AN ENCOURAGING ASSESSMENT OF METHODS TO INFORM PRIORITIES FOR UPDATING SYSTEMATIC REVIEWS

Alex J. Sutton[1], Sarah Donegan[2], Yemisi Takwoingi[3], Paul Garner[4], Carol Gamble[2], Alison Donald[4]

1 University of Leicester
2 University of Liverpool
3 University of Birmingham
4 Liverpool School of Tropical Medicine
5 GlaxoSmithKline

Corresponding Author:
Alex Sutton
Reader in Medical Statistics
Department of Health Sciences
University of Leicester
2nd Floor (Room 214e), Adrian Building
University Road
Leicester
LE1 7RH
UK

Tel: [+44/0] 116 229 7268
Fax: [+44/0] 116 229 7250
Email: ajs22@le.ac.uk

Abstract

Background: Methods to identify where new trials may change the interpretation of an existing systematic review will facilitate efficient review updating.

Objectives: To consider the use of statistical methods that aim to prioritise the updating of a collection of systematic reviews based on preliminary literature searches.

Methods: A new simulation based method estimating statistical power and the ratio of the weights assigned to the predicted new and old evidence and the existing Barrowman *n* approach are considered. Using only information on the numbers of subjects randomised in the "new" trials, these were applied retrospectively, by removing recent studies, to existing systematic reviews from the Cochrane Infectious Diseases Group.

Results: Twelve systematic reviews were included. When the removed studies were reinstated, inferences changed in five of them. These reviews were ranked, in order of update priority, 1, 2, 3 , 4 and 11 and 1, 2, 3, 4 and 12 by the Barrowman *n* and simulation based power approaches respectively. The low ranking of one significant meta-analysis by both methods was due to unexpectedly favourable results in the reinstated study.

Conclusions: This study demonstrates the feasibility of the use of analytical methods to inform update prioritisation strategies. Under conditions of homogeneity, Barrowman's *n* and simulated power were in close agreement. We encourage further, prospective, evaluation of these methods.

Keywords: meta-analysis, systematic review, updating, power, methodology, simulation

Running title: Priorities for updating systematic reviews

Word count of main text: 5,963 words

1. Introduction

Systematic review is an accepted method for synthesising information in many research fields, including medicine. As a component of such reviews, meta-analysis is used extensively to synthesise quantitative information from multiple studies. Meta-analysis of (well conducted) randomised controlled trials (RCT) evaluating the effectiveness of interventions are considered to provide the highest level of evidence for informing treatment decisions. As such, systematic review and meta-analysis underpin much of evidence based medicine.(1)

The use of meta-analysis has greatly increased in medicine over the past two decades. In 2006 alone there were 2042 articles tagged with the publication type "meta-analysis" in Medline. Just considering meta-analyses of RCTs, by October 2007, the Cochrane database alone included 3,298 completed systematic reviews. The Cochrane Collaboration policy is that all reviews should be updated every 2 years; that's 4.5 reviews a day! Now systematic reviews and meta-analyses exist for many interventions, the challenge in the future will be updating these and keeping the evidence base relevant.

However, systematic reviews can become rapidly out of date as new research evidence emerges. In a recent study, 23% of a sample of 100 systematic reviews became out-of-date within two years of their publication, and 15% within one year. (2) The Cochrane Collaboration updates their reviews as new trials become available, but the task is considerable. Indeed, the problem exists for all reviews, and it is just as valuable to know when reviews published in journals are out-of-date and new updated systematic reviews should be made available. (3)

Updating systematic reviews is often a continual and a very time consuming process (it has been suggested an update is as costly and time consuming as conducting the original review (4)) and resources are usually limited; much like revising any academic paper, the whole document needs reappraising and updating. This includes lots of detailed information on results of searching, data extraction of primary and secondary outcomes, quality assessment, construction of forest plots, re-interpretation

of data, updates to the discussion, and amendments to reference lists etc. The identification of new studies is a small fraction of the whole systematic review process and resulting document.

This has lead to editors of Cochrane review groups having to make difficult choices between removing reviews from the database if they are deemed too out-of-date, channel limited resources into updating them or, as done in a recent pilot study, (5) 'close a review' following a judgement that there is not likely to be new evidence in the future, or the research question is no longer of interest. As part of the groundwork that goes into making this assessment, review groups may (as the Cochrane Infectious Disease one currently does) keep literature searches up-to-date and identify all new potentially relevant RCTs for inclusion in each review they currently maintain. At this point, basic details of the new trials, including the number of participants randomised, can be extracted relatively quickly from abstracts supplied by electronic databases. This paper considers how, with only minimal information on the new studies, update prioritisation strategies can be devised for a collection of existing systematic reviews; such as those of review groups within the Cochrane Collaboration. It is hoped use of such strategies can help keep clinical recommendations as up-to-date and accurate as possible, for a limited resource, and, offer improvements over less strategic update strategies such as updating of all reviews according to a perpetual (unordered) rota. Indeed, others have observed that the arbitrary Cochrane strategy may result in an inefficient use of resources in slowly developing fields or delayed incorporation of new knowledge in rapidly evolving fields. (4)

In order to use the methods described in this paper, it is envisaged that persons who manage collections of systematic reviews (e.g. Cochrane Editors) would keep details of their primary outcome meta-analyses and sample sizes of relevant new studies in one spreadsheet. With the sending of a single analysis command, an update prioritisation list is generated for all reviews in question. Crucially, the assessment can be repeated at any time as and when new studies are published and there are resources free to work on review updates. We have developed and make available on request (e-mail 1st author) Stata (6) macros to implement all analyses presented in the paper.

The remainder of this paper is structured as follows. Section 2 outlines an existing and a newly adapted method for assessing whether the conclusions of a meta-analysis are likely to change when the new evidence is included. Section 3 describes the retrospective application of these methods to the reviews containing fixed effect meta-analyses in the Cochrane Infectious Disease Systematic Review Database. Section 4 further considers the application of the methods to meta-analyses where heterogeneity is present. Section 5, the discussion, concludes the paper.

2. The effect of new evidence on meta-analysis results

Methods considered to prioritise reviews for updating are outlined below.

2.1 Barrowman's *n*

This method was initially presented by Barrowman et al (7) as a way of identifying null (i.e. non statistically significant) meta-analyses that are ripe for updating. Barrowman's *n* can be interpreted as the number of additional subjects required (on average) to obtain a statistically significant result from a null meta-analysis. They derive the following equation for *n*:

$$n = N\left(Z_C^2 / Z^2 - 1\right), \qquad [1]$$

where Z_C is the critical value for the *Z* statistic used to determine a significant result (i.e. in this paper Z_C =1.96, corresponding to a two-sided Type I error rate of 5%). *Z* is the z-statistic from the current meta-analysis derived by dividing the estimated pooled treatment effect by its standard error and *N* is the total number of participants included in the studies in the meta-analysis.

Using equation 1, the "new participant:*n* ratio", i.e. the ratio of the actual number of participants in new studies to *n* (i.e. the predicted number required to obtain statistical significance) can be calculated. The higher this ratio, the more chance the updated meta-analysis would have a statistically significant result.

The original exposition of this method (7) stressed it was only intended to be a quick approximate estimate of the amount of new evidence required (although it may be sufficiently accurate for purposes of update prioritisation). For example, it does not take into account factors such as sampling error in the existing estimate of the treatment effect, varying baseline risk between studies, the number of new studies, and between study heterogeneity.

2.2 The power of an updated meta-analysis using simulation

While closed-form expressions for calculating power exist for some simple statistical tests, this is not the case for other more complex procedures such as the updating of an existing meta-analysis. Fortunately, power of virtually any statistical test can be calculated through simulation methods in which data are repeatedly simulated stochastically under the alternative hypothesis and analysed. (8) We adapt such methods here to calculate the probability of producing a statistically significant result when adding further studies (in which the results are consistent with those that already exist) to an existing meta-analysis.

Such a method was initially described in the literature (9) in the context of designing new studies with enough power to change inferences of the updated meta-analysis in which they would ultimately be included. However, the method can be adapted to use in an update prioritisation context. When designing new studies, power is typically fixed (often at around 80%) and the necessary sample size is derived. In the present context, the sample size of the new study(s) is known and it is estimation of power which is required for such a sample size.

A non-technical summary of this simulation approach to calculating power when adding new studies to an existing meta-analysis is given below, and, full technical details can be found elsewhere.(9) For simplicity, this summary assumes only one new study is to be added to the existing meta-analysis, but the approach readily generalises to simulating data for multiple new studies.

1. From a meta-analysis of the existing studies, a distribution for the effect in a new study is derived. An effect size from this distribution is then sampled representing the predicted effect in the new study.
2. Data representing the new study are generated stochastically according to the effect sampled in step 1 for a specified sample size.
3. This simulated study data are then added to the existing meta-analysis, which is then re-meta-analysed.
4. The hypothesis test on which decisions are to be based is then considered and whether the null is retained or rejected in favour of the alternative hypothesis established at a specified level of statistical significance (often 5%) is noted.
5. Steps 2 to 4 are repeated a large number of times (*K*) noting the outcome of the hypothesis test each time.
6. Power is estimated by calculating the proportion of the *K* simulations in which the null hypothesis is rejected at the specified statistical level.

The number of replications (*K*) used is arbitrary, with the accuracy of the estimation of power increasing with *K*. This general approach can be used with any outcome measure and with any meta-analysis model.

For illustration, a worked example for simulating a new study using a fixed effect meta-analysis model on the relative risk (RR) scale is given. Details of the calculations for other scenarios are available elsewhere. (9) Consider the meta-analysis in Figure 1a concerning the outcome failure by day 28 treating uncomplicated malaria with chloroquine or amodiaquine combined with sulfadoxine-pyrimethamine. This meta-analysis is considered in more detail below but at present it is used to illustrate how the simulation approach works. This meta-analysis has a pooled log RR of -1.01 with a variance of 0.322. Assuming the log RR to be Normally distributed, an effect size is sampled and exponentiated to produce an estimate of the RR in the new study (Stage 1). For simplicity, the event rate in the control group of the new study is assumed to be fixed at the un-weighted average of the event rates in the three existing studies; which is 0.12 in this instance (although an alternative approach to estimating this quantity, and the procedure could be made more elaborate to take into account the uncertainty in this quantity). Since RR = Probability of an event in the treatment group / Probability of an event in the control

group, an estimate of the event rate in the treatment group can be derived using this equation and the estimates of the other two quantities. Data for the new study can now be simulated (Stage 2) by simulating the number of events from binomial distributions for both treatment and control groups for a sample sizes equal to that for the new study and probabilities derived above. Stages 3 to 6 can then be conducted.

The specific equations used are dependent on the outcome measure and the meta-analysis model utilised (10) but the general approach is applicable to any meta-analysis scenario. (9) In particular, the calculations are more complex in a random effects context since the prediction equations need to take into account between study heterogeneity. The approach used to prediction is described elsewhere (9) and implemented here using an approximate formulae suggested previously (11).

Once power is calculated for each meta-analysis that could potentially be updated, they can be ranked in order of priority by their estimated powers (descending order). An advantage this approach has over the Barrowman *n* is that these powers give a direct estimate of the likelihood that inferences would change in any particular meta-analysis. For example, the decision may be taken to only update reviews in which the primary meta-analysis had power of 60% or greater etc.

2.3 Other measures used as criteria for updating meta-analyses

Using the simulation framework described, it is possible to consider power for virtually any hypothesis test, or predict any statistic related of the updated meta-analysis, and these could be used as alternative bases (or in addition) to the power of the standard hypothesis test presented above for update prioritisation. For example it would be possible to assess the likelihood the effect size of the updated meta-analysis lay inside, outside, or, across pre-specified limits representing clinical equivalence (9) and could be used to help distinguish identifying statistically significant from clinically significant treatment effects. Alternatively, the ratio of the standard error of the predicted new estimate of effect to the existing, or, the ratio of the sum of the weights (10) allotted to the predicted new and existing studies in the updated meta-analysis could be used as criterion. These would allow prioritisation based on the relative *magnitude* of new to old evidence (opposed to the *results* of the studies). All these potential indicators or ‘signals’, and several more, including the qualitative

'signals' used elsewhere for determining when systematic reviews are out of date (2), are incorporated into Stata (6) macros which were used to conduct all analyses presented in this paper (including generating the priority rankings), and are available from the first author on request. The next section provides an application of the use of the methods to a systematic review database.

3. Retrospective application of the methods

Here we consider the feasibility and performance of the described methods to assess the likelihood a meta-analysis' conclusions will change given only the number of patients included in new studies not included in the existing meta-analysis. A retrospective design was chosen since a prospective study would take an excessive amount of time to complete for little or no perceivable benefit. Of course, in practice such methods would be used prospectively as tools to assist the decisions regarding the timing and prioritisation of updating collections of systematic reviews and meta-analyses, such as those maintained by the Cochrane Collaboration.

Methods: All 67 completed (as of April 2006) systematic reviews by the Cochrane Infectious Diseases Review Group in the Cochrane Collaboration Systematic Review Database were considered for potential inclusion. A systematic review was included if it met the following criteria:

- Contained a meta-analysis and stated a primary outcome
- The primary outcome was binary
- The meta-analysis of the primary outcome contained no cluster randomised trials
- The meta-analysis of the primary outcome had to contain studies which were published over a duration of four or more years (i.e. the date of publication of the most recent study included has to be at least three years greater than the first study)

For the systematic reviews which did meet criteria 1-4, all trials which were published within 3 years of the most recent study included in the primary endpoint meta-analysis were removed (e.g. if the most recent study was published in 2003 then all

studies published in 2001, 2002 & 2003 were removed). A further criterion was then applied:

- A meta-analysis of the primary endpoint on the remaining studies had to be non-significant at the 5% significance level.

The above process retrospectively identifies meta-analyses which would have been non-significant 3 years prior to the inclusion of the most recent study and assumed treatment effects were homogeneous between studies.

The two methods described in Section 2 were then applied to these reduced meta-analysis datasets, to predict the likelihood the updated meta-analysis, including the removed studies, would produce a significant treatment effect (at the 5% level). In addition, the simulation based approach considered the weights of the predicted (updated) to old studies in the updated meta-analysis.

The only information the methods require is the total number of persons randomised to the treatment and control groups of the removed studies. Barrowman's *n* and corresponding new participant:*n* ratio were calculated and systematic reviews were ranked using the latter.

For the simulation method, for simplicity, it is assumed the control group event rate in all the new studies will be equal to the average in the existing studies (estimated by a simple unweighted average) for that meta-analysis. Analysis was conducted using the outcome measure (i.e. OR or RR) and meta-analysis model (i.e. Mantel-Haenszel, Peto or random effects (10)) used in the original Cochrane review. The power and ratio of weight given to the predicted new and existing studies in the updated meta-analysis was then estimated using 5000 simulation replications and systematic reviews ranked accordingly.

The results of using Barrowman's *n* and the prioritisation rankings based on power and % weight of the new studies were compared and contrasted with each other and with the actual results of the updated meta-analyses.

Results: Of the 67 reviews in the Cochrane Collaboration Infectious Disease Review Group Database, 12 met our inclusion criteria. The meta-analyses used a fixed effect model in 10 of these reviews and a random effects model was used in the remaining two. In total, 55 reviews were excluded: 5 reviews did not contain any trials or did not conduct meta-analysis; the primary outcome of 4 reviews was not binary; 2 reviews analysed cluster randomised trials; and the results were still significant after the recent studies had been omitted in 14 reviews. In the remaining 30 cases, the small number of trials in the meta-analyses prevented the required retrospective analysis. While full details of the included systematic reviews are available elsewhere, (12-23) a brief description of the topic of each review is provided in the second column of Table 1.

Applying Barrowman's *n*: In Table 1 the Barrowman *n* and new participant:*n* ratio for the meta-analysis of the primary endpoint within each review is provided. The reviews are ranked on the magnitude of the ratio in descending order. Five of the 12 systematic reviews have ratios greater than one indicating more new evidence has accumulated than the Barrowman's *n* method estimated to be required to obtain statistical significance. The actual *p*-value for the updated meta-analysis is provided in the last column of Table 1. It can be seen that five of the 12 meta-analyses were actually statistically significant at the 5% level when updated. These five meta-analyses were ranked 1, 2, 3, 4 and 11 in order of update priority. Hence, if the five meta-analyses with a new participant ratio greater than one had been identified for update, four out of the five meta-analyses would have changed inferences at the 5% level. However, the change in inferences in review ranked 11 (drugs for treating tapeworm infection of the brain (20)) would have been missed. Other than for review 11, there is good agreement between the rank priority given to each review and the *p*-value of the updated meta-analysis.

Applying power by simulation: The estimated power of each of the updated meta-analyses calculated by simulation is displayed, ranked in descending order, in column 3 of Table 2. The first thing to note is that only one of the updated meta-analyses achieved power over 80% - a level often considered acceptable when carrying out primary research - and six out of the 12 do not achieve 50% power. The five meta-

analyses in which inferences did change were ranked 1, 2, 3, 4 and 12. There was close agreement between the ranks given to this and the Barrowman *n* method.

Ratio of the weights given to the new and old evidence: The estimate of the ratio of the weights are given to the new and old evidence in the simulated updated meta-analysis and associated rankings are provided in the final two columns of Table 2. While the systematic review ranked highest (Rotavirus vaccine for preventing diarrhoea (23)) is consistent with previous criteria, much less agreement is seen beyond this.

Detailed examination of two meta-analyses: For illustration, two of the meta-analyses are presented in more detail. Firstly, the review ranked second for updating by Barowman:*n* ratio and power and ranked fourth based on the weight ratio - Chloroquine or amodiaquine combined with sulfadoxine-pyrimethamine for treating uncomplicated malaria (21). The primary outcome in the meta-analysis is failure by day 28. Figure 1a presents the meta-analysis used to base the calculation on with the two most recent studies removed (references to individual trials in each meta-analysis are available from the original Cochrane reviews). This meta-analysis suggests the treatment is potentially efficacious (RR = 0.37 (95% CI 0.12 to 1.11)) but not statistically significant at the 5% level ($p = 0.07$). When the two removed studies are re-instated into the meta-analysis (Figure 1b), although the pooled RR increases considerably to 0.64, it is now highly statistically significant ($p = 0.01$). Examination of the weightings (10) given to each study in the meta-analysis in Figure 1b indicates that the two newest studies, initially removed, are together given over 80% of the weight in the meta-analysis (resulting in an actual weight ratio of 4.5), reflecting the considerable "new" evidence that was available and hence the potential benefit in updating the review. Note this observed weight ratio is higher than that predicted by the simulation model (1.6) due to the much higher event rates observed in the two 'new' studies compared to the existing studies from which predictions on the new studies were made.

Secondly, we consider the review of drugs for treating tapeworm infection of the brain (20) which was ranked 11 and 12 by the Barrowman *n* and simulation approaches respectively but (unpredictably) produced a statistically significant result

when updated. The meta-analysis on which calculations were based, with one study removed is displayed in figure 2a. Here three studies exists, two of which are much more precise than the third. These two studies both have point estimates close to the null (RR of 1.04 and 0.86). The pooled point estimate is also close to the null (RR = 0.92) with relatively high precision (95% CI 0.78 to 1.09). The excluded study has a "surprising" result with a large statistically significant beneficial effect (RR 0.49 (95% CI 0.25 to 0.96)). This has an "unpredictable" effect on the meta-analysis of making the treatment effect statistically significant overall. Hence, all methods "failed" to give high priority to this review for updating because the new evidence was inconsistent from what was predicted given the previous three studies (although the weight ratio prediction was reasonably accurate). It is interesting to note that, following inclusion of the later study with the large beneficial effect, it is probably not reasonable to consider the study effects as homogeneous, which invalidates the fixed effect assumption made using the simulation method.

Relationship between methods: As a way of assessing the relationship between the Barrowman *n* and power by simulation methods, the power of updating each meta-analysis was calculated by the simulation method for a study with size equal to that calculated as the Barrowman *n*. The results are presented in the final column of Table 2. The estimated power ranges from 32% to 58% with an average of 47% (power could not be determined for the mycobacterium vaccine for TB review (19) due to the extreme nature of the associated Barrowman *n*). Since Barrowman's *n* is a calculation of the size a study would have to be to obtain a significant result in a meta-analysis *on average*, or put another way, 50% of the time, these estimated powers all around 50% are not surprising and suggest close agreement between methods. An interesting observation is that the estimated power from the two meta-analyses which use a random effects model are furthest away from 50% suggesting potentially poorer agreement between methods when the simulation approach models between study heterogeneity.

The next section considers in further detail the application of the methods in a heterogeneous treatment effects scenario.

4. Use of methods with heterogeneous treatment effects

In this section we consider a further meta-analysis which has heterogeneous treatment effects. For illustration, the event data for the most recent study are (retrospectively) removed and we then apply the Barrowman *n* method and the simulation approach to predict the power and weight ratio of new to old studies using both fixed and random effect meta-analysis models.

The meta-analysis comes from the Cochrane review of chemotherapy alone versus endocrine therapy alone for metastatic breast cancer (24) and the outcome is the tumour response rate of endocrine therapy versus chemotherapy. In the original review a fixed effect analysis (despite large heterogeneity being estimated) on the relative risk scale was carried out. We use the odds ratio scale here (although this makes little difference to the calculations) and consider both fixed and random effect analyses and compare the results.

Figures 3a and b present the fixed and random effect meta-analysis models for the earliest 6 studies included in the meta-analysis respectively. It can be seen that choice of meta-analysis model changes the pooled effect considerably with the fixed effects model estimating the treatment effect as 1.36 (95% CI 0.99 to 1.86) and the random effect as 1.17 (0.53 to 2.60). The I^2 statistic (25) (which estimates the proportion of total variation in point estimates attributable to heterogeneity rather than sampling error) for this dataset is 73% (i.e. reasonably large), and the test for heterogeneity has a p-value of <0.001. The excluded study from these plots is published in 1992 (Dixon et al.) that randomised approximately thirty patients each to the two treatment groups.

For the fixed effect analysis, the Barrowman *n* is 26, hence the new participant ratio is 2.29 for a study of the size actually conducted. The power of the new study to change inferences of the meta-analysis by simulation under a fixed effect model is estimated to be 52.7%. Hence, using both methods one would conclude there would be at least a

reasonable chance inferences changed if the meta-analysis were updated despite the weight ratio being a low 0.08.

For the random effects analysis, the Barrowman *n* was 2846 - several orders of magnitude larger than for the fixed effects - and with a low new participant ratio of 0.01. The power by simulation is estimated to be 0% (i.e. none of the simulations produced a significant result) with a weight ratio of 0.15 (this is larger than under the fixed effect scenario since studies get more equal weighting under random effects models). Hence, using any of the indicators, one would conclude there would be very little chance that inferences would change if the meta-analysis were updated. This implies the meta-analysis model choice can be critical to updating decisions.

Figures 3c and d indicate the actual meta-analysis results when the most recent study was reinstated. Under the fixed effects model the treatment effect did become significant (OR = 1.38 (1.02 to 1.88) but not under a random effects model (1.26 (0.62 to 2.56)) which coincides well with the predictions both methods make.

Given the potential difference in results between fixed and random effect model predictions, it is important to consider how the decision of which method to use is made, in a particular context. This is made more difficult by knowledge that the test for heterogeneity has low power when the number of studies is small; (26) which they typically will be for many medical interventions.

5. Discussion

A recent systematic review (4) suggests that the whole process of updating systematic reviews is under-researched and that the importance of updating has not been well recognised. We have begun to rectify this situation by exploring the use of methods based on both approximate closed form solutions and simulation methods to help indicate when systematic reviews need updating, and to create a prioritisation order for updating across a collection of such reviews. Such prioritisation will become increasingly important over time as the number of published reviews, and, hence burden of updating, increases. While the focus of this paper has been on meta-

analyses of interventions the methods readily translate to other areas of application (e.g. environmental health etc).

We believe the results presented in this paper are encouraging enough to justify pursuing their further evaluation. We hope a prospective pilot can be conduced to assess the added value of using the methods over current updating practices.

Barrowman's *n* and simulated power approaches produce results in close agreement under conditions of homogeneity, and when the new studies are consistent with those meta-analysed previously. However, we believe, the simulation approach will generally produce more accurate results over a broader range of scenarios, including when between study heterogeneity is present, due to a more realistic modelling of the meta-analysis data.

In reality, there would be no need to restrict the update priority assessment to only "significant" meta-analyses if the simulation approach is used. (Barrowman's *n* was intended for use only in situations where the existing meta-analysis is non-significant.) For presently "significant" meta-analyses it would be possible to calculate power to become "non-significant" – i.e. the criteria are reversed but the principle is the same (this possibility is programmed in the developed software).

Restricting attention to methods which only consider changes in statistical significance have important limitations, due to their arbitrary nature (27). Of course, important changes in effect sizes may happen that are not associated with changes in statistical significance. We consider a further measure in the ratio of the predicted total weightings given to the predicted new and existing evidence. We like this measure since it takes into account the relative magnitude of new evidences as well as the impact of (any) between study heterogeneity on the analysis (big studies get relatively less weight the more heterogeneity there is present). The further possibility of using clinical limits of equivalence has been discussed above and in more detail elsewhere. (9)

In fact, the simulation method could be used to calculate prioritisation based on any quantitative measure relating to the updated meta-analysis. Recent work (2) compiled

a list of further potential ‘signals’ which were used to identify potentially out of date systematic reviews. These included both quantitative and qualitative measures. While we have implemented all the quantitative methods in the software, further work is needed to evaluate the relationships between signals and evaluate the possibility of using more complicated composite updating criteria based on multiple measures or ‘signals’. Using multiple measures, it may also be possible to define criteria to establish when enough evidence has been accrued in a given topic and the systematic review can be deemed decisive. A further extension would be to place the whole process in a fully decision theoretic framework, and use value of information methods, to equate cost of updating to the benefit (in monetary terms) of the update. (28) A further area of potentially fruitful research would be to assess how good predictions of new evidence are using only the information in the existing systematic review. If this were possible, it would mean the methods presented here could be adapted to be used even when literature searches are not kept up-to-date.

When the new evidence is not consistent with the existing meta-analysis then none of the approaches considered will work well. Reasons why new evidence may be inconsistent with previous evidence include pipeline reporting biases (29), “drift” in the effectiveness of the intervention, or chance. It is hard to envisage a complete solution to this problem, but systematic differences in effects between studies could be accounted for by including covariates in the meta-analysis models. Further, perhaps, when searching for new trials, special status could be given for those with results which are extreme compared to the trials in the existing meta-analysis.

Below, we make some observations relating to the details of the simulation methods we have used. Unlike previous, related, work (9) predicting future study results using meta-analysis, we do not use Bayesian methods in this paper or the developed software. We took the decision that the extra computational burden and software requirements would make the software less user-friendly. We felt this outweighed the benefits of a slightly more accurate prediction model afforded by Bayesian methods (for the random effect context) for the priority updating purposes considered in the paper. This would not necessarily be the case if one were designing a future study for an individual meta-analysis context (as considered in the previous paper) where accuracy is arguably more crucial and the computational burden will be much lower.

A related technical issue under random effect modelling is that standard meta-analysis methods ignore the uncertainty in the estimation of the between study heterogeneity parameter, which can be large when there are few studies in the meta-analysis (a context in which such methods may often be applied). Classical (30, 31) and Bayesian (32) solutions to allow for such uncertainty are available but rarely used in practice. However, as explained above, we traded simplifications in modelling for lower computational burden.

In predicting future studies we make the assumption that the underlying event rate in new studies is the same as the average of the control rates in the observed studies. Adding further sophistication to the simulation process would allow this to be relaxed in favour of imputing a stochastic estimate derived from the observed distribution of baseline rates. In fact, to do this it would be necessary to modify the standard meta-analysis models since these make no assumptions about the baselines (i.e. each one is estimated independently as 'nuisance' parameters and they are not assumed to be related to each other in any way). Incorporating random effects for baseline effects (as well as treatment effects) (33) would be one approach allowing stochastic baseline predictions to be made.

We fully acknowledge that further information relating to a systematic review not contained in the meta-analysis of the primary outcome may be pertinent to updating. For example, new information on adverse effects, or the development of competing interventions may be critical. Also, it may be advantageous to consider the prevalence and severity of the disease in question when setting update priorities. Also, methodology may have improved since the original systematic review was done and hence updates using the new methodology may be desirable. The first analyses identifying predictors of the speed a systematic review will go out of date have been published. (2) We support the development of a broader approach to update prioritisation which could incorporate this information and which the methods presented in this paper could feed into, but acknowledge that this is some way off.

An alternative approach to updating reviews would be to update the meta-analysis of the primary outcome in all reviews but only progress with the updating of the total

systematic review if inferences actually change. However, such an approach would induce bias in the published literature since effect sizes would be over-estimated because of the updating process being of a "data dredging" nature. Conversely, usinga strategic approach to updating may actually reduce the problem of multiple testing (34) in updated meta-analysis which means (N.B. even if no treatment effect exists, the null hypothesis will be rejected eventually if the meta-analysis is updated enough times). Further, methods are emerging to control for this when updating meta-analysis and these could be used in conjunction with the methods presented here. (35)

There may be merit in integrating an assessment of how much further evidence is required on a topic into the publication of the initial systematic review. Such information would be useful to inform the design of new studies as well as providing guidance on when an update would be worthwhile. If such an endeavour was seen as a requirement of systematic review, it would help bring coherence to the way new research was conducted. Of course, if all new studies were designed to adequately power an updated meta-analysis (9) then it would always be worthwhile updating the meta-analysis and thus reducing the need for a priority setting process such as the ones considered here, but as this study shows this is often not the case.

Acknowledgements: The authors would like to thank David R Jones and Kaveh Shojania for useful discussions relating to this paper and two anonymous referees for their insightful and constructive comments which considerably improved this paper.

Conflicts of interest: None declared.

Table 1 Results of applying Barrowman *n* calculations to the 12 primary outcome meta-analyses

Rank by Barrowman new participant ratio	Review	Total number of studies (number removed)	Number of subjects in existing (non-removed) studies	Barrowman *n*	Actual total number of new subjects in removed studies	New participant ratio*	*P*-value of the updated meta-analysis
1	Rotavirus vaccine for preventing diarrhoea (23)	7 (5)	410	365	7397	20.28	0.01
2	Chloroquine or amodiaquine combined with sulfadoxine-pyrimethamine for treating uncomplicated malaria (21)	5 (2)	172	38	289	7.6	0.01
3	Steroids for tuberculous meningitis (12)	6 (3)	344	66	251	3.8	0.02
4	Artemisinin derivatives for treating severe malaria (22)	11 (3)	1379	300	700	2.33	0.02
5	Reduced osmolarity oral rehydration solution for cholera (13)	5 (3)	223	194	393	2.03	0.26
6	Antibiotics for treating salmonella gut infections (14)	5 (1)	339	67	46	0.69	0.07
7	Routine anticovulsants for cerebral malaria (15)	3 (1)	233	599	340	0.57	0.36
8	High first dose quinine regimen for severe malaria (17)	3 (1)	72	191	72	0.38	0.43
9	Interventions for melioidosis (16)	4 (2)	182	1157	314	0.27	0.69
10	Drugs to prevent malaria-related illness in pregnant women and death in the newborn (18)	4 (2)	1431	7186	1459	0.2	0.90
11	Drugs for tapeworm of the brain (20)	4 (1)	242	743	63	0.08	0.04
12	Mycobacterium vaccae immunotherapy for treating tuberculosis (19)	5 (3)	476	3718118	1265	0	0.60

* The ratio of the actual total number of participants in new studies to the predicted number required to obtain statistical significance

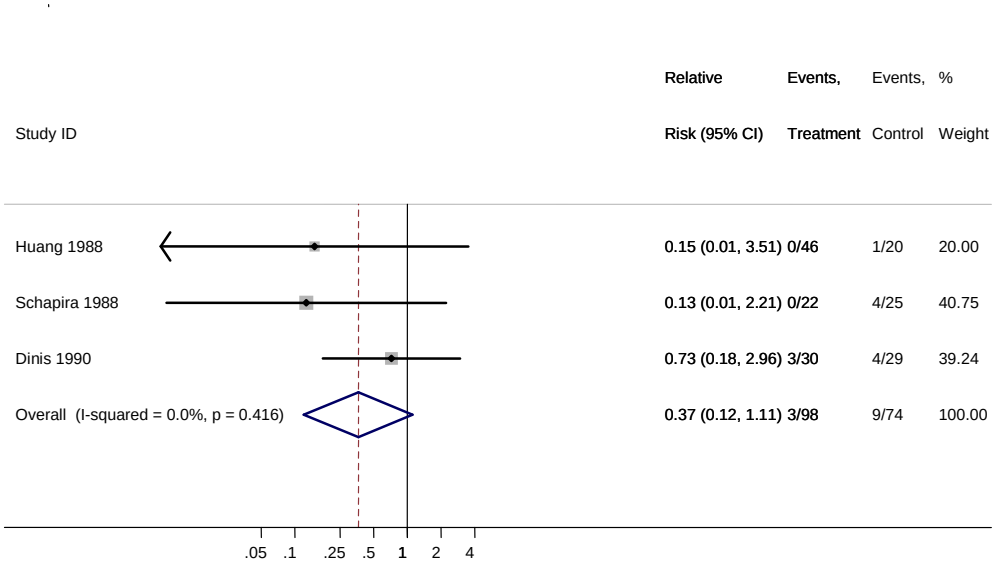
Table 2 Results of applying power by simulation to the 12 primary outcome meta-analyses

Rank by simulated power	Review	Estimated power when adding new studies	*P*-value of the updated meta-analysis	Rank by Barrowman new participant ratio	Estimated power when adding new study of Barrowman *n* size	Meta-analysis model used	Ratio of weight given to new and old studies in meta-analysis	Rank by weight ratio estimate
1	Rotavirus vaccine for preventing diarrhoea (23)	89%	0.01	1	32%	Random	15.7	1
2	Chloroquine or amodiaquine combined with sulfadoxine-pyrimethamine for treating uncomplicated malaria (21)	79%	0.01	2	44%	Fixed	1.6	4
3	Steroids for tuberculous meningitis (12)	67%	0.02	3	46%	Fixed	0.72	9
4	Artemisinin derivatives for treating severe malaria (22)	64%	0.02	4	58%	Random	0.45	10
5	Reduced osmolarity oral rehydration solution for cholera (13)	62%	0.26	5	48%	Fixed	1.7	3
6	Antibiotics for treating salmonella gut infections (14)	53%	0.07	6	54%	Fixed	0.19	12
7	Routine anticovulsants for cerebral malaria (15)	34%	0.36	7	46%	Fixed	1.3	6
8	Interventions for melioidosis (16)	27%	0.69	9	52%	Fixed	1.5	5
9	Drugs to prevent malaria-releated illness in pregnant women and death in the newborn (18)	20%	0.90	10	54%	Fixed	1.0	7
10	Mycobacterium vaccae for treating tuberculosis (19)	19%	0.60	12	NA*	Fixed	2.4	2
11	High first dose quinine regimen for severe malaria (17)	17%	0.43	8	46%	Fixed	1.0	7
12	Drugs for tapeworm of the brain (20)	6%	0.04	11	42%	Fixed	0.28	11

* Barrowman *n* too large for this study to calculate power

Figure 1 Primary outcome from Cochrane Systematic review of chloroquine or amodiaquine combined with sulfadoxine-pyrimethamine for treating uncomplicated malaria: Failure by day 28

a) Two most recent studies removed



b) Two most recent studies reinstated

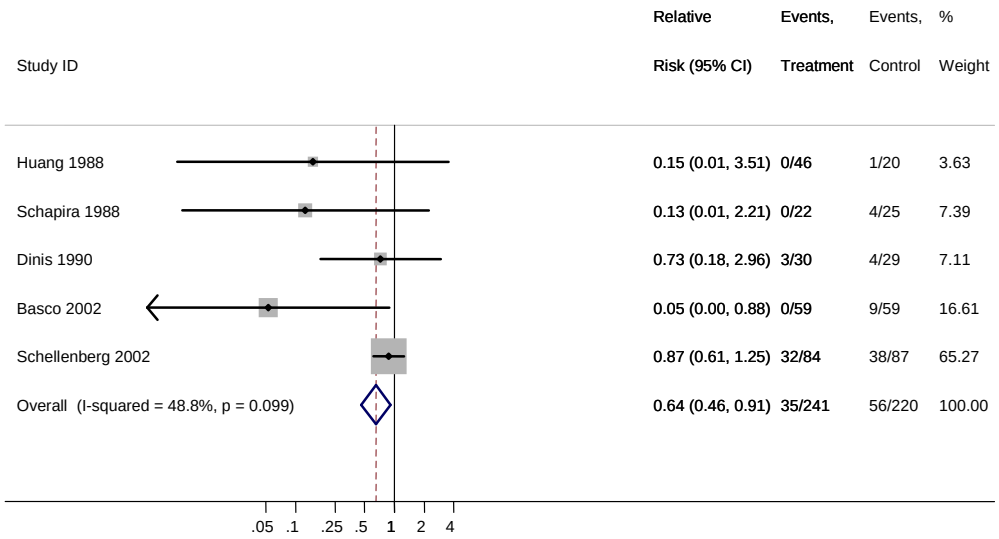
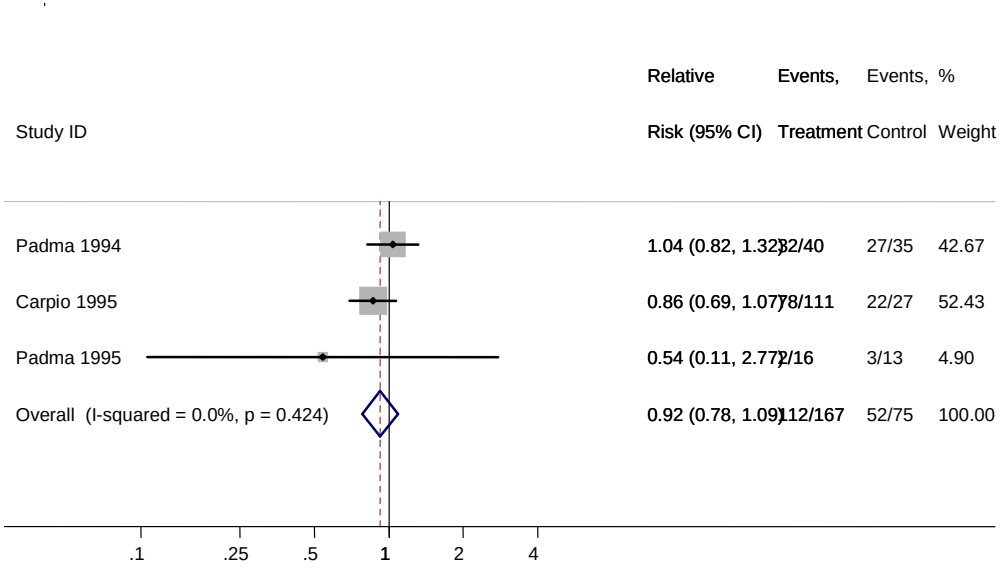


Figure 2 Primary outcome from Cochrane Systematic review on drugs for treating tapeworm infection of the brain: Cyst persistence

a) Most recent study removed



b) Most recent study reinstated

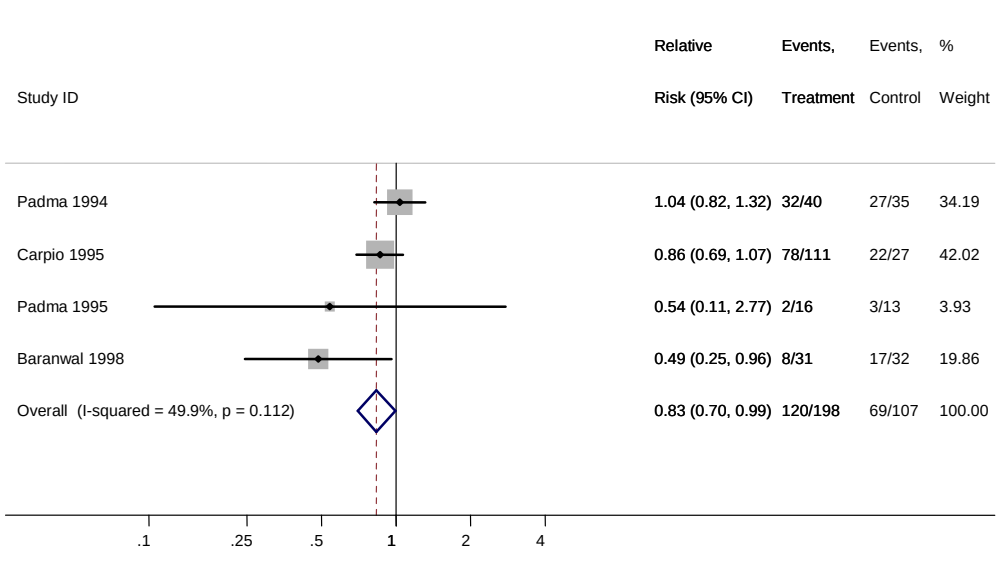
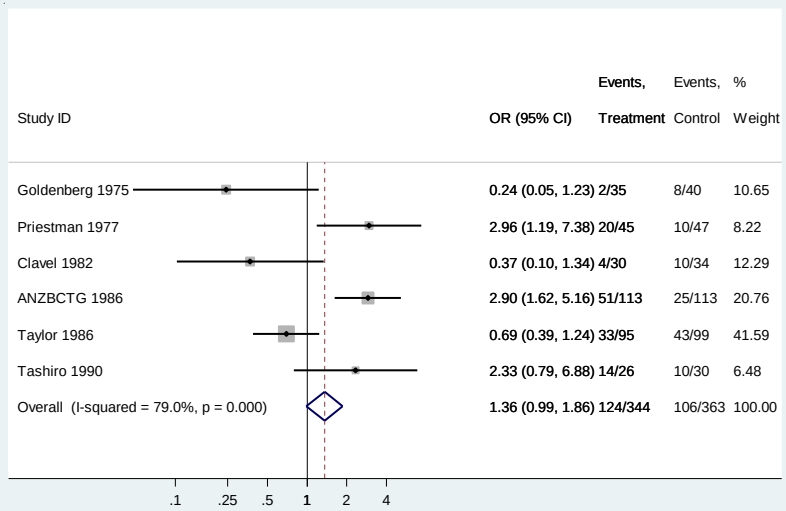
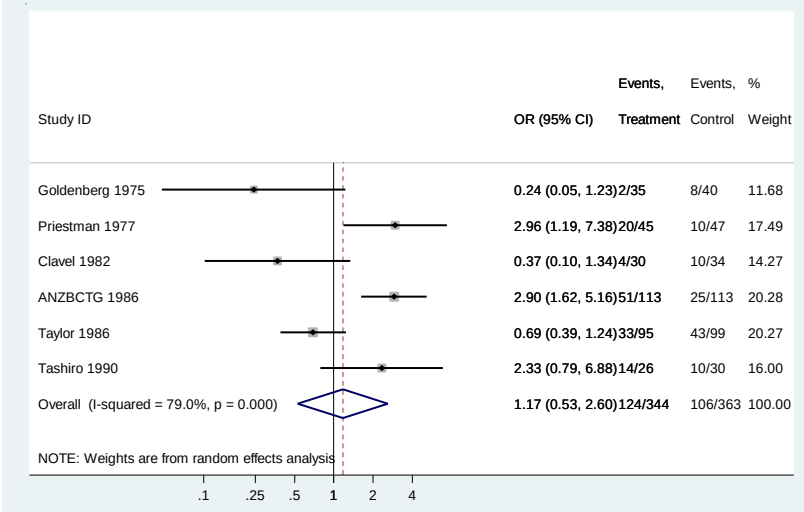


Figure 3 Chemotherapy versus endocrine therapy for metastatic breast cancer: Tumour response rate

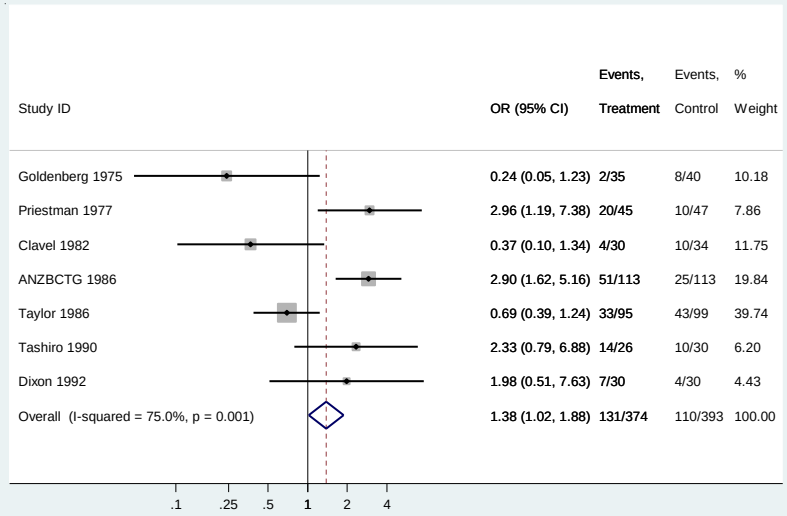
a) Fixed effect analysis – most recent study removed



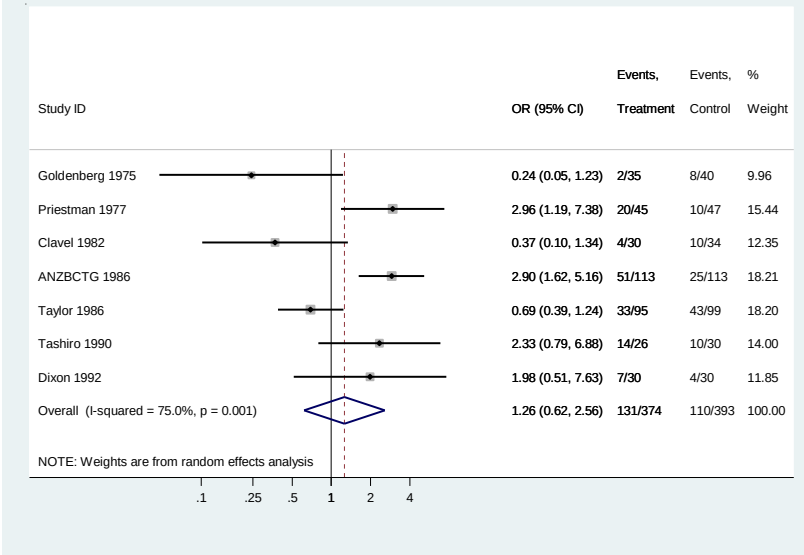
b) Random effect analysis – most recent study removed



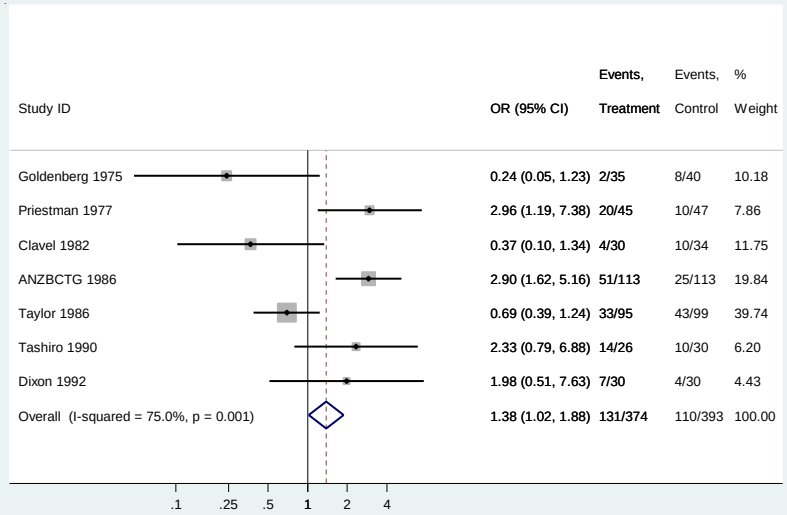
c) Fixed effect analysis – most recent study reinstated



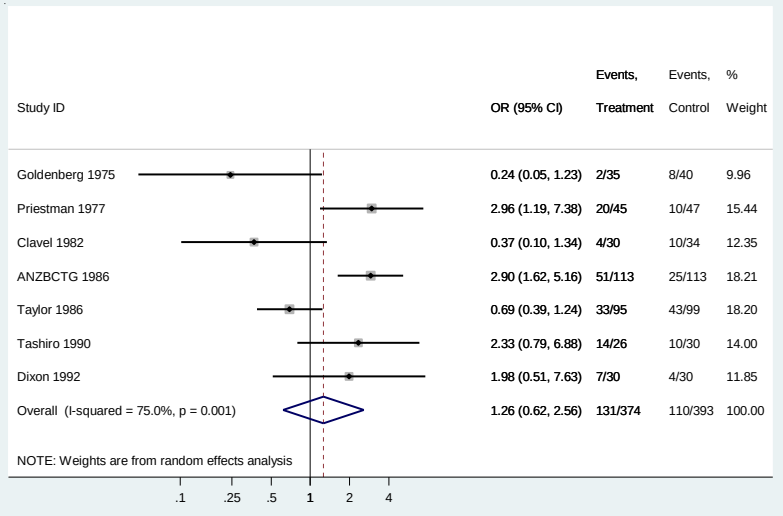
d) Random effect analysis – most recent study reinstated



c) Fixed effect analysis – most recent study reinstated



d) Random effect analysis – most recent study reinstated



References:

1. Egger M, Davey Smith G, Altman.D.G. (Eds.) Systematic revews in health care: Meta-analysis in context. London: BMJ Books; 2000.
2. Shojania KG, Sampson M, Ansari MT, Jun J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. Annals of Internal Medicine. 2007;147:224-33.
3. Moher D, Tsertsvadze A. Systematic reviews: when is an update an update? Lancet. 2006;367:881-3.
4. Moher D, Tsertsvadze A, Tricco A, Eccles M, Grimshaw J, Sampson M, Barrowman N. A systematic review identified few methods and strategies describing when and how to update systematic reviews. Journal of Clinical Epidemiology. 2007;60:1095-104.
5. Lutje V. Closing Cochrane reviews. Cochrane News (http://wwwcochraneorg/newslett/CochraneNews38highrescorrectedpdf);December 2006.
6. Stata. Stata Corp. 2006. Statistical Software: Release 9.0. College Station, TX. Stata Corporation.
7. Barrowman NJ, Fang M, Sampson M, Moher D. Identifying null meta-analyses that are ripe for updating. BMC Medical Research Methodology. 2003.
8. Feiveson AH. Power by simulation. The Stata Journal. 2002;2:107-24.
9. Sutton AJ, Cooper NJ, Jones DR, Lambert PC, Thompson JR, Abrams KR. Evidence-based sample size calculations based upon updated meta-analysis. StatMed. 2007;26:2479-50.
10. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Methods for meta-analysis in medical research. John Wiley: London; 2000.
11. Higgins JPT, Spiegelhalter DJ, Thompson SG. Random effects meta-analysis: important questions and Bayesian solutions. (Submitted)
12. Prasad K, Volmink J, Menon GR. Streroids for treating tuberculosus meningitis. Cochrane Database of Systematic Reviews. 2006;CD002244. PMID:10908535.
13. Murphy C, Hahn S, Volmink J. Reduced osmolarity oral rehydration solution for treating cholera. Cochrane Database of Systematic Reviews. 2006;CD003754. PMID: 15495063.
14. Sirinavin S, Garner P. Antibiotics for treating salmonella gut infections. Cochrane Database of Systematic Reviews. 2006;CD001167. PMID: 10796610.
15. Meremikwu M, Marson AG. Routine anticonvulsants for treating cerebral malaria. Cochrane Database of Systematic Reviews. 2006;CD002152. PMID: 12076440.
16. Samuel M, Ti TY. Interventions for treating melioidosis. Cochrane Database of Systematic Reviews. 2006;CD001263. PMID: 12519552.
17. Lesi A, Meremikwu M. High first dose quinine regimen for treating severe malaria. Cochrane Database of Systematic Reviews. 2006;CD003341. PMID: 15266481.
18. Garner P, G λμεζογλυ AM. Drugs for preventing malaria-related illness in pregnant women and death in the newborn. Cochrane Database of Systematic Reviews. 2006;CD000169. PMID: 12535391.
19. de Bruyn G, Garner P. Mycobacterium vaccae immunotherapy for treating tuberculosis. Cochrane Database of Systematic Reviews. 2006;CD001166. PMID: 12535403.
20. Salinas R, Prasad K. Drugs for treating neurocysticercosis (tapeworm infection of the brain). Cochrane Database of Systematic Reviews. 2006;CD000215. PMID: 10796510.
21. McIntosh HM, Jones KL. Chloroquine or amodiaquine combined with sulfadoxine-pyrimethamine for treating uncomplicated malaria. Cochrane Database of Systematic Reviews. 2006;CD000386. PMID: 16235276.
22. McIntosh HM, Olliaro P. Artemisinin derivatives for treating severe malaria. Cochrane Database of Systematic Reviews. 1998, Issue 2. Art. No.: CD000527.

23. Soares-Weiser K, Goldberg E, Tamimi G, Pitan OC, Leibovici L. Rotavirus vaccine for preventing diarrhoea. Cochrane Database of Systematic Reviews. 2004, issue 1. Art. No.:CD002848.
24. Wilcken N, Hornbuckle J, Ghersi D. Chemotherapy alone versus endocrine therapy alone for metastatic breast cancer. Cochrane Database of Systematic Reviews. 2006;CD002747. PMID: 14651858.
25. Higgins JPT, Thompson SG. Quantifying heterogeneity in meta-analysis. Statistics in Medicine. 2002;21:1539-58.
26. Hardy RJ, Thompson SG. Detecting and describing heterogeneity in meta-analysis. StatMed. 1998;17:841-56.
27. Sterne JAC, Davey Smith G, Cox DR. Sifting the evidence - what's wrong with significance tests? Another comment on the role of statistical methods. BMJ. 2001;322:226-31.
28. Ades AE, Lu G, Claxton K. Expected value of sample information calculations in medical decision modeling. Medical Decision Making. 2004;24:207-27.
29. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. BrMedJ. 1997;315:640-5.
30. Hardy RJ, Thompson SG. A likelihood approach to meta-analysis with random effects. StatMed. 1996;15:619-29.
31. Biggerstaff BJ, Tweedie RL. Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis. StatMed. 1997;16:753-68.
32. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: A comparative study. Stat Med. 1995;14:2685-99.
33. Van Houwelingen HC, Zwinderman KH, Stijnen T. A bivariate approach to meta-analysis. StatMed. 1993;12:2273-84.
34. Berkey CS, Mosteller F, Lau J, Antman EM. Uncertainty of the time of first significance in random effects cumulative metaanalysis. Controlled Clinical Trials. 1996;17:357-71.
35. Higgins JPT, Whitehead A, Simmonds M. Sequential methods for random-effects meta-analyses. (*Submitted*).