness to the left. When skewness is small, the skewness $\gamma \propto r - r^{-1}$. Thus, r is a useful measure of skewness. The formula for excess kurtosis shows that for small kurtosis, $\kappa \propto \alpha^{-4}$, so that α^{-1} is a useful measure of kurtosis. Hence, this alternative parameterisation helps us to interpret the model fit. As $\alpha^{-1} \to 0$ we regain the normal random effects model. However, when skewness is extreme, so that one of a, b is very large and the other very small, to avoid numerical problems, $1/a$ and $1/b$ should be used when computing the likelihood. Table 10 shows this situation for the paroxetine dataset. We recommend the routine use of $1/a$ and $1/b$ over r,a for robustness; if desired, one can then refit using r,a.

2.2.2. The four parameter skew marginal model. The estimation of five parameters is a very demanding task in typical meta-analysis datasets that only contain a few studies; we have already criticised the model of Beath (2014) on the grounds that it requires the estimation of four parameters. Furthermore, the full five parameter model is difficult to interpret, despite the reparameterisation suggested in the previous section. It is therefore desirable to eliminate p from the model in a similar way as before, and so we take $p = \frac{u^2}{u^2 + a^{-2} + b^{-2}}$, where a^{-2} and b^{-2} are the variances of the exponential distributions. As before, because u_i^2 varies from one study to the next, p becomes p_i when making this simplification. Alternatively, we can write $p = \frac{s_0^2 + \tau^2}{s_0^2 + \tau^2 + a^{-2} + b^{-2}}$, where s_0 is defined in (3).

Also as before, we can reparamerise and write $a = \alpha/r$, so that we can interpret the parameters in the way explained in the preceding texts.

For illustrative purposes, Figure 1 shows the pdf with $\mu = 0, u^2 = 1$, $\alpha = 0.5$ and $r = 10$. As the exponential tails become longer (a, b decrease) the mixing probability tends to zero, so that the bulk of the distribution is normal, but with a very long tail, that can be skew to one side. Figure 1 gives an indication of how the proposed four parameter skew marginal model can very effectively describe outliers.

2.2.3. Computing the log-likelihood. Even though all our proposed marginal models avoid numerical integration, if evaluated directly from (4) and (5), and using the definition of $\phi(\cdot)$, the pdf and hence the likelihood can still present numerical problems. This is because, without due care, the computation may require multiplying together
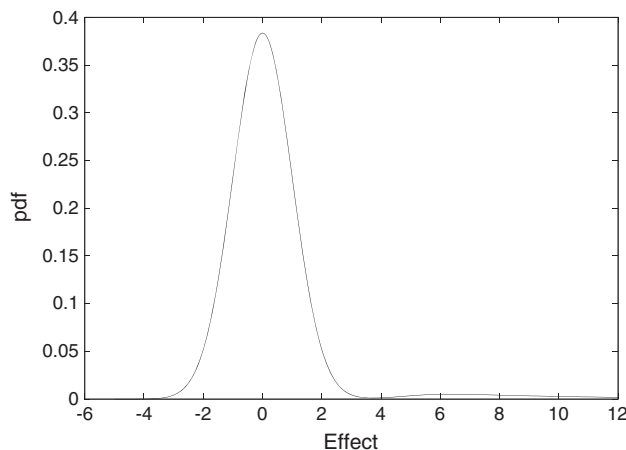

Figure 1. The four-parameter mixed skew distribution with $u = 1, \mu = 0, \alpha = 0.5$ and $r = 10$. Note the bimodality

very large and very small values, which results in numerical instability. Recall that a powerful motivation for our models is that they avoid numerical problems and so are 'safe' for routine use.

To avoid such numerical problems, we suggest computing the contribution to the log-likelihood from a study as follows: We compute the pdf $f(x)$ from (4) and (5) and then compute $\ln f(x)$, so that the full data likelihood is evaluated as the sum of the $\ln f(x)$. Writing

$$f(x) = (1 - p)\frac{\exp\left(-(x - \mu)^2/2u^2\right)}{\sqrt{2\pi}u} + p\frac{ab}{a + b}(T_1 + T_2), \tag{6}$$

where

$$T_1 = \exp\left(a^2u^2/2 - a(x - \eta)\right)\Phi\left((x - \eta)/u - ua\right)$$

and

$$T_2 = \exp\left(b^2u^2/2 + b(x - \eta)\right)\Phi\left(-(x - \eta)/u - ub\right)$$

the difficulty becomes clear: if the arguments in the $\Phi(\cdot)$ functions in T_1 and T_2 are very large and negative then $\Phi(\cdot)$ takes very small values but the other terms in T_1 and T_2 may be very large. If this occurs then we obtain the product of very small and large numbers and hence potential numerical instability. To overcome this problem we suggest using the first term of the asymptotic expansion

$$\Phi(x) \simeq \frac{\exp(-x^2/2)}{|x|\sqrt{2\pi}}$$

for $x < -8$. This enables us to absorb the $\Phi(\cdot)$ components of T_1 and T_2 into the exponents in these terms and so avoid this issue. This means that in practice we evaluate T_1 as given in the preceding texts if $(x - \eta)/u - ua > -8$, else we use the approximation

$$T_1 = \frac{\exp\left(-(x - \eta)^2/2u^2\right)}{\sqrt{2\pi}|(x - \eta)/u - ua|}.$$

Similarly, we evaluate T_2 as in the preceding texts if $-(x - \eta)/u - ub > -8$, else we use the approximation

$$T_2 = \frac{\exp\left(-(x - \eta)^2/2u^2\right)}{\sqrt{2\pi}((x - \eta)/u + ub)}.$$

3. Computational details

The model was fitted using a Fortran program that invoked a Numerical Algorithms Group function minimiser (e.g. E04UCF) and computed standard errors on fitted model parameters by numerical differentiation using E04XAF. Random restarts were used to ensure that the global likelihood maximum had been found, although this did not seem to be necessary. The program ensured a robust determination of standard errors, by allowing negative eigenvalues of the estimated covariance matrix to be identified, and further computations were then performed to improve the estimated matrix. We also provide R code in the web supplementary for two of our main proposals: the three parameter symmetric marginal model (section 2.1.2) and the four parameter skew marginal model (section 2.2.2). These are our main proposals because they allow heavy tails and skew distributions using as few extra parameters as possible.

3.1. Meta-regression

Meta-regressions are sometimes used to model study level covariate effects. This type of model adds to the computational complexity very slightly: to carry out a meta-regression it is only necessary to allow μ to differ from one study to the next and so define a separate μ_i for each study. We then regress μ_i on m covariates x_i, e.g. $\mu_i = \mu_0 + \sum_{k=1}^{m}\beta_k x_k$ and can perform maximum likelihood estimation as before. As shown in the succeeding texts, the pravastatin dataset shows an effect that fades with time, so a meta-regression with annual slope β was carried out. The corresponding covariate, the year of publication, was centred so that μ_0

is the effect size in the centre of the period spanned by the studies. Introducing regression parameters in this way requires sufficient numbers of studies to identify them and meta-regression models may be poorly identified in practice.

4. Results

4.1. Monte Carlo studies

We performed two simulation studies to examine the use of the proposed methodology. In order to provide a severe test of our methods, we explored the use of the full five parameter skew marginal model (section 2.2.1). However, we recognise that reasonably large datasets are required to identify our models in practice, so datasets with a sample size of 40 studies were simulated. We fitted three models to each simulated dataset: the standard random effects model (two parameters), the full five parameter skew marginal model and finally a reduced form of this full model, not discussed up to this point, where we constrain $a = b$ (four parameters). This four parameter model was fitted in order to explore the ease or difficulty of identifying the correct model. In our simulation study we produce some datasets where the conventional random-effects model is true ($a = b = \infty$), some datasets where this reduced $a = b$ model is true and some where the full model is true. We use the notation $x/y/z$ where x, y and z are the number of two-, four- or five-parameter model fits, where the minimum Akaike information criterion estimation (MAICE) of the three models is chosen as the best fitting model. The root mean square error (RMSE) of the fitted means was computed, and also the median absolute error, which should be more robust, for both the full and standard random effects models.

In the first simulation study we simulated data where the within-study variances are correctly assumed to be fixed and known. We randomly generated these variances as twice a random chi-squared variable, rejecting this if the variance was less than 0.072 or more than 4.8. This gives a variance with a mean a little below unity. This is about four times what is often seen for log-odds ratios (Brockwell and Gordon, 2001), but it was convenient to scale everything up like this. The effects were then simulated from model (4).

Table 1 shows the results of the first simulation study, using 1000 simulations for each of 16 different sets of parameter values. The skewness and kurtosis quoted are from the true model, but with all studies having the average information, so these are only a rough guide. It can be seen that when there is no skewness or kurtosis ($a = b = \infty$), the conventional two-parameter random effects model is nearly always fitted, and the error increases only slightly on allowing the four- and five-parameter models to be fitted. However, when there are indeed outliers in the dataset (finite a and b), the error incurred on using the normal random effect model can be enormous. The 'right' model is often not fitted however, reflecting the fact that the models can be difficult to identify but their use gives a huge improvement over just ignoring the possibility of outliers and using the conventional random-effects model. This simulation study therefore provides clear evidence that our methodology is useful in datasets where outliers are present.

Table 1. Monte Carlo simulation results where data are simulated as continuous with fixed and known within-study variances.

τ	a	b	p	γ	κ	Model	RMSE	RMSE (2)	m. er.	m. er. (2)
0.2	∞	∞	N/A	0	0	996/2/2	0.129	0.126	0.083	0.087
0.2	0.1	0.1	0.2	0	26.3	3/773/224	0.151	1.02	0.103	0.662
0.2	0.1	0.1	0.2	0	27.0	0/765/235	0.149	10.1	0.097	6.18
0.2	0.1	0.01	0.2	−4.4	41.4	0/709/291	0.141	7.12	0.097	4.6
0.5	∞	∞	N/A	0	0	996/1/3	0.163	0.165	0.112	0.110
0.5	0.1	0.1	0.2	0	26.1	7/741/252	0.208	1.04	0.131	0.633
0.5	0.01	0.01	0.2	0	27.0	0/809/191	0.190	9.99	0.125	5.93
0.5	0.1	0.01	0.2	−4.4	41.4	1/699/300	0.189	7.57	0.123	4.72
0.2	∞	∞	N/A	0	0	1000/0/0	0.129	0.129	0.085	0.085
0.2	0.1	0.1	0.1	0	53.7	76/589/375	0.141	0.732	0.093	0.423
0.2	0.01	0.01	0.1	0	57.6	8/680/312	0.142	6.94	0.092	3.97
0.2	0.1	0.01	0.1	−6.2	85.7	20/620/360	0.140	4.99	0.094	3.08
0.5	∞	∞	N/A	0	0	993/4/3	0.161	0.161	0.104	0.105
0.5	0.1	0.1	0.1	0	54.0	69/574/357	0.176	0.748	0.112	0.436
0.5	0.01	0.01	0.1	0	57.0	23/677/300	0.175	6.89	0.119	4.14
0.5	0.1	0.01	0.1	−6.2	85.7	23/611/366	0.174	5.22	0.126	2.95

γ and κ are model skewness and kurtosis. The 'model' column shows numbers of simulations fitting the two-, four- or five-parameter models. Finally, the root mean squared error and median absolute error are shown for the MAICE fit (RMSE and m. er.) and the standard random effects two-parameter model fit (RMSE (2) and m. er. (2)).

A second simulation was performed which takes into account the fact that the within-study variances are estimated in practice. Here we simulated two by two tables as in comparative trials for a binary outcome. Here for each study we first generated the log-odds *ratio* from the five-parameter skew model using model (4) with the within-study variances set to zero. Next, the number of patients in each arm was generated as uniformly distributed between 50 and 500, and the probability of success in the control arm was generated as uniform between 0.1 and 0.3. The probability of an event in the treatment arm is then trivially calculable. Next, the two by two tables were generated using the binomial distribution. Finally, the estimated log-odds ratios and their within-study variances were computed and used in analysis.

Table 2 shows that using the long-tailed and skew models still gives very much better accuracy than using the normal model when outliers are present. Similar overall conclusions are obtained in this second simulation study which provides further evidence that our methodology is useful in datasets where outliers are present. One might also ask what happens with smaller sample sizes. With sample size 10 we have found the improvement from using our new models to be even greater but performing extensive simulation studies with small sample sizes is challenging because then it is more difficult to identify the models.

4.2. Fits to real data

The models described in the preceding texts are illustrated by fits to four datasets: the effect of fluoride toothpaste on dental caries, CDP-choline, pravastatin (Gehr *et al.*, 2006) and paroxetine (Julious and Whitehead, 2011).

The fluoride toothpaste dataset consists of 70 studies of the effectiveness of fluoride toothpaste compared with a placebo; the 'prevented fraction' is given as control minus treatment, so a negative value is beneficial. The CDP-choline dataset comprises 10 trials of cytidinediphosphocholine analysis for cognitive and behavioural disturbances associated with chronic cerebral disorders in the elderly, and a standardised difference between mean measurements in the treatment and control groups was used; positive values show benefit. The pravastatin dataset consists of 64 studies of the effect of pravastatin for lowering LDL-cholesterol. The value is the percentage reduction relative to a control. The paroxetine dataset consists of 23 studies of measurements on the HAMD (Hamilton depression rating scale) for paroxetine versus placebo; a positive difference is beneficial.

Fits using *t*-distributed random-effects and skew-*t* random-effects are also given in the tables; these models use the more conventional hierarchical modelling approach that we otherwise avoid here and assume that the μ_i follows these distributions. The use of *t*-distributed random effects is described in Baker and Jackson (2008). The skew *t*-distribution is described in Fernandez and Steel (1998), and its use in meta-analysis as the random-effects distribution is described by Lee and Thompson (2008). The 's_0' models, where s_0 is as defined in (3), are also fitted.

Although we present AIC statistics for all models in order to determine which models best describe the data, for brevity we only provide parameter estimates from the models that are of particular interest in each example.

Table 2. Monte Carlo simulation results where data are simulated using binomial distributions.

τ	a	b	p	γ	κ	Model	RMSE	RMSE (2)	m. er.	m. er. (2)
0.2	∞	∞	N/A	0	0	997/0/3	0.049	0.048	0.031	0.033
0.2	0.1	0.1	0.2	−4.0	21.0	2/767/231	0.061	0.380	0.041	0.220
0.2	0.1	0.1	0.2	0.	21.0	0/773/227	0.058	0.465	0.039	0.124
0.2	0.1	0.01	0.2	−3.9	32.5	1/508/491	0.060	0.747	0.041	0.498
0.5	∞	∞	N/A	0	0	985/6/9	0.107	0.088	0.056	0.060
0.5	0.1	0.1	0.2	−.4	20.9	0/769/231	0.125	0.372	0.079	0.219
0.5	0.01	0.01	0.2	0	21.0	0/773/227	0.111	0.513	0.074	0.315
0.5	0.1	0.01	0.2	−3.9	32.5	1/613/386	0.114	0.780	0.071	0.593
0.2	∞	∞	N/A	0	0	994/2/4	0.050	0.049	0.033	0.036
0.2	0.1	0.1	0.1	−.30	9.0	1/771/228	0.086	0.690	0.056	0.468
0.2	0.01	0.01	0.1	0	9.0	9.780/211	0.136	0.926	0.053	0.654
0.2	0.1	0.01	0.1	−2.8	14.8	9/397/594	0.245	1.530	0.051	1.380
0.5	∞	∞	N/A	0	0	988/4/8	0.094	0.089	0.061	0.056
0.5	0.1	0.1	0.1	−.29	9.0	61/745/194	0.502	0.890	0.110	0.625
0.5	0.01	0.01	0.1	0	9.0	61/745/194	0.502	0.890	0.110	0.625
0.5	0.1	0.01	0.1	−2.8	14.8	116/580/304	0.680	1.569	0.121	1.430

γ and κ are model skewness and kurtosis, the 'model' column shows numbers of simulations fitting the two, four or five-parameter models. Finally, the root mean squared error and median absolute error are shown for the MAICE fit (RMSE and m. err.) and the standard random-effects two-parameter model fit (RMSE (2) and med. err. (2)).

4.3. Example one: fluoride toothpaste

Table 3 shows the models fitted to the fluoride toothpaste dataset, with numbers of parameters, (minus) log-likelihood and Akaike information criterion (AIC). It can be seen that the new models perform better than the t-distribution for the random effect used by Baker and Jackson (2008). Further, the models with fewer parameters, where the mixing probability is not taken as an extra parameter, perform best. The skew model of this type gives a dramatic reduction in AIC, and this is the best model by the minimum-AIC criterion (the MAICE model). It outperforms the skew-t model fit.

Table 4 shows parameter values with standard errors for some of the fitted models. We see that $\hat{r} < 1$ so the fitted distribution is skew to the left, i.e. in this case, where the effect is a reduction of caries, in the direction of increasing fluoride effect. This confirms the results of Lee and Thompson (2007) who used the skewed t-distribution (Fernandez and Steel, 1998) in their Bayesian analysis. The mean effect size decreases by nearly 10% on adopting the skew model, a slightly higher reduction than from the symmetric long-tailed model. This shows that modelling the random effect better can have a significant impact on effect estimation.

4.4. Example two: CDP-choline

Tables 5 and 6 show the corresponding results for the CDP-choline dataset. Again, the four-parameter skew marginal model gives the best fit, but the three parameter symmetric marginal model performs almost as well. Both these models reduce the estimated effect size drastically, roughly halving it. The resulting model fits in Table 6 contrast considerably to those from the previous example: here the between-study variation comes purely from the mixture distribution because $\hat{\tau}^2 = 0$. The residuals are skew purely to the right, again in the direction of increasing effect size.

Here it would probably be better to work with the two exponential tail means $1/a$ and $1/b$ rather than with α and r, and this would avoid the failure of the standard error computation. This bears out our earlier point that the $1/a, 1/b$ parameterisation is more robust, whereas the r, α parameterisation is easier to relate to. In Table 6 we show that both the estimate of $1/a = \alpha r$ and its standard error are more easily computed than those of α and r separately.

4.5. Example three: pravastatin

For this example there is evidence of a time trend, so we include year of study as a covariate in meta-regression models as explained in section 4. Tables 7 and 8 show comparable results for prevastatin, where '+ slope' indicates

Table 3. Models fitted to the fluoride toothpaste dataset.			
Model	Params.	$-\ell$	AIC
Fixed effect	1	20.823	43.646
Random effects	2	1.233	6.466
t-distributed random effects	3	−13.071	−20.143
Three-parameter symmetric	3	−17.148	−28.297
Three-parameter symmetric s_0	3–12.06	−18.12	
Four-parameter symmetric	4	−14.636	−21.271
Skew t-random effects	4	−15.181	−23.181
Four-parameter skew s_0	4	−16.943	−25.887
*Four-parameter skew	4	−21.914	−35.828
Five-parameter skew	5	−17.199	−24.399

The column $-\ell$ shows minus the log-likelihood function. The optimum (MAICE) model is marked with an asterisk.

Table 4. Parameter values for models fitted to the fluoride toothpaste dataset, with standard errors.			
Model	Parameter	Estimate	SE
Random effect	μ	−0.300	0.0193
	τ	0.119	0.0217
Three-parameter symmetric	μ	−0.282	0.0166
	τ	0.911	0.0161
	v	0.932	0.275
Four-parameter skew	μ	−0.273	0.0178
(MAICE)	τ	0.0809	0.0177
	α	0.229	0.132
	r	0.437	0.231

The standard random effects model and the MAICE model are shown.

Table 5. Models fitted to the CDP dataset.

Model	Params.	$-\ell$	AIC
Fixed effect	1	9.759	21.519
Random effects	2	8.199	20.397
t-distributed random effects	3	3.943	13.886
Three-parameter symmetric	3	2.847	11.694
Three-parameter symmetric s_0	3	3.954	13.908
Four-parameter symmetric	4	3.007	14.014
Skew t-random-effects	4	2.906	13.811
Four-parameter skew s_0	4	2.307	12.614
*four-parameter skew	4	1.403	10.806
Five-parameter skew	5	1.929	13.858

The column $-\ell$ shows minus the log-likelihood function. The optimum (MAICE) model is marked with an asterisk.

Table 6. Parameter values for models fitted to the CDP dataset, with standard errors.

Model	Parameter	Estimate	SE
Random effects	μ	0.389	0.156
	τ	0.383	0.174
Three-parameter symmetric	μ	0.194	0.068
	τ	0	0
	v	1.221	0.568
Four-parameter skew	μ	0.192	0.068
(MAICE)	τ	0.0	0.0
	α	0.707×10^{-5}	-
	r	150444	-
	$\alpha r = 1/a$	1.064	0.505

The standard random effects model and the MAICE model are shown.

Table 7. Models fitted to the pravastatin dataset.

Model	Params.	$-\ell$	AIC
Fixed effect	1	187262.496	374526.993
Fixed effect + slope	2	163401.645	326807.290
Random effects	2	181.451	366.902
*Random effects + slope	3	170.610	347.220
t-distributed random-effects+slope	4	170.609	349.217
Three-parameter symmetric + slope	4	170.610	349.220
Three-parameter symmetric + slope s_0	4	170.610	349.220
Four-parameter symmetric + slope	5	170.610	351.220
Skew t random effects + slope	5	169.962	349.925
Four-parameter skew s_0 + slope	4	170.610	351.220
Four-parameter skew + slope	5	168.880	347.761
Five-parameter skew + slope	6	170.610	353.220

The column $-\ell$ shows minus the log-likelihood function. The optimum (MAICE) model is marked with an asterisk.

Table 8. Parameter values for models fitted to the pravastatin dataset, with standard errors.

Model	Parameter	Estimate	SE
Random effects + slope	μ	29.506	0.435
(MAICE)	τ	3.475	0.307
	β	−0.643	0.127
Four-parameter skew + slope	μ	29.726	0.450
	τ	3.519	0.0915
	α	0.361	0.011
	r	0.819	0.0330
	β	−0.647	0.0235

The standard random effects model and the model with the next lowest AIC are shown.

that the covariate effect of year is included in the model. Here there is a marked falloff in effect with time, at an annual rate given by the parameter β. This time, the preferred model is simply the standard meta-regression model with a normally distributed random effect. Some skew models perform nearly as well (they fit better but the AIC is higher). There is slight evidence of skewness to the left, but $\alpha \ll \tau$, showing that the long-tailed behaviour is relatively unimportant. The skew model shifts the estimated effect μ up only slightly, by roughly half its standard error. For this example, and in contrast to the previous two, the fitting of heavy-tailed and skew models satisfies us that the usual Gaussian random effect model is adequate.

4.6. Example four: paroxetine

Finally, Tables 9 and 10 show results for the paroxetine dataset from Julious and Whitehead (2011). Again, the four parameter skew model fits best. The skewness of the distribution of the study estimates, which tends to be larger for smaller studies, has caused the random effects effect estimate to be larger than the fixed effects estimate, because the larger studies now less relative weight in a random effects meta-analysis. However, allowing for the skewness in the data has reduced the mean effect estimate even below the fixed effects estimate. We conclude with this example because it serves to emphasise that the careful modelling of the between-study heterogeneity can have considerable implications for the resulting statistical inference. As mentioned earlier, here again the $1/a, 1/b$ parameterisation works, whereas because the left exponential mean $1/b$ is zero, the r, α parameterisation gives an infinity.

4.7. Identifying outliers

This is sometimes desired as a sensitivity analysis, even when using long-tailed or skew random effects distributions. The approach that we suggest here is to fit a flexible distribution to the data, such as one of the marginal distributions we propose here, and then to compute a p-value for each study to have come from its fitted distribution. It is better if possible to omit each study in turn from the dataset and to compute the p-values based on the models fitted to all other studies. This idea differs from the approach of Beath (2014) who used the mixture distribution to estimate the probability that studies come from the long-tailed component of the mixture, and thus are outliers. Beath's approach has the difficulty that the meaning of 'long-tailed' depends on the dataset, so that a study deemed to be an outlier might not differ drastically from other observations. In contrast, the 'p-value approach' (for want of a better name) flags up outliers that are unlikely to come from the same distribution of effects as the others.

However, a difficulty when using our proposal is that one then has to cope with the multiple-testing problem to avoid spurious identification of outliers. We recommend applying a simple and useful multiple testing criterion,

Table 9. Models fitted to the paroxetine dataset.			
Model	Parameters.	$-\ell$	AIC
Fixed effect	1	100.830	203.66
Random effects	2	48.566	101.151
t-distributed random-effects	3	48.575	103.149
Three-parameter symmetric	3	48.576	103.151
Three-parameter symmetric s_0	3	48.576	103.151
Four-parameter symmetric	4	48.576	105.151
Skew t-random-effects	4	46.808	101.615
Four-parameter skew s_0	4	46.015	100.030
*Four-parameter skew	4	44.955	97.909
four-parameter skew	5	45.145	100.29

The column $-\ell$ shows minus the log-likelihood function. The optimum (MAICE) model is marked with an asterisk.

Table 10. Parameter values for models fitted to the paroxetine dataset, with standard errors.			
Model	Parameter	estimate	s.e.
Fixed effect	μ	2.917	0.131
Random effects	μ	3.360	0.508
	τ	1.805	0.453
Four-parameter skew	μ	2.223	0.223
(MAICE)	τ	0.457	0.243
	$1/a$	1.370	0.306
	$1/b$	0	0

false discovery rate (Benjamini and Hochberg, 1995). Here the p-values are ordered from smallest to largest, and if the kth smallest is the largest of the independent p-values for which $p_{(k)} < k\alpha/n$, where α is the test size, we take the k studies with the lowest p-values as outliers. One can use this criterion as a rough guide, bearing in mind the caveat that the p-values are not quite independent.

We exemplify this methodology using the paroxetine dataset, using the four parameter skew marginal model (section 2.2.2) as the flexible distribution required. Table 11 shows the two-sided p-values for each study, computed by fitting the model to all other studies, i.e. a 'leave-one-out method'. Julious and Whitehead (2011) note that studies 4 and 13 look like outliers, but their methodology does not in the end identify them as such. A similar conclusion follows from Table 11. Given 23 studies, it is not surprising that two of them (4 and 13) have p-values of around 2%. Hence, we also conclude that the estimates from studies 4 and 13 are unusual but not outlying.

We also applied this idea to the fluoride data using the four parameter skew marginal model (results not shown) This produced one study (number 8) with a very small p-value of 0.0042; this study shows a huge reduction in caries. There was also another outlier (number 57) with $p = 0.0096$, this time with very low reduction in caries. These would be candidates for deletion or for applying Gumedze and Jackson's method (2011). However, the probability that the lowest of n independent p-values is below p is $1 - (1 - p)^n \simeq np$, so we would require the most extreme p-value to be below $0.05/70 \approx 0.00071$, which it is not. Applying our recommended multiple-testing method, these two studies would not be regarded as outliers under the four parameter skew marginal model.

It is interesting that Gumedze and Jackson (2011) examined this dataset and identified three outliers, which the present approach would not have performed. In part this is because the skewness of effect in the fitted model persists even when an apparent outlier is removed from the dataset. However, this is also in part because of the fact that Gumedze and Jackson (2011) look for outliers under the usual random-effects model. Their method gives -0.28 for the treatment effect which is not far from the results in table 4, so the inferences for the average effect are highly robust to the distributional assumptions made.

5. Discussion

Some new marginal distributions for meta-analysis have been developed, and their use has been exemplified using four datasets. For those who do not wish to identify specific outliers, but instead wish to model and downweight them by modelling the random effect distribution adequately, this methodology offers an advance on what has gone before. The new distributions are mathematically tractable, so that convolving the within and between-study variation does not require numerical integration. This is needed with, for example, the t-distribution and its skewed version. Hence, the computations required here are much simpler; the only special function that is required is the standard normal cumulative distribution function and this is very widely available.

Table 11. Effect sizes, standard errors and p-values for the paroxetine dataset.			
Study number	Effect	SE	p-value
1	3.00	1.70	0.7331
2	1.90	2.23	0.4062
3	1.63	1.17	0.3481
4	7.18	0.55	0.0270
5	1.12	0.96	0.1730
6	3.50	1.13	0.8617
7	5.50	1.06	0.2757
8	5.50	0.92	0.2264
9	4.70	0.85	0.3387
10	0.82	0.88	0.0886
11	4.09	0.89	0.4973
12	5.04	0.93	0.3109
13	8.34	0.93	0.0206
14	5.17	0.92	0.2826
15	2.30	0.42	0.8661
16	1.50	0.92	0.2939
17	1.80	0.63	0.4206
18	3.33	0.82	0.7015
19	1.70	0.42	0.3005
20	2.60	0.42	0.8206
21	1.30	0.40	0.0318
22	2.40	0.39	0.9823
23	2.93	0.38	0.4191

Here the computation was performed using Fortran, but we also provide R code in the Supporting Information to implement two of our proposed models.

Computing the cumulative distribution functions is also straightforward; this is needed for goodness of fit plots and tests (Baker and Jackson, 2008). Generating random numbers is also easy; this could be used for parametric bootstrap computations used to obtain standard errors. We provide full details of these further and potentially useful mathematical details in the appendices.

For those who wish to identify specific studies as outliers, and for example model them using Gumedze and Jackson's method (2011) or simply remove them, these new marginal distributions are also useful. One can fit the model to all studies except the one of interest, and then compute a p-value, i.e.the probability of obtaining such as extreme observation or one more so, given the model fitted to the other observations. Here again the tractable marginal distributions offer an advantage because the distribution function is easy to compute.

Further work in this area might explore the use of further variants of the new marginal models presented. A useful rule of thumb might be to insist upon having a minimum number of studies for every parameter included in the model presented here, and further work could explore what this minimum number might be. The use of the proposed models in Bayesian statistics could also be explored, where we can anticipate that introducing informative priors would be another way to help identify them.

Appendix A:

This appendix describes further properties of the distribution from the marginal model in section 2.1, to facilitate its use in both further methodological work and application. These properties are most easily derived directly from the mixture distribution in equation (1). The variance is

$$\text{var}(Y) = (1-p)u^2 + p(u^2 + v^2) = u^2 + pv^2,$$

For the three parameter symmetric marginal model (section 2.1.2), we have $p = u^2/(u^2 + v^2)$ so that the variance becomes

$$\text{var}(Y) = \frac{u^2(u^2 + 2v^2)}{u^2 + v^2},$$

so that as $v^2 \to \infty$ the variance doubles from u^2 to $2u^2$.

The excess kurtosis is

$$\kappa = \frac{3p(1-p)v^4}{(u^2 + pv^2)^2}$$

For the three-parameter symmetric marginal model (section 2.1.2), we have $p = u^2/(u^2 + v^2)$ so that the excess kurtosis becomes

$$\kappa = \frac{3v^6}{u^2(u^2 + 2v^2)^2}.$$

For the three-parameter distribution, as $v \to \infty$, $p \to 0$ and $\kappa \to \infty$.

The cumulative distribution function is

$$F(y) = (1-p)\Phi((y-\mu)/u) + p\Phi\left((y-\mu)/\sqrt{u^2 + v^2}\right).$$

Quantiles could be found by using $F(y)$ and Newton–Raphson iteration starting from μ. The generation of random numbers is also straightforward. To simulate one random number, begin by generating a standard normal random variate. Then with probability p multiply it by $\sqrt{u^2 + v^2}$ and with probability $1-p$ multiply it by u. Finally, add the mean μ.

Appendix B:

This appendix provides an outline of the derivation and moment generating function of the lagged-normal distribution. This is given to make the paper self-contained, and because surprisingly little has been written about this distribution. We will derive the distribution of $\varepsilon_1 - \varepsilon_2 + \varepsilon_3$, where $\varepsilon_1 \sim \text{Exp}(1/a)$, $\varepsilon_2 \sim \text{Exp}(1/b)$ and $\varepsilon_3 \sim N(0, u^2)$. We denote the resulting probability density function as $f_l(x)$, so that the probability density function of R_2 is $f_l(x - \eta)$.

Writing the exponential pdfs as $f_1(x) = a\,\exp(-ax)$ for $x \geq 0$, else zero, and $f_2(x) = b\,\exp(-bx)$ if $x \geq 0$, else zero, the pdf of their difference, $\varepsilon_1 - \varepsilon_2$, is the convolution

$$f(x) = \int_{-\infty}^{\infty} f_1(x+y) f_2(y)\,dy \tag{7}$$

The convolution in (7) is a general expression for the pdf of the difference between two random variables. This can be obtained after a change of variables in the integration in the usual convolution for the pdf of the sum of two random variables; replacing the sign of the second random variable to obtain the difference means that we change the sign of argument in the corresponding density function. The functions $f_1(x)$ and $f_2(x)$ are zero for negative x, so that the presence of $f_1(x+y)$ in (7) means that the lower limit in the integration depends on the sign of x. We have

$$f(x) = \int_0^{\infty} f_1(x+y) f_2(y)\,dy$$

for $x > 0$ and

$$f(x) = \int_{-x}^{\infty} f_1(x+y) f_2(y)\,dy$$

for $x < 0$. Then for $x > 0$,

$$f(x) = ab\int_0^{\infty} \exp(-ax - (a+b)y)\,dy = \frac{ab}{a+b}\,\frac{\exp(-ax)}{}.$$

For $x < 0$, note that the distribution is continuous at $x = 0$ and is a mixture of exponentials; thus defining I_+ as 1 for $x > 0$ (i.e. it is the Heaviside step function), else zero, and $I_- = 1 - I_+$, we have that

$$f(x) = \frac{ab}{a+b}\{I_+\exp(-ax) + I_-\exp(bx)\},$$

which is a mixture distribution. The pdf of the sum of the three random variates, $\varepsilon_1 - \varepsilon_2 + \varepsilon_3$ is now found by convolving:

$$f(x) = \frac{ab}{(a+b)\sqrt{2\pi u^2}}\left\{\int_0^{\infty}\exp\left\{-(x-z)^2/2u^2 - az\right\}dz + \int_{-\infty}^0 \exp\left\{-(x-z)^2/2u^2 + bz\right\}dz\right\}$$

where this integration can can be performed by completing the square in the exponents, to yield (5).

The moment generating function $M(t)$ of the lagged-normal distribution can also be found as $\int_{-\infty}^{\infty} f(y)\exp(ty)\,dy$. This integral is the Laplace transform of the convolution of three pdfs, i.e. the product of the three transforms, or

$$\int_{-\infty}^{\infty} f(y)\exp(ty)\,dy = \frac{ab\,\exp(t^2u^2/2)}{(a-t)(b+t)}.$$

This yields the mgf for moments about the mean as

$$M(t) = \frac{ab\,\exp(t^2u^2/2 - (1/a - 1/b)t)}{(a-t)(b+t)},$$

which now on expanding in a power series in t yields zero mean and the higher moments about the mean; all moments exist. The moment generating function of the mixed distribution is $(1-p)\exp(u^2t^2/2) + pM(t)$. Moments for the random effect can be found by setting $u^2 = \tau^2$.

Appendix C:

This appendix describes further properties of the five parameter skew distribution introduced in section 2.2.1. The moments are relatively straightforward using the fact that $R_2 - \mu$ is the sum of three random variables, $N(0, \mu^2)$, $\mathrm{Exp}(a) - 1/a$ and $-(\mathrm{Exp}(b) - 1/b)$, each of which has zero mean, and a little algebra yields

$$\mathrm{var}(X) = u^2 + p\left(a^{-2} + b^{-2}\right),$$

with skewness

$$\gamma = \frac{2p\left(a^{-3} - b^{-3}\right)}{\left\{u^2 + p\left(a^{-2} + b^{-2}\right)\right\}^{3/2}}.$$

In terms of the parameters r and α introduced in the main text,

$$\gamma = \frac{2p\alpha^{-3}(r - r^{-1})(r^2 + 1 + r^{-2})}{\{u^2 + p\alpha^{-2}(r^2 + r^{-2})\}^{3/2}}.$$

The (excess) kurtosis is

$$\kappa = \frac{3p\left\{2(a^{-4} + b^{-4}) + (1 - p)(a^{-2} + b^{-2})^2\right\}}{\{u^2 + p(a^{-2} + b^{-2})\}^2},$$

or

$$\kappa = \frac{3p\alpha^{-4}\left\{2(r^4 + r^{-4}) + (1 - p)(r^2 + r^{-2})^2\right\}}{\{u^2 + p\alpha^{-2}(r^2 + r^{-2})\}^2}.$$

For the four-parameter distribution, as $p \to 0$, if $a \to 0$ the skewness tends to infinity, and as a or b tends to zero, the kurtosis tends to infinity. Thus, these distributions can model very long-tailed and skew data.

The cumulative probability distribution function corresponding to $f_l(y)$ can be obtained by integrating this function to yield

$$F(y) = \Phi(y/u) + \frac{1}{a + b}\left\{a \exp(b^2 u^2/2 + by)\,\Phi(-y/u - ub) - b\exp(a^2 u^2/2 - ay)\,\Phi(y/u - ua)\right\}.$$

This means that the cumulative distribution function $G(x)$ of the five-parameter skew marginal distribution is

$$G(x) = (1 - p)\,\Phi((x - \mu)/u) + pF(x - \eta).$$

Quantiles can be found by Newton–Raphson iteration from μ, using $G(x)$.

The distribution will be bimodal if one of a, b is small and the other large, giving a very skew distribution, otherwise it will be unimodal. There is no simple expression for the mode.

The fact that a random variable from the lagged-normal distribution is the sum of three independent random variables can be exploited. Rather than using a method based on the distribution function, random numbers can be generated more quickly and easily by generating a standard normal r.v. X_3 and two standard exponential r.vs X_1 and X_2. Then the random variable X from the lagged-normal distribution with mean μ is

$$X = uX_3 + X_1/a - X_2/b - 1/a + 1/b + \mu.$$

To generate a random number from the four- or five-parameter mixed distribution: with probability *1-p* generate a random number from a normal distribution with mean μ and variance u^2, and with probability p generate X as described.

References

Baker RD, Jackson D. 2013. Meta-analysis inside and outside particle physics: two traditions that should converge? *Research Synthesis Methods* **4**(2): 109–124.

Baker RD, Jackson D. 2008. A new approach to outliers in metaanalysis. *Healthcare Management Science* **11**: 121–131.

Bassingthwaighte JB, Ackerman FH, Wood EH. 1966. Applications of the lagged normal density curve as a model for arterial dilution curves. *Circulation Research* **18**(4): 398415.

Beath K. 2014. A finite mixture method for outlier detection and robustness in meta-analysis. *Research Synthesis Methods* **5**(4): 285–293.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B* **57**(1): 289–300.

Brockwell SE, Gordon IR. 2001. A comparison of statistical methods for meta-anlaysis. *Statistics in Medicine* **20**: 825–840.

DerSimonian R, Laird N. 1986. Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**: 177–188.

Fernandez C, Steel M. 1998. On Bayesian modelling of fat tails and skewness. *Journal of the American Statistical Association 93, 359371. samples of measures of departure from normality, Proc. Roy. Soc. A* **130**: 16–28.

Gehr BT, Weiss C, Porzsolt F. 2006. The fading of reported effectiveness. A meta-analysis of randomised controlled trials. *BMC Medical Research Methodology* **6**: 25.

Gumedze FN, Jackson D. 2011. A random effects variance shift model for detecting and accomodating outliers in meta-analysis. *BMC Medical Research Methodology* **11**: 19.

Hoaglin DC. 2015. We know less than we should about methods of meta-analysis. *Research Synthesis Methods*, **6**: 287–289.

Higgins JPT, Thompson SG. 2002. Quantifying heterogeneity in a meta-analysis, Statistics in Medicine 21, 1539-1558. skewness and kurtosis. *The Statistician* **47**: 183–189.

Johnson NL, Kotz S, Balakrishnan N. 1995. Continuous Univariate Distributions vol 2, Wiley, New York.

Julious SA, Whitehead A. 2011. Investigating the assumption of homogeneity of treatment effects in clinical studies with application to meta-analysis. *Pharmaceutical Statistics* **11**: 49–56.

Lee KL, Thompson SG. 2008. Flexible parametric models for random-effect distributions. *Statistics in Medicine* **27**: 418–434.

Sterne JAC. 2001. Investigating and dealing with publication and other biases in meta-analysis. *British Medical Journal* **323**: 101.

Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. 2012. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *International Journal of Epidemiology* **41**: 818–827.

Viechtbauer W, Cheung MW-L. 2010. Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods* **1**(2): 112–125.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web site.