

Research article

Open Access

Identifying null meta-analyses that are ripe for updating

Nicholas J Barrowman^{*1,2,3}, Manchun Fang¹, Margaret Sampson¹ and David Moher^{1,2}

Address: ¹Chalmers Research Group, Children's Hospital of Eastern Ontario Research Institute, Ottawa, Canada, ²Department of Pediatrics, University of Ottawa, Ottawa, Canada and ³School of Mathematics & Statistics, Carleton University, Ottawa, Canada

Email: Nicholas J Barrowman^{*} - nbarrowman@cheo.on.ca; Manchun Fang - mfang@cheo.on.ca; Margaret Sampson - msampson@uottawa.ca; David Moher - dmoher@uottawa.ca

^{*} Corresponding author

Published: 23 July 2003

Received: 03 April 2003

BMC Medical Research Methodology 2003, **3**:13

Accepted: 23 July 2003

This article is available from: <http://www.biomedcentral.com/1471-2288/3/13>

© 2003 Barrowman et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: As an increasingly large number of meta-analyses are published, quantitative methods are needed to help clinicians and systematic review teams determine when meta-analyses are not up to date.

Methods: We propose new methods for determining when non-significant meta-analytic results might be overturned, based on a prediction of the number of participants required in new studies. To guide decision making, we introduce the "*new participant ratio*", the ratio of the actual number of participants in new studies to the predicted number required to obtain statistical significance. A simulation study was conducted to study the performance of our methods and a real meta-analysis provides further evidence.

Results: In our three simulation configurations, our diagnostic test for determining whether a meta-analysis is out of date had sensitivity of 55%, 62%, and 49% with corresponding specificity of 85%, 80%, and 90% respectively.

Conclusions: Simulations suggest that our methods are able to detect out-of-date meta-analyses. These quick and approximate methods show promise for use by systematic review teams to help decide whether to commit the considerable resources required to update a meta-analysis. Further investigation and evaluation of the methods is required before they can be recommended for general use.

Background

Systematic reviews, particularly of randomized trials, have gained an important foothold within the evidence-based movement. One of their advantages is easy to understand: clinicians have little time to read individual reports of randomized trials. A more efficient use of their time is to review the totality of evidence pertaining to the question under consideration. A systematic review meets this need particularly if it is based on up-to-date evidence. How can

it be determined whether the evidence is up to date? Simply examining the publication date of a systematic review is not a dependable guide. Some fields evolve rapidly while others progress at a more gradual pace. Delays in the editorial process of traditional print journals can even render a systematic review outdated from the moment it is published.

While the advantage of a published systematic review clearly starts to diminish as new eligible studies are identified, the precise trajectory of decreasing utility is unknown [1]. We believe that quantitative methods are needed to help clinicians and systematic review teams determine when a systematic review can be considered up to date. In this paper, we address a narrower objective. First, we restrict our attention to systematic reviews that feature quantitative synthesis of evidence, or meta-analysis. Second, we focus only on meta-analyses whose pooled results are not statistically significant ("null" meta-analyses). Such results are not infrequent: estimates from two databases showed 38% [2] and 25% [3] of meta-analytic results were non-significant. Our goal is to identify meta-analyses whose pooled results would become statistically significant if they were updated.

As the body of published systematic reviews grows, the problem of outdatedness is attracting more attention. Several groups require that systematic reviews be kept up to date; for example reviews published in the Cochrane Library must be updated at least every 2 years[4]. However the Cochrane Reviewer's Handbook notes that "How often reviews need updating will vary depending on the production of valid new research evidence." Rather than pre-specifying a fixed updating frequency, Ioannidis et al. [5] proposed "recalculating the results of a cumulative meta-analysis with each new or updated piece of information." However, in addition to raising statistical issues of multiple testing[6], this would require continuous monitoring and evaluation of the literature. The resource implications of either of these approaches could be substantial, and the Cochrane Reviewer's Handbook recommends the establishment, for each review, of "guides addressing when new research evidence is substantive enough to warrant a major update or amendment." A conceptual model for assessing the current validity of clinical practice guidelines was developed by Shekelle et al..[7] Their approach used focused literature searches together with expert input.

Here we propose a quantitative approach based on the notion that it may be possible to predict the amount of additional information required to change a non-significant meta-analytic result to a significant one. We measure this additional information in terms of the number of participants in primary studies. Our prediction can then be used to obtain a "diagnostic test" for out-of-date meta-analyses.

This test presupposes that a primary meta-analytic result can be identified in each systematic review (analogous to a primary analysis in a clinical trial). Published systematic reviews often feature multiple meta-analyses and the

authors do not always make it clear which one is of greatest interest.

We begin by deriving and explaining the prediction and then illustrating it using a meta-analysis of trials of intravenous streptokinase for acute myocardial infarction [8]. Next, we describe how the prediction is used to formulate a diagnostic test. A computer simulation is then used to study the performance of the test in three different configurations. Receiver operator characteristic (ROC) curves are derived to illustrate the results.

Methods

Predicting the number of additional participants required to obtain significance

Consider a meta-analysis involving a total of N participants that yields a pooled estimate $\hat{\theta}$ with associated standard error SE_N . The z -statistic is defined to be the ratio

$$Z = \hat{\theta} / SE_N. \quad (1)$$

When the absolute value of Z exceeds a critical value Z_c (for example $Z_c = 1.96$, corresponding to a two-sided Type I error rate of 5%), the pooled estimate is judged to be statistically significant. Failure to obtain statistical significance may be due to the absence of any clinically meaningful effect or to a Type II error resulting from limited power. Here we focus on the latter case, namely pooled estimates that are non-significant due to limited power.

Although the power of a meta-analysis may depend on several factors, here we focus on sample size alone. Suppose that a total of N participants were included in the meta-analysis and that $|Z| < Z_c$. Now suppose that studies involving an additional n participants subsequently become available. Assuming there is no secular trend in the results, how large a value of n would be needed to obtain a statistically significant pooled estimate? In the appendix, we derive the following prediction, by extrapolating the results of the original meta-analysis:

$$n = N(Z_c^2 / Z^2 - 1). \quad (2)$$

Note that the closer $|Z|$ is to Z_c , the smaller the predicted n . As an illustration, consider Lau *et al.*'s cumulative meta-analysis of trials of intravenous streptokinase for acute myocardial infarction [8]. To definitively establish statistical significance, we will fix the probability of Type-I error (α) at 1%. If a meta-analysis of the first five trials (up to and including the second trial published in 1971) had been performed, the result would not have been

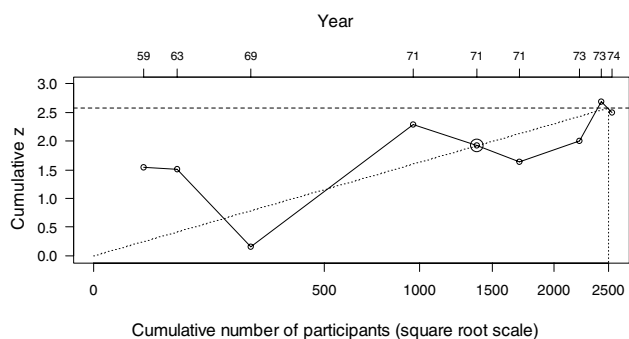


Figure 1
Cumulative meta-analysis of 9 trials of intravenous streptokinase for acute myocardial infarction [8] showing the application of the prediction (equation 2) using the results of the first 5 trials. The cumulative z-statistic is plotted versus the cumulative number of participants, shown on a square root scale (see Appendix). The horizontal dashed line indicates the two-sided critical value of the z-statistic at $\alpha = 0.01$. The diagonal dotted line emanates from the origin and passes through the point representing the meta-analysis of the first 5 trials (double circle). The location where this line intersects the horizontal dashed line is marked by a vertical dotted line projecting down to the horizontal axis and indicating the predicted number of participants required to obtain statistical significance.

statistically significant. Figure 1 illustrates the application of the prediction following this meta-analysis.

One way to evaluate the quality of the prediction is to compare it with the total number of participants when statistical significance was actually obtained. This is not a perfect gauge because of the discrete sample sizes of the trials, but it can still provide insight. In Figure 1, we used the middle of three trials published in 1971 as the last trial in a hypothetical "original" meta-analysis. In that case, the predicted number of participants required to obtain significance (2506) is very close to the number when significance was actually obtained (2432). In fact, the marginal impact of each of these 1971 trials on the pooled estimate was substantial, which influences the predicted number of cases needed to demonstrate an effect. When the first of the three 1971 trials is instead used as the last trial in the original meta-analysis, the prediction (1232) substantially underestimates the number of participants required. On the other hand, when the last of the 1971 trials is used as the last trial in the original meta-analysis, the prediction (4230) is a substantial overestimate.

A diagnostic test

Suppose some time has elapsed since the completion of a meta-analysis (again, assumed to have shown a non-significant pooled result), and now additional evidence is available. Suppose the results of additional studies of the same clinical question are now available, involving a total of m participants. Is the meta-analysis out of date?

If the meta-analysis were in fact updated and the pooled result were statistically significant, the earlier meta-analysis could be deemed "out of date". Conversely, if the updated result were to remain non-significant, the earlier meta-analysis could be deemed "not out of date". Note that these designations should be understood as referring only to the narrow issue of statistical significance.

A simple decision is to "diagnose" the meta-analysis as being out of date if and only if $m > n$, or equivalently if and only if $m/n > 1$. For convenience we refer to m/n as the "new participant ratio", representing the ratio of the observed number of additional participants to the predicted number of additional participants to obtain statistical significance. Thus, the new participant ratio being greater than 1 provides a diagnostic test for whether a meta-analysis is out of date. Like any diagnostic test, false-positive and false-negative errors are possible, as illustrated in Table 1.

The sensitivity of the test is given by $a/(a + c)$, while the specificity is given by $d/(b + d)$. To obtain higher specificity, it is necessary to reduce the frequency of false positives. One way to do this is to tighten the criterion for declaring a meta-analysis to be out of date. For example, a meta-analysis could be diagnosed as out of date if and only if the new participant ratio is greater than q , for some $q > 1$. Conversely, to obtain higher sensitivity, a value $q < 1$ could be chosen. Any desired level of sensitivity (or specificity) could be obtained by choosing an appropriate value of q , with a corresponding trade-off in specificity (or sensitivity). By varying q in this way, an ROC curve[9] is easily constructed. The area under the ROC curve[10] can be used to gauge the performance of the test.

A computer simulation
 To assess the performance of the test, a series of simulations was conducted. The individual studies in the simulation were taken to be RCTs with binary outcomes. For simplicity, the log odds ratio was used to measure intervention effectiveness, and the Mantel-Haenszel estimator provided effect estimates. Three configurations were explored in the simulation study (Table 2), and are explained in detail below.

A computer simulation

Since many meta-analyses involve relatively few trials [11], the maximum number of trials per meta-analysis, denoted M , was taken to be either 10 or 20. Since the goal was to model meta-analyses with non-significant pooled estimates, the size of effects and the number of

Table 1: Diagnostic test result versus true status of a meta-analysis

Test indicates	Meta-analysis is	
	Out of date	Not out of date
Out of date	True positive ^a	False positive ^b
Not out of date	False negative ^c	True negative ^d

Table 2: Simulation configurations

Configuration	Maximum trials per meta-analysis, <i>M</i>	Expected participants per study arm, <i>E</i>	Control arm event rate	Odds ratio
A	10	25	1 – 50%	1/2
B	20	25	1 – 50%	1/2
C	10	25	1 – 50%	2/3

participants per arm were carefully chosen. Relatively strong effects were modelled (odds ratios of 1/2 and 2/3), so the expected number of participants per study arm, denoted *E*, could be relatively small.

For each configuration, the basic simulation procedure was as follows. In each step of the simulation, a number of trials were simulated. Events in the control arm of each trial were simulated using a binomial distribution with event rate sampled from a uniform distribution between 1% and 50%. Events in the treatment arm were simulated from an independent binomial distribution with event rate determined by the event rate in the control arm and the odds ratio. A meta-analysis of the trials was then performed. Our methods require a non-significant meta-analysis as a starting point, therefore if a simulated meta-analysis showed a significant pooled estimate, it was discarded and a new one was simulated. This was continued until a usable starting point was obtained. This process was repeated until a non-significant pooled estimate was obtained. The predicted number of additional participants required to obtain statistical significance, *n*, was then computed and recorded. Additional trials were then simulated and the total number of additional participants, *m*, was computed and recorded. The meta-analysis was then performed again, this time using the complete set of trials, and the statistical significance of the pooled estimate recorded. This was repeated 10000 times for each configuration. All tests of statistical significance were two-sided, since this is customary in meta-analyses, and a significance level of $\alpha = 0.01$ was used.

The number of trials and participants in the simulations were chosen as follows. The number of trials in the original meta-analysis was determined by randomly choosing with equal probability a number between 2 and *M*. The number of participants per study arm was determined by randomly sampling from a Poisson distribution with expected value *E*. The number of additional trials was determined by randomly choosing with equal probability a number between 1 and 1/2 *M*. (In our simulations, *M* was always even.) Again, the number of participants per study arm was determined by randomly sampling from a Poisson distribution with expected value *E*.

Results

By design, the *z*-statistics of the simulated meta-analyses were non-significant before updating. Figure 2 shows their distributions before and after updating.

Of the three simulation configurations, configuration C shows the lowest percentage of significant meta-analyses after updating (11%). This is because the odds ratio (2/3) is close to 1 and the number of trials in the original meta-analysis (10) is small.

One way to gauge the quality of predictions is to compare the new participant ratios of meta-analyses found to be out of date versus those not out of date. Figure 3 displays this comparison for simulation configuration C.

The new participant ratios for the meta-analyses that are out of date are typically larger than those for the meta-analyses that are not out of date. This suggests that the meta-analyses with the largest new participant ratios may

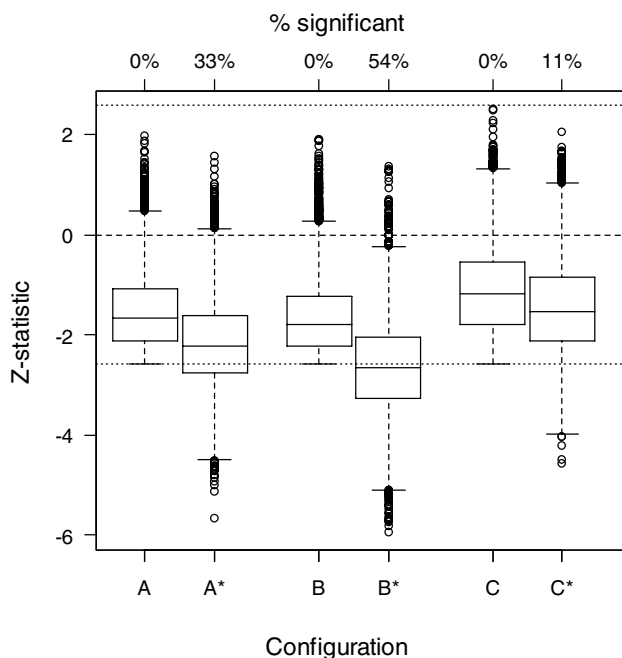


Figure 2
Boxplots showing distribution of z-statistics for each simulation configuration before (A,B,C) and after (A*,B*,C*) updating. The midline of each box marks the median. The bottom and top of the box represent the 25th and 75th percentiles, respectively; thus the height of the box is the interquartile range (IQR). Values that lie more than 1.5 IQRs beyond the top or bottom of the box (outliers) are shown individually. Lines project from the top and bottom of the box to the most extreme values not more than 1.5 IQRs beyond the box.

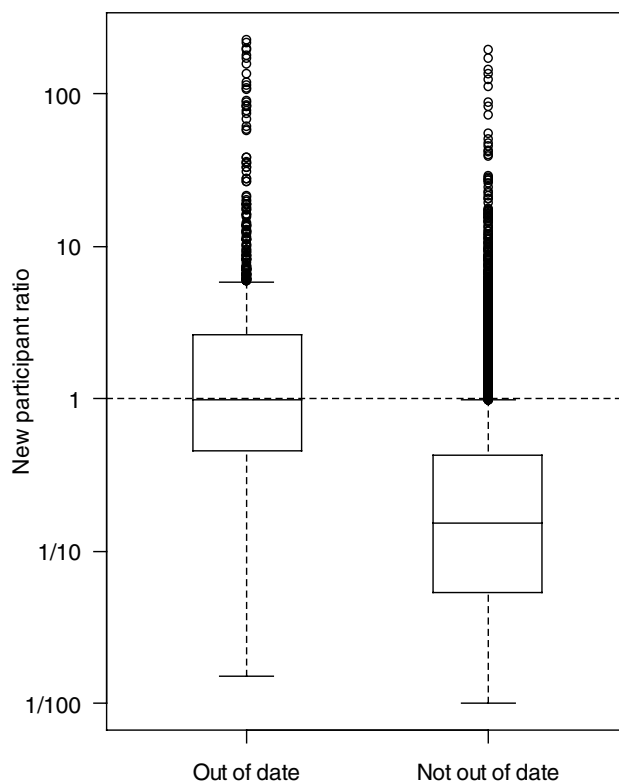


Figure 3
Boxplots comparing the new participant ratio for meta-analyses found to be out of date versus those not out of date in simulation configuration C. See caption of Figure 3 for interpretation of boxplots. Not shown are cases where the new participant ratio is infinite (7 cases out of 10000, 5 [71%] of which were in the out-of-date group) or where the new participant ratio < 1/100 (1361 cases out of 10000, 6 [0.4%] of which were in the out-of-date group).

be the best candidates for updating. Indeed, of the 10 meta-analyses showing the largest new participant ratios in simulation configuration C, 8 were out of date, whereas overall 11% of the meta-analyses in that configuration were out of date.

If the goal is instead to decide whether a given non-significant meta-analysis is up to date, we can diagnose the meta-analysis as being out of date when the new participant ratio is greater than some threshold q , say $q = 1$. The results of this diagnostic test applied to simulation configuration C are shown in Table 3.

The sensitivity in configuration C is thus $521 / (521 + 539) = 49\%$, while the specificity is $8053 / (8053 + 887) = 90\%$. By selecting different thresholds for the new participant ratio, the sensitivity and specificity can be varied, producing an ROC curve for each configuration (Fig. 4). The area

under each ROC curve was computed as a gauge of the performance of the diagnostic test, by numerical integration using 10000 intervals of equal width. The areas were 0.81, 0.80, and 0.85 for configurations A, B, and C respectively. Hosmer and Lemeshow[12] suggest that for areas under the ROC curve between 0.7 and 0.8 the discrimination can be considered acceptable, while for areas between 0.8 and 0.9 the discrimination can be considered excellent.

Discussion
 The determination that a meta-analysis is not up to date or the decision to update a meta-analysis may be based on a variety of considerations. We believe that one of these should be the quantity of evidence that is newly available.

Table 3: Diagnostic test result with threshold $q = 1$ versus true status of meta-analyses in simulation configuration C.

Test indicates	Meta-analysis is	
	Out of date	Not out of date
Out of date	521	887
Not out of date	539	8053

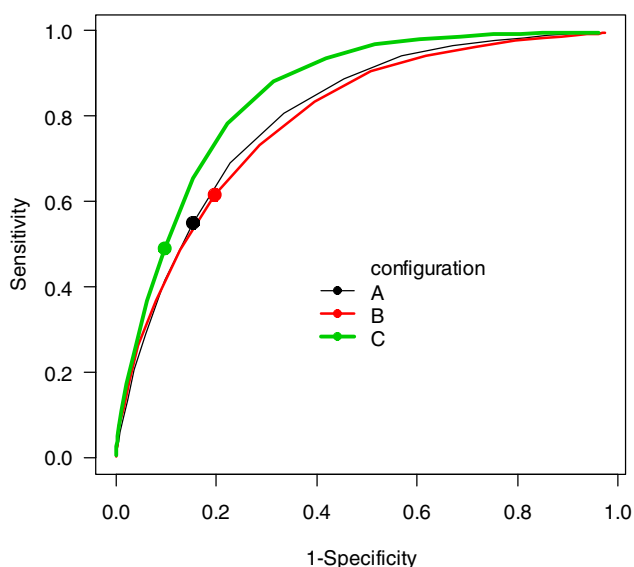


Figure 4
ROC curves for the diagnostic test for each configuration. The dot on each curve indicates the sensitivity and 1-specificity when the threshold for the new participant ratio is $q = 1$. For configuration A, when $q = 1$, the sensitivity is 55% and the specificity is 85%. For configuration B these values are 62% and 80%, and for configuration C they are 49% and 90%. Increasing q corresponds to moving counter-clockwise along the curve. Note that a Bernoulli random decision with probability p that a review is out of date has sensitivity p and specificity $1 - p$, corresponding to an ROC curve consisting of a 45-degree line. ROC curves above the 45-degree line indicate performance superior to chance.

The prediction and diagnostic rule developed here use a metric of the quantity of new information relative to the information previously synthesized. This allows for the fact that some fields evolve at an extremely rapid pace while others progress more gradually.

Our methods depend on very limited input, requiring only the number of participants included in the original meta-analysis, the z-statistic for the pooled estimate, and

the number of additional participants in new studies. Furthermore, these quantities are easily obtained. The z-statistic is often reported or is easily computed from reported quantities such as p -values, while the numbers of participants are often available in abstracts. It may also be possible to extract the numbers of participants automatically using pattern recognition software. An example of automatic identification and extraction of numerical values from text is the use of natural language queries to extract drug dosage information [13]. Another strength of our methods is that z-statistics are easily computed regardless of the measure of effectiveness used in the meta-analysis (e.g. odds ratio, standardized mean difference).

Rather than arbitrarily specifying a fixed frequency for updating meta-analyses, our methods provide empirical estimates of when meta-analyses need updating. Rosenthal[14] proposed that when meta-analysts obtain a statistically significant pooled estimate they also report "the tolerance for future null results" in the form of the number of non-significant *studies* that would be required to overturn their result. In this spirit, our prediction of the number of additional *participants* required to obtain statistical significance could be incorporated in the report of a meta-analysis with a non-significant pooled estimate as guidance for future studies or updates. Unlike approaches based on expert input, our methods are purely quantitative. Application of our methods may also help to avoid multiple testing issues associated with continuous updating of meta-analyses.

Simulations suggest that our methods are able to detect meta-analyses that would become significant if they were updated. In our three simulation configurations, our diagnostic test had sensitivity of 55%, 62%, and 49% with corresponding specificity of 85%, 80%, and 90% respectively. The test is easily modified to obtain higher sensitivity or specificity.

The new participant ratio can be used to identify meta-analyses that are the best candidates for updating. In one simulation, of the 10 meta-analyses showing the largest new participant ratios, 8 were out of date, whereas overall 11% of the meta-analyses were out of date. Given a

collection of non-significant meta-analyses, a practical strategy might be to update the meta-analysis with the largest new participant ratio first, then move to the meta-analysis with the next largest new participant ratio, and so forth.

For health policy decision makers, such as the National Health Service, faced with having to make decisions with competition for diminishing budgets, our approach might help indicate which meta-analyses are more 'informative' in terms of what an updated meta-analysis would contribute.

Our methods show promise as a rough gauge of whether a meta-analysis is up to date. If a quick literature search uncovers several apparently relevant studies published subsequent to a meta-analysis, one might have reason to doubt the currency of the review. Calculation of the "new participant ratio" introduced here can make such judgments more precise and objective. Based on our results, a new participant ratio greater than 5 may be taken as strong evidence that a meta-analysis is out of date; conversely, a new participant ratio less than 1/5 suggests that the weight of evidence accumulated since the meta-analysis was published is unlikely to overturn its findings. Such judgments are, of course, conditional on the thoroughness of the literature search and relevance assessment of any additional studies. For example, a large new participant ratio could be misleading if, upon closer examination, several of the new studies were not in fact relevant to the clinical question.

Systematic review groups may choose to conduct a thorough literature search and relevance evaluation even before committing to update a review. In such cases, less stringent thresholds for the new participant ratio may be acceptable. For example, a new participant ratio greater than 2 could be taken as strong evidence that a meta-analysis needs to be updated; conversely, a new participant ratio less than 1/2 would suggest that the meta-analysis does not need to be updated.

An alternative approach to our method is possible using considerations of clinical significance. Equation (9) in the Appendix could be modified to use an effect judged to be clinically important rather than the effect observed in the original meta-analysis. This would lead to a different prediction of the number of additional participants required to obtain statistical significance. A meta-analyst might wish to perform an update when this number of participants had accrued.

It is important to note that our methods have some limitations. First, they are based on the assumption that the variance of the pooled estimates shrinks at a rate inversely

proportional to the total number of participants in all studies. Since the variance of pooled estimates is in general determined not only by sample size, but also by factors such as baseline risk and statistical heterogeneity, this assumption may not hold exactly. Second, our methods are based on the assumption that there is no secular trend in the results. In particular cases this assumption may be violated. Third, particularly when the number of participants in the studies included in the original meta-analysis is small, our methods may give highly variable results. Fourth, while simulations suggest that our diagnostic test has good overall performance at determining whether an update is warranted for a given meta-analysis, it is not clear what an appropriate choice of the q parameter would be. If the true shape of the ROC curve were known then q could be selected to obtain a desired level of sensitivity or specificity. Finally, a real-world decision to update a meta-analysis would likely be guided not only by the quantitative methods proposed here, but also by clinical and scientific considerations.

It might be argued that our methods place too much emphasis on the narrow issue of statistical significance. Indeed, there has recently been a shift away from hypothesis testing towards estimation, with an accompanying change of emphasis from p -values to confidence intervals[15]. Given large enough sample sizes, it is always possible to obtain statistical significance. Nevertheless, acceptance of new therapeutic interventions continues to depend at least in part on statistical significance, and we believe that this is not without merit. A logical next step following from our methods would be a framework for incorporating considerations of clinical significance into the decision-making process.

Conclusions

This work has focused on meta-analyses that fail to show a significant effect, but would do so if updated. The methods we have developed show promise for use by individual clinicians to help determine whether meta-analyses are up to date, and for use by systematic review teams to help decide whether to commit the considerable resources required to update a meta-analysis. Further investigation and evaluation of the methods is required before they can be recommended for general use.

Competing interests

None declared.

Authors' contributions

MS initially conceived of the study and NJB derived the prediction rule. NJB and MS secured funding for the project. NJB, MF, MS, and DM participated in the design of the study. NJB coordinated the project and wrote all manuscript drafts. DM reviewed simulation results, and

participated in the drafting and revision of the manuscript. MS participated in the drafting and revision of the manuscript. MF wrote the simulation and data analysis programs and produced the figures. All authors read and approved the final manuscript.

Appendix

Recall that the standard error of a sample mean is inversely proportional to the square root of the sample size. Many estimators have the same property, and we will therefore suppose that the standard error SE_N of the pooled estimate $\hat{\theta}$ is inversely proportional to the square root of the number of participants, N . Say for some $\sigma > 0$,

$$SE_N = \sigma / \sqrt{N}. \quad (5)$$

While this has the same form as the standard error of a sample mean where σ represents the standard deviation, it should be noted that here σ is simply a constant of proportionality. From equations (1) and (5),

$$\sigma = \sqrt{N} \hat{\theta} / Z. \quad (6)$$

Suppose additional studies become available yielding n additional participants. From equation (5), the standard error of the pooled estimate based on all $N + n$ participants is

$$SE_{N+n} = \sigma / \sqrt{N+n}. \quad (7)$$

Substituting equation (6) into equation (7) gives

$$SE_{N+n} = \frac{\sqrt{N} \hat{\theta}}{\sqrt{N+n} Z}. \quad (8)$$

Furthermore, suppose there is no secular trend in the results, i.e. the pooled estimate based on all $N + n$ participants remains the same, but now it just reaches statistical significance, i.e.

$$\left| \hat{\theta} / SE_{N+n} \right| = Z_c. \quad (9)$$

Substituting equation (8) into equation (9) gives

$$Z_c = \sqrt{N+n} \frac{|Z|}{\sqrt{N}} \quad (10)$$

Solving for n gives

$$n = N(Z_c^2 / Z^2 - 1). \quad (11)$$

This provides a prediction of the number of additional participants required to warrant updating a meta-analysis.

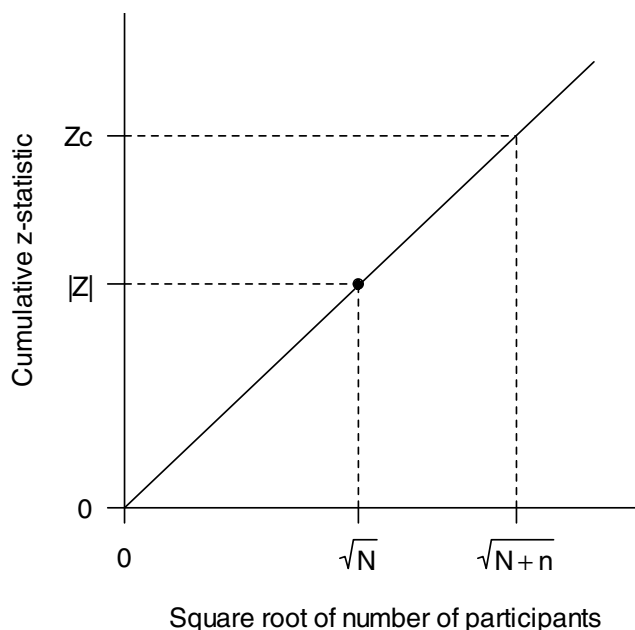


Figure 5
Geometrical illustration of the prediction. The number of additional participants, n , is determined by extrapolating a line segment starting at the origin (0,0) and passing through the point $(\sqrt{N}, |Z|)$ to the point where it intersects with a horizontal line at Z_c .

Note that the prediction depends only on N (the number of participants in the original meta-analysis), Z (the z-statistic from the original meta-analysis), and Z_c (the critical z-value chosen for statistical significance). Crucially, note that the results of additional trials are *not* required.

The form in which Z enters equation (11) has important consequences. First, note that as $Z \rightarrow 0$ from either the right or the left, $n \rightarrow \infty$, and strictly speaking when $Z = 0$, n is not defined. With discrete data (e.g. data from randomized controlled trials with binary outcomes), a pooled z-statistic of zero can occur with non-zero probability. For convenience, we therefore define $n = \infty$ when $Z = 0$. Second, note that since Z enters equation (11) as Z^2 , the sign of Z is irrelevant. In a meta-analysis of randomized controlled trials, for example, equation (11) gives the same prediction regardless of whether the results favour the intervention or control group. It is therefore not necessary to interpret the sign of the z-statistic.

Geometrical interpretation

A geometrical interpretation provides a more intuitive way to understand the prediction. Equation (10) is a linear equation in $\sqrt{N+n}$ with slope $|Z| / \sqrt{N}$ and intercept 0. In other words, on a graph of the cumulative z-sta-

tistic versus the cumulative sample size shown on a square-root scale, the total number of participants predicted is the intersection of a horizontal line at Z_c with a line that passes through the origin (0,0) and the point defined by the original meta-analysis (Fig. 5). From equation (6), the slope of the line can be expressed as $\hat{\theta} / \sigma$. The straight-line extrapolation reflects the assumptions that both the pooled estimate, $\hat{\theta}$, and the constant of proportionality for the standard error, σ , remain unchanged as the number of participants increases.

Acknowledgements

We gratefully acknowledge the helpful feedback provided by Paul Shekelle, Howard Schachter, and Robert Platt. We are also grateful to Khaled El Emam, who contributed to the conceptualization of the project. The CHEO Research Institute and the Natural Sciences and Engineering Research Council of Canada provided financial support.

References

1. Stead LF, Lancaster T and Silagy CA: **Updating a systematic review – what difference did it make? Case study of nicotine replacement therapy** *BMC Med Res Methodol* 2001, **1**:10.
2. McAuley L, Pham B, Tugwell P and Moher D: **Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses?** *Lancet* 2000, **356**:1228-1231.
3. Sampson M, Barrowman NJ, Moher D, Klassen TP, Pham B, Platt R, St. John PD, Viola R and Raina P: **Should systematic reviewers search Embase in addition to Medline?** *Journal of Clinical Epidemiology* 2003.
4. **Cochrane Reviewers Handbook 4.1.6 [updated January 2003]** In *The Cochrane Library Issue 1* Edited by: Clarke M, Oxman AD. Oxford: Update Software; 2003.
5. Ioannidis JP, Contopoulos-Ioannidis DG and Lau J: **Recursive cumulative meta-analysis: a diagnostic for the evolution of total randomized evidence from group and individual patient data** *J Clin Epidemiol* 1999, **52**:281-291.
6. Pogue JM and Yusuf S: **Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis** *Control Clin Trials* 1997, **18**:580-593.
7. Shekelle PG, Ortiz E, Rhodes S, Morton SC, Eccles MP, Grimshaw JM and Woolf SH: **Validity of the Agency for Healthcare Research and Quality clinical practice guidelines: how quickly do guidelines become outdated?** *JAMA* 2001, **286**:1461-1467.
8. Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F and Chalmers TC: **Cumulative meta-analysis of therapeutic trials for myocardial infarction** *N Engl J Med* 1992, **327**:248-254.
9. Hanley JA: **Receiver operating characteristic (ROC) methodology: the state of the art** *Crit Rev Diagn Imaging* 1989, **29**:307-335.
10. Hanley JA and McNeil BJ: **The meaning and use of the area under a receiver operating characteristic (ROC) curve** *Radiology* 1982, **143**:29-36.
11. Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, Pham B and Klassen TP: **Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses** *Health Technol Assess* 1999, **3**:i-98.
12. Hosmer DW and Lemeshow S: *Applied Logistic Regression* 2nd edition. New York: Wiley; 2000.
13. Croft WB, Callan JP and Aronow DB: **Effective access to distributed heterogeneous medical text databases** *Medinfo* 1995, **8(Pt 2)**:1719.
14. Rosenthal R: **The "file drawer problem" and tolerance for null results** *Psychological Bulletin* 1979, **86**:638-641.
15. Gardner MJ and Altman DG: **Confidence intervals rather than P values: estimation rather than hypothesis testing** *Br Med J (Clin Res Ed)* 1986, **292**:746-750.

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2288/3/13/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

