Research White Paper

Comparison of Effects as Evidence Evolves From Single Trials to High-Quality Bodies of Evidence

AHRQ
Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

Research White Paper

Comparison of Effects as Evidence Evolves From Single Trials to High-Quality Bodies of Evidence

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-2012-00008-I

Prepared by:
RTI International–University of North Carolina Evidence-based Practice Center
Research Triangle Park, NC

Investigators:
Gerald Gartlehner, M.D., M.P.H.[1,2]
Andreea Dobrescu, M.D.[3]
Tammeka Swinson Evans, M.O.P.[2]
Kylie Thaler, M.D., M.P.H.[1]
Barbara Nussbaumer,MSc[1]
Isolde Sommer, PhD, M.P.H[1]
Kathleen N. Lohr, Ph.D.[2]

[1] Austrian Cochrane Branch, Danube University, Krems, Austria
[2]RTI International, Research Triangle Park, NC
[3]Victor Babes University of Medicine and Pharmacy, Timisoara, Romania

AHRQ Publication No. 15-EHC010-EF
March 2015

This report is based on work conducted by the Research Triangle Institute International – University of North Carolina Chapel Hill (RTI-UNC) Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-2012-00008-I). The findings and conclusions in this document are those of the author(s), who are responsible for its content, and do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

AHRQ or U.S. Department of Health and Human Services endorsement of any derivative products that may be developed from this report, such as clinical practice guidelines, other quality enhancement tools, or reimbursement or coverage policies may not be stated or implied.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact EffectiveHealthCare@ahrq.hhs.gov.

Drs. Gartlehner and Thaler are members of the GRADE Working Group. Drs. Gartlehner and Lohr are co-authors of the AHRQ guidance for grading the strength of evidence. None of the other authors have any affiliations or financial involvement disclosures to report.

Suggested citation:
Gartlehner G, Dobrescu A, Swinson Evans T, Thaler K, Nussbaumer B, Sommer I, Lohr KN. Comparison of Effects as Evidence Evolves From Single Trials to High-Quality Bodies of Evidence. Research White Paper. (Prepared by the RTI–UNC Evidence-based Practice Center under Contract No. 290-2012-00008-I.) AHRQ Publication No. 15-EHC010-EF. Rockville, MD: Agency for Healthcare Research and Quality. March 2015.
www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-Based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To improve the scientific rigor of these evidence reports, AHRQ supports empiric research by the EPCs to help understand or improve complex methodologic issues in systematic reviews. These methods research projects are intended to contribute to the research base in and be used to improve the science of systematic reviews. They are not intended to be guidance to the EPC program, although may be considered by EPCs along with other scientific research when determining EPC Program methods guidance.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality.

We welcome comments on this Methods Research Project. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by email to epc@ahrq.gov.

Richard G. Kronick, Ph.D.
Director
Agency for Healthcare Research and Quality

David Meyers, M.D.
Acting Director, Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director, EPC Program, Task Order Officer
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Peer Reviewers

Prior to publication of the white paper, we sought input from independent Peer Reviewers without financial conflicts of interest. However, the conclusions and synthesis of the scientific literature presented in this report does not represent the views of individual reviewers.

Peer Reviewers must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential non-financial conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential non-financial conflicts of interest identified.

The list of Peer Reviewers follows:

Nancy Berkman, Ph.D.
Senior Research Analyst
RTI International
Research Triangle Park, NC

Philippe Ravaud, M.D.
Senior Lecturer of Epidemiology
Columbia University's Mailman School of Public Health
New York, NY
Professor of Epidemiology, Paris Descartes University
Director, the French Cochrane Center
Paris, France

Paul Shekelle, M.D., Ph.D.
Director, Southern California Evidence-Based Practice Center, RAND Corporation
Santa Monica, CA

Jonathan Treadwell, Ph.D.
Associate Director, ECRI Institute Evidence-based Practice Center
Plymouth Meeting, PA

Thomas Trikalinos, M.D., Ph.D.
Associate Professor of Health Services, Policy and Practice
Director of Evidence-Based Medicine, Brown University
Providence, RI
Alex Ellis, Ph.D. Candidate
Brown University
Providence, RI

Comparison of Effects as Evidence Evolves From Single Trials to High-Quality Bodies of Evidence

Structured Abstract

Objective. The objective of our methods project was to use a diverse sample of medical interventions to assess empirically whether first trials rendered substantially different treatment effect estimates than reliable, high-quality bodies of evidence.

Study design and setting. We employed a meta-epidemiological study design using 100 bodies of evidence from Cochrane reports that had been graded as high quality of evidence. To determine the concordance of effect estimates between first and subsequent trials, we applied both quantitative and qualitative approaches. For quantitative assessment, we used Lin's concordance correlation and calculated z-scores; to determine the magnitude of differences of treatment effects, we calculated standardized mean differences (SMDs) and ratios of relative risks. We determined qualitative concordance based on a 2-tiered approach incorporating changes in statistical significance and magnitude of effect.

Results. First trials both over- and under-estimated the true treatment effects in no discernible pattern. Nevertheless, depending on the definition of concordance, effect estimates of first trials were concordant with pooled subsequent studies in at least 33 percent but up to 50 percent of comparisons. The pooled magnitude of change as bodies of evidence advanced from single trials to high-quality bodies of evidence was 0.16 SMD (95% confidence interval [CI], 0.12 to 0.21). In 80 percent of comparisons the difference in effect estimates was smaller than 0.5 SMDs. In first trials with large treatment effects (>0.5 SMD), however, estimates of effect substantially changed as new evidence accrued (mean change 0.68 SMD, 95% CI, .50 to 0.86)

Conclusion. Results of first trials often change but the magnitude of change, on average, is small. Exceptions are first trials that present large treatment effects which often dissipate as new evidence accrues.

Contents

Introduction.....1

Methods.....2
Empirical Data.....2
Assessing Concordance in Effect Estimates.....2
Qualitative Assessment of Concordance.....3
Quantitative Assessment of Concordance.....4

Results.....5
Qualitative Concordance Analyses.....5
Quantitative Concordance and Differences in Magnitudes of Effect Estimates.....12
Stratified Analyses.....14

Discussion.....16

Equations.....18

References.....19

Tables

Table 1. Characteristics and concordance of results of first trials and pooled subsequent trials 6
Table 2. Results of stratified analyses based on sample size, risk of bias, and magnitude of effect estimates of first trials.....15

Figures

Figure 1. Illustration of assessing the concordance in effect estimates between a first trial and the pooled subsequent trials 3
Figure 2. Overlap of confidence intervals of first trials and pooled subsequent trials 5
Figure 3. Meta-analysis of differences of effect estimates between first and pooled subsequent trials.....13
Figure 4. Modified Bland-Altman diagram comparing the point estimates of the first trials with the point estimates of pooled subsequent trials 14

Introduction

GRADE (Grading of Recommendations Assessment, Development and Evaluation) has become a widely adopted approach for conveying the uncertainty associated with findings and conclusions present in systematic reviews.[1] GRADE is thought to reflect, explicitly and transparently, the confidence that researchers have in an available body of evidence.[2] Its conceptual framework uses information about risk of bias, imprecision, inconsistency, indirectness, and publication bias to communicate the confidence that systematic reviewers have in estimates of treatment effects; GRADE calls this "quality of evidence" (QOE).[3] In the context of systematic reviews, GRADE defines QOE as the extent of the confidence that the estimates of an effect are correct. A grade of high QOE, for example, means that reviewers are confident that the effect estimate is close to the true effect and that new studies will not change the conclusions.

Assessing bodies of evidence that are limited to single studies is a particular challenge for systematic reviewers for two reasons: (1) consistency of results with other studies cannot. be determined, and (2) whether the first study published is also the first study conducted (or whether the first study published is just the most favorable study conducted) is usually unclear. Both GRADE guidance and related guidance for the U.S. Agency for Healthcare Research and Quality (AHRQ) Evidence-based Practice Center program recommend using the same approach for single- and multiple-study bodies of evidence.[3,4] Nevertheless, because reviewers cannot be certain that a single study, no matter how large, presents a definitive picture of the benefits and harms of a given treatment,[4] in practice they often tend to grade single-study bodies of evidence as low or very low.

A QOE of low or very low reflects a general lack of confidence in effect estimates. Whether effect estimates substantially change as a body of evidence advances from a single trial to a reliable body of evidence that is rated as high QOE, has not been examined yet. Several previous studies examined how research evidence evolves over time.[5-7] Results indicate that direction and magnitude of effects change, particularly if first studies report very large treatment effects.[5-8] None of these studies, however, compared the estimates of effects of first studies to reference points that other researchers determined to be reliable and of high certainty with respect to the correctness of effect estimates.

The objective of our methods project was to assess empirically whether first trials of a diverse sample of medical interventions rendered substantially different treatment effect estimates than reliable, high-quality bodies of evidence.

Methods

To address our objective, we used a meta-epidemiological approach based on large, systematically appraised bodies of evidence that Cochrane authors had graded as high QOE. We used effect estimates of high QOE studies as reference points because a grade of high QOE implies that investigators were very confident that the effect estimate is close to the truth and that new studies are unlikely to change conclusions. The basic assumption for our study was that these bodies of evidence had been graded correctly and can serve as reference points to evaluate whether the first published trial of each body of evidence showed concordant or discordant effect estimates compared with the remaining body of evidence.

Empirical Data

We systematically searched the Cochrane Library from 2010 to July 2014 to find Cochrane reports that presented the QOE in "summary-of-findings" tables. According to the Methodological Expectations of Cochrane Intervention Reviews (MECIR), summary-of-findings tables are highly desirable but not mandatory. Consequently, not all Cochrane reviews present them. We used "quality of evidence" OR "summary of findings" as search terms. Out of 1325 records, 1311 presented summary of findings tables. Of those, 293 reported at least one high QOE rating. From these, we randomly selected 100 reports that had to fulfill the following criteria: (1) include a body of evidence comprising more than three RCTs on therapeutic interventions two Cochrane authors graded independently as high QOE; (2) present meta-analytic outcomes authors reported as relative risks or odds ratios for binary outcomes, and as weighted mean differences or standardized mean differences (SMDs) for continuous outcomes; and (3) provide data to reproduce the meta-analyses.

If a review reported more than one high-QOE outcome, we randomly selected one outcome per review.

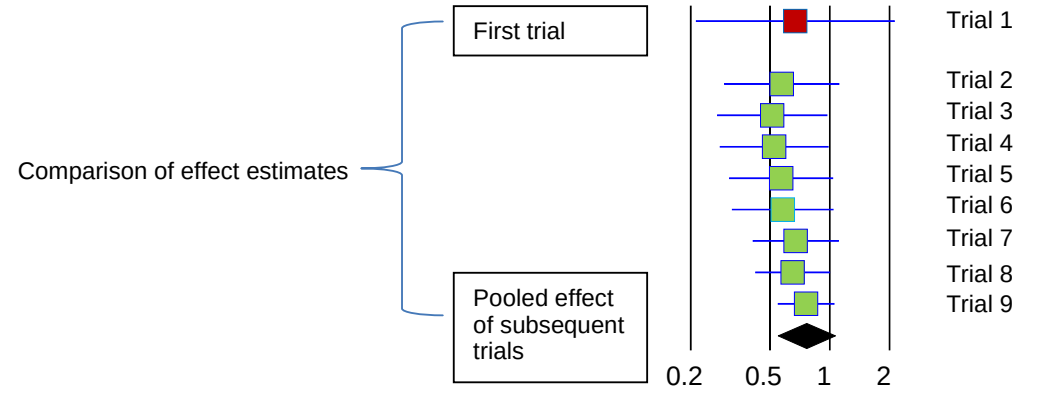
Assessing Concordance in Effect Estimates

To assess the concordance or discordance of effect estimates between the first published trial of a body of evidence (termed "first trial") and the pooled effect estimate of all trials that were published subsequently (termed "pooled subsequent trials"), we employed qualitative and quantitative methods. For all 100 bodies of evidence, we calculated the effect estimate of the first trial and the pooled estimate of the subsequent trials. For continuous outcomes we used SMDs (Cohen's d) as effect measures; for dichotomous outcomes we calculated relative risks.

To compare two independent samples of data, we did not include the first trial in the meta-analysis of the subsequent trials. For all meta-analyses we used DerSimonian-Laird random effects models.

Figure 1 illustrates the approach. The first row in the forest plot represents the first trial published (trial 1). The subsequent rows present the trials that were published later, following the first trial (trials 2-9). The pooled estimate in Figure 1 is the meta-analysis of trials 2 through 9 (the pooled subsequent trials).

Figure 1. Illustration of assessing the concordance in effect estimates between a first trial and the pooled subsequent trials



Qualitative Assessment of Concordance

We determined qualitative concordance based on a two-tiered approach incorporating changes in statistical significance and magnitude of effect. This approach was originally proposed to detect signals for updating systematic reviews;[9] we modified it for our purposes. We deemed the first trial and the pooled subsequent trials as concordant when:

1. Statistical significance did not change, *and*
2. The magnitude of treatment effects remained similar.

The first criterion captures whether a result that was statistically significant in the first trial was *not* statistically significant in the pooled subsequent trials or vice versa—a previously nonsignificant result has become statistically significant. To avoid counting trivial or 'borderline' changes in statistical significance, we required that at least one of the two results have a p-value outside the range of 0.04 to 0.06. In other words, we did not consider cases in which a p-value changed statistical significance within this range (e.g., a change from p=0.041 to p=0.059 did not count as a change in statistical significance, nor did the converse change from p=0.059 to p=0.041).

The second criterion assessed whether the magnitude of change of the effect estimate between first trial and pooled subsequent trials was smaller than a predefined threshold. We employed three different thresholds to determine the similarity of treatment effects.

1. A relative risk increase or reduction of less than 25 percentage points for dichotomous outcomes and less than 0.20 SMDs for continuous outcomes;
2. A relative risk increase or reduction of less than 50 percentage points for dichotomous outcomes and less than 0.5 SMDs for continuous outcomes;
3. A staggered approach based on the magnitude of treatment effect in the first trial and the type of outcome
    a. For first trials with regular treatment effects (relative risk 0.5 to 2.00, or SMD <0.8) a change in relative risk increase/reduction of less than 25 percentage points or less than 0.20 SMD

b. For first trials with large treatment effects (relative risk 0.5 and >2.00, or SMD>0.8), a change in relative risk increase/reduction of less than 50 percentage points or less than 0.5 SMD.
c. For outcomes that can be considered extremely patient-relevant (e.g., mortality, stroke, myocardial infarction) a change in relative risk of less than 10 percentage points.

To determine concordance between first trials and pooled subsequent trials, we programmed and tested a Microsoft Excel spreadsheet that employed the above criteria.

Quantitative Assessment of Concordance

To assess quantitatively the concordance between effect estimates of first trials and pooled effects of subsequent trials, we employed three strategies (an explanation of the equations is provided in the Equations section for accessibility purposes).

First, to test for concordance, we used Lin's concordance correlation coefficient.[10] To be able to use a common metric for all studies, we converted dichotomous outcomes to SMDs using the following formula: $d = \log OddsRatio * \frac{\sqrt{3}}{\pi}$

Lin's concordance correlation determines how closely related two variables are in a linear fashion and how they correspond to each other. A value of +1 indicates perfect concordance, a value of -1 reflects perfect discordance, and a value of 0 denotes the absence of either concordance or discordance. We determined the concordance for point estimates and z-scores of first trials and pooled subsequent trials.

Second, to test for statistically significant differences between effect estimates, we calculated z-scores comparing the effect estimates of the first and the pooled subsequent trials: $z = \frac{effect_{first\ trial} - effect_{pooled\ trials}}{\sqrt{variance_{first\ trial} + variance_{pooled\ trials}}}$

Third, to determine differences in magnitudes of effect estimates between first and pooled subsequent trials, we calculated differences in SMDs (Cohen's d) for all comparisons and ratios of relative risks for dichotomous outcomes. We used DerSimonian-Laird random-effects meta-analyses to calculate summary differences in magnitudes of effects.

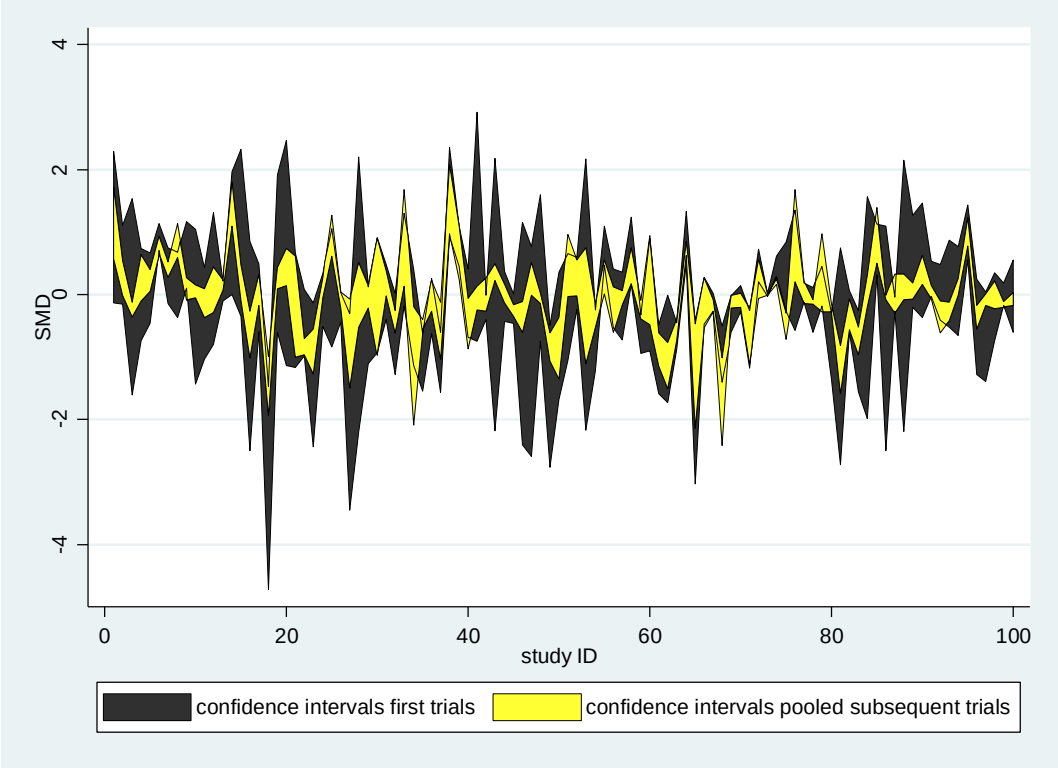
To determine the influence of risk of bias, sample sizes of first and pooled subsequent trial, and the magnitude of effect estimates of the first trial on results, we employed meta-regression analyses.

We conducted statistical analyses using Stata IC 13 (Stata Corporation, TX, USA), Comprehensive Meta-Analysis V2 (Biostat Inc, Englewood, NJ, USA), and Microsoft Excel 2010 (Microsoft Corporation, Redmond, WA, USA).

Results

The 100 included bodies of evidence dealt with a wide array of clinical topics employing dichotomous (n=79) and continuous (n=21) outcomes. Sample sizes of the first trials varied greatly (from 20 to 46,500 participants; median 112) and often led to wide, indeterminate confidence intervals. The pooled subsequent trials included samples from 55 to more than 280,000 participants (median 1071). Figure 2 depicts the overlap of confidence intervals between first trials and pooled subsequent trials. In 37 percent (n=37) of the comparisons statistical significance changed between first trials and pooled subsequent trials, i.e., a result that was statistically significant in the first trial was *not* statistically significant in the pooled subsequent trials or vice versa.

Figure 2. Overlap of confidence intervals of first trials and pooled subsequent trials



Qualitative Concordance Analyses

Table 1 summarizes characteristics and effect estimates of the 100 pairs of first and pooled subsequent trials. Based on our criteria for concordance (see Methods), effect estimates between first and pooled subsequent trials were concordant in 36 percent of comparisons for threshold 1 which employed the strictest criteria. In other words, in 64 percent of comparisons either the statistical significance changed between first and pooled subsequent trials or the difference in point estimates was larger than a 25 percent relative risk increase or reduction (or larger than 0.2 SMD). By comparison, 50 percent of first trials were concordant with pooled subsequent trials when we applied threshold 2. Threshold 2 was more lenient than threshold 1 with respect to the magnitude of change of point estimates (50 percent relative risk increase or reduction). Forty-four percent of first trials were concordant with pooled subsequent trials when we applied threshold 3 (staggered approach).

Table 1. Characteristics and concordance of results of first trials and pooled subsequent trials

	Inter-vention and Outcome	First Trial		Pooled Subsequent Trials		Combined		
		N	Effect Estimate (confidence interval)	N	Effect Estimate (confidence interval)	Effect Estimate (confidence interval)	P-value of Difference	Qualitative Concor-dance (using 3 thresholds)
Akl et al., 2011[11]	Sympto-matic VTE	84	RR: 0.33 (0.01 to 7.95)	2180	RR: 0.55 (0.37 to 0.83)	RR: 0.55 (0.37 to 82)	0.78	Discordant[a] Discordant[b] Discordant[c]
Alejandria et al., 2013[12]	All-cause mortality	38	RR: 0.69 (0.18 to 2.64)	645	RR: 0.80 (0.27 to 2.32)	RR: 1.02 (0.81 to 1.29)	0.88	Concordant[a] Concordant[b] Discordant[c]
Amato et al., 2010[13]	Discon-tinuation of treatment	142	RR: 0.69 (0.03 to 16.66)	839	RR: 1.10 (0.75 to 1.62)	RR: 1.09 (0.74 to 1.60)	0.79	Discordant[a] Concordant[b] Discordant[c]
Amato et al., 2011[14]	Retention in treatment	74	RR: 0.98 (0.86 to 1.11)	3,050	RR: 1.03 (0.99 to 1.07)	RR: 1.02 (0.99 to 1.06)	0.47	Concordant[a] Concordant[b] Concordant[c]
Amato et al., 2011[15]	Adverse events	37	RR: 0.98 (0.26 to 3.79)	91	RR: 1.03 (0.43 to 2.47)	RR: 1.11 (0.62 to 1.98)	0.96	Concordant[a] Concordant[b] Concordant[c]
Amato et al., 2013[16]	Completion of treatment	72	RR: 1.67 (1.07 to 2.6)	1,309	RR: 1.06 (0.96 to 1.18)	RR: 1.08 (1.01 to 1.16)	0.05	Discordant[a] Discordant[b] Discordant[c]
Bailey et al., 2013[17]	More than 50% pain relief	199	RR: 1.97 (1.46 to 2.66)	447	RR: 1.38 (1.21 to 1.58)	RR: 1.47 (1.29 to 1.68)	0.03	Discordant[a] Discordant[b] Discordant[c]
Bauer et al., 2011[18]	Response to treatment	727	RR: 1.13 (1.07 to 1.19)	518	RR: 1.08 (1.01 to 1.16)	RR: 1.11 (1.06 to 1.16)	0.31	Concordant[a] Concordant[b] Concordant[c]
Bird et al., 2014[19]	Pain-free response	381	RR: 4.07 (2.02 to 8.19)	5,444	RR: 2.94 (2.47 to 3.50)	RR: 2.99 (2.52 to 3.53)	0.38	Discordant[a] Discordant[b] Discordant[c]
Boyle et al., 2012[20]	Systemic reaction to an insect sting	30	RR: 0.10 (0.01 to 0.68)	175	RR: 0.14 (0.04 to 0.44)	RR: 0.13 (0.05 to 0.34)	0.80	Concordant[a] Concordant[b] Concordant[c]
Braithwaite et al., 2014[21]	Iris or retinal neo-vasculari-zation	98	RR: 0.32 (0.06 to 1.84)	838	RR: 0.23 (0.10 to 0.53)	RR: 0.24 (0.11 to 0.52)	0.75	Discordant[a] Discordant[b] Discordant[c]
Brito et al., 2011[22]	All-cause mortality	326	RR: 2.12 (0.26 to 17.32)	6434	RR: 0.90 (0.78 to 1.04)	RR: 0.90 (0.79 to 1.02)	0.43	Discordant[a] Discordant[b] Discordant[c]
Buchleitner et al., 2012[23]	Mortality	141	RR: 0.96 (0.02 to 47.65)	1,224	RR: 1.19 (0.89 to 1.58)	RR: 1.19 (0.89 to 1.58)	0.92	Concordant[a] Concordant[b] Concordant[c]
Cates et al., 2012[24]	Serious adverse events	204	RR: 0.20 (0.01 to 4.04)	6442	RR: 1.56 (0.99 to 2.45)	RR: 1.49 (0.96 to 2.33)	0.19	Discordant[a] Discordant[b] Discordant[c]
Cates et al., 2013[25]	Final rise in FEV	53	SMD: -0.05 (-0.59 to 0.49)	254	SMD: 0.07 (-0.17 to 0.32)	SMD: 0.05 (-0.17 to 0.28)	0.69	Concordant[a] Concordant[b] Concordant[c]
Charoenkwan et al., 2014[26]	Febrile morbidity	100	RR: 0.72 (0.17 to 3.06)	471	RR: 2.21 (0.95 to 5.17)	RR: 1.66 (0.80 to 3.45)	0.19	Discordant[a] Discordant[b] Discordant[c]
Chauhan et al., 2014[27]	Exacer-bations requiring cortico-steroids	429	RR: 0.3 (0.08 to 1.09)	5,494	RR: 0.89 (0.78 to 1.02)	RR: 0.88 (0.77 to 1.01)	0.10	Discordant[a] Discordant[b] Discordant[c]

Table 1. Characteristics and concordance of results of first trials and pooled subsequent trials (continued)

	Inter-vention and Outcome	First Trial		Pooled Subsequent Trials		Combined		Qualitative Concor-dance (Using 3 Thresholds)
		N	Effect Estimate (Confidence Interval)	N	Effect Estimate (Confidence Interval)	Effect Estimate (Confidence Interval)	P-value of Difference	
Chin et al., 2013[28]	Supple-mental local anesthetic blocks	40	RR: 1,06 (0.84 to 1.34)	821	RR: 1,02 (0.95 to 1.1)	RR: 1.03 (0.96 to 1.10)	0.77	Concordant[a] Concordant[b] Concordant[c]
Cho et al., 2014[29]	Urine volume	90	SMD: 0.18 (-0.23 to 0.60)	430	SMD: 0.27 (-0.02 to 0.55)	SMD: 0.25 (0.02 to 0.49)	0.71	Concordant[a] Concordant[b] Concordant[c]
Chong et al., 2013[30]	Exacer-bations	700	RR: 1.01 (0.82 to 1.24)	4,828	RR: 0.83 (0.78 to 0.88)	RR: 0.84 (0.79 to 0.89)	0.07	Discordant[a] Discordant[b] Discordant[c]
Clarke et al., 2014[31]	At least 50% of maximum pain relief	125	RR: 6.08 (2.09 to 17.70)	673	RR: 5.04 (2.95 to 8.62)	RR: 5.08 (3.20 to 8.05)	0.77	Discordant[a] Discordant[b] Discordant[c]
Clifford et al., 2012[32]	Left-ventricular ejection fraction	40	SMD: 0.48 (-0.15 to 1.11)	839	SMD: 0.27 (-0.01 to 0.54)	SMD: 0.28 (0.02 to 0.54)	0.55	Discordant[a] Concordant[b] Concordant[c]
Costi et al., 2014[33]	Agitation	133	RR: 0.07 (0.01 to 0.49)	1114	RR: 0.39 (0.29 to 0.51)	RR: 0.37 (0.28 to 0.50)	0.09	Discordant[a] Concordant[b] Concordant[c]
Eftimov et al., 2013[34]	Significant improveme nt in disability score	28	RR: 1.16 (0.32 to 4.24)	170	RR: 2.37 (1.69 to 3.33)	RR: 2.26 (1.63 to 3.15)	0.30	Discordant[a] Discordant[b] Discordant[c]
Feagan et al., 2012[35]	Failure to maintain remission	57	RR: 1.09 (0.39 to 3.04)	1,598	RR: 1.12 (0.98 to 1.28)	RR: 1.12 (0.99 to 1.27)	0.96	Concordant[a] Concordant[b] Concordant[c]
Fernandes et al., 2013[36]	Length of hospital stay	29	SMD: -0.3 (-1.03 to 0.43)	614	SMD: -0.14 (-0.37 to 0.09)	SMD: -0.15 (-0.36 to 0.06)	0.69	Concordant[a] Concordant[b] Concordant[c]
Fransen et al., 2014[37]	Pain	68	SMD: -0.73 (-1.22 to -0.24)	451	SMD: -0.33 (-0.52 to -0.15)	SMD: -0.38 (-0.55 to -0.21)	0.12	Discordant[a] Concordant[b] Concordant[c]
Fullerton et al., 2014[38]	Manifes-tation of neuropathy	539	RR: 0.29 (0.13 to 0.66)	664	RR: 0.37 (0.23 to 0.59)	RR: 0.35 (0.23 to 0.53)	0.62	Concordant[a] Concordant[b] Concordant[c]
Gafter et al., 2012[39]	Mortality	52	RR: 1.27 (0.49 to 3.24)	219	RR: 0.69 (0.56 to 0.86)	RR: 0.71 (0.57 to 0.88)	0.22	Discordant[a] Discordant[b] Discordant[c]
Gogtay et al. 2013[40]	Remaining parasitae-mic after 24 hours	180	RR: 0.20 (0.08 to 0.50)	1,472	RR: 0.43 (0.38 to 0.50)	RR: 0.42 (0.36 to 0.49)	0.11	Concordant[a] Concordant[b] Concordant[c]
Gowing et al., 2009[41]	Withdrawal from treatment	72	RR: 1.12 (0.77 to 1.61)	409	RR: 1.69 (1.35 to 2.1)	RR: 1.61 (1.31 to 1.99)	0.06	Discordant[a] Discordant[b] Discordant[c]
Griffiths et al., 2013[42]	Hospital admission	31	RR: 0.94 (0.06 to 13.82)	1,967	RR: 0.74 (0.64 to 0.85)	RR: 0.735 (0.64 to 0.85)	0.87	Discordant[a] Discordant[b] Discordant[c]
Gurion et al., 2012[43]	Mortality	388	RR: 1.09 (0.99 to 1.2)	3,017	RR: 1.01 (0.96 to 1.06)	RR: 1.02 (0.98 to 1.07)	0.17	Concordant[a] Concordant[b] Concordant[c]

Table 1. Characteristics and concordance of results of first trials and pooled subsequent trials (continued)

	Inter-vention and Outcome	First Trial		Pooled Subsequent Trials		Combined	P-value of Difference	Qualitative Concor-dance (Using 3 Thresholds)
		N	Effect Estimate (Confidence Interval)	N	Effect Estimate (Confidence Interval)	Effect Estimate (Confidence Interval)		
Hahn et al., 2014[44]	Adverse events due to hyper-tension	181	RR: 0.77 (0.41 to 1.45)	1160	RR: 0.97 (0.66 to 1.43)	RR: 0.93 (0.69 to 1.26)	0.55	Concordant[a] Concordant[b] Concordant[c]
Häuser et al., 2013[45]	With-drawals due to adverse events	207	RR: 1.62 (0.81 to 3.26)	1,734	RR: 1.84 (1.52 to 2.22)	RR: 1.83 (1.53 to 2.18)	0.74	Discordant[a] Discordant[b] Discordant[c]
Hayward et al., 2012[46]	Reduction in sore throat pain at 24 hours	58	SMD: 0.70 (0.17 to 1.24)	559	SMD: 0.47 (0.18 to 0.75)	SMD: 0.49 (0.24 to 0.75)	0.44	Discordant[a] Concordant[b] Concordant[c]
Hemmingsen et al., 2013[47]	Hypo-glycemia	153	RR: 2.6 (0.52 to 12.99)	27,974	RR: 1.98 (1.36 to 2.86)	RR: 2.01 (1.41 to 2.86)	0.76	Discordant[a] Discordant[b] Discordant[c]
Hoare et al., 2010[48]	Quality of movement	29	SMD: 0.56 (-0.18 to 1.30)	55	SMD: 0.91 (0.13 to 1.68)	SMD: 0.81 (0.31 to 1.31)	0.52	Discordant[a] Discordant[b] Discordant[c]
Hodson et al., 2013[49]	Cytome-galovirus infections	104	RR: 0.26 (0.09 to 0.72)	1,005	RR: 0.43 (0.35 to 0.52)	RR: 0.42 (0.32 to 0.56)	0.34	Concordant[a] Concordant[b] Concordant[c]
Holme et al., 2013[50]	All-cause Mortality	46551	RR: 1.0 (0.97 to 1.02)	28,3091	RR: 1.0 (0.99 to 1.01)	RR: 1.00 (0.99 to 1.01)	1.00	Concordant[a] Concordant[b] Concordant[c]
Horey et al., 2013[51]	Preferred and actual mode of birth	1192	RR: 1.01 (0.96 to 1.08)	729	RR: 1.02 (0.92 to 1.13)	RR: 1.02 (0.96 to 1.07)	0.88	Concordant[a] Concordant[b] Concordant[c]
Howe et al., 2011[52]	Change in bone mineral density	49	SMD: 0.1 (0.46 to 0.67)	766	SMD: 0.22 (0.05 to 0.40)	SMD: 0.22 (0.06 to 0.38)	0.70	Discordant[a] Discordant[b] Discordant[c]
Hughes et al., 2014[53]	Discontinu ation of treatment	147	RR: 0.04 (0.0 to 0.65)	348	RR: 0.25 (0.08 to 0.78)	RR: 0.20 (0.06 to 0.65)	0.30	Discordant[a] Discordant[b] Discordant[c]
Irving et al., 2012[54]	Clinically apparent hepatitis A	1037	RR: 0.19 (0.08 to 0.42)	40,393	RR: 0.05 (0.01 to 0.17)	RR: 0.10 (0.03 to 0.29)	0.11	Concordant[a] Concordant[b] Concordant[c]
Itchaki et al., 2013[55]	Disease control	143	RR: 0.90 (0.59 to 1.36)	616	RR: 0.90 (0.65 to 1.25)	RR: 0.89 (0.69 to 1.14)	1.00	Concordant[a] Concordant[b] Concordant[c]
Karner et al., 2014[56]	Improve-ment in quality of life	921	RR: 1.64 (1.36 to 1.98)	15,866	RR: 1.23 (1.18 to 1.29)	RR: 1.26 (1.19 to 1.33)	0.00	Discordant[a] Concordant[b] Discordant[c]
Katalinic et al., 2010[57]	Joint mobility	28	SMD: 0.00 (-0.74 to 0.74)	109	SMD: 0.27 (-0.1 to 0.64)	SMD: 0.24 (-0.08 to 0.57)	0.52	Discordant[a] Concordant[b] Concordant[c]
Kew Kayleigh et al., 2013[58]	Exacer-bations requiring oral steroids	2135	RR: 0.83 (0.67 to 1.03)	3,975	RR: 0.74 (0.63 to 0.88)	RR: 0.79 (0.69 to 0.91)	0.42	Discordant[a] Discordant[b] Discordant[c]

Table 1. Characteristics and concordance of results of first trials and pooled subsequent trials (continued)

	Inter-vention and Outcome	First Trial		Pooled Subsequent Trials		Combined	P-value of Difference	Qualitative Concor-dance (Using 3 Thresholds)
		N	Effect Estimate (Confidence Interval)	N	Effect Estimate (Confidence Interval)	Effect Estimate (Confidence Interval)		
Kew Kayleigh et al., 2014[59]	Withdrawals	281	RR: 0.69 (0.40 to 1.18)	7,946	RR: 0.87 (0.84 to 0.91)	RR: 0.87 (0.84 to 0.91)	0.41	Discordant[a] Discordant[b] Discordant[c]
Koretz et al., 2013[60]	Variceal bleeding	1,050	RR: 0.52 (0.13 to 2.05)	662	RR: 0.14 (0.03 to 0.73)	RR: 0.30 (0.10 to 0.88)	0.22	Discordant[a] Discordant[b] Discordant[c]
Kramer et al., 2014[61]	Parasito-logical failure-	59	RR: 0.18 (0.08 to 0.39)	805	RR: 0.45 (0.33 to 0.62)	RR: 0.42 (0.29 to 0.59)	0.03	Discordant[a] Concordant[b] Discordant[c]
Kruis et al., 2013[62]	Hospital days per patient	50	SMD: 0.29 (-0.27 to 0.84)	691	SMD: -0.51 (-0.72 to -0.30)	SMD: -0.42 (-0.68 to -0.15)	0.01	Discordant[a] Discordant[b] Discordant[c]
Kumar et al., 2011[63]	Duration of photo-therapy-hours	69	SMD: -0.06 (-0.53 to 0.41)	223	SMD: -0.24 (-0.60 to 0.12)	SMD: -0.21 (-0.52 to 0.10)	0.52	Concordant[a] Concordant[b] Concordant[c]
La Mantia et al., 2012[64]	Sustained EDSS increase	502	RR: 0.83 (0.69 to 0.99)	604	RR: 1.0 (0.73 to 1.35)	RR: 0.90 (0.75 to 1.08)	0.31	Discordant[a] Discordant[b] Discordant[c]
Lai et al., 2013[65]	Adverse effects	49	RR: 2.04 (0.74 to 5.61)	2,954	RR: 1.08 (0.93 to 1.25)	RR: 1.09 (0.94 to 1.27)	0.23	Discordant[a] Discordant[b] Discordant[c]
Law et al., 2013[66]	Pain after 2 hours	576	RR: 3.09 (2.34 to 4.07)	2,819	RR: 2.62 (2.18 to 3.14)	RR: 2.69 (2.29 to 3.15)	0.33	Discordant[a] Concordant[b] Concordant[c]
Lazzerini et al., 2013[67]	Adverse events	50	RR: 1.0 (0.02 to 48.49)	2290	RR: 1.56 (1.32 to 1.85)	RR: 1.56 (1.31 to 1.85)	0.83	Discordant[a] Discordant[b] Discordant[c]
Lemiengre et al., 2012[68]	Treatment failure	192	RR: 1.41 (0.11 to 1.54)	2,175	RR: 0.55 (0.41 to 0.76)	RR: 0.55 (0.40 to 0.74)	0.68	Discordant[a] Discordant[b] Discordant[c]
Lewis et al., 2013[69]	Vomiting	183	RR: 0.75 (0.52 to 1.07)	883	RR: 0.73 (0.57 to 0.94)	RR: 0.74 (0.60 to 0.91)	0.91	Discordant[a] Discordant[b] Discordant[c]
Li et al., 2014[70]	Withdrawal due to adverse effects.	501	RR: 0.62 (0.40 to 0.96)	802	RR: 0.84 (0.74 to 0.96)	RR: 0.77 (0.63 to 0.95)	0.19	Concordant[a] Concordant[b] Concordant[c]
Liakopoulos et al., 2012[71]	Length of stay in hospital	40	SMD:0.04 (-0.41 to 0.48)	837	SMD: -0.35 (-0.61 to -0.1)	SMD: -0.31 (-0.55 to -0.07)	0.11	Discordant[a] Discordant[b] Discordant[c]
Lopez-Olivo et al., 2014[72]	Adverse events	35	RR: 3.17 (0.14 to 72.77)	578	RR: 2.00 (1.25 to 3.19)	RR: 2.02 (1.27 to 3.20)	0.79	Discordant[a] Discordant[b] Discordant[c]
Main et al., 2013[73]	Stroke	2,763	RR: 1.11 (0.85 to 1.44)	30,434	RR: 1.30 (1.12 to 1.50)	RR: 1.25 (1.10 to 1.42)	0.31	Discordant[a] Discordant[b] Discordant[c]
Manheimer et al., 2010[74]	Function	284	SMD: -0.20 (-0.44 to 0.03)	1,114	SMD: -0.09 (-0.21 to 0.04)	SMD: -0.11 (-0.22—0.01)	0.38	Concordant[a] Concordant[b] Concordant[c]
Massel et al., 2013[75]	Thrombo-embolism	163	RR: 0.35 (0.12 to 1.05)	3,959	RR: 0.48 (0.35 to 0.65)	RR: 0.48 (0.36 to 0.64)	0.60	Discordant[a] Discordant[b] Discordant[c]

Table 1. Characteristics and concordance of results of first trials and pooled subsequent trials (continued)

	Inter-vention and Outcome	First Trial		Pooled Subsequent Trials		Combined	P-value of Difference	Qualitative Concor-dance (Using 3 Thresholds)
		N	Effect Estimate (Confidence Interval)	N	Effect Estimate (Confidence Interval)	Effect Estimate (Confidence Interval)		
Middeldorp et al., 2014[76]	Incidence of recurrent VTE	214	RR: 0.11 (0.01 to 0.83)	3,322	RR: 0.22 (0.12 to 0.40)	RR: 0.21 (0.12 to 0.37)	0.56	Concordant[a] Concordant[b] Concordant[c]
Moja et al., 2012[77]	Congestive heart failure	1,804	RR: 3.64 (1.76 to 7.53)	8,477	RR: 5.43 (2.42 to 12.17)	RR: 4.88 (2.64 to 9.04)	0.48	Discordant[a] Discordant[b] Discordant[c]
Mössler et al., 2011[78]	Mental state: Negative symptoms	76	SMD: -1.08 (-1.56 to -0.60)	164	SMD: -0.57 (-1.03 to -0.12)	SMD: -0.71 (-1.11 to -0.32)	0.09	Discordant[a] Discordant[b] Discordant[c]
Musini et al., 2014[79]	Systolic blood pressure	90	SMD: -1.08 (-1.54 to -0.62)	2,555	SMD: -0.48 (-0.56 to -0.40)	SMD: -0.50 (-0.58 to -0.42)	0.06	Discordant[a] Discordant[b] Discordant[c]
Musini et al., 2009[80]	Cardiovascular morbidity and mortality	81	RR: 0.41 (0.22 to 0.76)	23,013	RR: 0.75 (0.66 to 0.85)	RR: 0.74 (0.65 to 0.84)	0.01	Discordant[a] Concordant[b] Concordant[c]
Nannini et al., 2013[81]	Mortality	3,067	RR: 0.79 (0.66 to 0.93)	4,456	RR: 1.02 (0.58 to 1.79)	RR: 0.80 (0.68 to 0.95)	0.40	Discordant[a] Discordant[b] Discordant[c]
Nelson et al., 2011[82]	Persistence of the anal fissure	28	RR: 1.0 (0.02 to 47.08)	308	RR: 0.98 (0.42 to 2.33)	RR: 0.98 (0.42 to 2.28)	0.99	Concordant[a] Concordant[b] Concordant[c]
Nelson et al., 2012[83]	Healing	24	RR: 0.2 (0.01 to 3.76)	955	RR: 0.24 (0.15 to 0.4)	RR: 0.24 (0.15 to 0.40)	0.91	Discordant[a] Discordant[b] Discordant[c]
Nelson et al., 2014[84]	Surgical wound infection	83	RR: 0.57 (0.29 to 1.13)	2,372	RR: 0.34 (0.28 to 0.41)	RR: 0.35 (0.29 to 0.41)	0.15	Discordant[a] Discordant[b] Discordant[c]
Nüesch et al., 2010[85]	Withdrawal because of adverse events	115	RR: 3.61 (1.95 to 6.71)	2,247	RR: 4.21 (3.03 to 5.84)	RR: 3.96 (3.02 to 5.17)	0.68	Discordant[a] Discordant[b] Discordant[c]
Nussbaum et al., 2012[86]	Leaving the study early- any reason	247	RR: 0.59 (0.46 to 0.75)	1936	RR: 0.79 (0.71 to 0.87)	RR: 0.75 (0.66 to 0.86)	0.03	Concordant[a] Concordant[b] Concordant[c]
Pandian et al., 2013[87]	Multiple pregnancy rate	53	RR: 5.78 (0.75 to 44.76)	1,411	RR: 6.94 (2.39 to 20.13)	RR: 6.55 (2.61 to 16.45)	0.89	Discordant[a] Discordant[b] Discordant[c]
Pani et al., 2011[88]	Antide-pressants and alcohol abstinence	20	RR: 1.5 (0.87 to 2.59)	922	RR: 1.25 (0.88 to 1.79)	RR: 1.28 (0.95 to 1.73)	0.60	Discordant[a] Concordant[b] Discordant[c]
Paul et al., 2013[89]	Mortality	104	RR: 0.96 (0.4 to 2.29)	1,614	RR: 0.89 (0.75 to 1.05)	RR: 0.89 (0.76 to 1.05)	0.88	Discordant[a] Concordant[b] Discordant[c]
Perez et al., 2009[90]	Mortality	38	RR: 0.3 (0.01 to 6.94)	84,273	RR: 0.93 (0.88 to 0.98)	RR: 0.93 (0.88 to 0.98)	0.51	Discordant[a] Discordant[b] Discordant[c]
Rajaram et al., 2013[91]	Hospital length of stay	88	SMD: 0.04 (-0.40 to 0.48)	415	SMD: 0.18 (-0.02 to 0.38)	SMD: 0.15 (-0.03 to 0.34)	0.56	Concordant[a] Concordant[b] Concordant[c]

Table 1. Characteristics and concordance of results of first trials and pooled subsequent trials (continued)

	Inter-vention and Outcome	First Trial		Pooled Subsequent Trials		Combined		
		N	Effect Estimate (Confidence Interval)	N	Effect Estimate (Confidence Interval)	Effect Estimate (Confidence Interval)	P-value of Difference	Qualitative Concor-dance (Using 3 Thresholds)
Rehman et al., 2011[92]	Inconti-nence	142	RR: 0.74 (0.31 to 1.75)	292	RR: 1.00 (0.81 to 1.24)	RR: 0.98 (0.80 to 1.21)	0.52	Discordant[a] Concordant[b] Discordant[c]
Richards et al., 2013[93]	Physical activity as dichoto-mous outcome	329	RR: 1.18 (0.80 to 1.76)	2,948	RR: 1.35 (0.82 to 2.2)	RR: 1.25 (0.95 to 1.65)	0.69	Concordant[a] Concordant[b] Concordant[c]
Rubinstein et al., 2011[94]	Functional status	54	SMD: -0.18 (-0.73 to 0.36)	1,364	SMD: -0.05 (-0.16 to 0.06)	SMD: -0.06 (-0.16 to 0.05)	0.66	Concordant[a] Concordant[b] Concordant[c]
Santa Cruz et al., 2013[95]	Mortality before hospital discharge	549	RR: 0.91 (0.69 to 1.21)	1,750	RR: 0.90 (0.80 to 1.02)	RR: 0.90 (0.81 to 1.01)	0.95	Concordant[a] Concordant[b] Concordant[c]
Shepperd et al., 2011[96]	Dying at home	310	RR: 1.39 (1.16 to 1.67)	342	RR: 1.19 (0.92 to 1.55)	RR: 1.32 (1.14 to 1.54)	0.34	Discordant[a] Discordant[b] Discordant[c]
Sinclair et al., 2011[97]	Rate of cholera	5,582	RR: 0.30 (0.20 to 0.44)	23,423	RR: 0.36 (0.26 to 0.48)	RR: 0.34 (0.27 to 0.43)	0.48	Concordant[a] Concordant[b] Concordant[c]
Soares-Weiser et al., 2012[98]	Severe rotavirus diarrhea	215	RR: 0.22 (0.05 to 1.0)	40,416	RR: 0.13 (0.07 to 0.26)	RR: 0.14 (0.08 to 0.26)	0.54	Concordant[a] Concordant[b] Concordant[c]
Solomon et al., 2014[99]	Loss of fewer than 15 letters visual acuity	22	RR: 1.19 (0.84 to 1.68)	2,424	RR: 1.0 (0.98 to 1.02)	RR: 1.00 (0.98 to 1.02)	0.33	Concordant[a] Concordant[b] Concordant[c]
Spinks et al., 2013[100]	Incidence of otitis media	506	RR: 0.23 (0.01 to 4.67)	3,254	RR: 0.33 (0.17 to 0.62)	RR: 0.32 (0.17 to 0.61)	0.83	Discordant[a] Discordant[b] Discordant[c]
Swinnen et al., 2011[101]	Weight gain	319	SMD: -0.23 (-0.46 to 0.01)	1,931	SMD: -0.26 (-0.35 to -0.16)	SMD: -0.25 (-0.34 to -0.17)	0.82	Discordant[a] Discordant[b] Discordant[c]
Tacklind et al., 2012[102]	Peak urine flow	85	SMD: -0.08 (-0.51 to 0.34)	582	SMD: 0.11 (-0.05 to 0.27)	SMD: 0.09 (-0.06 to 0.24)	0.40	Concordant[a] Concordant[b] Concordant[c]
Tangsiri-watthana et al., 2013[103]	Pain	52	SMD: -0.13 (-0.68 to 0.41)	329	SMD: -0.47 (-0.87 to -0.06)	SMD: -0.40 (-0.73 to -0.07)	0.29	Discordant[a] Discordant[b] Discordant[c]
Thomas et al., 2014[104]	Caesarean section	80	RR: 0.50 (0.21 to 1.20)	966	RR: 0.97 (0.75 to 1.26)	RR: 0.92 (0.72 to 1.18)	0.15	Discordant[a] Concordant[b] Discordant[c]
Torrego et al., 2014[105]	Hospital admissions	109	RR: 1.96 (0.38 to 10.28)	320	RR: 4.97 (1.37 to 18.08)	RR: 3.50 (1.26 to 9.68)	0.39	Discordant[a] Discordant[b] Discordant[c]
Venekamp et al., 2013[106]	Vomiting, diarrhoea or rash	149	RR: 3.21 (0.34 to 30.14)	1,874	RR: 1.36 (1.17 to 1.57)	RR: 1.34 (1.16 to 1.55)	0.46	Discordant[a] Discordant[b] Discordant[c]

Table 1. Characteristics and concordance of results of first trials and pooled subsequent trials (continued)

	Inter-vention and Outcome	First Trial		Pooled Subsequent Trials		Combined		
		N	Effect Estimate (Confidence Interval)	N	Effect Estimate (Confidence Interval)	Effect Estimate (Confidence Interval)	P-value of Difference	Qualitative Concor-dance (Using 3 Thresholds)
Webster et al., 2013[107]	Catheter-related blood stream infection	206	RR: 1.0 (0.02 to 49.92)	4,600	RR: 0.73 (0.14 to 3.89)	RR: 0.77 (0.17 to 3.57)	0.89	Discordant[a] Concordant[b] Discordant[c]
Wilhelmus et al., 2010[108]	Healing of herpes simplex virus keratitis	70	RR: 1.20 (0.96 to 1.50)	331	RR: 2.1 (1.44 to 3.08)	RR: 1.86 (1.35 to 2,54)	0.01	Discordant[a] Discordant[b] Discordant[c]
Yeoh et al., 2012[109]	Blood lead level	95	SMD: -0.03 (-0.43 to 0.37)	720	SMD: 0.02 (-0.13 to 0.17)	SMD: 0.01 (-0.12 to 0.15)	0.82	Concordant[a] Concordant[b] Concordant[c]
Yuan et al., 2014[110]	Mortality	194	RR: 1.00 (0.21 to 4.83)	439	RR: 1.50 (0.43 to 5.23)	RR: 1.28 (0.49 to 3.39)	0.70	Discordant[a] Discordant[b] Discordant[c]

EDSS = Expanded Disability Status Scale; FEV = forced expiratory volume; RR = relative risk; SMD = standardized mean difference; VTE = venous thromboembolism

Concordance was defined as:

[a]No change in statistical significance; relative change in magnitude of effects <25 percent (<0.2 SMDs for continuous outcomes)

[b]No change in statistical significance; relative change in magnitude of effects <50 percent (<0.5 SMDs for continuous outcomes)

[c]No change in statistical significance and,

--relative change in magnitude of effects <25 percent (<0.2 SMDs for continuous outcomes) for small treatment effects, or

--relative change in magnitude of effects <50 percent (<0.2 SMDs for continuous outcomes) for large treatment effects, or

--relative change in magnitude of effects <10 percent for extremely patient-relevant outcomes.

Quantitative Concordance and Differences in Magnitudes of Effect Estimates

The concordance correlation between the point estimates of first trials and point estimates of the pooled subsequent trials was high (rho 0.78, where 1 indicates perfect concordance). When we took the variance of estimates into consideration (z-scores), the correlation coefficient decreased; however, estimates from first and pooled subsequent trials were still strongly correlated (rho 0.64).

We also set out to determine differences in the magnitudes of effect estimates between first trials and pooled subsequent trials using ratios of relative risks and SMDs as the outcome measures. Seven percent of comparisons revealed statistically significant differences between effect estimates of first trials and those of the pooled subsequent trials (see Table 1).

Based on a random effects meta-analysis, the mean difference between estimates of first trials and pooled subsequent trials was 0.16 SMD (95% CI, 0.12 to 0.21; Figure 3); the median was 0.21 SMD (interquartile range: 0.13 to 0.45).

Figure 3. Meta-analysis of differences of effect estimates between first and pooled subsequent trials

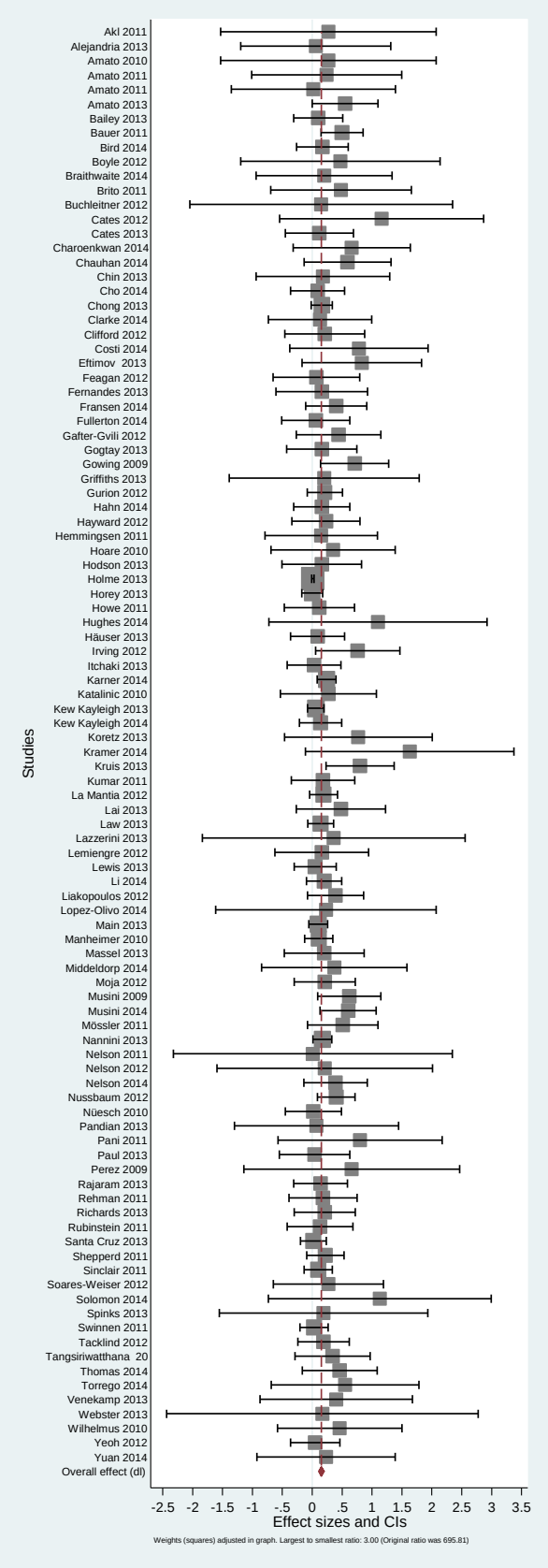
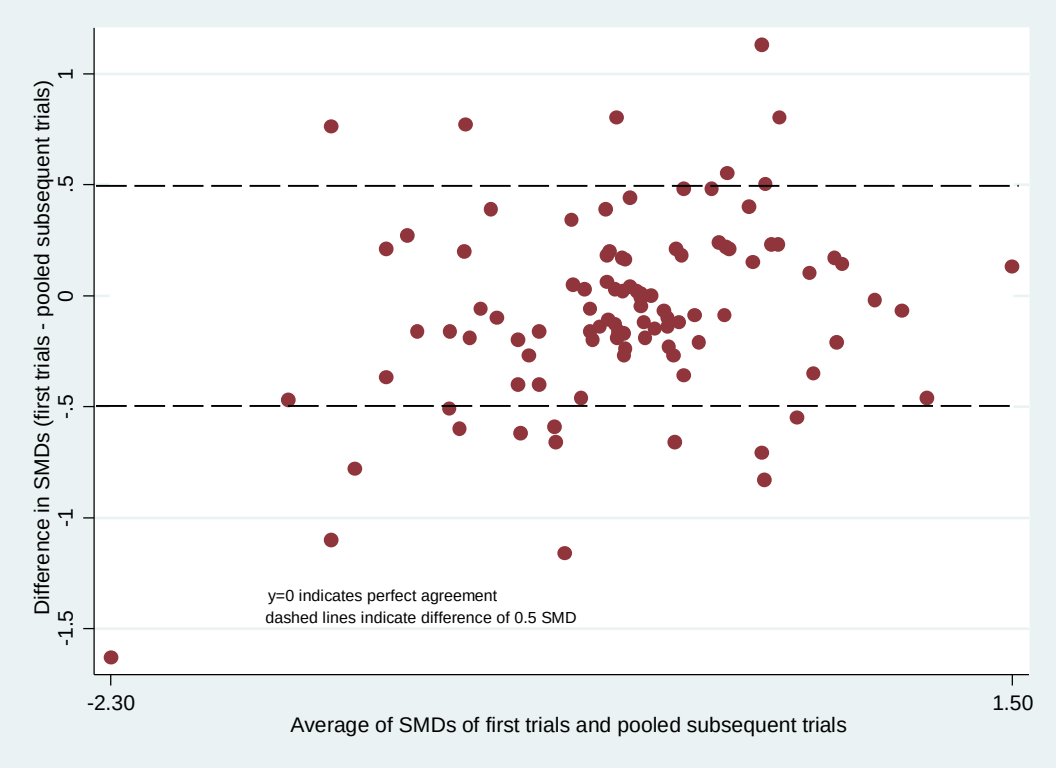


Figure 4 displays a modified Bland-Altman diagram comparing point estimates of first trials with point estimates of pooled subsequent trials. Bland-Altman diagrams plot the

relationship between each pair of estimates (i.e., first trial vs. pooled subsequent trials). The average of the two measurements, $\frac{(effect_{first\ trial} + effect_{pooled\ trials})}{2}$, is plotted on the x-axis; the difference between the measurements, $effect_{first\ trial} - effect_{pooled\ trials}$, is plotted on the y-axis. In 80 of 100 comparisons, the difference between the point estimate of the first trial and the pooled subsequent trials was smaller than 0.5 SMD (indicated by the dotted lines in Figure 4).

Figure 4. Modified Bland-Altman diagram comparing the point estimates of the first trials with the point estimates of pooled subsequent trials



SMD = standardized mean difference

To determine whether first trials, on average, favor the experimental intervention, we reordered data for dichotomous outcomes so that a ratio of relative risks of greater than 1 indicated that the first trial presented a more beneficial treatment effect of the experimental arm (either greater benefit or less harm, depending on the outcome) than the pooled subsequent trials. Overall, the pooled estimate of all ratios of relative risks indicated that first trials, on average, did not favor treatment effects of the experimental intervention (ratio of relative risks: 1.03, 95% CI, 0.98 to 1.08).

Stratified Analyses

In stratified analyses and meta-regression we explored the effects of sample sizes of first and pooled subsequent trials, risk of bias ratings, and magnitude of effects (effects of ≤0.5 SMDs vs. effect >0.5 SMDs) of first trials on the differences in effect estimates between first trials and pooled subsequent trials. Only magnitude of effect estimates of first trials had a statistically significant association ($p<0.001$) with differences in effect estimates between first trials and subsequent pooled trials. In first trials with treatment effects ≥0.5 SMD, the mean change of estimates of effect between first and pooled subsequent trials was 0.68 SMD

(95% CI, .50 to 0.86). Table 2 summarizes Lin's concordance correlations and differences in effect estimates stratified by different characteristics of first trials.

Table 2. Results of stratified analyses based on sample size, risk of bias, and magnitude of effect estimates of first trials

	Lin's Concordance Correlation (rho) of Point Estimates	Pooled Difference in SMDs Between First Trial and Pooled Subsequent Trials (95% CI)
Overall	0.75	0.16 (0.12 to 0.21)
First trials <= 300 participants	0.77	0.25 (0.18 to 0.33)
First trials > 300 participants	0.83	0.14 (0.08 to 0.20)
First trials with low RoB ratings	0.78	0.13 (0.07 to 0.20)
First trials with unclear RoB ratings	0.74	0.30 (0.20 to 0.39)
First trials with high RoB ratings	0.87	0.14 (0.07 to 0.21)
First trials with treatment effects <=0.5 SMDs	0.87	0.04 (0.02 to 0.05)
First trials with treatment effects >0.5 SMDs	0.87	0.68 (0.50 to 0.86)

CI = confidence interval; RoB = risk of bias; SMD = standardized mean difference

Discussion

To our knowledge, our study was the first attempt to use high-quality bodies of evidence as reference points to assess the magnitude of changes of effect estimates as bodies of evidence advance from a first trial to a reliable body of evidence graded as high QOE. The premise of our study was the GRADE approach and its definitions. GRADE links QOE grades to the degree of confidence that estimates are close to the true effect (and thus will remain stable as new evidence accrues). This concept can be criticized from a philosophical perspective because quantifiable entities (grades of QOE) are linked to an abstract concept (the truth) that can never be verified. Nevertheless, we purposely took GRADE definitions at face value. GRADE is used by more than 70 international organizations; most decisionmakers conceivably accept and rely on GRADE assessments and their current definitions.

Results of our meta-epidemiological research indicate that first trials both over- and under-estimate the true effects; we could find no discernable pattern. The average magnitude of difference, however, is relatively small. In 80 percent of comparisons the difference in effect estimates was smaller than 0.5 SMD, which is often viewed as a moderate treatment effect.[111,112] Depending on how we defined concordance of effects, effect estimates of first trials were concordant with pooled subsequent studies in at least 33 percent but up to 50 percent of these comparisons. The largest change in effect estimates occurred when first trials reported large treatment effects.

Our findings empirically confirm current GRADE and AHRQ guidance that systematic reviewers should not grade single-study bodies of evidence as very low (or insufficient for AHRQ EPCs) simply by default. Risk of bias, precision and magnitude of effect estimates, publication bias, or issues of directness vary substantially across single-study bodies of evidence; reviewers should take these factors into account to achieve a nuanced QOE assessment. Reviewers can, of course, still assign a grade of very low when warranted by problems such as imprecision, indirectness of evidence, or other problems such as publication bias.

A highly cited article by Pereira et al.[5] compared very large treatment effects (i.e., treatment effects with an odds ratio $\geq$5) of first trials with effect estimates of subsequent trials.[5] Those authors concluded that most large treatment effects become much smaller as new studies emerge. These findings are compatible with ours, although a substantial decrease in very large treatment effects of first trials as new evidence emerges is probably attributable mostly to regression to the mean. Trikalinos et al. and Dechartres et al. also report changes in treatment effects without a discernible pattern as evidence evolves.[6,8] Other studies focused mainly on genetic associations, which are difficult to compare with clinical interventions.[7,113] The main difference between our research and other studies is that our investigation focused on high-quality bodies of evidence as reference points with which we compared the effect estimates of first trials. Other studies included any meta-analysis, regardless of QOE assessments.[7,8,113]

Our study has several limitations. First, the size of our sample was limited. We included 100 bodies of evidence that had been graded as high QOE. Because of the consistency of our findings across a broad array of clinical topics, however, we believe that the overall conclusion would not change substantially with a larger sample of bodies of evidence.

Second, how representative our sample is remains unclear. Because we wanted to use a reference standard for which researchers had high confidence that effect estimates are correct (close to the true effect), we focused on high QOE evidence. A remaining question is whether bodies of evidence that will never progress to high QOE would have the same degree of concordance between first trials and subsequent trials as our sample.

Third, we relied on QOE grades of Cochrane authors. Because author groups differed across these systematic reviews, some heterogeneity in approaches regarding QOE grades is likely. Studies have shown that the inter-rater reliability of QOE grades is limited.[4,114] Nevertheless, such heterogeneity reflects a real-world situation because most guideline developers or other decisionmakers that use Cochrane reports to support decisions would not reassess QOE. In addition, Cochrane reports go through rigorous international peer review, and the methodological quality usual is high.

Finally, how to determine the effect of publication bias on results of our study is unclear. Realistically, publication bias could play a role in any systematic review. We were not able to assess whether a first trial in a meta-analysis of the Cochrane report was actually the first trial conducted. First trials without statistically significant results may have never been published. Also, the body of subsequent investigations may not represent that totality of evidence conducted for a given comparison. How these biases affect the difference in magnitudes of effect estimates between first and pooled subsequent trials is unclear.

Equations

Equation 1.	Lin's concordance correlation	Equation 1 is a linear equation for converting dichotomous outcomes to SMDs. The equation is as follows: d=log⁡ ⟦OddsRatio* √3/π
Equation 2.	Z-scores	Equation 2 is a linear type of equation for calculating z scores. The equation is as follows: z=(⟦effect⟧_(first trial)-⟦effect⟧_ (pooled trials))/(√(⟦⟦variance⟧_(first trial)⟧^ +⟦⟦variance⟧_(pooled trials)⟧^))
Equation 3.	Bland-Altman Diagram	Equation 3 is a graphical type of equation for plotting the relationship between a pair of estimates. The equation is as follows: ((⟦effect⟧_ (first trial)+ ⟦effect⟧_(pooled trials)))/2
Equation 4.	Bland-Altman Diagram	Equation 4 is a graphical type of equation for plotting the relationship between a pair of estimates. The equation is as follows: ⟦effect⟧_(first trial)-⟦effect⟧_(pooled trials)

References

1. Gartlehner G, Sommer I, Swinson Evans T, et al. Grades for quality of evidence are associated with distinct likelihoods that treatments effects will remain stable. J Clinical Epidemiology. 2014.
2. Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clinical Epidemiology. 2011 Apr;64(4):383-94. Epub: 2011/01/05. PMID: 21195583.
3. Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clinical Epidemiology. 2011;64(4):401-6.
4. Berkman ND, Lohr KN, Ansari M, et al. Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update. Methods Guide for Comparative Effectiveness Reviews (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 290- 2007-10056-I). AHRQ Publication No. 13(14)-EHC130-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.
5. Pereira TV, Horwitz RI, Ioannidis JP. Empirical evaluation of very large treatment effects of medical interventions. JAMA. 2012 Oct 24;308(16):1676-84. Epub: 2012/10/25. PMID: 23093165.
6. Trikalinos TA, Churchill R, Ferri M, et al. Effect sizes in cumulative meta-analyses of mental health randomized trials evolved over time. J Clinical Epidemiology. 2004 Nov;57(11):1124-30. PMID: 15612138.
7. Ioannidis JP, Trikalinos TA. Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials. J Clinical Epidemiology. 2005 Jun;58(6):543-9. PMID: 15878467.
8. Dechartres A, Trinquart L, Boutron I, et al. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. BMJ. 2013;346:f2304. PMID: 23616031.
9. Shojania KG, Sampson M, Ansari MT, et al. Updating Systematic Reviews. Technical Review No. 16. AHRQ Publication No. 07-0087. Rockville, MD: Agency for Healthcare Research and Quality; 2007.
10. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics. 1989 Mar;45(1):255-68. Epub: 1989/03/01. PMID: 2720055.
11. Akl Elie A, Gunukula S, Barba M, et al. Parenteral anticoagulation in patients with cancer who have no therapeutic or prophylactic indication for anticoagulation. Cochrane Database of Systematic Reviews. 2011(4). PMID: CD006652.
12. Alejandria Marissa M, Lansang Mary Ann D, Dans Leonila F, et al. Intravenous immunoglobulin for treating sepsis, severe sepsis and septic shock. Cochrane Database of Systematic Reviews. 2013(9). PMID: CD001090.
13. Amato L, Minozzi S, Vecchi S, et al. Benzodiazepines for alcohol withdrawal. Cochrane Database of Systematic Reviews. 2010(3). PMID: CD005063.
14. Amato L, Minozzi S, Davoli M, et al. Psychosocial combined with agonist maintenance treatments versus agonist maintenance treatments alone for treatment of opioid dependence. Cochrane Database of Systematic Reviews. 2011(10). PMID: CD004147.
15. Amato L, Minozzi S, Pani Pier P, et al. Dopamine agonists for the treatment of cocaine dependence. Cochrane Database of Systematic Reviews. 2011(12). PMID: CD003352.
16. Amato L, Davoli M, Minozzi S, et al. Methadone at tapered doses for the management of opioid withdrawal. Cochrane Database of Systematic Reviews. 2013(2). PMID: CD003409.

17. Bailey E, Worthington Helen V, van Wijk A, et al. Ibuprofen and/or paracetamol (acetaminophen) for pain relief after surgical removal of lower wisdom teeth. Cochrane Database of Systematic Reviews. 2013(12). PMID: CD004624.

18. Bauer K, Skoetz N, Monsef I, et al. Comparison of chemotherapy including escalated BEACOPP versus chemotherapy including ABVD for patients with early unfavourable or advanced stage Hodgkin lymphoma. Cochrane Database of Systematic Reviews. 2011(8). PMID: CD007941.

19. Bird S, Derry S, Moore RA. Zolmitriptan for acute migraine attacks in adults. Cochrane Database of Systematic Reviews. 2014(5). PMID: CD008616.

20. Boyle Robert J, Elremeli M, Hockenhull J, et al. Venom immunotherapy for preventing allergic reactions to insect stings. Cochrane Database of Systematic Reviews. 2012(10). PMID: CD008838.

21. Braithwaite T, Nanji Afshan A, Lindsley K, et al. Anti-vascular endothelial growth factor for macular oedema secondary to central retinal vein occlusion. Cochrane Database of Systematic Reviews. 2014(5). PMID: CD007325.

22. Brito V, Ciapponi A, Kwong J. Factor Xa inhibitors for acute coronary syndromes. Cochrane Database of Systematic Reviews. 2011(1). PMID: CD007038.

23. Buchleitner Ana M, Martínez-Alonso M, Hernández M, et al. Perioperative glycaemic control for diabetic patients undergoing surgery. Cochrane Database of Systematic Reviews. 2012(9). PMID: CD007315.

24. Cates Christopher J, Cates Matthew J. Regular treatment with formoterol for chronic asthma: serious adverse events. Cochrane Database of Systematic Reviews. 2012(4). PMID: CD006923.

25. Cates Christopher J, Welsh Emma J, Rowe Brian H. Holding chambers (spacers) versus nebulisers for beta-agonist treatment of acute asthma. Cochrane Database of Systematic Reviews. 2013(9). PMID: CD000052.

26. Charoenkwan K, Kietpeerakool C. Retroperitoneal drainage versus no drainage after pelvic lymphadenectomy for the prevention of lymphocyst formation in patients with gynaecological malignancies. Cochrane Database of Systematic Reviews. 2014(6). PMID: CD007387.

27. Chauhan BF, Ducharme FM. Addition to inhaled corticosteroids of long-acting beta2-agonists versus anti-leukotrienes for chronic asthma. Cochrane Database Syst Rev. 2014;1:CD003137. Epub: 2014/01/25. PMID: 24459050.

28. Chin KJ, Alakkad H, Adhikary SD, et al. Infraclavicular brachial plexus block for regional anaesthesia of the lower arm. Cochrane Database Syst Rev. 2013;8:CD005487. Epub: 2013/08/30. PMID: 23986434.

29. Cho Y, Johnson David W, Craig Jonathan C, et al. Biocompatible dialysis fluids for peritoneal dialysis. Cochrane Database of Systematic Reviews. 2014(3). PMID: CD007554.

30. Chong J, Leung B, Poole P. Phosphodiesterase 4 inhibitors for chronic obstructive pulmonary disease. Cochrane Database of Systematic Reviews. 2013(11). PMID: CD002309.

31. Clarke R, Derry S, Moore RA. Single dose oral etoricoxib for acute postoperative pain in adults. Cochrane Database of Systematic Reviews. 2014(5). PMID: CD004309.

32. Clifford David M, Fisher Sheila A, Brunskill Susan J, et al. Stem cell treatment for acute myocardial infarction. Cochrane Database of Systematic Reviews. 2012(2). PMID: CD006536.

33. Costi D, Cyna Allan M, Ahmed S, et al. Effects of sevoflurane versus other general anaesthesia on emergence agitation in children. Cochrane Database of Systematic Reviews. 2014(9). PMID: CD007084.

34. Eftimov F, Winer John B, Vermeulen M, et al. Intravenous immunoglobulin for chronic inflammatory demyelinating polyradiculoneuropathy. Cochrane Database of Systematic Reviews. 2013(12). PMID: CD001797.

35. Feagan Brian G, MacDonald John K. Oral 5-aminosalicylic acid for maintenance of remission in ulcerative colitis. Cochrane Database of Systematic Reviews. 2012(10). PMID: CD000544.

36. Fernandes RM, Bialy LM, Vandermeer B, et al. Glucocorticoids for acute viral bronchiolitis in infants and young children. Cochrane Database Syst Rev. 2013;6:CD004878. Epub: 2013/06/05. PMID: 23733383.

37. Fransen M, McConnell S, Hernandez-Molina G, et al. Exercise for osteoarthritis of the hip. Cochrane Database of Systematic Reviews. 2014(4). PMID: CD007912.

38. Fullerton B, Jeitler K, Seitz M, et al. Intensive glucose control versus conventional glucose control for type 1 diabetes mellitus. Cochrane Database of Systematic Reviews. 2014(2). PMID: CD009122.

39. Gafter-Gvili A, Fraser A, Paul M, et al. Antibiotic prophylaxis for bacterial infections in afebrile neutropenic patients following chemotherapy. Cochrane Database Syst Rev. 2012;1:CD004386. Epub: 2012/01/20. PMID: 22258955.

40. Gogtay N, Kannan S, Thatte Urmila M, et al. Artemisinin-based combination therapy for treating uncomplicated Plasmodium vivax malaria. Cochrane Database of Systematic Reviews. 2013(10). PMID: CD008492.

41. Gowing L, Ali R, White JM. Buprenorphine for the management of opioid withdrawal. Cochrane Database Syst Rev. 2009(3):CD002025. Epub: 2009/07/10. PMID: 19588330.

42. Griffiths B, Ducharme Francine M. Combined inhaled anticholinergics and short-acting beta2-agonists for initial treatment of acute asthma in children. Cochrane Database of Systematic Reviews. 2013(8). PMID: CD000060.

43. Gurion R, Belnik-Plitman Y, Gafter-Gvili A, et al. Colony-stimulating factors for prevention and treatment of infectious complications in patients with acute myelogenous leukemia. Cochrane Database of Systematic Reviews. 2012(6). PMID: CD008238.

44. Hahn D, Cody June D, Hodson Elisabeth M. Frequency of administration of erythropoiesis-stimulating agents for the anaemia of end-stage kidney disease in dialysis patients. Cochrane Database of Systematic Reviews. 2014(5). PMID: CD003895.

45. Hauser W, Urrutia G, Tort S, et al. Serotonin and noradrenaline reuptake inhibitors (SNRIs) for fibromyalgia syndrome. Cochrane Database Syst Rev. 2013;1:CD010292. Epub: 2013/02/27. PMID: 23440848.

46. Hayward G, Thompson Matthew J, Perera R, et al. Corticosteroids as standalone or add-on treatment for sore throat. Cochrane Database of Systematic Reviews. 2012(10). PMID: CD008268.

47. Hemmingsen B, Lund Søren S, Gluud C, et al. Targeting intensive glycaemic control versus targeting conventional glycaemic control for type 2 diabetes mellitus. Cochrane Database of Systematic Reviews. 2013(11). PMID: CD008143.

48. Hoare Brian J, Wallen Margaret A, Imms C, et al. Botulinum toxin A as an adjunct to treatment in the management of the upper limb in children with spastic cerebral palsy (UPDATE). Cochrane Database of Systematic Reviews. 2010(1). PMID: CD003469.

49. Hodson Elisabeth M, Ladhani M, Webster Angela C, et al. Antiviral medications for preventing cytomegalovirus disease in solid organ transplant recipients. Cochrane Database of Systematic Reviews. 2013(2). PMID: CD003774.

50. Holme Ø, Bretthauer M, Fretheim A, et al. Flexible sigmoidoscopy versus faecal occult blood testing for colorectal cancer screening in asymptomatic individuals. Cochrane Database of Systematic Reviews. 2013(9). PMID: CD009259.

51. Horey D, Kealy M, Davey M-A, et al. Interventions for supporting pregnant women's decision-making about mode of birth after a caesarean. Cochrane Database of Systematic Reviews. 2013(7). PMID: CD010041.

52. Howe Tracey E, Shea B, Dawson Lesley J, et al. Exercise for preventing and treating osteoporosis in postmenopausal women. Cochrane Database of Systematic Reviews. 2011(7). PMID: CD000333.

53. Hughes Richard AC, Swan Anthony V, van Doorn Pieter A. Intravenous immunoglobulin for Guillain-Barré syndrome. Cochrane Database of Systematic Reviews. 2014(9). PMID: CD002063.

54. Irving Greg J, Holden J, Yang R, et al. Hepatitis A immunisation in persons not previously exposed to hepatitis A. Cochrane Database of Systematic Reviews. 2012(7). PMID: CD009051.

55. Itchaki G, Gafter-Gvili A, Lahav M, et al. Anthracycline-containing regimens for treatment of follicular lymphoma in adults. Cochrane Database of Systematic Reviews. 2013(7). PMID: CD008909.

56. Karner C, Chong J, Poole P. Tiotropium versus placebo for chronic obstructive pulmonary disease. Cochrane Database of Systematic Reviews. 2014(7). PMID: CD009285.

57. Katalinic Owen M, Harvey Lisa A, Herbert Robert D, et al. Stretch for the treatment and prevention of contractures. Cochrane Database of Systematic Reviews. 2010(9). PMID: CD007455.

58. Kew Kayleigh M, Karner C, Mindus Stephanie M, et al. Combination formoterol and budesonide as maintenance and reliever therapy versus combination inhaler maintenance for chronic asthma in adults and children. Cochrane Database of Systematic Reviews. 2013(12). PMID: CD009019.

59. Kew Kayleigh M, Seniukovich A. Inhaled steroids and risk of pneumonia for chronic obstructive pulmonary disease. Cochrane Database of Systematic Reviews. 2014(3). PMID: CD010115.

60. Koretz Ronald L, Pleguezuelo M, Arvaniti V, et al. Interferon for interferon nonresponding and relapsing patients with chronic hepatitis C. Cochrane Database of Systematic Reviews. 2013(1). PMID: CD003617.

61. Kramer Christine V, Zhang F, Sinclair D, et al. Drugs for treating urinary schistosomiasis. Cochrane Database of Systematic Reviews. 2014(8). PMID: CD000053.

62. Kruis Annemarije L, Smidt N, Assendelft Willem JJ, et al. Integrated disease management interventions for patients with chronic obstructive pulmonary disease. Cochrane Database of Systematic Reviews. 2013(10). PMID: CD009437.

63. Kumar P, Chawla D, Deorari A. Light-emitting diode phototherapy for unconjugated hyperbilirubinaemia in neonates. Cochrane Database of Systematic Reviews. 2011(12). PMID: CD007969.

64. La Mantia L, Vacchi L, Di Pietrantonj C, et al. Interferon beta for secondary progressive multiple sclerosis. Cochrane Database of Systematic Reviews. 2012(1). PMID: CD005181.

65. Lai Nai M, Chaiyakunapruk N, Lai Nai A, et al. Catheter impregnation, coating or bonding for reducing central venous catheter-related infections in adults. Cochrane Database of Systematic Reviews. 2013(6). PMID: CD007878.

66. Law S, Derry S, Moore RA. Sumatriptan plus naproxen for acute migraine attacks in adults. Cochrane Database of Systematic Reviews. 2013(10). PMID: CD008541.

67. Lazzerini M, Ronfani L. Oral zinc for treating diarrhoea in children. Cochrane Database of Systematic Reviews. 2013(1). PMID: CD005436.

68. Lemiengre MB, van Driel ML, Merenstein D, et al. Antibiotics for clinically diagnosed acute rhinosinusitis in adults. Cochrane Database Syst Rev. 2012;10:CD006089. Epub: 2012/10/19. PMID: 23076918.

69. Lewis Sharon R, Nicholson A, Cardwell Mary E, et al. Nonsteroidal anti-inflammatory drugs and perioperative bleeding in paediatric tonsillectomy. Cochrane Database of Systematic Reviews. 2013(7). PMID: CD003591.

70. Li Edmond CK, Heran Balraj S, Wright James M. Angiotensin converting enzyme (ACE) inhibitors versus angiotensin receptor blockers for primary hypertension. Cochrane Database of Systematic Reviews. 2014(8). PMID: CD009096.

71. Liakopoulos Oliver J, Kuhn Elmar W, Slottosch I, et al. Preoperative statin therapy for patients undergoing cardiac surgery. Cochrane Database of Systematic Reviews. 2012(4). PMID: CD008493.

72. Lopez-Olivo Maria A, Siddhanamatha Harish R, Shea B, et al. Methotrexate for treating rheumatoid arthritis. Cochrane Database of Systematic Reviews. 2014(6). PMID: CD000957.

73. Main C, Knight B, Moxham T, et al. Hormone therapy for preventing cardiovascular disease in post-menopausal women. Cochrane Database of Systematic Reviews. 2013(4). PMID: CD002229.

74. Manheimer E, Cheng K, Linde K, et al. Acupuncture for peripheral joint osteoarthritis. Cochrane Database of Systematic Reviews. 2010(1). PMID: CD001977.

75. Massel David R, Little Stephen H. Antiplatelet and anticoagulation for patients with prosthetic heart valves. Cochrane Database of Systematic Reviews. 2013(7). PMID: CD003464.

76. Middeldorp S, Prins Martin H, Hutten Barbara A. Duration of treatment with vitamin K antagonists in symptomatic venous thromboembolism. Cochrane Database of Systematic Reviews. 2014(8). PMID: CD001367.

77. Moja L, Tagliabue L, Balduzzi S, et al. Trastuzumab containing regimens for early breast cancer. Cochrane Database of Systematic Reviews. 2012(4). PMID: CD006243.

78. Mössler K, Chen X, Heldal Tor O, et al. Music therapy for people with schizophrenia and schizophrenia-like disorders. Cochrane Database of Systematic Reviews. 2011(12). PMID: CD004025.

79. Musini Vijaya M, Nazer M, Bassett K, et al. Blood pressure-lowering efficacy of monotherapy with thiazide diuretics for primary hypertension. Cochrane Database of Systematic Reviews. 2014(5). PMID: CD003824.

80. Musini Vijaya M, Tejani Aaron M, Bassett K, et al. Pharmacotherapy for hypertension in the elderly. Cochrane Database of Systematic Reviews. 2009(4). PMID: CD000028.

81. Nannini LJ, Poole P, Milan SJ, et al. Combined corticosteroid and long-acting beta2-agonist in one inhaler versus inhaled corticosteroids alone for chronic obstructive pulmonary disease. Cochrane Database Syst Rev. 2013;8:CD006826. Epub: 2013/08/31. PMID: 23990350.

82. Nelson Richard L, Chattopadhyay A, Brooks W, et al. Operative procedures for fissure in ano. Cochrane Database of Systematic Reviews. 2011(11). PMID: CD002199.

83. Nelson Richard L, Thomas K, Morgan J, et al. Non surgical therapy for anal fissure. Cochrane Database of Systematic Reviews. 2012(2). PMID: CD003431.

84. Nelson Richard L, Gladman E, Barbateskovic M. Antimicrobial prophylaxis for colorectal surgery. Cochrane Database of Systematic Reviews. 2014(5). PMID: CD001181.

85. Nuesch E, Rutjes AW, Husni E, et al. Oral or transdermal opioids for osteoarthritis of the knee or hip. Cochrane Database Syst Rev. 2009(4):CD003115. Epub: 2009/10/13. PMID: 19821302.

86. Nussbaum Abraham M, Stroup TS. Paliperidone palmitate for schizophrenia. Cochrane Database of Systematic Reviews. 2012(6). PMID: CD008296.

87. Pandian Z, Marjoribanks J, Ozturk O, et al. Number of embryos for transfer following in vitro fertilisation or intra-cytoplasmic sperm injection. Cochrane Database of Systematic Reviews. 2013(7). PMID: CD003416.

88. Pani Pier P, Trogu E, Vecchi S, et al. Antidepressants for cocaine dependence and problematic cocaine use. Cochrane Database of Systematic Reviews. 2011(12). PMID: CD002950.

89. Paul M, Dickstein Y, Schlesinger A, et al. Beta-lactam versus beta-lactam-aminoglycoside combination therapy in cancer patients with neutropenia. Cochrane Database of Systematic Reviews. 2013(6). PMID: CD003038.

90. Perez Marco I, Musini Vijaya M, Wright James M. Effect of early treatment with anti-hypertensive drugs on short and long-term mortality in patients with an acute cardiovascular event. Cochrane Database of Systematic Reviews. 2009(4). PMID: CD006743.

91. Rajaram Sujanthy S, Desai Nayan K, Kalra A, et al. Pulmonary artery catheters for adult patients in intensive care. Cochrane Database of Systematic Reviews. 2013(2). PMID: CD003408.

92. Rehman H, Bezerra CC, Bruschini H, et al. Traditional suburethral sling operations for urinary incontinence in women. Cochrane Database Syst Rev. 2011(1):CD001754. Epub: 2011/01/21. PMID: 21249648.

93. Richards J, Hillsdon M, Thorogood M, et al. Face-to-face interventions for promoting physical activity. Cochrane Database of Systematic Reviews. 2013(9). PMID: CD010392.

94. Rubinstein Sidney M, van Middelkoop M, Assendelft Willem JJ, et al. Spinal manipulative therapy for chronic low-back pain. Cochrane Database of Systematic Reviews. 2011(2). PMID: CD008112.

95. Santa Cruz R, Rojas Juan I, Nervi R, et al. High versus low positive end-expiratory pressure (PEEP) levels for mechanically ventilated adult patients with acute lung injury and acute respiratory distress syndrome. Cochrane Database of Systematic Reviews. 2013(6). PMID: CD009098.

96. Shepperd S, Wee B, Straus Sharon E. Hospital at home: home-based end of life care. Cochrane Database of Systematic Reviews. 2011(7). PMID: CD009231.

97. Sinclair D, Abba K, Zaman K, et al. Oral vaccines for preventing cholera. Cochrane Database of Systematic Reviews. 2011(3). PMID: CD008603.

98. Soares-Weiser K, MacLehose H, Bergman H, et al. Vaccines for preventing rotavirus diarrhoea: vaccines in use. Cochrane Database of Systematic Reviews. 2012(11). PMID: CD008521.

99. Solomon Sharon D, Lindsley K, Vedula Satyanarayana S, et al. Anti-vascular endothelial growth factor for neovascular age-related macular degeneration. Cochrane Database of Systematic Reviews. 2014(8). PMID: CD005139.

100. Spinks A, Glasziou Paul P, Del Mar Chris B. Antibiotics for sore throat. Cochrane Database of Systematic Reviews. 2013(11). PMID: CD000023.

101. Swinnen Sanne G, Simon Airin CR, Holleman F, et al. Insulin detemir versus insulin glargine for type 2 diabetes mellitus. Cochrane Database of Systematic Reviews. 2011(7). PMID: CD006383.

102. Tacklind J, MacDonald R, Rutks I, et al. Serenoa repens for benign prostatic hyperplasia. Cochrane Database of Systematic Reviews. 2012(12). PMID: CD001423.

103. Tangsiriwatthana T, Sangkomkamhang Ussanee S, Lumbiganon P, et al. Paracervical local anaesthesia for cervical dilatation and uterine intervention. Cochrane Database of Systematic Reviews. 2013(9). PMID: CD005056.

104. Thomas J, Fairclough A, Kavanagh J, et al. Vaginal prostaglandin (PGE2 and PGF2a) for induction of labour at term. Cochrane Database of Systematic Reviews. 2014(6). PMID: CD003101.

105. Torrego A, Solà I, Munoz Ana M, et al. Bronchial thermoplasty for moderate or severe persistent asthma in adults. Cochrane Database of Systematic Reviews. 2014(3). PMID: CD009910.

106. Venekamp Roderick P, Sanders S, Glasziou Paul P, et al. Antibiotics for acute otitis media in children. Cochrane Database of Systematic Reviews. 2013(1). PMID: CD000219.

107. Webster J, Osborne S, Rickard Claire M, et al. Clinically-indicated replacement versus routine replacement of peripheral venous catheters. Cochrane Database of Systematic Reviews. 2013(4). PMID: CD007798.

108. Wilhelmus Kirk R. Antiviral treatment and other therapeutic interventions for herpes simplex virus epithelial keratitis. Cochrane Database of Systematic Reviews. 2010(12). PMID: CD002898.

109. Yeoh B, Woolfenden S, Lanphear B, et al. Household interventions for preventing domestic lead exposure in children. Cochrane Database of Systematic Reviews. 2012(4). PMID: CD006047.

110. Yuan Y, Zeng X, Hu Y, et al. Omentoplasty for oesophagogastrostomy after oesophagectomy. Cochrane Database of Systematic Reviews. 2014(10). PMID: CD008446.

111. Cohen J. Statistical Power Analysis for the Behavioral Sciences (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

112. Norman GR, Sloan JA, Wyrwich KW. The truly remarkable universality of half a standard deviation: confirmation through another look. Expert Rev Pharmacoecon Outcomes Res. 2004 Oct;4(5):581-5. PMID: 19807551.

113. Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG, et al. Establishment of genetic associations for complex diseases is independent of early study findings. Eur J Hum Genet. 2004 Sep;12(9):762-9. Epub: 2004/06/24. PMID: 15213707.

114. Mustafa RA, Santesso N, Brozek J, et al. The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. J Clinical Epidemiology. 2013 Jul;66(7):736-42 e5. PMID: 23623694.