

Sequential methods for random-effects meta-analysis

Julian P. T. Higgins,^{a*†} Anne Whitehead^b and Mark Simmonds^c

Although meta-analyses are typically viewed as retrospective activities, they are increasingly being applied prospectively to provide up-to-date evidence on specific research questions. When meta-analyses are updated account should be taken of the possibility of false-positive findings due to repeated significance tests. We discuss the use of sequential methods for meta-analyses that incorporate random effects to allow for heterogeneity across studies. We propose a method that uses an approximate semi-Bayes procedure to update evidence on the among-study variance, starting with an informative prior distribution that might be based on findings from previous meta-analyses. We compare our methods with other approaches, including the traditional method of cumulative meta-analysis, in a simulation study and observe that it has Type I and Type II error rates close to the nominal level. We illustrate the method using an example in the treatment of bleeding peptic ulcers. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: meta-analysis; sequential methods; cumulative meta-analysis; prospective meta-analysis; prior distributions

1. Introduction

The meta-analysis of results from multiple, similar studies is traditionally considered a retrospective activity. However, many meta-analyses are updated over time as new studies are undertaken or are identified from the previously hidden literature. In particular, systematic reviews in the *Cochrane Database of Systematic Reviews* [1] are expected to be updated approximately every two years [2]. Furthermore, genuinely prospective meta-analyses are increasingly being performed. For example, meta-analyses have been planned to combine evidence from existing, ongoing clinical trials [3, 4], and clinical trials have been designed with the aim of meta-analysing their results [5]. In the field of genetic epidemiology, potentially important genetic variants are examined prospectively in multiple existing collections of DNA samples until sufficient evidence of their role has been determined [6, 7].

If a meta-analysis is conducted repeatedly on the addition of new studies, without any allowance for multiple testing, the overall risk of a false-positive finding will increase with the number of meta-analyses performed. Indeed, Berkey *et al.* note that when there is no underlying effect the process of continually updating a meta-analysis using standard significance tests will lead eventually to a false-positive result [8]. One way to address this problem is to exploit formal sequential methods. Although sequential methods are well established in the analysis of individual randomized trials, they have received much less attention in the meta-analysis context. This is an area where opinions differ. Some argue that, because meta-analysts do not (typically) have control over the generation of new evidence, they are not in a position to act on stopping rules and so sequential methods should not be applied.

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

^bMedical and Pharmaceutical Statistics Research Unit, Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, U.K.

^cWolfson Institute of Preventive Medicine, Bart's and The London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, U.K.

*Correspondence to: Julian P. T. Higgins, MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 0SR, U.K.

†E-mail: julian.higgins@mrc-bsu.cam.ac.uk

In contrast, our view is that since meta-analysts frequently make decisions on whether to recommend further research or whether to recommend that there is convincing evidence for the presence or absence of an effect, sequential methods can be similarly important to meta-analyses as to individual primary studies. The intention is to facilitate a justifiable recommendation rather than to control future research directly.

Whitehead [9] describes the use of a stopping boundaries approach for fixed and random-effects meta-analyses based on the sequential trial methodology of Whitehead [10]. Pogue and Yusuf suggest the use of monitoring boundaries based on alpha-spending functions and stochastic curtailment, with application to fixed-effect meta-analysis [11], and these have been implemented in practice [12]. These methods are based on desirable sample sizes in terms of numbers of participants across all studies. Lan *et al.* apply the law of the iterated logarithm that ‘penalizes’ the usual test statistic to account for multiple tests, equivalent to using a particular open ended boundary in the framework of Whitehead, although the approach does not control Type II error [13, 14].

Heterogeneity of effects is a key characteristic of many meta-analyses, and estimation of the among-study variance may create problems in the early stages of a sequential meta-analysis. The methods of Pogue have been adapted to account for heterogeneity in a (retrospective) cumulative meta-analysis scenario, by adjusting the desired sample size based on a function of I^2 , a measure of inconsistency across the observed studies’ findings [15, 16]. The adjustment is derived essentially from a comparison of the (assumed known) variances of fixed-effect and random-effects meta-analysis estimates. In prospective meta-analyses these will not be known, although anticipated values of I^2 might be used. Much of the present paper is devoted to the problem of accounting prospectively for heterogeneity, as it is the principal obstacle to the simple application of sequential trial methodology.

In this paper we consider the use of frequentist sequential approaches to meta-analysis with the aim of controlling the risk of drawing incorrect conclusions. We take an approach based on ‘group-sequential’ trial methodology, with a stopping rule. A previously proposed method is reviewed and extended to better incorporate random effects. We compare six methods (including a sequential fixed-effect method) through a simulation study and compare them in an example in the treatment of bleeding peptic ulcers. Our emphasis is on the determination of a straightforward approach that has good empirical properties, recognizing that even more formal methods are available through the use of dedicated software such as PEST [17].

2. A sequential approach to meta-analysis

2.1. Random-effects meta-analysis

Before introducing the sequential methods we review a general parametric approach to meta-analysis on which the sequential methods are based [18]. Suppose we have an estimate, y_i , of a true effect θ_i with estimated variance v_i , from study i , $i = 1, \dots, k$. The effect might be measured as a log odds ratio or a difference in means for a comparative study, or a simple mean or logit proportion for a single group study; numerous other options are possible. We make the usual assumption that the estimates and variances are uncorrelated, although this will not always be the case (for example, if larger studies tend to be done in populations in which a treatment is less effective). Alternative approaches are available that do not rely on this assumption. For example, the raw data from the studies could be modelled, which may require individual participant data. Alternatively, all studies could be given equal weight [19], although recent proposals to adopt this practice have not been enthusiastically received [20–22].

A random-effects analysis involves inference on the distribution of the θ_i across studies. A standard assumption is that this distribution is normal with mean μ and variance τ^2 . The variance describes the extent of heterogeneity. A simple estimate of μ is given by the weighted average

$$\hat{\mu} = \frac{\sum_{i=1}^k w_i^* y_i}{\sum_{i=1}^k w_i^*},$$

where $w_i^* = (v_i + \hat{\tau}^2)^{-1}$ represents the weight attributed to study i . An approximate variance of this estimate is given by

$$\text{var}(\hat{\mu}) = \frac{1}{\sum_{i=1}^k w_i^*}.$$

The heterogeneity variance τ^2 may be estimated simply by $\hat{\tau}_{DL}^2$, a method of moments estimator described by DerSimonian and Laird [23], where

$$\hat{\tau}_{DL}^2 = \max \left(0, \frac{Q - (k - 1)}{\sum_{i=1}^k w_i - \left(\sum_{i=1}^k w_i^2 / \sum_{i=1}^k w_i \right)} \right),$$

$w_i = 1/v_i$ and

$$Q = \sum_{i=1}^k w_i \left(y_i - \left(\sum_{i=1}^k w_i y_i / \sum_{i=1}^k w_i \right) \right)^2$$

is the standard statistic used to test for heterogeneity across studies. In many circumstances, inferences from a random-effects meta-analysis should be based on the full distribution of the θ_i across studies rather than on μ alone. This can be achieved using a predictive interval for the effect in a new study, which takes into account both the mean μ and variance τ^2 of effects [24]. In the methods we describe below, we assume that primary inference is to be made on the mean effect across studies, i.e. on μ .

If $\hat{\tau}_{DL}^2$ is equal to zero, then w_i^* is equal to w_i and the resulting calculations lead to a fixed-effect meta-analysis.

The estimates y_i and v_i are typically calculated from either Wald statistics or score statistics. For the former, $y_i = \hat{\theta}_i$ and $v_i = \text{var}(\hat{\theta}_i)$, where $\hat{\theta}_i$ denotes the maximum likelihood estimate of θ_i and $\text{var}(\hat{\theta}_i)$ its variance. For the latter, we may write $y_i = Z_i/V_i$ and $v_i = 1/V_i$, where Z_i is the efficient score for θ_i and V_i is Fisher's information, both evaluated at $\theta_i = 0$. When sample sizes are large and θ_i is reasonably near 0, there will be little difference between the results from the two approaches.

2.2. A monitoring boundary approach to sequential random-effects meta-analysis

To follow the progress of a meta-analysis over time we require a measure of the amount of relevant information contained within it. One approach is to use the total number of study participants [11, 15, 16]. Pogue *et al.* introduce a desirable total that they refer to as the 'optimal information size', based on standard sample size calculations [11]. Wetterslev *et al.* adapt this to account for observed heterogeneity in a retrospective cumulative meta-analysis [16], although in our applications this heterogeneity is yet to be observed. In contrast, Whitehead [9] and Lan *et al.* [13, 14] use statistical information in the form of the inverse variance of $\hat{\mu}$ or, equivalently, the sum of the weights (specifically, Whitehead uses Fisher's information). We choose the statistical information approach, since it relates directly to the precision of the meta-analysis estimate and hence the amount of evidence contained in the collection of studies. In particular, if highly heterogeneous studies are combined sequentially over time, the precision of a random-effects meta-analytic estimate may decrease while the number of participants increases. Our approach may be used when study estimates y_i and w_i are calculated either from Wald statistics (as in Lan *et al.*) or score statistics (as in Whitehead).

A boundaries approach to sequential meta-analysis involves monitoring information using the sum of weights, $V_j = \sum_{i=1}^{t_j} w_i^*$ after t_j studies have been included, $j = 1, 2, \dots$. A plot of V_j against the cumulative effect size, $Z_j = \sum_{i=1}^{t_j} w_i^* y_i$, at each update of the meta-analysis provides a visualization of the path of the meta-analysis over time. A line from the origin to a point (Z_j, V_j) has slope equal to the simple random-effects meta-analytic estimate at the corresponding stage. Note that as information about τ^2 changes over time, the individual study weights, w_i^* , are recalculated on each update.

In the absence of heterogeneity, and when τ^2 is assumed to be zero (as in a fixed-effect meta-analysis), this approach may be justified formally [9]. It is equivalent to the group-sequential approach to monitoring clinical trials described by Whitehead [10]. Monitoring boundaries may be used to stop the meta-analysis when there is sufficient evidence of an effect or lack of an effect based on a pre-specified significance level, power and clinically important effect.

Whitehead [9] describes the implementation of a triangular design in the context of a prospectively planned meta-analysis, and Bollen *et al.* implement a double-triangular design [25]. Here we focus on the restricted procedure of Whitehead, equivalent to an O'Brien and Fleming stopping rule [26]. The O'Brien and Fleming design produces a rectangular stopping boundary in the (Z, V) plane, with symmetric upper and lower boundaries at $Z = \pm H$ and a vertical boundary at $V = V_{\max}$ reflecting the

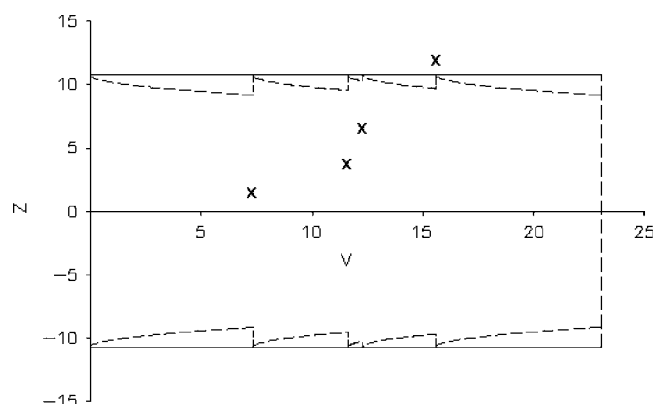


Figure 1. Sequential fixed-effect meta-analysis of peptic ulcers data.

Table I. Monitoring boundaries in the (Z, V) plane for the O'Brien and Fleming design.				
$1 - \beta$	Boundary	α		
		0.001	0.01	0.05
0.8	Horizontal ($\pm H$)	± 14.576	± 9.779	± 6.457
	Vertical (V_{\max})	17.535	12.138	8.299
0.9	Horizontal ($\pm H$)	± 16.120	± 11.029	± 7.461
	Vertical (V_{\max})	21.447	15.438	11.079
0.95	Horizontal ($\pm H$)	± 17.394	± 12.061	± 8.288
	Vertical (V_{\max})	24.972	18.461	13.673

Values presented correspond to a reference effect size of $\mu_R = 1$. Boundaries for other values of μ_R are obtained by dividing H by μ_R , and V_{\max} by μ_R^2 .

maximum amount of information to be collected (Figure 1). Table I provides monitoring boundaries for a selection of choices of two-sided significance level (α) and power ($1 - \beta$), obtained using the software PEST 4.0 [17]. To allow for the possibility of the path crossing the boundary between updates, a 'Christmas tree correction' is applied, bringing the vertical boundaries in at update j to values $H'_j = H - 0.583\sqrt{V_j - V_{j-1}}$ and $-H'_j = -H + 0.583\sqrt{V_j - V_{j-1}}$, where $V_0 = 0$ [10].

As soon as $|Z_j| \geq H'_j$ or $V_j \geq V_{\max}$, the meta-analysis will be stopped with the conclusion that the mean effect is in one direction or the other (for the former criterion) or that the mean effect is zero (if only the latter criterion is true). We refer to the values of Z_j and V_j at this point as Z_{end} and V_{end} . If the path stops with $V_{\text{end}} > V_{\max}$, then we extend the Christmas tree correction to V_{end} , and consider an effect to be present if $|Z_{\text{end}}| \geq H'_{\text{end}}$.

The methods we describe can be extended to address incremental information within studies as well as across studies. For instance, if at time j , study i has interim data yielding an estimate y_{ij} and weight w_{ij}^* (based on the current estimate of τ^2 across currently available data), then we can calculate

$$V_j = \sum_{i=1}^{t_j} w_{ij}^* \quad Z_j = \sum_{i=1}^{t_j} w_{ij}^* y_{ij}$$

This situation was addressed by Whitehead [9] but we do not pursue it further in the present paper.

2.3. Repeated confidence intervals

The sequential procedure, described in the previous section in terms of monitoring boundaries, may also be presented using the repeated confidence intervals approach introduced by Jennison and Turnbull [27]. This enables us to present sequential meta-analyses using forest plots, the conventional way of illustrating meta-analyses. Lower and upper limits, L_j and U_j , for repeated confidence intervals after

the inclusion of t_j studies can be obtained by inverting the sequential design, yielding

$$L_j = \frac{Z_j - H'_j}{V_j}, \quad U_j = \frac{Z_j + H'_j}{V_j}. \quad (1)$$

The stopping rules based on the horizontal monitoring boundaries, that is stop the meta-analysis if $Z_j \leq -H'_j$ or if $Z_j \geq H'_j$, is equivalent to stopping if the interval (L_j, U_j) excludes 0.

The repeated confidence intervals (L_j, U_j) have the property that all of them, calculated up until $V = V_{\max}$, contain the true value of the parameter μ with probability $1 - \alpha$, where α is the chosen two-sided significance level. When the amount of information is less than V_{\max} , the confidence interval series has probability greater than $1 - \alpha$ that the truth is contained in every interval. This means that at any time, an individual confidence interval has probability greater than $1 - \alpha$ of containing the truth, and is therefore wider than a confidence interval from a conventional meta-analysis.

The repeated confidence interval calculated at $V = V_{\max}$ is the last valid member of the sequence. Thus if the amount of information exceeds V_{\max} it is unclear how to proceed. For consistency with the previous section, we propose the calculation of repeated confidence intervals until and including the first time that V_{\max} is exceeded. Thereafter, this last confidence interval is used irrespective of how many additional data accumulate. This is likely to be conservative, but is based on the result that the coverage property of the valid sequence of repeated confidence intervals is also satisfied by the intersection of all of the repeated confidence intervals.

2.4. Formal versus ad hoc stopping rules

At each analysis the repeated confidence interval provides confidence limits that are valid and not dependent on any stopping rule which might be used. Thus information about the magnitude of μ can be obtained from the latest confidence interval. If the sequential meta-analysis follows a formal stopping rule, then a final analysis would be conducted after a stopping boundary had been crossed. A point estimate of μ is given by $Z_{\text{end}}/V_{\text{end}}$ and the repeated confidence interval can be calculated. However, this point estimate may be a biased estimate. Methods for obtaining adjusted estimates and confidence intervals are available based on sequential methodology (see, for example, Chapter 5 of Whitehead [10]), but these require specialized software, and we do not address them in detail here. The adjusted confidence interval will be narrower than that based on repeated confidence interval methodology. This is because the latter is not dependent on the stopping rule that has been used.

2.5. Estimation of heterogeneity variance

A key issue in sequential random-effects meta-analysis is the heterogeneity variance parameter, τ^2 . In the early stages of a sequential procedure, when very few studies are included, uncertainty surrounding any estimate of among-study variance will be considerable. Lan *et al.* prefer to over-estimate τ^2 by using the arithmetic mean of the y_i when there are five or fewer studies, mainly to avoid spurious estimates of zero [13, 14]. Whichever estimate is used, consecutive point estimates of the parameter are likely to be variable, and if $\hat{\tau}_j > \hat{\tau}_{j-1}$ it is possible that V_j will be smaller than V_{j-1} . As a consequence, the path (Z_j, V_j) of the sequential meta-analysis will go backwards. The situation also has implications for the Christmas tree correction, and if V_j is smaller than V_{j-1} we set $H'_j = H$, so that a correction is only made if the path goes forwards. Within a repeated confidence interval framework, a decrease from V_{j-1} to V_j yields a confidence interval wider than the previous one. We consider that the possibility that the confidence intervals can become wider (or that the path goes backwards)—and the actual widening of the confidence intervals—is a suitable reflection of the increased uncertainty resulting from identifying inconsistency among the studies.

In Section 3 we consider five methods for addressing the heterogeneity variance in a sequential meta-analysis: (i) ignoring it (a fixed-effect meta-analysis); (ii) implementing a standard random-effects method (using the DerSimonian–Laird estimate); (iii) accounting for uncertainty in τ^2 (using the approach of Biggerstaff and Tweedie [28]); (iv) Bayesian updating of uncertainty in τ^2 ; and (v) approximate Bayesian updating of uncertainty in τ^2 . The last two approaches are proposed to allow for prior information on τ^2 to be incorporated in the early stages of the sequential meta-analysis. We refer to the sequential approaches using Bayesian inference for τ^2 as ‘semi-Bayes’ to reflect the fact that inference on the parameter of interest (μ) is frequentist. We compare the five approaches, contrasting them with a naïve cumulative random-effects meta-analysis without correction for multiple looks, in a simulation study in Section 4.

3. Specific methods for sequential meta-analysis

3.1. Fixed-effect sequential meta-analysis using the general parametric approach

For a fixed-effect sequential meta-analysis, as described by Whitehead [9], we take

$$Z_j = \sum_{i=1}^{t_j} w_i y_i, \quad V_j = \sum_{i=1}^{t_j} w_i,$$

where y_i and w_i are as defined in Section 2.1. As with all methods described below, we apply the confidence interval formulae in Equation (1) to produce repeated confidence intervals.

3.2. Random-effects sequential meta-analysis using the general parametric approach

For a random-effects sequential meta-analysis, also described by Whitehead, we take

$$Z_j = \sum_{i=1}^{t_j} w_i^* y_i, \quad V_j = \sum_{i=1}^{t_j} w_i^*,$$

where y_i and w_i^* are as defined in Section 2.1 and w_i^* involves the DerSimonian–Laird estimator, $\hat{\tau}_{DL,j}^2$, applied to the studies accumulated so far in the sequential meta-analysis. The estimate of τ_j^2 is assumed known at each update, as in a conventional meta-analysis.

3.3. Accounting for uncertainty in the estimation of τ^2

Methods that incorporate uncertainty in $\hat{\tau}^2$ are available (Biggerstaff and Tweedie [28], Hardy and Thompson [29]). Biggerstaff and Tweedie consider an approach for accounting for uncertainty in the DerSimonian–Laird estimate based on a gamma approximation to the distribution of Q . This leads the authors to a revised weighting scheme for combining the effect estimates, y_i :

$$\hat{\mu}_{BTj} = \frac{\sum_{i=1}^{t_j} \tilde{w}_i y_i}{\sum_{i=1}^{t_j} \tilde{w}_i},$$

where \tilde{w}_i are obtained as the expected values of the DerSimonian–Laird weights, w_i^* , over the approximating distribution for $\hat{\tau}_{DL}^2$ (allowing for truncation at zero), obtained for example using numerical integration.

As the new weights, \tilde{w}_i , are no longer inverse variances, we define Z_j and V_j as

$$Z_j = \frac{\hat{\mu}_{BTj}}{\text{var}(\hat{\mu}_{BTj})}, \quad V_j = \frac{1}{\text{var}(\hat{\mu}_{BTj})}.$$

Incorporating uncertainty in estimates of τ^2 into the meta-analysis should slow the progress of the (Z, V) path towards the boundary, since the uncertainty is reflected in a smaller value for V . Early repeated confidence intervals are likely to be wide.

3.4. Incorporating prior information about heterogeneity: Bayesian updating of τ^2

A natural approach to making inferences about τ^2 in the presence of a small number of studies is to draw on external evidence about its likely value. An informative prior distribution may contribute heavily to producing a realistic estimate in the early stages of a sequential meta-analysis. We may therefore expect the estimates of τ^2 in the meta-analysis to be less erratic, and the path to proceed forward towards the boundaries. We propose to use Bayes' theorem to update the value of τ^2 , and to insert this into the frequentist sequential approach outlined above.

A suitable prior distribution may be developed following the ideas of Higgins and Whitehead [30], who gathered a collection of meta-analyses in the same therapeutic area and produced an empirical distribution from the observed degrees of heterogeneity for a similar type of outcome. A point summary from the prior distribution may be used in the analysis of the first study in the sequential meta-analysis, and point summaries from updated posterior distribution for τ^2 may be used in subsequent updates of the sequential meta-analysis.

Let $p(\tau^2)$ represent the prior distribution for τ^2 , which here will be considered to be an inverse gamma distribution with shape parameter η and scale parameter λ . At the j th meta-analysis the likelihood function for the data $L(\mu, \tau^2; \{y_1, \dots, y_{t_j}\}, \{v_1, \dots, v_{t_j}\})$ is a product of normal distributions with means equal to μ and variances equal to $(v_i + \tau^2)$. The joint posterior distribution of μ and τ^2 at this meta-analysis, assuming an independent prior distribution, $p(\mu)$, for μ , is given by

$$\begin{aligned} p(\mu, \tau^2 | \{y_1, \dots, y_{t_j}\}, \{v_1, \dots, v_{t_j}\}) \\ \propto p(\mu)p(\tau^2) \times L(\mu, \tau^2; \{y_1, \dots, y_{t_j}\}, \{v_1, \dots, v_{t_j}\}) \\ \propto p(\mu)(\tau^2)^{-\eta-1} \exp\left(-\frac{\lambda}{\tau^2}\right) \times \prod_{i=1}^{t_j} \sqrt{\frac{1}{2\pi(v_i + \tau^2)}} \exp\left(-\frac{1}{2} \sum_{i=1}^{t_j} \left[\frac{(y_i - \mu)^2}{v_i + \tau^2}\right]\right). \end{aligned}$$

Inference on τ^2 would normally be performed by integration or simulation. However, μ is unknown and in our frequentist analysis is not associated with a probability distribution, so cannot be integrated out. A solution is to replace μ by its estimate from the $(j - 1)$ th meta-analysis. This makes τ^2 the only unknown parameter in the posterior formula, and a simple one-dimensional numerical integration can be performed to obtain its posterior mean, $\hat{\tau}_{B,j}^2$ as follows.

$$\hat{\tau}_{B,j}^2 = \frac{\int (\tau^2)^{-\eta} \exp\left(-\frac{\lambda}{\tau^2}\right) \times \prod_{i=1}^{t_j} \sqrt{\frac{1}{2\pi(v_i + \tau^2)}} \exp\left(-\frac{1}{2} \sum_{i=1}^{t_j} \left[\frac{(y_i - \hat{\mu}_{j-1})^2}{v_i + \tau^2}\right]\right) d\tau^2}{\int (\tau^2)^{-\eta-1} \exp\left(-\frac{\lambda}{\tau^2}\right) \times \prod_{i=1}^{t_j} \sqrt{\frac{1}{2\pi(v_i + \tau^2)}} \exp\left(-\frac{1}{2} \sum_{i=1}^{t_j} \left[\frac{(y_i - \hat{\mu}_{j-1})^2}{v_i + \tau^2}\right]\right) d\tau^2}.$$

If the first meta-analysis ($j = 1$) contains several studies (i.e. $t_1 > 1$), then a suitable value for $\hat{\mu}_{j-1}$ might be provided by the estimate from a standard random-effects meta-analysis based on the first $(t_1 - 1)$ studies. If the first meta-analysis contains exactly two studies ($t_1 = 2$) we would set $\hat{\mu}_{j-1}$ equal to y_1 .

We opt for the mean rather than the median or mode since not only is it straightforward to calculate, but it will tend to overestimate τ^2 in the early stages, producing an analysis that is less likely to yield a false positive result. As information accumulates, and the posterior distribution of τ^2 develops a stronger peak, the mean will move towards the median and the mode.

For a random-effects sequential meta-analysis, we take

$$Z_j = \sum_{i=1}^{t_j} w_{B,i}^* y_i, \quad V_j = \sum_{i=1}^{t_j} w_{B,i}^*,$$

where $w_{B,i}^* = (v_i + \hat{\tau}_{B,j}^2)^{-1}$. In principle, the approach could be extended to incorporate additional uncertainty in τ^2 by applying the ideas of Biggerstaff and Tweedie. Specifically, the integrations in the calculation of $\hat{\tau}_{B,j}^2$ could be performed for the weights $w_{B,i}^*$ rather than for $\hat{\tau}_{B,j}^2$. We have not implemented this extension.

3.5. An approximation to the Bayesian approach for updating τ^2

We now propose a simpler approach that avoids the numerical integration. Suppose the true effects in the studies to date were known exactly, so that $y_i = \theta_i$ and $v_i = 0$. Then the inverse gamma prior distribution for τ^2 is the conjugate prior distribution for the unknown variance of a normal distribution, and the posterior distribution for τ^2 is then an inverse gamma posterior distribution with parameters $(\eta + (t_j/2))$ and $(\lambda + (t_j \tau_j^2/2))$, where $\tau_j^2 = \sum_{i=1}^{t_j} (\theta_i - \mu)^2 / t_j$ is the heterogeneity variance among the studies up to t_j . We may estimate this variance using the DerSimonian–Laird estimate $\hat{\tau}_{DL,j}^2$. As the mean of an inverse gamma distribution with parameters η and λ is given by $\lambda/(\eta - 1)$, it follows that the posterior mean for τ^2 is given by

$$\hat{\tau}_{AB,j}^2 = \frac{2\lambda + t_j \hat{\tau}_{DL,j}^2}{2\eta + t_j - 2} = \frac{2(\eta - 1)\hat{\tau}_{AB,0}^2 + t_j \hat{\tau}_{DL,j}^2}{2(\eta - 1) + t_j},$$

where $\hat{\tau}_{AB,0}^2$ is the mean of the prior inverse gamma distribution. The posterior mean can therefore be viewed as a weighted average of the prior estimate of τ^2 and the usual estimate from the studies in the meta-analysis. As the number of studies increases, the effect of the prior information decreases.

For a random-effects sequential meta-analysis, our approximate semi-Bayes approach takes

$$Z_j = \sum_{i=1}^{t_j} w_{AB,i}^* y_i, \quad V_j = \sum_{i=1}^{t_j} w_{AB,i}^*$$

where $w_{AB,i}^* = (v_i + \hat{\tau}_{AB,j}^2)^{-1}$. R code for this method is included in the Appendix.

In practice, the approximations here might be expected to yield adequate results when the studies are large (in which case $y_i \approx \theta_i$ and $v_i \approx 0$), and to overestimate τ^2 when the studies are small, due to the overdispersion in the distribution of the estimates, y_i , compared with the distribution of the true effects, θ_i .

4. Simulation study

4.1. Methods of the simulation study

We conducted a simulation study to examine the behaviour of the five sequential meta-analysis methods described in Section 3 under a wide range of conditions. These were also compared with a traditional cumulative meta-analysis approach that simply performs a conventional random-effects meta-analysis on the addition of each study. We chose an O'Brien and Fleming design with 90 per cent power to detect an effect μ_R of 0.5 with a two-sided significance level of 5 per cent. This value for μ_R corresponds to a moderately sized standardized mean difference [31], or to an odds ratio, hazard ratio or risk ratio of 1.65. The boundaries for the design, obtained from the package PEST 4.0, are given by $H = 14.92$ and $V_{\max} = 44.32$ (see Table I).

The simulation study was conducted with true effect sizes of $\mu = 0, 0.25, 0.5$, and 1, that is ranging from no effect to an effect substantially larger than μ_R . Values for the heterogeneity variance τ^2 were 0, 0.0625, and 0.25. We simulated studies such that the expected number of studies under the null hypothesis under a fixed-effect analysis would be $t = 5, t = 10$, and $t = 20$, as described below. This corresponds to an approximate doubling of sample size for each case we investigated. In total we examined 36 scenarios.

Within each simulated sequential meta-analysis the true effect size from each study, θ_i , was sampled from a $N(\mu, \tau^2)$ distribution. The mean within-study variance of the effect estimates, y_i , across all studies was taken to be $\sigma^2 = t/V_{\max}$, ensuring that there would be approximately t studies in a sequential fixed-effect meta-analysis if the vertical boundary was crossed at V_{\max} and $\tau^2 = 0$. The within-study variance of the effect estimate for each study was sampled from a uniform distribution in the range $(\sigma^2 - 0.75\sigma^2, \sigma^2 + 0.75\sigma^2)$. This ensures that the studies have different degrees of precisions, reflecting a realistically wide range of study sizes within each meta-analysis. As an approximate interpretation, for the case of one-group studies estimating means with a between-person standard deviation equal to 4 in every study, this ensures that, for $t = 5$, sample sizes would be distributed between 81 and 567 rather than all being equal to 142. The prior distribution for τ^2 for the semi-Bayes methods was an inverse gamma distribution with parameters $\eta = 1.5$ and $\lambda = 0.08$. This corresponds to a prior estimate for τ^2 of 0.16. This is less than the maximum simulated heterogeneity, and, if the true effect size were 0.5, allows for some studies to have negative effect sizes. Taking $\eta = 1.5$ is equivalent to having prior information with the weight of one study, so the prior estimate contributes less than 10 per cent towards the posterior estimate of heterogeneity after ten studies have been included in the meta-analysis.

For each of the 36 scenarios 5000 sequential meta-analyses were simulated and analysed using each of the six approaches, using the procedures described in Section 2. The stopping rule for the naive cumulative meta-analysis method was to stop (conclude an effect exists) if the naive 95 per cent confidence interval excludes 0; or to stop (conclude no effect) if V_j exceeds V_{\max} ; or otherwise to continue. The simulation was conducted in R (the program is available upon request).

From the simulations, for each approach, we recorded (i) the mean number of studies taken to reach a stopping boundary of the design; (ii) the proportion of simulations stopping having crossed the upper boundary (in order to evaluate Type I error under the null hypothesis); (iii) the percentage of simulations in which all calculated two-sided 95 per cent confidence intervals contained the true effect size μ (that is, the coverage of all confidence intervals); (iv) the percentage of simulations in which the confidence interval at the time when a stopping boundary is crossed (based on Z_{end} and V_{end}) contained the true effect size μ (coverage of the last confidence interval); (v) estimates of the heterogeneity variance, τ^2 , at the time when a stopping boundary is crossed and (vi) mean value of V_{end} .

4.2. Results of the simulation study

Tables II–IV give a selection of the results when $t=5$ (few, large studies), $t=10$ (more, average sized studies) and $t=20$ (many, small studies). The results of the simulations should be interpreted in comparison with the fixed-effect sequential method as well as in isolation. The fixed-effect sequential method should in theory have good properties in the absence of heterogeneity. However, the results do not agree with the expectations due to a combination of (i) the inadequacy of the Christmas tree correction; (ii) our random sampling of within-study variances from a uniform distribution, which means that we cannot guarantee to hit the V_{\max} boundary with the r th study and (iii) the *ad hoc* decision to consider an effect present when $V_{\text{end}} > V_{\max}$ and $|Z_{\text{end}}| \geq H'_{\text{end}}$, a region in which sequential theory does not formally apply. The third issue is expected to increase the probability of declaring that an effect exists, with a larger increase the further away V_{end} is from V_{\max} . When $\mu=0$, the vertical

Table II. Results from simulation study with approximately five studies.

μ	τ^2		Naïve random-effects	Fixed-effect	Random-effects	BT	Approx. semi-Bayes	Semi-Bayes
0	0	Studies at stopping	6.1	5.5	6.4	6.4	7.5	8.6
		$P(\text{benefit} \mu)$	0.048	0.039	0.034	0.033	0.017	0.007
		Coverage, all CIs	90.2	92.2	93.1	93.2	96.5	98.4
		Coverage, last CI	90.2	92.2	93.1	93.2	96.5	98.4
		τ^2 median	0.014	0	0.015	0.015	0.035	0.060
		τ^2 mean	0	0	0	0	0.021	0.057
		V at stopping	48.6	48.7	49.0	49.0	48.5	47.4
	0.0625	Studies at stopping	7.1	5.3	7.7	7.8	8.8	9.2
		$P(\text{benefit} \mu)$	0.102	0.095	0.067	0.071	0.045	0.027
		Coverage, all CIs	79.9	81.5	86.2	86.3	91.2	94.6
		Coverage, last CI	79.9	81.5	86.2	86.3	91.2	94.6
		τ^2 median	0.040	0	0.044	0.044	0.063	0.074
		τ^2 mean	0.0072	0	0.015	0.14	0.041	0.066
		V at stopping	47.7	48.5	48.3	48.3	48.0	47.2
	0.25	Studies at stopping	10.8	4.8	13.1	13.1	14.3	12.7
		$P(\text{benefit} \mu)$	0.181	0.225	0.115	0.114	0.084	0.083
		Coverage, all CIs	64.0	55.6	77.7	77.7	83.7	83.9
		Coverage, last CI	64.0	55.6	77.7	77.7	83.7	83.9
		τ^2 median	0.15	0	0.16	0.17	0.19	0.16
		τ^2 mean	0.12	0	0.15	0.15	0.18	0.12
		V at stopping	45.5	46.3	46.1	46.1	46.2	46.2
0.5	0	Studies at stopping	3.3	3.7	4.3	4.3	5.2	5.9
		$P(\text{benefit} \mu)$	0.958	0.948	0.958	0.959	0.968	0.974
		Coverage, all CIs	89.3	91.6	92.5	92.5	95.8	98.2
		Coverage, last CI	96.0	96.3	97.0	96.9	98.4	99.2
		τ^2 median	0.03	0	0.018	0.010	0.045	0.068
		τ^2 mean	0	0	0	0	0.032	0.065
		V at stopping	22.2	32.0	32.0	32.0	31.1	30.3
	0.0625	Studies at stopping	3.8	3.7	5.2	5.2	6.0	6.2
		$P(\text{benefit} \mu)$	0.942	0.909	0.938	0.939	0.947	0.948
		Coverage, all CIs	80.8	80.9	86.7	86.7	91.3	94.6
		Coverage, last CI	92.9	91.5	94.5	94.5	96.7	97.9
		τ^2 median	0.062	0	0.046	0.046	0.012	0.073
		τ^2 mean	0.0014	0	0.0046	0.0046	0.043	0.081
		V at stopping	21.0	31.4	30.7	30.7	30.6	29.8
	0.25	Studies at stopping	5.3	3.7	8.3	8.3	9.1	8.0
		$P(\text{benefit} \mu)$	0.924	0.840	0.909	0.909	0.908	0.901
		Coverage, all CIs	63.8	56.4	77.9	78.1	83.1	83.5
		Coverage, last CI	86.7	79.5	91.8	91.9	94.9	94.6
		τ^2 median	0.17	0	0.15	0.15	0.15	0.14
		τ^2 mean	0.10	0	0.11	0.11	0.18	0.11
		V at stopping	18.1	30.6	28.4	28.4	28.5	28.2

BT, Biggerstaff and Tweedie.

Table III. Results from simulation study with approximately 10 studies.

μ	τ^2		Naïve random-effects	Fixed- effect	Random- effects	BT	Approx. semi-Bayes	Semi- Bayes
0	0	Studies at stopping	10.9	10.5	11.9	11.9	12.5	13.9
		$P(\text{benefit} \mu)$	0.075	0.032	0.028	0.027	0.023	0.015
		Coverage, all CIs	84.7	93.2	93.9	93.9	95.1	96.9
		Coverage, last CI	84.7	93.2	93.9	93.9	95.1	96.9
		τ^2 median	0.022	0	0.025	0.022	0.033	0.067
		τ^2 mean	0	0	0	0	0.015	0.065
	V at stopping	46.3	46.4	46.6	46.6	46.7	46.1	
	0.0625	Studies at stopping	11.6	10.3	13.4	13.4	14.0	14.5
		$P(\text{benefit} \mu)$	0.108	0.057	0.046	0.046	0.037	0.027
		Coverage, all CIs	77.4	88.5	90.9	90.9	92.5	94.5
		Coverage, last CI	77.4	88.5	90.9	90.9	92.5	94.5
		τ^2 median	0.049	0	0.057	0.057	0.067	0.079
		τ^2 mean	0.0066	0	0.024	0.024	0.038	0.071
	V at stopping	46.0	46.4	46.5	46.5	46.5	46.0	
	0.25	Studies at stopping	15.1	9.7	19.5	19.5	19.9	17.5
		$P(\text{benefit} \mu)$	0.177	0.136	0.069	0.070	0.060	0.063
		Coverage, all CIs	64.9	71.8	85.9	85.8	87.8	87.7
		Coverage, last CI	64.9	71.8	85.9	85.8	87.8	87.7
τ^2 median		0.16	0	0.19	0.19	0.20	0.15	
τ^2 mean		0.13	0	0.18	0.18	0.19	0.12	
V at stopping	45.1	45.9	45.7	45.8	45.8	45.7		
0.5	0	Studies at stopping	5.1	7.0	7.9	7.9	8.5	9.4
		$P(\text{benefit} \mu)$	0.958	0.933	0.943	0.944	0.942	0.947
		Coverage, all CIs	85.9	93.6	94.6	94.6	95.5	97.4
		Coverage, last CI	96.3	97.7	98.1	98.1	98.2	99.0
		τ^2 median	0.043	0	0.031	0.031	0.048	0.073
		τ^2 mean	0	0	0	0	0.023	0.069
	V at stopping	16.1	30.0	29.7	29.7	29.6	29.5	
	0.0625	Studies at stopping	5.6	7.0	9.0	9.0	9.5	9.7
		$P(\text{benefit} \mu)$	0.945	0.903	0.922	0.922	0.926	0.932
		Coverage, all CIs	78.2	87.9	90.8	90.9	92.7	94.6
		Coverage, last CI	94.3	95.7	96.8	96.8	97.4	98.1
		τ^2 median	0.075	0	0.061	0.061	0.077	0.085
		τ^2 mean	0	0	0.013	0.014	0.037	0.076
	V at stopping	15.4	29.9	29.7	29.7	29.7	29.4	
	0.25	Studies at stopping	7.1	6.8	12.3	12.4	12.7	11.3
		$P(\text{benefit} \mu)$	0.938	0.861	0.909	0.909	0.909	0.901
		Coverage, all CIs	65.0	72.1	85.4	85.2	87.0	86.5
		Coverage, last CI	89.9	90.2	95.5	95.5	95.9	95.5
τ^2 median		0.18	0	0.18	0.18	0.19	0.14	
τ^2 mean		0.11	0	0.15	0.15	0.18	0.11	
V at stopping	13.5	29.1	28.1	28.1	28.3	28.4		

BT, Biggerstaff and Tweedie.

boundary will be crossed in the majority of simulations, and it can be seen from Table II ($\tau^2=0$) with approximately five studies to reach V_{\max} that the Type I error rate is 0.039 instead of 0.025. When there are more and smaller studies, for example in Table IV, V_{end} moves closer to V_{\max} and it can be seen that the Type I error rate reduces to 0.028. In addition, the probability that all confidence intervals up to and including the one at V_{end} will contain the true effect size will only be equal to 0.95 if $V_{\text{end}}=V_{\max}$. If $V_{\text{end}}<V_{\max}$, this probability will be greater than 0.95, and if $V_{\text{end}}>V_{\max}$ this probability will be less than 0.95. The closer that V_{end} is to V_{\max} , the closer will be the probability to 0.95. When $\mu=0$ and $\tau^2=0$, this probability is equal to 0.922 (Table II) and 0.939 (Table IV). The results for the other sequential meta-analysis approaches should be viewed in this light.

The Biggerstaff and Tweedie method is not noticeably different from the sequential random-effects meta-analysis. This is because it does not change the estimate of heterogeneity, but widens the confidence interval in line with the uncertainty in the estimation of heterogeneity. This difference in confidence

Table IV. Results from simulation study with approximately 20 studies.

μ	τ^2		Naïve random-effects	Fixed- effect	Random- effects	BT	Approx. semi-Bayes	Semi- Bayes		
0	0	Studies at stopping	19.7	16.3	22.5	22.5	22.8	24.2		
		$P(\text{benefit} \mu)$	0.097	0.028	0.023	0.023	0.021	0.016		
		Coverage, all CIs	80.7	93.9	94.7	94.8	95.2	96.3		
		Coverage, last CI	80.7	93.9	94.7	94.8	95.2	96.3		
		τ^2 median	0.035	0	0.038	0.038	0.044	0.073		
		τ^2 mean	0	0	0	0	0.0008	0.069		
		V at stopping	45.3	45.4	45.5	45.5	45.5	45.3		
		0.0625	Studies at stopping	20.3	20.3	24.1	24.1	24.4	24.8	
			$P(\text{benefit} \mu)$	0.122	0.040	0.036	0.032	0.031	0.025	
	Coverage, all CIs		76.1	91.5	93.1	93.2	93.6	95.0		
	Coverage, last CI		76.1	91.5	93.1	93.2	93.6	95.0		
	τ^2 median		0.063	0	0.071	0.071	0.076	0.084		
	τ^2 mean		0.0017	0	0.023	0.023	0.031	0.076		
	V at stopping		45.2	45.4	45.5	45.5	45.5	45.2		
	0.25		Studies at stopping	23.7	19.7	30.6	30.6	30.7	27.2	
			$P(\text{benefit} \mu)$	0.157	0.079	0.046	0.045	0.044	0.049	
		Coverage, all CIs	67.0	83.7	90.5	90.5	90.8	89.8		
		Coverage, last CI	67.0	83.7	90.5	90.5	90.8	89.8		
		τ^2 median	0.17	0	0.21	0.21	0.21	0.14		
		τ^2 mean	0.14	0	0.20	0.20	0.20	0.11		
		V at stopping	44.8	45.2	45.3	45.3	45.3	45.2		
		0.5	0	Studies at stopping	8.6	13.6	15.2	15.2	15.5	16.3
				$P(\text{benefit} \mu)$	0.952	0.910	0.918	0.918	0.920	0.929
	Coverage, all CIs			79.9	93.4	94.4	94.4	94.7	95.8	
Coverage, last CI	95.8			97.8	97.8	97.8	97.9	98.7		
τ^2 median	0.067			0	0.049	0.049	0.058	0.081		
τ^2 mean	0			0	0	0	0.016	0.076		
V at stopping	13.6			29.3	29.1	29.1	29.1	28.9		
0.0625	Studies at stopping			9.0	13.6	16.2	16.2	16.4	16.6	
	$P(\text{benefit} \mu)$			0.952	0.897	0.913	0.913	0.914	0.919	
	Coverage, all CIs		75.8	91.7	93.3	93.4	93.9	95.4		
	Coverage, last CI		95.0	97.6	98.1	98.1	98.0	98.6		
	τ^2 median		0.092	0	0.079	0.079	0.087	0.090		
	τ^2 mean		0	0	0.015	0.015	0.028	0.081		
	V at stopping		13.1	29.2	28.8	28.8	28.8	29.0		
	0.25		Studies at stopping	10.1	13.2	19.8	19.9	20.0	17.8	
			$P(\text{benefit} \mu)$	0.950	0.878	0.908	0.908	0.907	0.896	
Coverage, all CIs			66.0	84.0	90.8	90.8	90.7	89.6		
Coverage, last CI			91.7	94.8	97.1	97.1	97.1	96.8		
τ^2 median			0.20	0	0.20	0.20	0.21	0.11		
τ^2 mean			0.10	0	0.17	0.17	0.18	0.14		
V at stopping			11.3	28.0	27.8	27.8	27.9	28.1		

BT, Biggerstaff and Tweedie.

intervals appears to be small after the first few studies, so the Biggerstaff and Tweedie adjustment has little impact on the analysis.

The simulation results show that the sequential random-effects meta-analysis generally takes at least one study more to reach a stopping boundary than the sequential fixed-effect meta-analysis, the additional number of studies increasing with increasing heterogeneity. The semi-Bayes methods generally result in about one more study than the sequential random-effects meta-analysis. This shows that taking account of heterogeneity can reduce the risk of early stopping.

As the number of studies increases it can be seen that the Type I error rate for the naïve cumulative random-effects meta-analysis without correction for multiple looks also increases, in contrast to the other random-effects approaches which allow for the multiple looks. Type I error rates increase with increasing heterogeneity, this being most noticeable for the sequential fixed-effect meta-analysis, which makes no allowance for heterogeneity. The use of a sequential random-effects meta-analysis improves

the Type I error rate in the presence of heterogeneity, but it is still unacceptably high when there is a moderate amount of heterogeneity. Hence these methods may produce spurious findings if there is heterogeneity. The semi-Bayes methods produce values closer to the desired value, but they too show a slight increase with heterogeneity.

The semi-Bayes methods show considerably greater coverage than the sequential fixed-effect meta-analysis when there is no heterogeneity. This over-coverage is caused by their invalid prior assumption of non-zero heterogeneity. The coverage falls off rapidly with heterogeneity for the sequential fixed and random-effects meta-analyses. This reduction in coverage is smaller for the semi-Bayes methods.

There is some underestimation of heterogeneity in the sequential random-effects meta-analysis, which may lead to smaller confidence intervals and hence to early stopping. The semi-Bayes methods provide estimates of τ^2 that are closer to the true value provided that τ^2 is not equal to 0. If there is no heterogeneity the Bayesian methods give an overestimate.

In conclusion, the semi-Bayes and approximate semi-Bayes methods had the most desirable false-positive and coverage properties in our simulation study, despite the prior distribution being informative and derived independently of the simulated meta-analyses. Our findings may be dependent on the choice of prior distribution, and we return to this issue in the Discussion.

5. Application

Sacks *et al.* consider the data from 23 trials that compare endoscopic haemostasis with a control treatment for the treatment of bleeding peptic ulcers [32]. The outcome of interest is post-treatment bleeding and the treatment effect recorded here is the log-odds ratio of ‘no bleeding’. The trial estimates y_i and v_i are calculated from Wald statistics, that is

$$y_i = \log \left(\frac{S_1 F_2}{S_2 F_1} \right) \quad \text{and} \quad v_i = \frac{1}{S_1} + \frac{1}{S_2} + \frac{1}{F_1} + \frac{1}{F_2},$$

where S_1 and F_1 are the number of patients with ‘no bleeding’ and ‘bleeding’ in the endoscopic haemostasis group, and S_2 and F_2 are defined similarly for the control group. A positive value for y_i favours the experimental treatment. The data and log-odds ratio estimates and confidence intervals for the individual studies are presented in Table V. For trials in which there were no patients with ‘no

Table V. Peptic ulcers data: randomized trials of endoscopic hemostasis compared with control for bleeding peptic ulcers: Log-odds ratio of ‘no bleeding’.									
Trial, i	Hemostasis		Control		w_i	$w_i y_i$	Log-odds ratio (y_i) [odds ratio]	95 per cent CI ($y_i \pm 1.96/\sqrt{w_i}$)	
	Bled	Total	Bled	Total					
1. Vallon (1980)	20	68	23	68	7.32	1.50	0.20	[1.2]	-0.52, 0.93
2. Swain (1981)	11	36	17	40	4.29	2.22	0.52	[1.7]	-0.43, 1.47
3. Papp (1982)	1	16	13	16	0.68	2.83	4.17	[65]	1.79, 6.56
4. Rutgeerts (1982)	5	52	19	54	3.31	5.39	1.63	[5.1]	0.55, 2.71
5. MacLeod (1983)	6	21	8	24	2.38	0.53	0.22	[1.3]	-1.05, 1.49
6. Jensen (1984)	2	7	7	9	0.74	1.62	2.17	[8.8]	-0.10, 4.44
7. Kernohan (1984)	9	21	7	24	2.52	-1.51	-0.60	[0.5]	-1.83, 0.63
8. Goudie (1984)	7	21	5	25	2.15	-1.49	-0.69	[0.5]	-2.03, 0.64
9. Freitas (1985)	7	36	17	42	3.62	3.75	1.04	[2.8]	0.01, 2.07
10. Swain (1986)	7	69	27	68	4.54	8.00	1.76	[5.8]	0.84, 2.68
11. O’Brien (1986)	17	101	34	103	8.72	7.76	0.89	[2.4]	0.23, 1.55
12. Krejs (1987)	19	85	18	89	7.28	-0.92	-0.13	[0.9]	-0.85, 0.60
13. Brearley (1987)	6	20	8	21	2.27	0.82	0.36	[1.4]	-0.94, 1.66
14. Moreto (1987)	1	16	11	21	0.80	2.23	2.80	[16.5]	0.61, 5.00
15. Laine (1987)	0	10	12	14	0.39	1.81	4.65	[105]	1.51, 7.80
16. Panes (1987)	3	55	25	58	2.36	6.09	2.57	[13.1]	1.30, 3.85
17. Chung (1987)	0	34	34	34	0.25	2.09	8.47	[4761]	4.52, 12.4
18. Balanzo (1988)	7	36	15	36	3.43	3.72	1.08	[3.0]	0.03, 2.14
19. Fellerton (1989)	0	20	5	23	0.44	1.09	2.50	[12]	-0.46, 5.46
20. Angerinas (1989)	7	33	4	32	2.14	-1.36	-0.63	[0.5]	-1.97, 0.71
21. Rutgeerts (1989)	10	40	12	20	2.93	4.40	1.50	[4.5]	0.36, 2.65
22. Chiozzini (1989)	4	34	5	19	1.80	1.78	0.99	[2.7]	-0.47, 2.45
23. Laine (1989)	7	38	15	37	3.48	3.85	1.11	[3.0]	0.05, 2.16

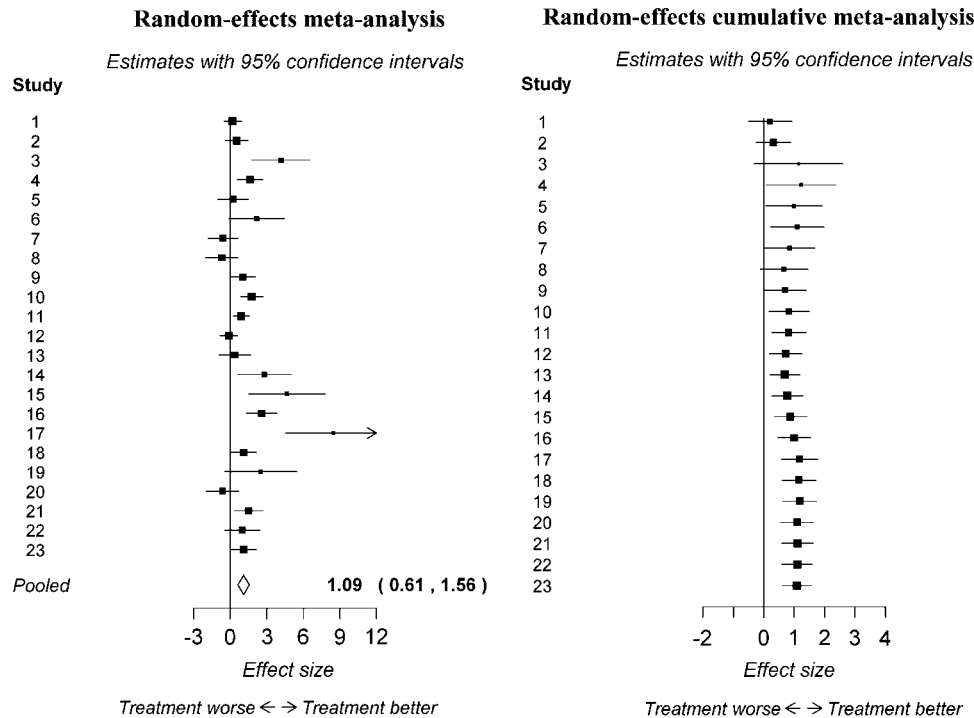


Figure 2. Random-effects meta-analysis and naïve random-effects cumulative meta-analysis of peptic ulcers data.

bleeding’ or no patients with ‘bleeding’ in at least one treatment group, 0.5 was added to all of the four cells of the 2×2 contingency table. Although this is common practice, for meta-analyses of rare events it is often preferable to use alternative approaches such as logistic regression [33]. However, such methods do not fall conveniently into the general sequential framework we describe. A conventional random-effects meta-analysis of the full data set (Figure 2) demonstrates a significant benefit of treatment, with a log-odds ratio of 1.09 (an odds ratio of 3.0), but with considerable heterogeneity ($\hat{\tau}_{DL} = 0.91$).

We suppose for the purposes of illustration that this meta-analysis had been planned with the aim of incorporating observations at the end of each trial as they became available. Given the anticipated heterogeneity of effects, we apply the random-effects methods described above. The analysis uses an O’Brien and Fleming design in order to avoid early stopping unless the odds ratio is very large, with a significance level of 5 and 90 per cent power to detect a log-odds ratio of 0.693 (an odds ratio of 2). From Table I, for $\mu_R = 0.693$ this gives $H = 10.77$ and $V_{\max} = 23.07$ (Figure 1).

The sequential random-effects meta-analysis and the approximate semi-Bayes and semi-Bayes methods are applied and compared with a sequential fixed-effect meta-analysis (Figure 3). To assess sensitivity to prior distributions, two inverse gamma prior distributions for τ^2 are used for the semi-Bayes methods: an $IG(1.5, 0.08)$ has a prior mean of 0.16 and an $IG(1.5, 1)$ has a prior mean of 2. These prior means are respectively much smaller and much larger than the DerSimonian–Laird estimate based on all of the trials. Table VI shows the number of studies included at the point when the stopping criterion is met, together with estimates of the treatment effect ($Z_{\text{end}}/V_{\text{end}}$) and heterogeneity, and the last repeated confidence interval. If the plan for this sequential meta-analysis had been to follow the formal stopping rule of the O’Brien and Fleming design, then it would be more appropriate to present a ‘final’ analysis which correctly adjusts for earlier looks. The results from conducting this ‘final’ analyses using the PEST package [17] are included in Table VI. Such an analysis corrects for the bias in the $Z_{\text{end}}/V_{\text{end}}$ estimate, and provides an appropriate P value and confidence interval. The confidence interval is narrower than the last repeated confidence interval, because the repeated confidence intervals have had no effect on the decision to conduct further studies.

These results show that a sequential fixed-effect meta-analysis, which ignores the heterogeneity, stops after four studies (see also Figure 1), leading to a smaller estimate of the treatment effect compared with that from all trials and a confidence interval that is narrow. The sequential random-effects method requires 11 studies, and produces a slightly larger estimate of the treatment effect and a wider confidence interval. When the prior distribution provides a much smaller underestimate of τ^2 than is apparent in

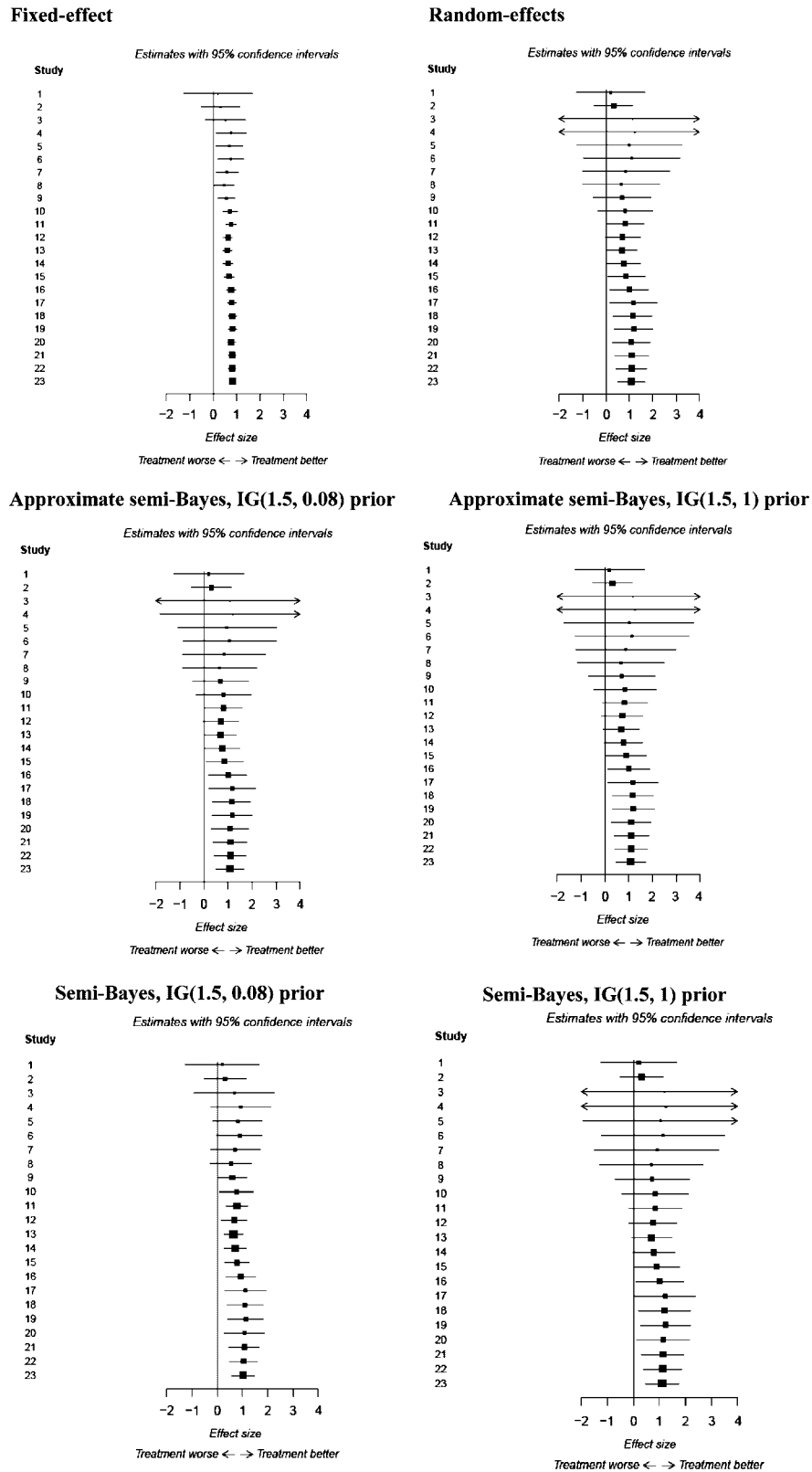


Figure 3. Sequential meta-analysis of peptic ulcers data.

the dataset, as with the $IG(1.5, 0.08)$ prior, the semi-Bayes method stops after a smaller number of studies (nine) although the approximate semi-Bayes method carries on for the same 11 studies as for the sequential random-effects method. The earlier stopping of the semi-Bayes $IG(1.5, 0.08)$ method

Table VI. Results from the sequential meta-analyses of the peptic ulcers data.

		Fixed	Random	Approx semi-Bayes IG(1.5,0.08)	Approx semi-Bayes IG(1.5, 1)	Semi-Bayes IG(1.5,0.08)	Semi-Bayes IG(1.5, 1)
At stopping	Number of trials	4	11	11	15	9	15
	Z_{end}/V_{end}	0.77	0.82	0.82	0.89	0.61	0.90
	Last confidence interval	0.14, 1.39	0.014, 1.63	0.042, 1.59	0.032, 1.75	0.015, 1.20	0.0054, 1.79
	$\hat{\tau}^2$	0	0.55	0.52	0.74	0.17	0.79
PEST analysis	<i>P</i> -value	0.0055	0.0052	0.0041	0.0048	0.0161	0.0039
	Median unbiased estimate of μ [odds ratio]	0.74	0.81	0.82	0.81	0.60	0.84
	Confidence interval for μ	[2.1]	[2.2]	[2.3]	[2.2]	[1.8]	[2.3]
		0.22, 1.24	0.24, 1.38	0.26, 1.37	0.25, 1.37	0.11, 1.09	0.27, 1.41

results in smaller estimates of treatment effect and heterogeneity. The use of an IG(1.5, 1) prior with its large prior estimate of heterogeneity causes the approximate semi-Bayes and semi-Bayes methods to stop after 15 trials. The estimates of treatment effect and heterogeneity from the sequential approaches are smaller than those from the overall random effects analysis of 23 studies. This is mainly due to the seventeenth trial, which has a large, beneficial effect of treatment, increasing the heterogeneity.

In our example, we have assumed that a sequential approach to the meta-analysis had been planned before the trials had reported their results. The example is instructive in illustrating some of the limitations of applying our generic statistical methods prospectively in a meta-analysis situation. We have remarked already that 2 × 2 contingency tables may be better analysed using binomial likelihoods. This would have avoided the need to make continuity corrections by adding 0.5 *s* to cell counts. Furthermore, the use of maximum likelihood estimates of odds ratios also creates a correlation between the y_i and v_i . Indeed, a commonly applied test for association between the estimates and standard errors produces $P = 0.005$ [34], although a more appropriate test based on efficient score and Fisher information statistics, which overcomes artefactual correlation between the estimates and weights, produces $P = 0.44$ [35]. Also, some of the studies display extreme treatment effects; for example, in the Chung study, 0 and 100 per cent of the patients bled in the treatment and control groups, respectively. This study is not very informative about the magnitude of the odds ratio, and receives very little weight in the meta-analysis, despite providing convincing evidence that the treatment has a large effect on the absolute risk of bleeding. Using risk differences rather than odds ratios as the effect size would have given this study more weight, but risk differences are generally discouraged as a metric for meta-analysis [36].

6. Discussion

This paper has considered the use of sequential methods for meta-analyses that incorporate random effects to allow for heterogeneity between studies. Our emphasis has been on determining a straightforward approach that has good empirical properties. We have developed and compared five methods, including a direct extension of the standard random-effects meta-analysis method, an extension of the approach of Biggerstaff and Tweedie, a semi-Bayes method involving Bayesian updating of the heterogeneity variance, and an approximation to the semi-Bayes method. The choice of an O'Brien and Fleming sequential design with a Christmas tree correction for discrete monitoring leads to a simple procedure for calculating repeated confidence intervals, which enables us to present sequential meta-analyses using forest plots, the conventional way of illustrating meta-analysis results. When either a repeated confidence interval excludes zero or V_{max} is reached, then the recommendation can be made that no further research is required to inform the question addressed by the meta-analysis.

Some issues are worthy of discussion, including philosophical, practical, and technical considerations. First, should sequential methods be applied to meta-analysis? Second, if sequential methods are to be used, what methods should be implemented in practice? Third, what technical problems remain that require further research?

The use of sequential methods for meta-analysis is contentious. Whereas the steering committee for a prospective study has control over the recruitment of participants, researchers undertaking meta-analyses may have no direct control over whether further studies are performed. Whether sequential methods have a role for updating of meta-analyses that are unconnected with primary researchers (such as those in Cochrane reviews) remains an open question. Chalmers and Lau [37] question the need to correct for multiple looks within a cumulative meta-analysis because of the lack of direct control. However, we would remark that many systematic reviews contain recommendations for practice and for research, and decisions over the content of these recommendations are similar to decisions over whether a primary research study should continue to recruit participants. Formal sequential methods for this process allow the wider field of evolving knowledge to be subject to the same rigorous considerations of Type I and Type II errors that individual randomized trials have traditionally received.

Meta-analyses are now well established as part of the research process, and many have argued for the need for primary studies to both be informed by meta-analyses of existing evidence, and be rounded off with an updated meta-analysis including the newly generated evidence. Thus primary research is increasingly viewed as part of a wider sequential process, and the methods described here may have application in this context. For instance, there may well be more enthusiasm for conducting a new study if the effect estimate is close to an upper or lower stopping boundary in a scheme such as Figure 1. Finally, we believe that sequential meta-analyses can play an important role in the actual design of individual studies, since the amount of further information that would be required to meet a pre-specified V_{\max} can be determined. However, there are many considerations in the design of a new study, and it is rare for them to be designed primarily around an updated meta-analysis. Related issues are discussed by Sutton *et al.* [38].

One area of potential application of sequential methods is the monitoring of adverse effects of pharmacological interventions. However, as we remark in Section 5, the particular general parametric approach we describe does not lend itself well to rare events. Furthermore, it is not clear that strict control over false-positive findings is important in this context, since a small, non-statistically significant, signal should still be investigated when the adverse effect is major. More research is required to explore appropriate methods for this issue.

If sequential methods are to be implemented, can the methods we have described be recommended? Certainly we would recommend against a fixed-effect approach. However, in a random-effects framework, a key problem in the early stages of a sequential meta-analysis is the poor estimation of the heterogeneity variance. We argue that prior information is therefore necessary. Of our two semi-Bayes methods, the approximate semi-Bayes approach is straightforward to implement, and this is the one we recommend from among those we investigated. From the simulation results, it appears that the method has reasonable false-positive and coverage properties, although these are likely to be dependent on the choice of prior distribution for the among-study variation. Our application demonstrated that the early stages of the sequential meta-analysis can be affected by the prior distribution. Further empirical research is needed to characterize the degree of heterogeneity that can be anticipated in a meta-analysis with particular clinical and methodological features, so that realistic informative prior distributions can be formulated. An alternative approach would be a sensitivity analysis in which the sequential meta-analysis is repeated for a variety of plausible values of τ^2 , but an explicit stopping rule does not follow from such an approach.

Our recommendations are interim, since more work is required in this area. We have commented on methods based on the law of the iterated logarithm proposed by Lan *et al.* [13, 14]. Their open-ended boundaries have the advantage of allowing studies to continue being conducted as long as the null hypothesis is not rejected, but they do not ensure that a pre-specified power is achieved. Open-ended boundaries with specified Type 2 error may offer preferable methods. Some technical issues with the proposed methods might also be investigated. For example, we have assumed that primary inference is to be made on the mean effect across studies. This does not take into account the extent of heterogeneity. It may be more appropriate to base decisions on alternative characteristics of the random-effects distribution, such as the probability that an individual study will have an effect of a particular size. Such inferences can be drawn from the predictive distribution of the effect in a new study [24]. Another limitation is that we have proposed the calculation of repeated confidence intervals until and including the first time that V_{\max} is exceeded. When this last repeated confidence interval is calculated, there is a probability of slightly less than $1 - \alpha$ that the truth is contained in every interval. The reduction in the probability depends on the increase in V from V_{\max} , and will be greater if only a few sequential steps have been conducted. A better understanding of this issue would be useful. Finally,

the advantage of using the Christmas tree correction for discrete monitoring with the O'Brien and Fleming design is that this leads to a simple procedure for calculating the repeated confidence intervals, but there are also some disadvantages to this approach. First, Stallard and Facey [39] show that the Christmas tree correction is less accurate for the O'Brien and Fleming design than for other types of sequential design, particularly if there are only a few looks at the data. Second, the Christmas tree correction depends on the information at the current (V_j) and previous (V_{j-1}) meta-analysis update. If $(V_j - V_{j-1})$ is less than 0, there is no Christmas tree correction. However, if at the next update $(V_{j+1} - V_j)$ is greater than 0, there will be a Christmas tree correction. However, V_{j+1} may be less than V_{j-1} . Further work is needed to assess the effect that this might have on the validity of the results.

We have considered a frequentist sequential approach to meta-analysis, proposing semi-Bayes approaches to allow prior information about τ^2 to be incorporated. The entire meta-analysis could alternatively be undertaken within a fully Bayesian framework. The meta-analyst would then calculate a new credibility interval for every meta-analysis update. The credibility interval, unlike the frequentist repeated confidence interval, is not corrected for multiple looks. The posterior distributions of μ and τ^2 from the first meta-analysis would become the prior distribution for the second and so on. If τ^2 increases during the course of the sequential meta-analysis, the credibility interval may become wider, but this causes no problem. Bayesian stopping rules for individual clinical trials have been proposed by various authors (see, for example [40, 41]). However, as the Bayesian inference is not affected by the repeated updates, Bayesian monitoring procedures can have very poor frequentist properties, such as inflated Type I errors [42].

Appendix

```
# An R function to perform a sequential meta-analysis using the approximate
# Semi-Bayes method to update estimates of the heterogeneity variance.
# Input:
# y, var: effect estimate and variance from each trial
# Vmax, H: vertical and horizontal boundaries of O'Brien and Fleming design
# eta, lambda: parameters of IG(eta, lambda) prior.
# Requires 'rmeta' package (http://cran.r-project.org/web/packages/rmeta/)

seq.ma.appbayes <-
function(y, var, Vmax, H, eta, lambda) {

  # set up output data
  require(rmeta)
  yout <- varout <- Zout <- Vout <- tausqout <- lowerout <- upperout <- stopping
    <- rep(0, length(y))
  Vlast <- 1 / var[1]

  # loop for each trial in the sequential meta-analysis
  for (i in 1:length(y)) {

    # Compute approximate Bayes estimate of heterogeneity variance
    # from DerSimonian-Laird estimate and prior distribution
    tau.sq.DL <- meta.summaries(y[1:i], sqrt(var[1:i]), method="random")$tau2
    tausqout[i] <- tau.sq.AB <- max((2 * lambda + i * tau.sq.DL) / (2 * eta + i - 2), 0)
    # Compute cumulative Z and V and estimates of y and var(y)
    Vout[i] <- cumul.V <- sum(1 / (var[1:i] + tau.sq.AB))
    Zout[i] <- cumul.Z <- sum(y[1:i] / (var[1:i] + tau.sq.AB))
    yout[i] <- cumul.Z / cumul.V
    varout[i] <- 1 / cumul.V

    # Compute "Christmas tree" correction
    Vdiff <- max(0, cumul.V - Vlast)
    corrected.upper <- H - 0.583 * sqrt(Vdiff)
    corrected.lower <- -H + 0.583 * sqrt(Vdiff)
    Vlast <- cumul.V
    # Compute repeated confidence intervals
    lowerout[i] <- (cumul.Z - corrected.upper) / cumul.V
    upperout[i] <- (cumul.Z - corrected.lower) / cumul.V

    # Determine if a stopping boundary is crossed
    # 1 = boundary crossed, 0 = not crossed
```

```

stopping[i] <- ifelse((i>2) & (lowerout[i]>0 | upperout[i]<0 | cumul.V>Vmax),
  1, 0)
}

# output results
out <-
  data.frame(y=yout, var=varout, Z=Zout, V=Vout, lower=lowerout,
    upper=upperout, tausq=tausqout, stopping=stopping)
return(out)
}

```

Acknowledgements

This work was supported by an MRC studentship, by MRC Grant U.1052.00.011 and by a grant from the PHG Foundation.

References

1. The Cochrane Collaboration. Cochrane Database of Systematic Reviews. *The Cochrane Library*. Wiley: Chichester, U.K., 2010.
2. Higgins JPT, Green S, Scholten RJPM. Maintaining reviews: updates, amendments and feedback. In *Cochrane Handbook for Systematic Reviews of Interventions*, Higgins JPT, Green S (eds). Wiley: Chichester, U.K., 2008; 31–49.
3. Simes RJ. Prospective meta-analysis of cholesterol-lowering studies: the prospective pravastatin pooling project (PPP) and the cholesterol treatment trialists (CTT) collaboration. *American Journal of Cardiology* 1995; **76**:C122–C126. DOI: 10.1016/S0002-9149(99)80482-2.
4. Baigent C, Keech A, Kearney PM, Blackwell L, Buck G, Pollicino C, Kirby A, Sourjina T, Peto R, Collins R, Simes R. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90 056 participants in 14 randomised trials of statins. *The Lancet* 2005; **366**:1267–1278. DOI: 10.1016/S0140-6736(05)67394-1.
5. Writer WD, Stienstra R, Eddleston JM, Gatt SP, Griffin R, Gutsche BB, Joyce TH, Hedlund C, Heeroma K, Selander D. Neonatal outcome and mode of delivery after epidural analgesia for labour with ropivacaine and bupivacaine: a prospective meta-analysis. *British Journal of Anaesthesia* 1998; **81**:713–717.
6. Schunkert H, Gotz A, Braund P, McGinnis R, Tregouet DA, Mangino M, Linsel-Nitschke P, Cambien F, Hengstenberg C, Stark K, Blankenberg S, Tiret L, Ducimetiere P, Keniry A, Ghorji MJ, Schreiber S, El Mokhtari NE, Hall AS, Dixon RJ, Goodall AH, Liptau H, Pollard H, Schwarz DF, Hothorn LA, Wichmann HE, Konig IR, Fischer M, Meisinger C, Ouwehand W, Deloukas P, Thompson JR, Erdmann J, Ziegler A, Samani NJ. Repeated replication and a prospective meta-analysis of the association between chromosome 9p21.3 and coronary artery disease. *Circulation* 2008; **117**:1675–1684. DOI: 10.1161/CIRCULATIONAHA.107.730614.
7. Sims AM, Timms AE, Bruges-Armas J, Burgos-Vargas R, Chou CT, Doan T, Dowling A, Fialho RN, Gergely P, Gladman DD, Inman R, Kauppi M, Kaarela K, Laiho K, Maksymowych W, Pointon JJ, Rahman P, Reveille JD, Tuomilehto J, Vargas-Alarcon G, Wordsworth BP, Xu H, Brown MA. Prospective meta-analysis of interleukin 1 gene complex polymorphisms confirms associations with ankylosing spondylitis. *Annals of the Rheumatic Diseases* 2008; **67**:1305–1309. DOI: 10.1136/ard.2007.081364.
8. Berkey CS, Mosteller F, Lau J, Antman EM. Uncertainty of the time of first significance in random effects cumulative metaanalysis. *Controlled Clinical Trials* 1996; **17**:357–371. DOI: 10.1016/S0197-2456(96)00014-1.
9. Whitehead A. A prospectively planned cumulative meta-analysis applied to a series of concurrent clinical trials. *Statistics in Medicine* 1997; **16**:2901–2913. DOI: 10.1002/(SICI)1097-0258(19971230)16:24<2901::AID-SIM700>3.0.CO;2-5.
10. Whitehead J. *The Design and Analysis of Sequential Clinical Trials* (revised 2nd edn). Wiley: Chichester, U.K., 1997.
11. Pogue JM, Yusuf S. Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. *Controlled Clinical Trials* 1997; **18**:580–593. DOI: 10.1016/S0197-2456(97)00051-2.
12. Devereaux PJ, Beattie WS, Choi PT, Badner NH, Guyatt GH, Villar JC, Cina CS, Leslie K, Jacka MJ, Montori VM, Bhandari M, Avezum A, Cavalcanti AB, Giles JW, Schrickler T, Yang H, Jakobsen CJ, Yusuf S. How strong is the evidence for the use of perioperative beta blockers in non-cardiac surgery? Systematic review and meta-analysis of randomised controlled trials. *BMJ* 2005; **331**:313–321. DOI: 10.1136/bmj.38503.623646.8F.
13. Lan KKG, Hu M, Cappelleri JC. Applying the law of iterated logarithm to cumulative meta-analysis of a continuous endpoint. *Statistica Sinica* 2003; **13**:1135–1145.
14. Hu M, Cappelleri JC, Lan KKG. Applying the law of iterated logarithm to control type I error in cumulative meta-analysis of binary outcomes. *Clinical Trials* 2007; **4**:329–340. DOI: 10.1177/1740774507081219.
15. Brok J, Thorlund K, Gluud C, Wetterslev J. Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses. *Journal of Clinical Epidemiology* 2008; **61**:763–769. DOI: 10.1016/j.jclinepi.2007.10.007.
16. Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. *Journal of Clinical Epidemiology* 2008; **61**:64–75. DOI: 10.1016/j.jclinepi.2007.03.013.
17. MPS Research Unit. *PEST 4*. University of Reading: Reading, U.K., 2000.
18. Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomised clinical trials. *Statistics in Medicine* 1991; **10**:1665–1677. DOI: 10.1002/sim.4780101105.

19. Shuster JJ. Empirical vs natural weighting in random effects meta-analysis. *Statistics in Medicine* 2010; **29**:1259–1265.
20. Laird N, Fitzmaurice G, Ding X. Comments on ‘Empirical vs natural weighting in random effects meta-analysis’. *Statistics in Medicine* 2010; **29**:1266–1267.
21. Waksman JA. Comments on ‘Empirical vs natural weighting in random effects meta-analysis’. *Statistics in Medicine* 2010; **29**:1268–1269.
22. Thompson SG, Higgins JPT. Comments on ‘Empirical vs natural weighting in random effects meta-analysis’. *Statistics in Medicine* 2010; **29**:1270–1271.
23. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**:177–188.
24. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society Series A* 2009; **172**:137–159.
25. Bollen CW, Uiterwaal CS, van Vught AJ, van dT I. Sequential meta-analysis of past clinical trials to determine the use of a new trial. *Epidemiology* 2006; **17**:644–649. DOI: 10.1097/01.ede.0000239658.19288.22.
26. O’Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics* 1979; **35**:549–556.
27. Jennison C, Turnbull BW. Interim analyses: the repeated confidence interval approach. *Journal of the Royal Statistical Society, Series B* 1989; **51**:305–361.
28. Biggerstaff BJ, Tweedie RL. Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis. *Statistics in Medicine* 1997; **16**:753–768. DOI: 10.1002/(SICI)1097-0258(19970415)16:7<753::AID-SIM494>3.0.CO;2-G.
29. Hardy RJ, Thompson SG. A likelihood approach to meta-analysis with random effects. *Statistics in Medicine* 1996; **15**:619–629. DOI: 10.1002/(SICI)1097-0258(19960330)15:6<619::AID-SIM188>3.0.CO;2-A.
30. Higgins JPT, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine* 1996; **15**:2733–2749. DOI: 10.1002/(SICI)1097-0258(19961230)15:24<2733::AID-SIM562>3.0.CO;2-0.
31. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Earlbaum: New Jersey, 1988.
32. Sacks HS, Chalmers TC, Blum AL, Berrier J, Pagano D. Endoscopic hemostasis: an effective therapy for bleeding peptic ulcers. *Journal of the American Medical Association* 1990; **264**:494–499.
33. Bradburn MJ, Deeks JJ, Berlin JA, Localio AR. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Statistics in Medicine* 2006; **26**:53–77. DOI: 10.1002/sim.2528.
34. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; **315**:629–634.
35. Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Statistics in Medicine* 2006; **25**:3443–3457.
36. Deeks JJ, Altman DG. Effect measures for meta-analysis of trials with binary outcomes. In *Systematic Reviews in Health Care: Meta-analysis in Context*, Egger M, Davey Smith G, Altman DG (eds). BMJ Publication Group: London, 2001; 313–335.
37. Chalmers TC, Lau J. Meta-analytic stimulus for changes in clinical trials. *Statistical Methods in Medical Research* 1993; **2**:161–172. DOI: 10.1177/096228029300200204.
38. Sutton AJ, Cooper NJ, Jones DR, Lambert PC, Thompson JR, Abrams KR. Evidence-based sample size calculations based upon updated meta-analysis. *Statistics in Medicine* 2006; **26**:2479–2500. DOI: 10.1002/sim.2704.
39. Stallard N, Facey KM. Comparison of the spending function methods and the christmas tree correction for group sequential trials. *Journal of Biopharmaceutical Statistics* 1996; **6**:361–373. DOI: 10.1080/10543409608835149.
40. Berry DA. Interim analyses in clinical-trials: classical vs bayesian approaches. *Statistics in Medicine* 1985; **4**:521–526. DOI: 10.1002/sim.4780040412.
41. Freedman LS, Spiegelhalter DJ. Comparison of Bayesian with group sequential methods for monitoring clinical trials. *Controlled Clinical Trials* 1989; **10**:357–367. DOI: 10.1016/0197-2456(89)90001-9.
42. Jennison C, Turnbull BW. *Group Sequential Methods with Applications to Clinical Trials*. Chapman & Hall/CRC: Boca Raton, FL, 2000.