ELSEVIER

Journal of Clinical Epidemiology 65 (2012) 511−519

Journal of Clinical Epidemiology

Graphical augmentations to the funnel plot assess the impact of additional evidence on a meta-analysis

Dean Langan[a,*], Julian P.T. Higgins[b], Walter Gregory[a], Alexander J. Sutton[c]

[a]*Clinical Trials Research Unit (CTRU), University of Leeds, 71-75 Clarendon Road, Leeds, West Yorkshire, LS2 9JT, UK*
[b]*MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK*
[c]*Department of Health Sciences, University of Leicester, UK*

Accepted 25 October 2011; Published online 18 February 2012

Abstract

Objective: We aim to illustrate the potential impact of a new study on a meta-analysis, which gives an indication of the robustness of the meta-analysis.

Study Design and Setting: A number of augmentations are proposed to one of the most widely used of graphical displays, the funnel plot. Namely, 1) statistical significance contours, which define regions of the funnel plot in which a new study would have to be located to change the statistical significance of the meta-analysis; and 2) heterogeneity contours, which show how a new study would affect the extent of heterogeneity in a given meta-analysis. Several other features are also described, and the use of multiple features simultaneously is considered.

Results: The statistical significance contours suggest that one additional study, no matter how large, may have a very limited impact on the statistical significance of a meta-analysis. The heterogeneity contours illustrate that one outlying study can increase the level of heterogeneity dramatically.

Conclusion: The additional features of the funnel plot have applications including 1) informing sample size calculations for the design of future studies eligible for inclusion in the meta-analysis; and 2) informing the updating prioritization of a portfolio of meta-analyses such as those prepared by the Cochrane Collaboration.

Keywords: Study design; Meta-analysis; Graphical display; Evidence-based medicine; Funnel plot; Clinical trial

1. Introduction

Graphical displays have become an integral part of reporting in meta-analysis and can facilitate the communication of important features and results of the associated statistical analysis. For example, the variability of results between studies, often referred to as heterogeneity, or the influence of individual studies on the analysis can be conveyed effectively using graphical means. The reader is referred elsewhere for recent comprehensive reviews and critiques of graphical displays used in meta-analysis [1,2].

This article proposes augmentations to one of the most widely used of graphical displays, the funnel plot [3]. In particular, we describe several novel overlays to the funnel plot, which provide a visual illustration of the impact that new studies would have on a given meta-analysis. We argue that the additional features may help to 1) establish the current robustness of a meta-analysis; 2) inform sample size calculations for the design of future studies that might be added to the meta-analysis [4]; and 3) help decide from of a portfolio of meta-analyses (such as those managed by Review Groups within the Cochrane Collaboration) which should be prioritized for updating [5].

A funnel plot is simply a scatter plot of each study's effect estimate (usually on the *x*-axis) against some measure of the precision of the effect (usually on the *y*-axis). They were first used by Light and Pillemer [3] to detect publication bias. Because larger studies typically have a more precise estimate of effect, theoretically there should be less variability between such estimates than those from less precise estimates from smaller studies, which are located lower down the plot. In the absence of bias and heterogeneity, the plot should therefore appear funnel shaped with greatest variability at the bottom and least variability at the top.

If publication bias is present, the plot can appear asymmetric because it is often assumed that small studies with

* Corresponding author. Tel.: +44-113-343-1493.
E-mail address: D.P.Langan@leeds.ac.uk (D. Langan).

0895-4356/$ - see front matter
doi: 10.1016/j.jclinepi.2011.10.009

What is new?

- Augmentations to the funnel plot can illustrate and help to assess the potential impact of a new study on a meta-analysis.
- The funnel plot was previously only considered for illustrating bias and between-study heterogeneity.
- The presented funnel plot augmentations will help:
  i. Meta-analysts considering how robust current conclusions are to the inclusion of future evidence from a further study;
  ii. Primary researchers to assess the impact studies they design could have on an existing meta-analysis; and
  iii. Inform the update prioritization of portfolios of systematic reviews, as required by organizations such as the Cochrane Collaboration.
- Implementation of the funnel plot augmentations is made available via downloadable code for the "R" statistics package.

negative effects are suppressed by such mechanisms (e.g., see Fig. 1A). However, several factors other than publication bias can lead to a funnel plot appearing asymmetrical [6]. If within-study biases are more likely in smaller studies than in larger studies, then asymmetry can occur, for example, if points lower down the plot are shifted in a particular direction. Furthermore, if heterogeneity is present, and is caused by factors that are correlated with precision, then asymmetry will occur in the funnel plot. In practice, it is difficult to distinguish between these potential reasons for funnel plot asymmetry, or indeed to distinguish any of them from chance. A discussion of the most appropriate scales for both axes of a funnel plot is available elsewhere [7]; we will plot the standard error (SE) on the vertical scale, with SE = 0 at the top. For a fuller history of the origins of funnel plots, see Sterne et al. [8].

In Section 2, we review previous augmentations to the funnel plot and their uses to assess publication biases and heterogeneity. Then, in Section 3, we present our novel overlays and apply them to illustrative examples from the perspective of the meta-analyst assessing the robustness of the conclusions to further evidence. In Section 4, we consider the application of the overlays from the perspective of someone designing a study to assess the impact of that study on the existing evidence base. Section 5, the discussion, concludes the paper.

2. Existing augmentations to funnel plots

Some general features have been proposed to be added to the funnel plot. We illustrate some in Fig. 1A, B using a data set derived from a fixed-effect meta-analysis of studies looking at whether rapid smoking is effective for quitting smoking in terms of abstinence at long-term follow-up [9]. Commonly included are the line of no effect (the thin vertical line in Fig. 1A), and the summary effect from the meta-analysis (the bold vertical line in Fig. 1A). Additionally, a pseudo confidence interval can be used to indicate the region within which we would expect 95% of studies to lie if the studies are all estimating the same underlying effect [10] (the dotted sloping lines in Fig. 1A). Such boundaries are useful for assessing the presence of heterogeneity in a meta-analysis data set because in the presence of heterogeneity less than 95% of the studies

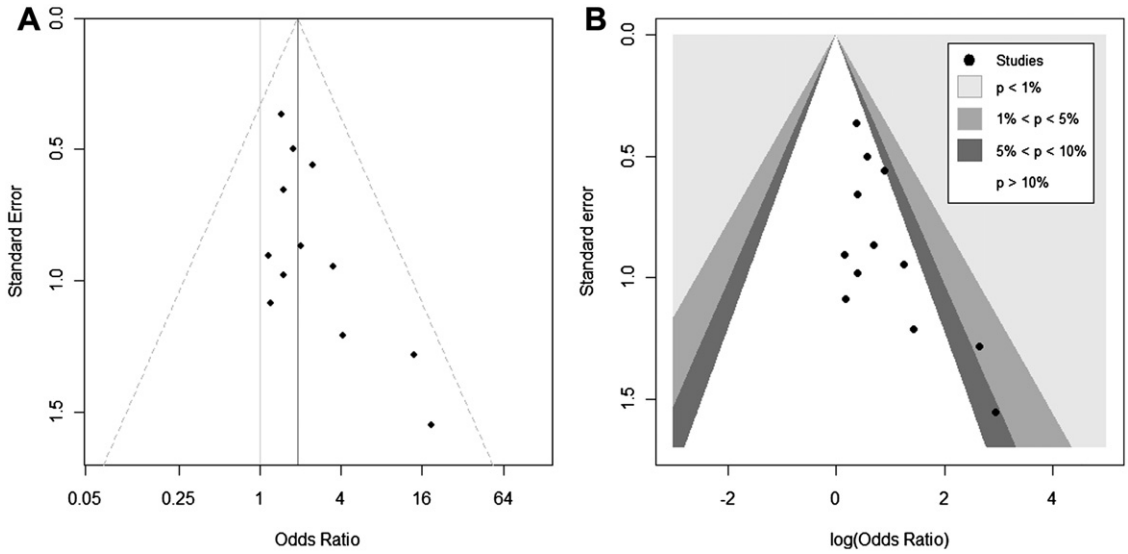

Fig. 1. A. Funnel plot including line of no effect, summary effect and pseudo 95% confidence interval derived from a review of rapid smoking for quitting smoking [9]. B. Contour enhanced funnel plot derived from a review of rapid smoking for quitting smoking [9].

would be inside the inverted triangle. The feature has since been widely used in meta-analysis [11,12] and implemented as a default setting into meta-analysis macros available for statistical packages such as Stata and R [10,13].

A more recently proposed additional feature is a series of contours of statistical significance for individual studies [14]. These contours, illustrated in Fig. 1B, highlight the regions in which studies would need to be in order to achieve a given level of significance; the contours are usually plotted for a range of traditionally used significance levels defined by *P*-values of 0.01, 0.05, and 0.1. These contours can help distinguish between publication biases and other factors as causes of funnel plot asymmetry. For instance, if the region where studies are perceived to be missing is an area of statistical nonsignificance (indicated by the white area in Fig. 1B) then this adds credence to the possibility that asymmetry is caused by publication bias, whereas if studies are missing in areas of high statistical significance (indicated by the darker regions on Fig. 1B) then publication bias might be a less plausible explanation.

In the next section, we consider further overlaying features, which assess the impact of further evidence on a meta-analysis.

3. Novel graphical methods for assessing the impact of potential future studies on a meta-analysis

We introduce two features that illustrate the potential impact of one additional study on a meta-analysis. Both are contours, which overlay the funnel plot and divide it into multiple regions in which an additional study might be located. Different regions represent different types of impact that the additional study would have on the meta-analysis. The potential impacts we consider are first, changes in statistical significance of the summary effect size and, second, the extent of between-study heterogeneity. The theory behind the contours is explained in Appendices 1 and 2 on the journal's Web site at www.jclinepi.com. We present the ideas through a series of examples, which highlight how such plots can reveal importantly different consequences for different meta-analysis data sets. Graphical features presented in this article are restricted to assessing the impact of one additional study; hence, the approach has limitations if multiple additional studies are of interest. A computer program in R (version 2.12.0; www.r-project.org) has been written to generate all graphs of the types presented in this article, and is available via the R package named extfunnel [15]. A second program is currently being written to produce the plots in Stata (StataCorp. 2011. Stata Statistical Software: Release 12. College Station, TX: StataCorp LP).

3.1. Significance contours

Contours of statistical significance show, in an additional study, which combinations of effect size and SE would be required to change or maintain the statistical significance of the summary estimate from the meta-analysis. Note that these contours relate to statistical significance of the meta-analysis, rather than of the individual studies as described in the previous section. The derivation of the statistical significance contours is documented in Appendix 1 on the journal's Web site at www.jclinepi.com.

Fig. 2A presents a fixed-effect meta-analysis of four clinical trials studying a change in the Epworth score for oral appliance vs. continuous positive airways pressure for treating obstructive sleep apnea [16]. The summary effect size represented by the solid diamond on the plot is 0.55 (95% confidence interval [CI]: −0.29, 1.38), implying that the summary result is not statistically significant at the 5% level. If an additional study existed, or were to be performed, that had an effect size and SE that placed it in the central region (nonshaded) of the graph, this would maintain the nonsignificance of the meta-analysis. However, if the additional study were to occupy the top left (light gray) or top right (dark gray) regions of the graph then the meta-analysis would become statistically significant with a difference in favor of the oral appliance or the continuous positive airways pressure, respectively. In this example, we might conclude that the meta-analysis is not robust to the addition of new evidence, because an additional study that would change the statistical significance of the meta-analysis is plausible. The results would seem particularly prone to moving toward a difference favoring the continuous positive airways pressure because one of the four existing trials is located in the area that would produce a statistically significant summary effect in this direction.

A contrasting example, presented in Fig. 2B, is taken from a meta-analysis looking at nonbenzodiazapines for acute lower back pain [17]. The random-effects meta-analysis is currently (just) statistically significant, as can be seen from the diamond at the top of the graph representing the summary risk ratio of 0.55 and 95% CI from 0.33 to 0.91. In this case, the meta-analysis could be considered reasonably robust to a single additional trial (but possibly not multiple trials) because it appears unlikely that a trial would be located in the two nonshaded regions representing a change to statistical nonsignificance: none of the four existing trials lie in this region.

An additional study occupying the (nonshaded) right hand region of Fig. 2B would demonstrate a RR (risk ratio) in the direction opposite to that currently observed in the meta-analysis. Adding this study, which contradicts the current meta-analysis, would change the meta-analysis to nonsignificant. However, an interesting feature of Fig. 2B is the small nonshaded region occupying the upper left portion of the graph. An additional study in this region (although unlikely in practice) has a very small RR in the same direction as the meta-analysis result, but would change the meta-analysis to have a nonsignificant summary RR despite concluding a RR of <1 in isolation. As a general rule,

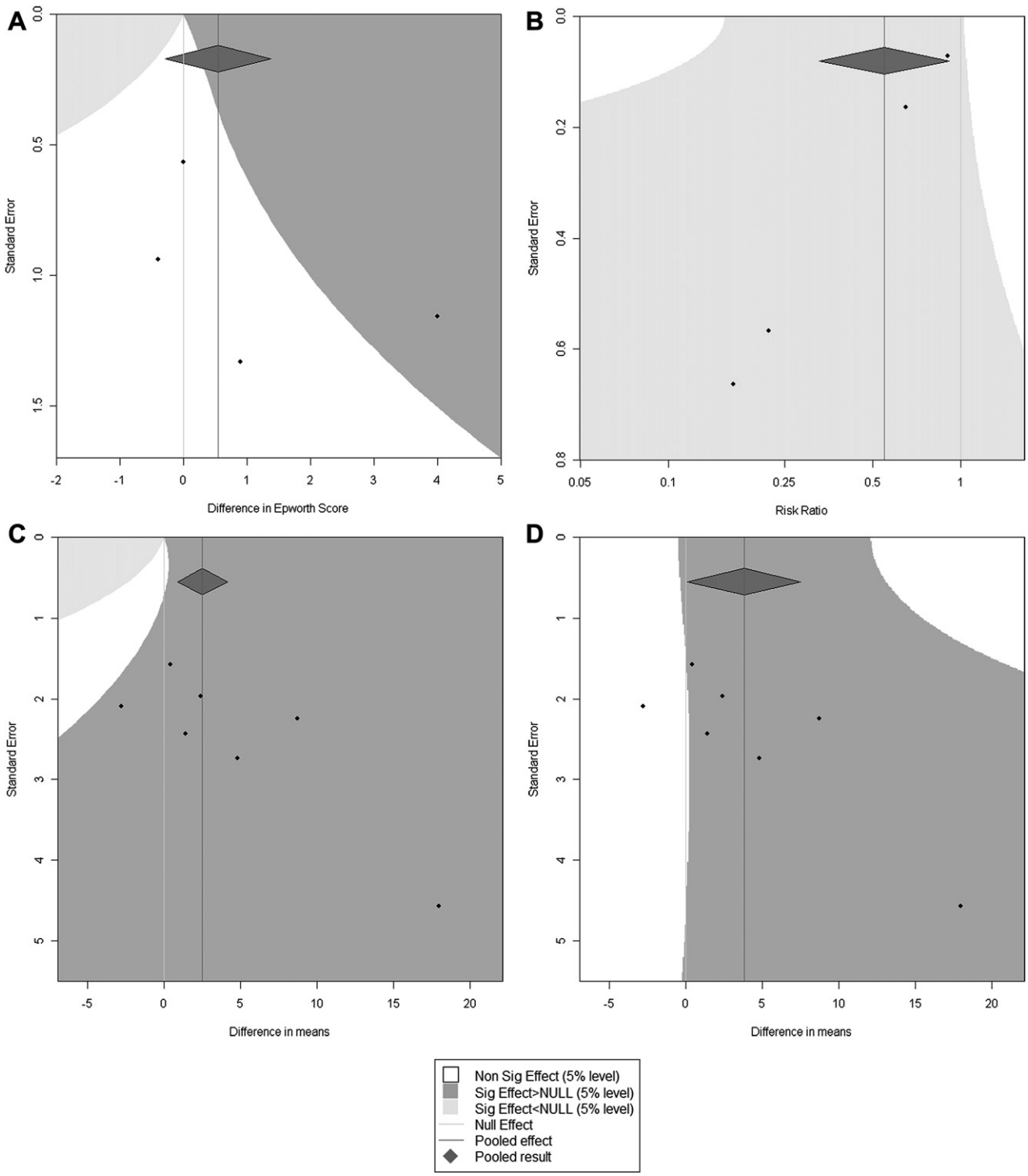

Fig. 2. A. Significance contours for the fixed-effect meta-analysis: oral appliance vs. continuous positive airways pressure with outcome the Epworth score [16]. B. Significance contours for the random-effects meta-analysis: comparison of nonbenzodiazapines vs. placebo for acute low back pain (2–4 days follow-up) [17]. C. Significance contours for the fixed-effect meta-analysis: comparison of kava vs. placebo for anxiety with outcome being improvement in the HAMA score [18]. D. Significance contours for the random-effects meta-analysis: comparison of kava vs. placebo for anxiety with outcome being improvement in the HAMA score [18].

an extreme region exists in either direction where a study would undoubtedly conclude some significant effect in isolation yet change the meta-analysis to become nonsignificant. This is because any shift in the summary estimates away from the null value caused by the extreme study is outweighed by the increase in between-study heterogeneity and subsequent increase in uncertainty around the mean of the random-effects distribution. Such apparently paradoxical findings will not occur under a fixed-effect meta-analysis model.

Fig. 2C, D illustrates a meta-analysis of clinical trials examining the effectiveness of kava for treating anxiety [18]. Fig. 2C shows statistical significance contours under a fixed-effect model, whereas Fig. 2D shows them for a random-effects model. These plots demonstrate that while the current summary estimates both have the same conclusion (i.e., kava is beneficial at the 5% significance level), the overlaid contours can look quite different for the two meta-analysis models. A further trial would have to be quite large under the fixed-effect assumption to change conclusions (i.e., have a SE less than approximately 2) while a trial of any size and a negative treatment difference could potentially change the conclusions of the random-effects model (and such a trial has already been observed). In neither instance would the conclusion appear particularly robust to the inclusion of a further trial. Note also that the nonshaded region in the top right of Fig. 2D shows the same phenomenon as Fig. 2B, that is, that an extreme positive effect in a new trial can make the random-effects meta-analysis nonsignificant.

3.2. Heterogeneity contours

Our second proposed augmentation to the funnel plot is a series of heterogeneity contours that indicate how the extent of between-study heterogeneity would change on the addition of a new study (see Appendix 2 on the journal's Web site at www.jclinepi.com). We consider the influence of the new study on the between-study variance parameter, τ^2, and the I^2 statistic [19]. These contours indicate the robustness of the extent of heterogeneity observed, giving an indication of the extent to which these measures could realistically change on the addition of a further study.

Fig. 3A considers a meta-analysis of Sanchi vs. control as a treatment for ischemic stroke [20]. The summary RR from a fixed-effect model is 0.33 and is represented in the Figure by the vertical solid line. The curved lines either side of this line are contours for particular values of τ^2, and represent the effect estimates and SEs that would be required of a new study for τ^2 to move to these specific values. Mental interpolation between contours can be used to get approximate values for unplotted values of τ^2. The current level of heterogeneity as measured by τ^2 is 0.187, this is represented in the figure by the solid black lines (i.e., second outer most curves on the figure). New studies lying on these contours would not affect the value of τ^2. To decrease the current value for τ^2, an additional study must occupy the region inside of these solid black contours. Similarly, an additional study occupying the region outside of these contours will increase τ^2.

The theoretical lower bound of 0 for τ^2 may not be attainable when heterogeneity is already present in a meta-analysis. The minimum possible value for τ^2 on the addition of a single new study may be obtained when the effect estimate in a new study is equal to the fixed-effect meta-analysis estimate from the existing studies. τ^2 will be minimized if this new study has a very small SE (see Appendix 2.1 on the journal's Web site at www.jclinepi.com). In the example given in Fig. 3A, the minimum τ^2 is 0.017. We plot contours for a range of τ^2 values chosen to span an area of plausible

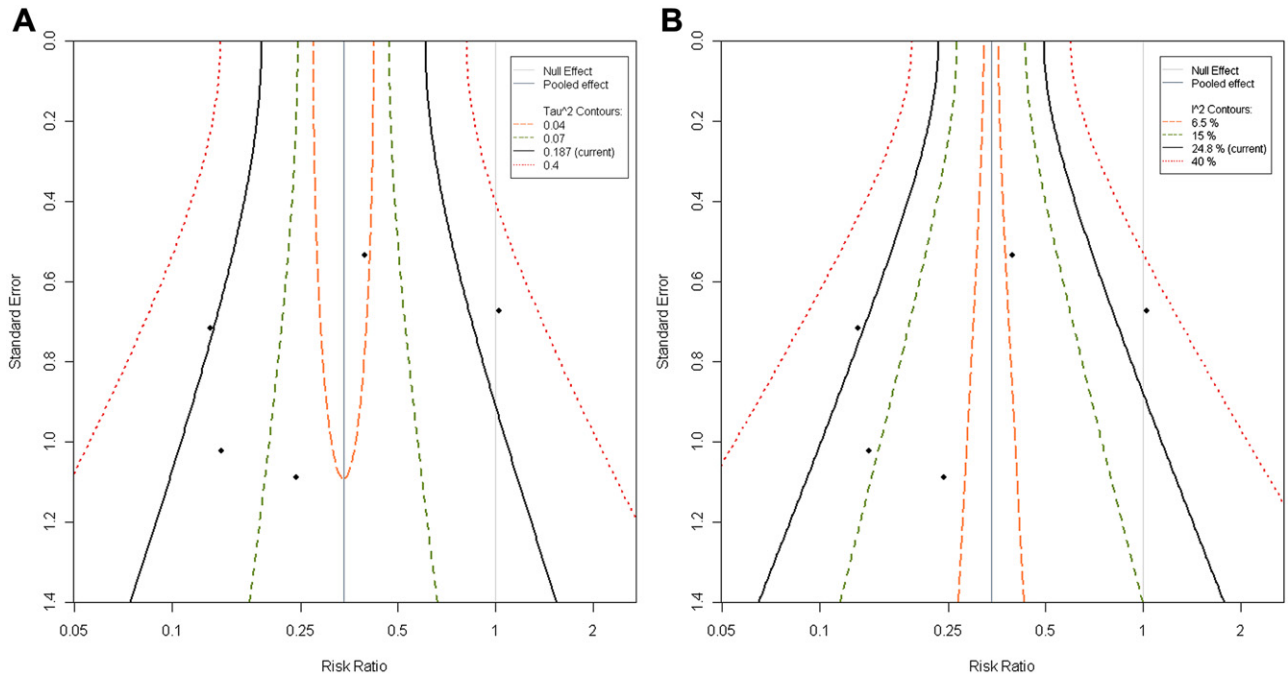

Fig. 3. A. Heterogeneity contours based on τ^2 for the meta-analysis: Sanchi vs. control for acute ischemic stroke with outcome as proportion of patients with no neurological improvement [20]. B. Same heterogeneity contours based on I^2 for the meta-analysis: Sanchi vs. control for acute ischemic stroke with outcome as proportion of patients with no neurological improvement [20].

estimates and SEs for a new study. For example, contours for $\tau^2 = 0.04$, 0.07, 0.187 (the observed τ^2), and 0.4 in Fig. 3A span the range of observed studies.

Fig. 3B illustrates similar contours for the heterogeneity statistic, I^2, using the same data. The value of I^2 for existing studies is 24.8%. The contours have different shapes from those of τ^2. For example, the contour for an I^2 value of 6.5% implies that in order to reduce I^2 to 6.5% a new study must have an effect estimate very close to the current summary RR of 0.33, closer than any other existing study. The SE of the additional study within this region has little impact on how much I^2 decreases. This is unlike the heterogeneity contours based on the τ^2 statistic seen in Fig. 3A. The contour for a τ^2 value of 0.04 implies that in order to reduce τ^2 to 0.04 a new study must have a SE larger than approximately 1.1 (i.e., the turning point for this contour). This matches intuition: larger additional studies should have more impact on the extent of heterogeneity.

4. Uses of the graphical overlays for planning future studies

Our proposed graphical methods can help to illustrate the robustness of an existing meta-analysis to the addition of a further study. Rather than taking the perspective of a meta-analyst summarizing existing evidence, we could instead take the perspective of a primary researcher designing a new study. The relevant question might then be ''What is the potential impact of an additional study I may conduct on the existing evidence base (as represented by the meta-analysis)?'' Specifically, it would be helpful for the researcher to know if their study was likely to change the statistical significance of the meta-analysis, and possibly how much their new study could potentially change the level of between-study heterogeneity. Of course, the interpretation of the plots in the previous section can be directly translated to this perspective. In this section, we consider the above plots for this use, together with some extra plot overlays specifically for this purpose.

Sutton et al. [21] and Goudie et al. [22] have suggested that meta-analysis can be a valuable approach to the design of future clinical trials, also proposing a formal methodical approach to sample size calculation of a trial for the incorporation of the trial into an existing meta-analysis using simulation methods [4]. Here, we illustrate how the plots described in Section 3 can complement the simulation approach described by Sutton et al. [4]. In particular, results from their simulations can be overlaid on the statistical significance contour plots to illustrate statistical power for a given sample size of a future study. Overlaying the simulation results on the heterogeneity contour plots is similarly possible. Following this, further exploratory features informing power and sample size, which will be useful before conducting detailed sample size calculations are considered (in this context, we use power to refer to the probability of changing the statistical significance of the meta-analysis and thus has a slightly different meaning than when used in more traditional context, e.g., the design and analysis of an individual trial on its own).

We consider as an example a meta-analysis of six clinical trials investigating the use of antibiotics vs. control to treat the common cold, specifically with regard to alleviating symptoms by 7 days. Sutton et al. [4] originally used this example to illustrate their simulation approach. A random-effects meta-analysis of the existing studies had a (nonsignificant) summary odds ratio of 0.77 (95% CI: 0.5, 1.21). Fig. 4A illustrates the statistical significance contour plot for this analysis. Only a trial located in the relatively small light gray area toward the top of the plot would produce a statistically significant and beneficial effect of antibiotic.

Sutton et al. [4] noted using their simulation approach that large sample sizes would be required for a new study to have any reasonable power to change the conclusions of the existing meta-analysis. The reasons for this become immediately apparent by looking at Fig. 4A, where it can be seen that no study with a SE greater than 0.2 can change the inferences of the meta-analysis. Thus, a first benefit of this plot is identifying situations like this in which much of the computational burden of a simulation approach can be avoided through the initial scoping of much of the parameter space allowed by the plot.

Fig. 4B presents a vertically stretched version of an area toward the top right of Fig. 4A. This is the critical region, within which a new study could lead to a change in the statistical significance of the meta-analysis. Onto this plot, we have overlaid 300 simulated trials generated from a meta-analytic predictive distribution as described by Sutton et al. [4]. Each of these trials has 2,000 (simulated) patients in each arm. The x-axis spread of these points represents the variability in the effect size, which would be observed in a future trial based on the meta-analysis of the current evidence. The proportion of the trials that are contained within the darker region provides an estimate of power of a future study to change the conclusions of the existing meta-analysis. In this instance, the power is relatively low because a large proportion of points lie outside the dark region. In fact, Sutton et al. observed the power to be 44%. The 95% prediction interval for the underlying treatment effect in a new study is also included in Fig. 4B [23] (i.e., the lines extending out from the pooled estimate ''diamond''). This is relatively wide because of the presence of heterogeneity among the existing trials. Note that the spread of simulated trials is wider than this interval because of the extra variability due to sampling error. The prediction interval can be thought of as the extent of variability in treatment effects that would be observed if sampling error were ignored (or the new studies were all infinitely large).

Fig. 4B demonstrates that power will not increase markedly for trials with sample sizes larger than 2,000 because the region of statistical significance is still relatively ''thin'' for smaller SEs compared with the range of

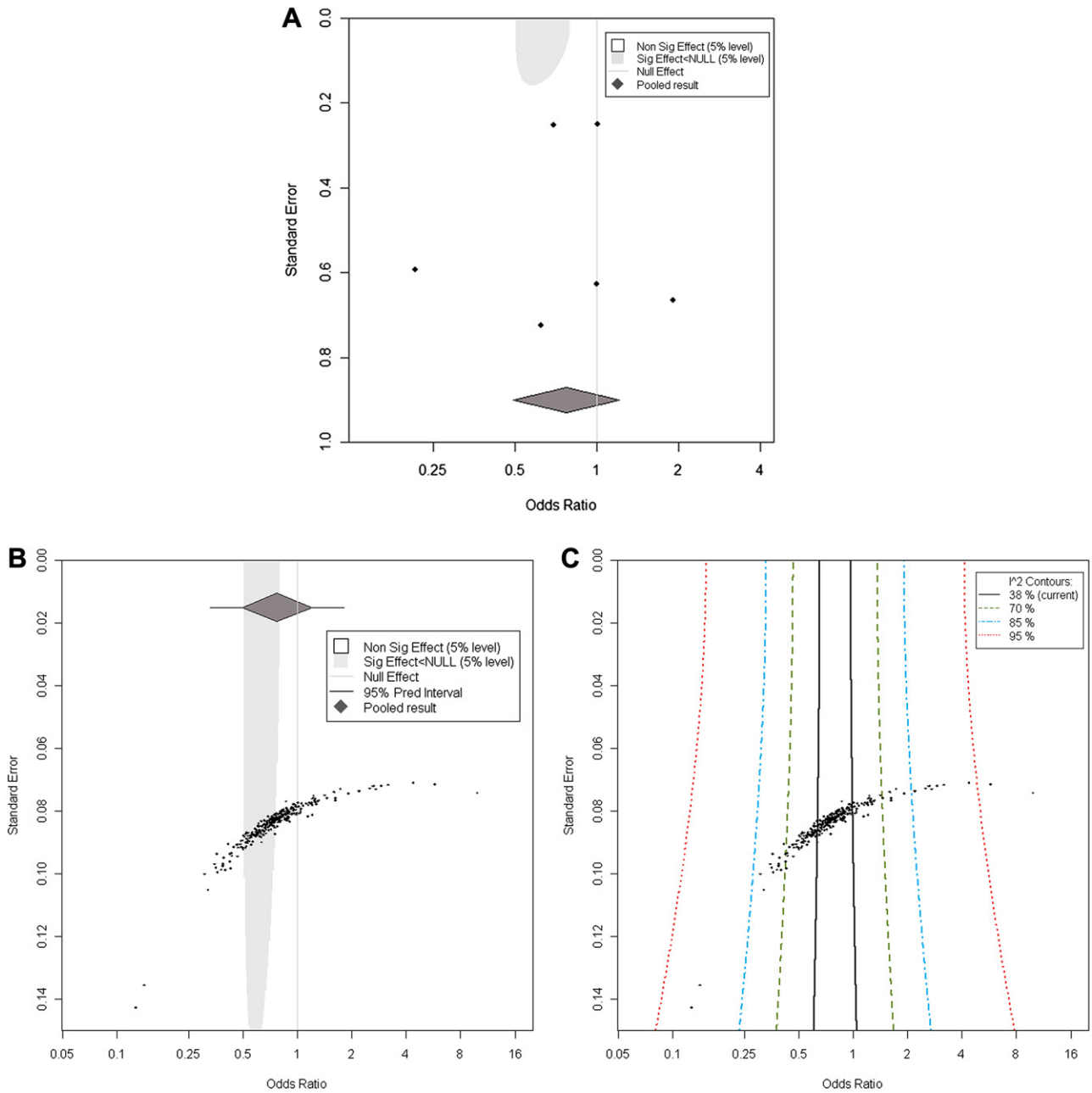

Fig. 4. A. Significance contours for the random-effects meta-analysis: antibiotics vs. control for the common cold to alleviate symptoms by 7 days [4]. B. Significance contours for the same data set as used in Fig. 4A, close up restricting *y*-axis scale range [4]. C. Same data set as Fig. 4A, B, plotting results from a simulated study overlaying the heterogeneity contours [4].

plausible odds ratios in a new trial. Similarly, a large proportion of the prediction interval is not included within the shaded region. Thus, in this example, the power of a new study will be low irrespective of its sample size; a common finding using this approach to sample size estimation.

Fig. 4C plots the same results from the simulation with the heterogeneity contours. This illustrates how the extent of heterogeneity might change given a new study with 2,000 patients in each arm. From the graph, we can see that most of the simulated studies fall within a region with I^2 less than 70%, indicating that the I^2 statistic would likely change from the current value of 38% to something between 25.6% (the minimum value derived elsewhere, see Appendix 2.2 on the journal's Web site at www.jclinepi.com) and 70%. In this case, the change in heterogeneity may not have a profound effect on the analysis strategy and clinical conclusions because heterogeneity is already reasonably high and a random-effects model is currently being used.

Fig. 5 shows results from a meta-analysis looking at an opioid antagonist (naltrexone) compared with placebo in

patients receiving nicotine replacement therapy for smoking cessation (with smoking abstinence as outcome) [24]. This meta-analysis contains three trials and the current fixed-effect meta-analysis estimate is 1.24, which is not statistically significant at the 5% level. The heterogeneity variance and I^2 are both estimated to be 0 for this data set. Fig. 5 presents a funnel plot of these data, including the statistical significance contours described in Section 3.1. Some additional features have been included, which help to assess the potential impact of further studies with varying sample sizes. The red lines, representing a pseudo 95% confidence interval for the meta-analysis under the fixed-effect assumption provide a region in which approximately 95% of new studies might lie if the fixed-effect model were true (as in Fig. 1A). Additional contours have been included to illustrate equivalent sample sizes corresponding to different risk ratio and SE combinations. These are generated assuming the risk of an event in the control arm in a new study is equal to the average of the risks within the existing studies and the allocation ratio between arms is 1:1. Contours are given for sample sizes of 50, 100, 200, 400, and 800 patients per arm. The curvature of these contours is because of the inherent correlation between an estimated risk ratio and its estimated SE.

Collectively, these features help to illustrate how likely a change in statistical significance might be given an additional study with fixed sample size. For example, a sample size of 50 patients in each arm has little chance of confirming that naltrexone is more effective than placebo because the dark gray region indicating a significant benefit for naltrexone falls outside the pseudo 95% confidence interval for this particular sample size. Hence, if the fixed-effect assumption is reasonable, then a trial with sample size 50 patients in each arm will have close to zero power. As the sample size of an additional study increases, the dark gray significance region falls within the pseudo 95% region, corresponding to an increase in power. Based on the graph, a sample size of approximately 800 per arm would be needed to obtain 50% power (i.e., approximately 50% of the pseudo 95% confidence interval is in the shaded area for this sample size). This sample size is much larger than any existing study. In summary, the graph suggests that although a statistically significant meta-analytic result is achievable, only a very large trial would have substantial power. The ability to draw such a conclusion from a single plot means that computationally expensive simulations to calculate power are not required.

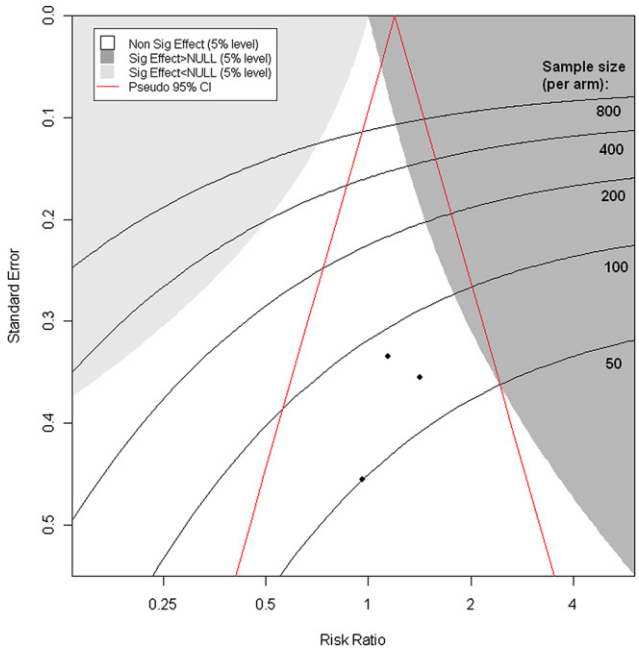

Fig. 5. Significance contours for fixed-effects meta-analysis: opioid antagonists for smoking cessation, naltrexone, and NRT (nicotine replacement therapy) is compared with a placebo and NRT—including additional sample size-related overlays [24].

The sample size calculations here are based on a fixed-effect meta-analysis model. Although it is theoretically possible to produce corresponding pseudo 95% intervals for a random-effects model, these are numerically complex to construct. When there is no within-study error, that is, for a theoretical study of infinite size, the width of the interval will be equal to the 95% prediction interval. Hence, this interval can be used as a guide in a random-effects context, although it ignores sampling error in the new study and hence is narrower than the "true" interval for finite sample sizes.

5. Discussion

We have proposed a flexible framework for illustrating the potential impact of a new study on a meta-analysis through the use of regions and contours on funnel plots. The framework has multiple applications, and choices around which to implement will be specific to the context. In addition to the contexts outlined in Section 4, our graphical ideas can be used to prioritize updates of meta-analyses based on the likelihood of a change in statistical significance. Indeed, the burden of continuously updating portfolios of meta-analyses, as currently undertaken by institutions including the Cochrane Collaboration, has been highlighted [5,25,26]. If statistical significance contour plots such as those in Section 3.1, were produced routinely for such portfolios of meta-analyses, and searches for relevant new studies were kept up to date, then a quick assessment could be made as to whether the new evidence is likely to change the conclusions of the existing review. An update prioritization strategy could be informed by these assessments, making the process of updating reviews more efficient. Therefore, we believe our plots are potentially useful to 1) meta-analysts producing systematic reviews; 2) trialists considering the design of future randomized controlled trials; and 3) editors of systematic review portfolios.

Our proposals have inherent limitations. First, the statistical significance contours do not currently account for a potential change in the statistical model (i.e., between fixed and

random effects) if an additional study increases or (perhaps more rarely) decreases the overall level of heterogeneity. For this reason, we recommend that statistical significance contours under the fixed-effect model are presented with caution; the random-effects model can be considered the more flexible model. Secondly, our proposals cannot currently be generalized to multiple additional studies. Generalizing the features would require one or more assumptions that are unlikely to hold, such as all additional studies having identical effects and SEs. Lastly, the graphical features are built on inverse-variance weighted average methods. Although these are probably the most commonly implemented, other common methods used for binary outcome data, such as Mantel−Haenszel methods or logistic regression, do not fit within the framework of a two-dimensional graph as they take into account the whole array of cell frequencies. Nevertheless, we believe that overlaying contours on funnel plots can provide valuable insights into the robustness and implications of a meta-analysis, particularly with regard to how a new study might impact on it.

Acknowledgments

The authors would like to thank the editor and three anonymous reviewers for their constructive comments, which improved the article.

Julian P.T. Higgins was funded by MRC Grant U.1052.00.011.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at 10.1016/j.jclinepi.2011.10.009.

References

[1] Bax L, Ikeda N, Fukui N, Yaju Y, Harukazu T, Moons KGM. More than numbers: the power of graphs in meta-analysis. Am J Epidemiol 2009;169:249−55. doi: 10.1093/aje/kwn340.

[2] Anzures-Cabrera J, Higgins JPT. Graphical displays for meta-analysis: an overview with suggestions for practice. Res Synthesis Methods 2010;1:66−80. doi: 10.1002/jrsm.6.

[3] Light RJ, Pillemer DB. Summing up. The science of reviewing research. Cambridge, MA: Harvard University Press; 1984.

[4] Sutton AJ, Cooper NJ, Jones DR, Lambert PC, Thompson JR, Abrams KR. Evidence-based sample size calculations based upon meta-analysis. Stat Med 2007;26:2479−500. doi: 10.1002/sim.2704.

[5] Sutton AJ, Donegan S, Takwoingi Y, Garner P, Gamble C, Donald A. An encouraging assessment of methods to inform priorities for updating systematic reviews. J Clin Epidemiol 2009;62:241−51. doi: 10.1016/j.jclinepi.2008.04.005.

[6] Lau J, Ioannidis JPA, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. BMJ 2006;333:597−600. doi: 10.1136/bmj.333.7568.597.

[7] Sterne JAC, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. J Clin Epidemiol 2001;54:1046−55. doi: 10.1016/S0895-4356(01)00377-8.

[8] Sterne JAC, Becker BJ, Egger M. The funnel plot. In: Rothstein HR, Sutton AJ, Borenstein M, editors. Publication bias in meta-analysis—prevention, assessment and adjustments. Chichester: Wiley; 2005.

[9] Hajek P, Stead LF. Aversive smoking for smoking cessation. Cochrane Database Syst Rev 2001;3. doi: 10.1002/14651858.CD000546.pub2. CD000546.

[10] Sterne J, Harbord R. Funnel plots in meta-analysis. Stata J 2004;4:127−41.

[11] Tong JL, Ran ZH, Shen J, Fan GQ, Xiao SD. Association between fecal bile acids and colorectal cancer: a meta-analysis of observational studies. Yonsei Med J 2008;49:792−803. doi: 10.3349/ymj.2008.49.5.792.

[12] Winkley K, Landau S, Eisler I, Ismail K. Psychological interventions to improve glycaemic control in patients with type 1 diabetes: systematic review and meta-analysis of randomised controlled trials. BMJ 2006;333:65−8. doi: 10.1136/bmj.38874.652569.55.

[13] Lumley T. rmeta: Meta-analysis. Available at http://cran.r-project.org/web/packages/rmeta. Accessed February 2, 2011.

[14] Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. J Clin Epidemiol 2008;61:991−6. doi: 10.1016/j.jclinepi.2007.11.010.

[15] Langan D. extfunnel: Additional funnel plot augmentations. Available at http://cran.r-project.org/web/packages/extfunnel. Accessed February 2, 2011.

[16] Lim J, Lasserson TJ, Fleetham J, Wright JJ. Oral appliances for obstructive sleep apnoea. [see figure 5]. Cochrane Database Syst Rev 2006;1. doi: 10.1002/14651858.CD004435.pub3. CD004435.

[17] Van Tulder MW, Touray T, Furlan AD, Solway S, Bouter LM. Muscle relaxants for non-specific low-back pain. [see analysis 2.4]. Cochrane Database Syst Rev 2003;4. doi: 10.1002/14651858.CD004252. CD004252.

[18] Pittler MH, Ernst E. Kava extract versus placebo for treating anxiety. Cochrane Database Syst Rev 2003;1. doi: 10.1002/14651858.CD003383.

[19] Higgins J, Thompson SG, Deeks J, Altman D. Measuring inconsistency in meta-analyses. BMJ 2003;327:557−60.

[20] Chen X, Zhou M, Li Q, Yang J, Zhang Y, Zhang D, et al. Sanchi for acute ischaemic stroke. [see analysis 1.5]. Cochrane Database Syst Rev 2008;4. doi: 10.1002/14651858.CD006305.pub2. CD006305.

[21] Sutton AJ, Cooper NJ, Jones DR. Evidence synthesis as the key to more coherent and efficient research. BMC Med Res Methodol 2009;9:29.

[22] Goudie AC, Sutton AJ, Jones DR, Donald A. Empirical assessment suggests existing evidence could be used more fully in designing randomised controlled trials. J Clin Epidemiol 2010;63:983−91. doi: 10.1016/j.jclinepi.2010.01.022.

[23] Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A Stat Soc 2009;172:137−59. doi: 10.1111/j.1467-985X.2008.00552.x.

[24] David SP, Lancaster T, Stead LF, Evins AE, Cahill K. Opioid antagonists for smoking cessation. Cochrane Database Syst Rev 2006;4. doi: 10.1002/14651858.CD003086.pub2. CD003086.

[25] Moher D, Tsertsvadze A. Systematic reviews: when is an update an update? Lancet 2006;367:881−3. doi: 10.1016/S0140-6736(06)68358-X.

[26] Moher D, Tsertsvadze A, Tricco A, Eccles M, Grimshaw J, Sampson M, et al. A systematic review identified few methods and strategies describing when and how to update systematic reviews. J Clin Epidemiol 2007;60:1095−104. doi: 10.1016/j.jclinepi.2007.03.008.