

## ORIGINAL ARTICLE

# Do trials reduce uncertainty? Assessing impact through cumulative meta-analysis of neonatal RCTs

SC Hay<sup>1,2</sup>, H Kirpalani<sup>3</sup>, C Viner<sup>4</sup>, R Soll<sup>5</sup>, D Dukhovny<sup>6</sup>, W-Y Mao<sup>1</sup>, J Profit<sup>7</sup>, SB DeMauro<sup>3</sup> and JAF Zupancic<sup>1,2</sup>

**OBJECTIVE:** To assess the impact of the latest randomized controlled trial (RCT) to each systematic review (SR) in Cochrane Neonatal Reviews.

**STUDY DESIGN:** We selected meta-analyses reporting the typical point estimate of the risk ratio for the primary outcome of the latest study ( $n = 130$ ), mortality ( $n = 128$ ) and the mean difference for the primary outcome ( $n = 44$ ). We employed cumulative meta-analysis to determine the typical estimate after each trial was added, and then performed multivariable logistic regression to determine factors predictive of study impact.

**RESULTS:** For the stated primary outcome, 18% of latest RCTs failed to narrow the confidence interval (CI), and 55% failed to decrease the CI by  $\geq 20\%$ . Only 8% changed the typical estimate directionality, and 11% caused a change to or from significance. Latest RCTs did not change the typical estimate in 18% of cases, and only 41% changed the typical estimate by at least 10%. The ability to narrow the CI by  $> 20\%$  was negatively associated with the number of previously published RCTs (odds ratio 0.707). Similar results were found in analysis of typical estimates for the outcomes of mortality and mean difference.

**CONCLUSION:** Across a broad range of clinical questions, the latest RCT failed to substantially narrow the CI of the typical estimate, to move the effect estimate or to change its statistical significance in a majority of cases. Investigators and grant peer review committees should consider prioritizing less-studied topics or requiring formal consideration of optimal information size based on extant evidence in power calculations.

*Journal of Perinatology* (2017) **37**, 1215–1219; doi:10.1038/jp.2017.126; published online 7 September 2017

## INTRODUCTION

Systematic reviews (SRs) guide clinical decision making via a structured and comprehensive synthesis of relevant randomized clinical trials (RCTs) studying the effects of a given intervention. Ideally, they provide a quick summary of available evidence, enabling clinicians to choose optimal therapies and researchers to focus their efforts on appropriate new trials.<sup>1,2</sup> Despite a steady stream of published RCTs, however, the answers to many questions posed in the neonatology SR literature remain equivocal. One potential contributor is the extent to which any single new trial can alter the conclusion of the SR to which it is added. Such impact is critical to whether resources for neonatal research are employed optimally.

We used the reviews of the Cochrane Neonatal Reviews Group, one of over 50 collaborative review groups in the Cochrane Collaboration, to assess the effect derived from the most recently added study to a SR. In particular, we aimed to describe the proportion of the most recently added studies that changed the magnitude, precision or statistical significance of the typical estimate of effect within neonatal meta-analyses. In addition, we examined which study characteristics influence the degree of impact on the conclusions of the SR.

## METHODS

We examined all current SRs in the Cochrane Database of Systematic Reviews as of 25 July 2016.<sup>3</sup> We selected for further study two overlapping

groups of SRs—those reporting meta-analytic typical estimates for the stated primary outcome of the most recently added study ( $n = 174$ ) and those reporting meta-analytic typical estimates for mortality ( $n = 128$ ). Where multiple primary outcomes or measures of mortality were reported, we chose for analysis the typical estimates that included the most trials at baseline, those for which the added trial was most recent or those with the narrowest ultimate confidence interval (CI).

For each included review, we generated a cumulative meta-analysis with RevMan,<sup>4</sup> Cochrane Collaboration software used in performing meta-analysis, by sequentially examining each of the included trials in chronological order. As each new trial was added, we calculated and recorded a new typical estimate of effect, summarizing all previously included trials. Each line of the resulting cumulative 'forest plot' thus summarized all available information from randomized trials of that therapy to that date.

We then characterized the impact of the most recently added study (which appeared as the last line of the cumulative forest plot) to each SR by calculating both the change in the typical estimate (in absolute units of relative risk, odds ratio or mean difference), and the change in its CI, between the penultimate and the most recently added RCT.

We employed multivariable logistic regression to determine characteristics of the most recently added study that affected its impact. Dependent variables included measures of precision (20% CI reduction), effect size (10% risk ratio reduction) and alteration in directionality (any change in significance, change to significance, change to insignificance or a 50% change in the z-score). These were changes considered to have potential impact on clinical practice for readers of SRs. Predictor variables included

<sup>1</sup>Department of Neonatology, Beth Israel Deaconess Medical Center, Boston, MA, USA; <sup>2</sup>Division of Newborn Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA; <sup>3</sup>Division of Neonatology, Children's Hospital of Philadelphia, Philadelphia, PA, USA; <sup>4</sup>Department of Obstetrics and Gynecology, New York University, New York, NY, USA; <sup>5</sup>Division of Neonatal Perinatal Medicine, University of Vermont College of Medicine, Burlington, VT, USA; <sup>6</sup>Division of Neonatology, Oregon Health and Science University, Portland, OR, USA and <sup>7</sup>Division of Neonatology, Stanford University, Stanford, CA, USA. Correspondence: Dr JAF Zupancic, Department of Neonatology, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Rose Building Room 318, Boston, MA 02215, USA.  
E-mail: jzupanci@bidmc.harvard.edu

Received 24 January 2017; revised 24 April 2017; accepted 15 May 2017; published online 7 September 2017

publication year, number of preceding included studies, study size and geographic location of the primary author in the United States versus elsewhere.

Finally a subgroup analysis was performed for studies after 1990. We chose this time-point to demarcate the division between the pre- and post-surfactant eras, as this change in practice effected one of the largest improvements in outcomes over the study period.

**RESULTS**

**Selection of reviews for inclusion**

We identified 339 candidate SRs from the Cochrane Database of Systematic Reviews (Figure 1).<sup>3</sup> Of these, 54 were excluded (48 had no studies meeting inclusion criteria of their respective systematic reviews, 5 had no analysis performed because of extreme heterogeneity of the included studies and 1 was retracted because of redundancy). Forest plots from the remaining 285 SRs were replotted as cumulative analyses for the stated primary outcome of the last published study, the outcome of mortality and the stated primary outcome reported as mean difference.

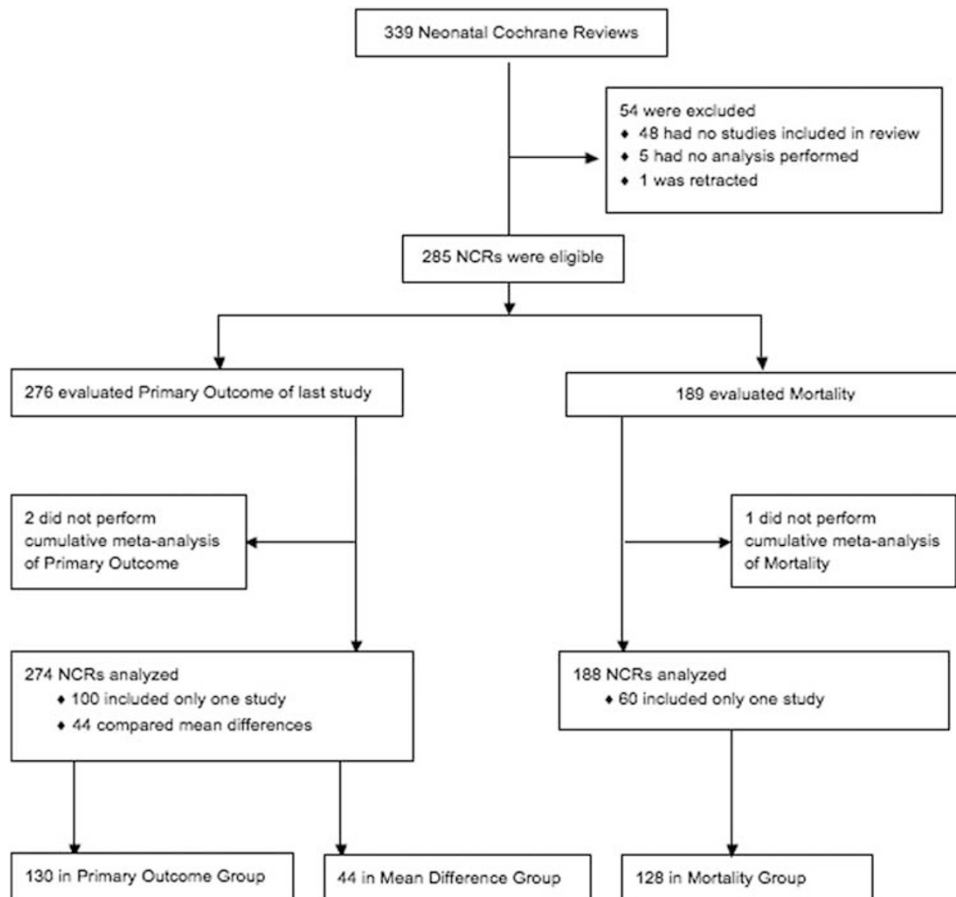
Characteristics of the included comparative SRs are represented in Table 1. Most SR forest plots included  $\leq 5$  studies, but 28% in the primary outcome group and 34% in the mortality group included  $\geq 6$  studies. Of these latest studies in the primary outcome group, 9% were published before 1990, 25% were published in the 1990s, 40% were published in the 2000s and 25% were published during or after 2010. The median number of years between the latest study and the penultimate study was 3

(interquartile range 1–5) with 32% having a distance of  $< 2$  years. The mortality group demonstrated a similar distribution of year of publication.

**Table 1.** Characteristics of included studies

Parameter	Number of included studies	Sample size of last study	Total patients in SR	Year of last study
<i>SRs reporting stated primary outcome of the added RCT</i>				
Median	3	101	377	2003
Interquartile range	2–6	50–214	177–1100	1997–2010
<i>SRs reporting mortality</i>				
Median	4	85	398	2003
Interquartile range	2–7	48–185	191–1234	1998–2010
<i>SRs reporting mean difference</i>				
Median	3	57	163	2009
Interquartile range	2–5	29–84	99–320	2003–2011

Abbreviations: RCT, randomized controlled trial; SR, systematic review. Characteristics of included comparative SRs.



**Figure 1.** Included studies. Description and classification of systematic reviews (SRs) of neonatology interventions from the Neonatal Review Group of the Cochrane Collaboration Database of Systematic Reviews.

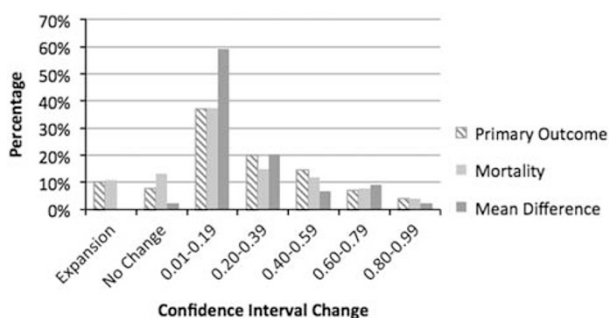
### Impact of most recently added trial

The majority of the most recently added studies across all outcomes (stated primary outcome, mortality, mean difference) failed to shrink the CI by  $\geq 20\%$  (Figure 2). For SRs reporting primary outcome, 18% of the latest RCTs either expanded or failed to narrow the CI. For SRs reporting either mortality or mean difference, failure to shrink the CI occurred in 24% and 2% of SRs, respectively.

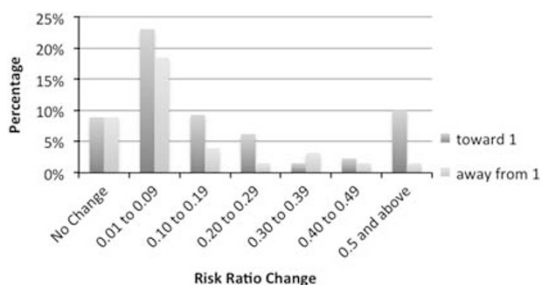
The typical estimate of the risk ratio for SRs reporting the stated primary outcome was unchanged with 18% of the latest RCT additions, whereas 30% moved the estimate away from the null, and the remainder moved the estimate toward the null (Figure 3). Of latest the RCTs, 41% moved the typical estimate by at least 10%. Latest RCTs changed the typical estimate from favoring the intervention to favoring the control, or vice versa, in 8% of cases. A change in significance occurred in 11% of cases.

For SRs reporting mortality, the typical estimate risk ratio was unchanged with 18% of the latest RCT additions, whereas 30% moved the estimate away from the null, and the remainder moved the estimate toward the null. Again, 38% of RCTs changed the typical estimate by at least 10%. Also, 6% of RCTs changed the typical estimate from favoring the intervention to favoring the control, or vice versa. Changes in statistical significance in the mortality group occurred in only 8% of cases.

Similarly, for SRs reporting mean difference, in which the impact of the most recently added RCT on the typical estimate was



**Figure 2.** Precision change. Change in width of the 95% confidence interval (CI) of the cumulative typical estimate of effect after addition of the most recent trial, according to type of study outcome reported in the forest plot. Numerical values on the x axis represent the proportion by which latest randomized controlled trials (RCTs) narrowed the CI.



**Figure 3.** Effect size change. Change in the cumulative typical estimate of effect after addition of the most recent trial for systematic reviews (SRs) reporting on the stated primary outcome of the last trial added. Dark gray bars represent movement toward 1 (toward the null), whereas light gray bars represent movement away from 1 (toward an effect). Numerical values on the x axis represent the absolute change in risk ratio after addition of the latest randomized controlled trial (RCT).

evaluated by a change in z-score, 2% produced no change, 48% increased the z-score and the remainder decreased the z-score. Of the RCTs, 52% changed the z-score by at least 50%. Of the latest studies in this group, 20% resulted in a change in statistical significance.

### Determinants of impact

In regression analysis of the primary outcome group, narrowing of the CI by  $>20\%$  was negatively associated with the number of previously published RCTs (odds ratio 0.707,  $P$ -value 0.0008) and study year (odds ratio 0.941,  $P$ -value 0.0140). Change of the typical estimate of at least 10% was negatively associated with the number of previously published studies only (odds ratio 0.805,  $P$ -value 0.0089). None of the evaluated factors demonstrated a significant association with the ability of a study to change the statistical significance of the effect estimate.

Analysis of the mortality and mean difference groups demonstrated similar significant correlations of 20% CI decrease with number of previously published RCTs. However, the significance of publication year was not replicated in either group. In the mean difference group, sample size of the latest RCT demonstrated a positive correlation with 20% CI decrease and a lead author located in the United States demonstrated a negative correlation with a change in the z-score of at least 10%.

### Postsurfactant era analysis (1991 and later)

The analysis was repeated with exclusion of studies published in 1990 and earlier. This left 117 studies in the primary outcome group, 118 in the mortality group and 40 in the mean difference group. We found no significant difference in study number, study size or review size in this subgroup compared with the full group of SRs. The results from this repeat analysis were largely congruent with the findings from the entire group in terms of impact on precision and effect size. Failure to shrink the CI by 20% occurred in 59%, 64% and 60% of SRs reporting stated primary outcome, mortality and mean difference, respectively. A change in statistical significance occurred in 10%, 8% and 20% of SRs reporting stated primary outcome, mortality and mean difference, respectively.

In regression analysis, we found similar significant correlations. The only notable difference in the regression analysis was the disappearance of a significant correlation with publication year and 20% CI decrease for the primary outcome group.

## DISCUSSION

Defining when a study question can be deemed ‘answered’ depends upon many factors. Of these, precision and distance from the null in meta-analyses reflect uncertainty or clinical significance and are important considerations for both clinicians devising treatment strategies and researchers planning new trials. Our study evaluates the impact of the most recent randomized trial over a broad range of neonatal clinical questions in the Cochrane Neonatal Reviews. We found that the most recently added RCT failed to substantially narrow the CI of the typical estimate of effect in a majority of cases or to change the statistical significance of the effect estimate. Moreover, in a substantial proportion, the typical estimate of effect itself remained unchanged or minimally changed. For all outcomes measured, narrowing of the CI of the typical estimate by an added study was negatively associated with the number of previously published RCTs.

For more than two decades, comprehensive systematic review and meta-analysis of the randomized clinical trial literature has been a core priority of the neonatology community.<sup>5,6</sup> This endeavor has had multiple beneficial effects, including earlier identification of efficacy and resulting adoption of useful therapies; earlier identification of lack of efficacy and resulting abandonment of useless or harmful therapies; more precise

estimates of effect size, particularly for less frequent but clinically important adverse outcomes; and acknowledgment of specific areas of uncertainty in the literature in order to guide future investigation.<sup>1</sup>

It is perhaps surprising, then, that we found so little benefit at the margin of adding trials to these meta-analyses. Why is this the case? The salutary power of systematic review and meta-analysis is predicated on its inputs, specifically the RCTs that comprise the reviews. If these RCTs are compromised in their internal validity—for example, if they are too small,<sup>7</sup> are focused on surrogate outcomes, are unblinded or have loss to follow-up<sup>8</sup>—or if they repeat questions that have been asked multiple times before,<sup>9</sup> then the impact of added studies will be unreliable and the utility of the systematic review process may be compromised. A more provocative finding by Ioannidis<sup>10</sup> suggests that similar factors (study power and bias, number of studies on a specific question, ratio of true to no relationships among those evaluated) may contribute to whether or not a study's findings are indeed true. A recent evaluation of 50 reports of cumulative meta-analyses by Clarke *et al.*<sup>11</sup> found several interesting themes, including instances where stable results would have been seen before addition of further trials and instances where further trials were insufficiently powered to resolve remaining uncertainties. Importantly, such deficiencies indicate that societal resources for research are being used less efficiently than is possible to answer important clinical questions.

We have focused on impact based on three types of measured outcomes—mortality, the stated primary outcome of the most recently added study and the mean difference for continuous variables. These outcomes provide clinically important and quantifiable information, utilizing data from the largest subsample of SRs. We recognize, however, that this approach may not fully measure the impact of the trial on meta-analytic estimates. Potentially, other outcomes may yield clinically important information with a greater effect on their own typical estimates.

Beyond the quantitative impact of our chosen outcomes on the meta-analyses, it is important to note that other factors have the potential to affect clinical uptake and recommendations by guideline organizations such as professional bodies. Some pivotal trials may guide clinical practice despite their weight in a meta-analysis. This occurs, for example, if they are deemed by practitioners to be of higher methodological quality, or if they examine an intervention that is slightly different from those in the SR but still meet inclusion requirements. Another factor is the context in which efficacy information is applied. In attempting to establish a theoretical basis for quality-improvement collaboratives, Eppstein *et al.*<sup>12</sup> have suggested that improvements in outcomes may be dependent on 'complex socio-technical environments' in which different interventions might interact, and that there are intrinsic limitations to the RCT approach in addressing these. Providers may elect to undervalue the results of meta-analyses if they believe that they are less applicable to their local clinical context.

Most notably missing from our analysis are considerations of potential bias. Our analysis rests on the RCT added most recently to the relevant SR forest plot that is assumed to be the latest temporally completed and published trial. It is possible, however, that publication bias resulted in later publication of earlier-completed studies; that later studies had not yet appeared in the updated SR review; or that negative results, or potentially results that confirm prior publications, had not appeared in the literature and therefore were never included in the SR. Delayed publication of trials with smaller sample sizes, lower funding levels and surrogate outcomes have been documented in other fields.<sup>13</sup> Moreover, the impact of the most recently added trial depends on which trials have already been included in the meta-analysis; SR conclusions have been shown to differ based on specific inclusion criteria<sup>14</sup> and the impact of the most recently added RCT would

also be affected in that case. In addition, trials may be planned concurrently or may be underway before the results of previous trials are available, and this is likely exemplified by the 32% of most recently added trials that were published < 2 years after the prior trial. This could argue for earlier sharing of results and coordination of primary outcomes.

In light of increasing constraints on the funding available for clinical research<sup>15</sup> it is important that funders, investigators and consumers understand whether or not current research is impactful. The neonatology community focuses its efforts not only on asking important questions, but also on answering those questions in the most methodologically rigorous way. Our study suggests that further deliberation could be included in the process, involving an explicit determination of how new experimental evidence might alter the landscape. For example, formal consideration of optimal information size using trial sequential analysis might be required by funders in power calculations for proposed studies.<sup>16</sup> Similarly, funding agencies might explicitly prioritize less-studied topics at the time of review, or generate requests for proposals on the basis of quantitative analysis of existing review databases such as Cochrane. Expansion of currently employed reporting checklists such as PRISMA and CONSORT, to document that such consideration had been undertaken prospectively, would aid in this process. We emphasize that these suggestions are not in conflict with the imperative of replication of findings.<sup>10</sup> Rather, they seek to strike a balance between underreplication and overreplication, and thus ensure a valid evidence base for practice that also makes optimal use of societal resources.

---

#### What is known about this topic

- Comprehensive systematic review and meta-analysis guides clinicians and researchers via earlier identification of therapy efficacy, estimates of effect size and acknowledgment of areas of uncertainty.

#### What this study adds

- We examine the impact of the latest studies added to meta-analyses across the full breadth of topics considered relevant by the Cochrane Collaboration.
  - The lack of substantial impact by a majority of these latest studies suggests a need for improved consideration when planning research.
- 

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### AUTHOR CONTRIBUTIONS

SCH: contributed to the study design, acquisition of the data, analysis and interpretation of data, drafting of the manuscript and revision based on the comments of the co-authors; HK: guided the study conception and design, and critical revision of the manuscript; CV: contributed to the acquisition of the data; RS: contributed to critical revision of the manuscript; DD: contributed to the study conception and design, analysis and interpretation of the data and critical revision of the manuscript; W-YM: contributed to the analysis and interpretation of data; JP and SBD: contributed to critical revision of the manuscript; JAFZ: guided the methodology, study conception and design, data checking and interpretation, drafting and critical revision of the manuscript. SCH and JAFZ had full access to all of the data in the study and are guarantors of the work. All authors approved the final manuscript as submitted.

#### REFERENCES

- 1 McGuire W, Fowlie PW, Soll RF. What has the Cochrane collaboration ever done for newborn infants? *Arch Dis Child Fetal Neonatal Ed* 2010; **95**(1): F2–F6.

- 2 Guyatt G, Rennie D, Meade M, Cook D. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. McGraw-Hill Education, American Medical Association: New York, NY, 2015.
- 3 *The Cochrane Library, Issue 7*. [Internet]. Wiley: Chichester, 2013.
- 4 *Review Manager (RevMan) [Computer program]*. 5.3 ed. The Nordic Cochrane Centre, The Cochrane Collaboration: Copenhagen, 2014.
- 5 Sinclair J, Bracken MB. *Effective Care of the Newborn Infant*. Oxford University Press: Oxford, 1992.
- 6 Sinclair JC, Haughton DE, Bracken MB, Horbar JD, Soll RF. Cochrane neonatal systematic reviews: a survey of the evidence for neonatal therapies. *Clin Perinatol* 2003; **30**(2): 285–304.
- 7 Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES *et al*. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 2013; **14**(5): 365–376.
- 8 DeMauro SB, Giaccone A, Kirpalani H, Schmidt B. Quality of reporting of neonatal and infant trials in high-impact journals. *Pediatrics* 2011; **128**(3): e639–e644.
- 9 Sinclair JC. Meta-analysis of randomized controlled trials of antenatal corticosteroid for the prevention of respiratory distress syndrome: discussion. *Am J Obstet Gynecol* 1995; **173**(1): 335–344.
- 10 Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005; **2**(8): e124.
- 11 Clarke M, Brice A, Chalmers I. Accumulating research: a systematic account of how cumulative meta-analyses would have provided knowledge, improved health, reduced harm and saved resources. *PLoS ONE* 2014; **9**(7): e102670.
- 12 Eppstein MJ, Horbar JD, Buzas JS, Kauffman SA. Searching the clinical fitness landscape. *PLoS ONE* 2012; **7**(11): e44901.
- 13 Gordon D, Taddei-Peters W, Mascette A, Antman M, Kaufmann PG, Lauer MS. Publication of trials funded by the National Heart, Lung, and Blood Institute. *N Engl J Med* 2013; **369**(20): 1926–1934.
- 14 Dechartres A, Altman DG, Trinquart L, Boutron I, Ravaud P. Association between analytic strategy and estimates of treatment outcomes in meta-analyses. *JAMA* 2014; **312**(6): 623–630.
- 15 Boadi K. Erosion of funding for the National Institutes of Health threatens U.S. leadership in biomedical research. 2014 Available from <https://www.americanprogress.org/issues/economy/reports/2014/03/25/86369/erosion-of-funding-for-the-national-institutes-of-health-threatens-u-s-leadership-in-biomedical-research/>.
- 16 Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. *J Clin Epidemiol* 2008; **61**(1): 64–75.