

## Scientific hypotheses can be tested by comparing the effects of one treatment over many diseases in a systematic review

Chen, Yen-fu; Hemming, Karla; Chilton, Peter J.; Gupta, Keshav K.; Altman, Douglas G.; Lilford, Richard J.

DOI:

[10.1016/j.jclinepi.2014.08.007](https://doi.org/10.1016/j.jclinepi.2014.08.007)

License:

Creative Commons: Attribution (CC BY)

*Document Version*

Publisher's PDF, also known as Version of record

*Citation for published version (Harvard):*

Chen, Y, Hemming, K, Chilton, PJ, Gupta, KK, Altman, DG & Lilford, RJ 2014, 'Scientific hypotheses can be tested by comparing the effects of one treatment over many diseases in a systematic review', *Journal of Clinical Epidemiology*, vol. 67, no. 12, pp. 1309-1319. <https://doi.org/10.1016/j.jclinepi.2014.08.007>

[Link to publication on Research at Birmingham portal](#)

### **Publisher Rights Statement:**

Eligibility for repository : checked 17/12/2014

### **General rights**

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

### **Take down policy**

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

# Scientific hypotheses can be tested by comparing the effects of one treatment over many diseases in a systematic review

Yen-Fu Chen<sup>a,b</sup>, Karla Hemming<sup>a</sup>, Peter J. Chilton<sup>a</sup>, Keshav K. Gupta<sup>c</sup>, Douglas G. Altman<sup>d</sup>, Richard J. Lilford<sup>a,b,\*</sup>

<sup>a</sup>School of Health and Population Sciences, University of Birmingham, Birmingham B15 2TT, UK

<sup>b</sup>Division of Health Sciences, University of Warwick, Coventry CV4 7AL, UK

<sup>c</sup>Faculty of Medicine, Imperial College London, London SW7 2AZ, UK

<sup>d</sup>Centre for Statistics in Medicine, University of Oxford, Oxford OX3 7LD, UK

Accepted 1 August 2014; Published online 2 October 2014

## Abstract

**Objectives:** To describe the use of systematic reviews or overviews (systematic reviews of systematic reviews) to synthesize quantitative evidence of intervention effects across multiple indications (multiple-indication reviews) and to highlight issues pertaining to such reviews.

**Study Design and Setting:** MEDLINE was searched from 2003 to January 2014. We selected multiple-indication reviews of interventions of allopathic medicine that included evidence from randomized controlled trials. We categorized the subject areas evaluated by these reviews and examined their methodology. Utilities and caveats of multiple-indication reviews are illustrated with examples drawn from published literature.

**Results:** We retrieved 52 multiple-indication reviews covering a wide range of interventions. The method has been used to detect unintended effects, improve precision by pooling results across indications, and examine scientific hypotheses across disease classes.

**Conclusion:** Systematic reviews of interventions are typically used to evaluate the effects of treatments, one indication at a time. Here, we argue that, with due attention to methodological caveats, much can be learned by comparing the effects of a given treatment across many related indications. © 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

**Keywords:** Multiple-indication reviews; Panoramic meta-analyses; Research methods; Detecting unintended effects; Evaluating effectiveness; Overviews; Assessing harms

## 1. Introduction

Systematic reviews of randomized controlled trials (RCTs) underpin the practice of evidence-based medicine, and the statistical technique of meta-analysis to pool

quantitative results over studies has become the most widely cited form of clinical research [1]. These methods have undoubtedly contributed to the advance of health care in the past two decades and have become the gold standard

Conflict of interest: None.

Funding: K.H., P.J.C., and R.J.L. acknowledge financial support for the submitted work from the National Institute for Health Research (NIHR) Collaborations for Leadership in Applied Health Research and Care (CLAHRC) for Birmingham and Black Country, and CLAHRC West Midlands; the Engineering and Physical Sciences Research Council (EPSRC) Multidisciplinary Assessment of Technology Centre for Healthcare (MATCH) program (EPSRC Grant GR/S29874/01); and the Medical Research Council Midland Hub for Trials Methodology Research program (MRC Grant G0800808). D.G.A. is supported by a program grant from Cancer Research UK (C5529). Y.-F.C. acknowledges financial support for the submitted work from the National Institute for Health and Care

Excellence (External Assessment Centre, contract number 572) and the NIHR Surgical Reconstruction and Microbiology Research Centre (SRMRC). The funders had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Provenance: The idea for this article stemmed from an interest in research methodology.

\* Corresponding author. Tel.: +44-(0)24765-75884; fax: +44-(0)24765-28375.

E-mail address: [r.j.lilford@warwick.ac.uk](mailto:r.j.lilford@warwick.ac.uk) (R.J. Lilford).

### What is new?

#### Key findings

- There can be benefits and new insights when using systematic reviews to compare and combine the effects of a treatment across a range of different indications—multiple-indication reviews.

#### What this adds to what was known?

- Multiple-indication reviews are increasingly being used to evaluate both desirable and unintended treatment effects, to improve the precision of effect estimates, and/or to explore potential effect modification by treatment indication.

#### What is the implication and what should change now?

- Producers and commissioners of systematic reviews and developers of clinical guidelines should consider using multiple-indication reviews instead of, or in addition to, reviews focusing on a single indication.
- Attention needs to be given to heterogeneity, potential confounding factors, and various biases at both trial and review level when undertaking a multiple-indication review.
- Suitable statistical methods can be used to examine and allow for variations both between individual studies and between different indications.

for synthesizing evidence to inform clinical and policy decisions. In a typical scenario of undertaking evidence synthesis for these purposes (such as a Cochrane systematic review and a health technology assessment report), a narrowly focused question or “decision problem” is formulated, which specifies the patient population, intervention(s), comparator(s), and outcome(s) to be covered. This approach allows evidence synthesis to be conducted within a clearly defined scope, which ensures that the task is manageable under a tight timeline and that the reviewed evidence is directly applicable to the question being set. Nevertheless, it is not uncommon that such an exercise produces inconclusive findings because of insufficient evidence. It also ignores theoretical and practical learning that can be achieved by taking a broader approach, such as comparisons of different treatments for a given indication [2], and “meta-epidemiologic studies” that compare the effect of study design on findings [3].

This article is concerned with a type of review that goes beyond the common approach of looking at a specific patient population (indication) and instead examines the effects of a given treatment across different indications. We

shall refer to such systematic reviews, where the effects of a given intervention are evaluated over multiple clinical indications or diseases, as “multiple-indication reviews.” In this article, we describe the use of such reviews in recent literature and highlight the potential contributions and methodological considerations of their use with some examples. We focus our attention on synthesis of RCT evidence on the effectiveness and harms of allopathic medicine and suggest other possible uses of the method in the discussion.

## 2. Systematic review

### 2.1. Method of review

#### 2.1.1. Search strategy

There is no standardized terminology for reviews of studies across multiple indications. As we shall see, a number of different terms are used to describe this type of review, and authors may conduct a multiple-indication review without giving it any particular moniker. After several iterations, we devised a search strategy that aims to capture systematic reviews and meta-analyses that have examined evidence for a treatment across different indications. As some of the multiple-indication reviews known to us were undertaken in the form of a review of systematic reviews (often termed “overviews”), the search strategy also specifically targets this type of study. The final search strategy is shown in [Appendix A](#) at [www.jclinepi.com](http://www.jclinepi.com). We searched the MEDLINE database for articles published in the English language between January 2003 and January 2014. In addition, Cochrane overviews were sought by searching The Cochrane Database of Systematic Reviews using the text word “overview.” The main purpose of this search was to capture a sufficiently wide range of multiple-indication reviews to describe the rationale and methods used.

#### 2.1.2. Study selection

Titles and abstracts of the records retrieved from the search were sifted by at least two of the authors independently. Full-text articles were obtained for records that were considered potentially relevant by at least one of the reviewers. Final decisions on inclusion and exclusion were made by consensus between two of the authors according to the criteria detailed in the following paragraphs. Discrepancies between authors were resolved through discussion. A flow diagram for the search and study selection process is shown in [Fig. 1](#).

Inclusion criteria—studies needed to meet all the following criteria to be included:

1. A systematic review of RCTs, an overview of systematic reviews of RCTs, or an analysis of individual patient data from RCTs in a drug development program. A systematic review is defined as having reported an explicit literature search strategy.

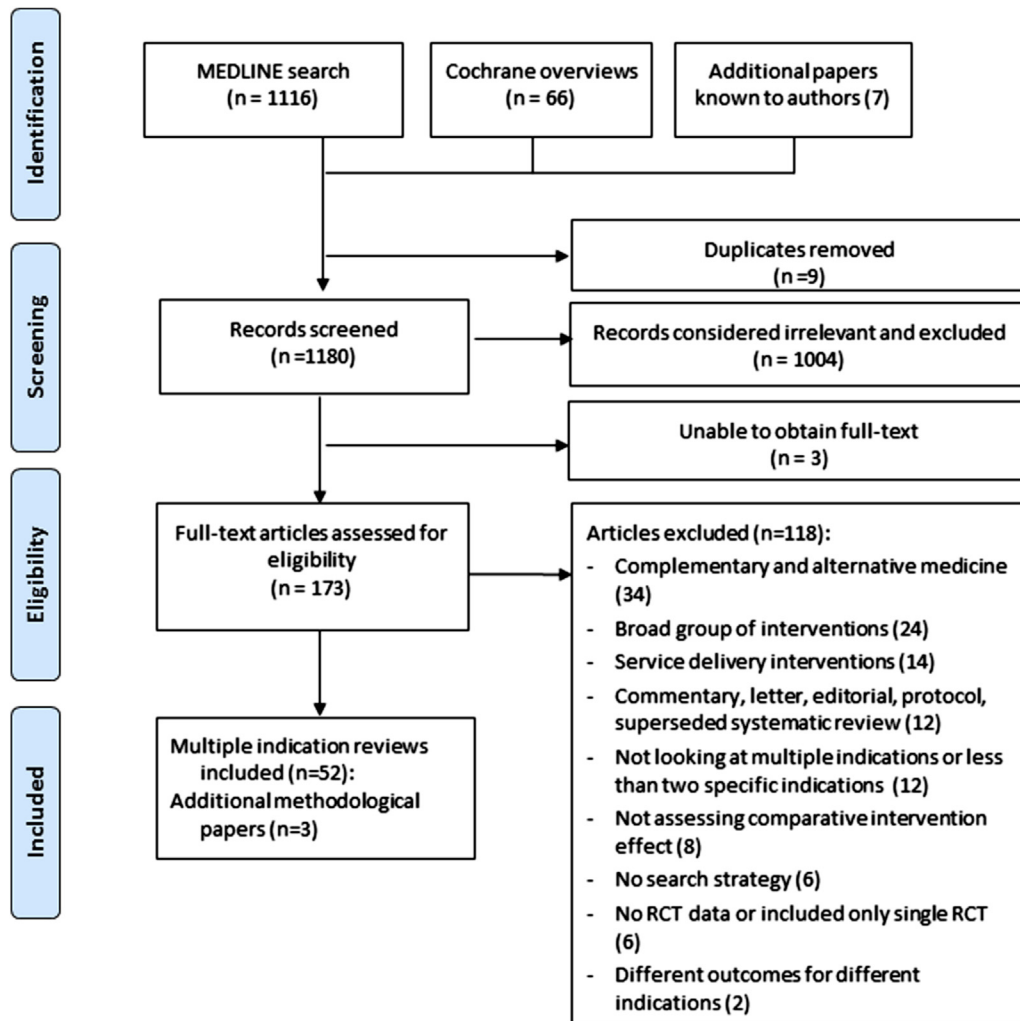


Fig. 1. Flow diagram of literature search and study selection process.

2. Evaluated quantitative evidence from more than one RCT.
3. Examined the effect (either intended or unintended) of an intervention for a given outcome in at least two named indications.
4. Published from 2003 onward.

Additionally, methodological articles that discussed the use of, and issues related to, multiple-indication reviews were also retained. These articles are described in Section 4.

Exclusion criteria—systematic reviews or overviews that met any of the following criteria were excluded:

1. Examined a broad group of heterogeneous interventions (such as Web-based interventions or telemedicine).
2. Focused on complementary and alternative medicines, most of which have been tried in a wide variety of (not necessarily related) disease conditions.

3. Focused on service delivery interventions that are not targeted at individual patients.
4. Looked at the effects of an intervention in two populations defined by the presence or absence of a disease condition (eg, management of obesity with or without diabetes).
5. Editorials, letters, and commentaries that do not present an original systematic review or overview.

Systematic reviews that were superseded by a more recent version and multiple-indication systematic reviews that had been covered in more recent multiple-indication overviews were also excluded.

### 2.1.3. Data extraction and synthesis

Included studies were examined, and data were extracted for the following attributes by one author using a structured coding framework (Table 1):

**Table 1.** Characteristics of included multiple-indication reviews ( $n = 52$ )

Characteristics	<i>n</i> (%)
Form of review	
Systematic review of individual randomized controlled trials (RCTs)	34 (65) <sup>a</sup>
Overviews (systematic review of systematic reviews)	15 (29)
Both (overviews that assessed both systematic reviews and individual RCTs)	3 (6)
Year of publication	
2003	0
2004	1
2005	2
2006	4
2007	5
2008	3
2009	6
2010	6
2011	10
2012	6
2013	7
2014 <sup>b</sup>	2
Type of intervention	
Pharmacotherapy (including immunotherapy)	32 (62)
Interventional procedure (including surgery and procedure involving medical devices)	9 (17)
Nutritional therapy	4 (8)
Physical therapy	4 (8)
Cognitive, behavior, and psychological therapy	3 (6)
Condition/indication included in the review	
Painful conditions	8 (15)
Surgery	7 (13)
Mental and neurodevelopmental disorders	6 (12)
Cardiovascular diseases	2 (4)
Cancer	2 (4)
Other specific group of conditions	3 (6)
Multiple groups of indications	24 (46)
Focus	
Effectiveness	19 (37)
Harm	13 (25)
Both	20 (38)
Inclusion of observational studies in addition to RCTs	
Yes	15 (29)
No	37 (71)
Risk of bias assessment undertaken	
Yes	36 (69)
No	16 (31)
Using results of risk of bias assessment	
Not applicable (risk of bias not assessed)	16 (31)
Results displayed or discussed but not directly used in quantitative analysis	21 (40)
Results used to exclude lower quality studies	4 (8)
Results used to perform subgroup/sensitivity analysis/meta-regression based on study quality/risk of bias	11 (21)
Presentation of results	
Description in text or tabulation	
Description of direction of effect/conclusion	6 (12)
Quantitative results—data from each arms, <i>P</i> values	2 (4)
Comparative effectiveness (point estimates and CIs)	10 (19)
Graphic (eg, forest plots)	34 (65)
Quantitative analysis (with respect to different indications)	
Pooled across different indications without considering between-indication variation	21 (40)
Pooled across indications for some outcomes (eg, adverse events) but not for the others (eg, effectiveness outcomes)	4 (8)
Quantitatively examined variation/heterogeneity between indications with or without pooling across indications	11 (21)
No pooling across different indications and no assessment of variation/heterogeneity between indications (quantitative results reported/displayed separately for each study/indication)	16 (31)

Abbreviation: CI, confidence interval.

<sup>a</sup> Three of the reviews involved analysis of individual patient data.

<sup>b</sup> Searched up to January 2014.

- The source material for the multiple-indication review (eg, systematic reviews of RCTs, individual RCT reports, or individual patient-level data).
- The intervention(s), indications, and main outcome(s) covered by the study.
- Whether there was an assessment of quality or risk of bias in the studies included in the multiple-indication reviews (eg, assessment of RCTs using the Cochrane risk of bias tool [4] in systematic reviews, or

assessment of systematic reviews using AMSTAR [5] in reviews of systematic reviews) and whether/how the results of the assessment were used to guide quantitative synthesis.

- How quantitative results were presented (in text, tables, or graphs such as forest plots).
- Whether an attempt was made to pool results across indications and whether potential variation in treatment effects between different indications was assessed.

The extracted data were checked by another author for accuracy, with discrepancies and disagreements resolved through discussion. We present the results in a table and select three cases to illustrate the potential use and caveats of multiple-indication reviews.

## 2.2. Findings of systematic review

Our search retrieved 1,180 unique records, of which 173 were considered potentially relevant. Fifty-two systematic reviews and overviews met the inclusion criteria after examination of the full text. Additionally, three relevant methodological articles were located [2,6,7]. A flow diagram of the study selection process is shown in Fig. 1, and a list of included and excluded studies can be found in Appendices B and C at [www.jclinepi.com](http://www.jclinepi.com), respectively.

Key features of the included multiple-indication reviews are summarized in Table 1. Use of the method has increased over the study decade. Two-thirds (65%) are systematic reviews of primary studies (three of which analyzed individual patient data) [8–10], whereas most of the remainder are overviews of systematic reviews. The reviews cover a wide range of interventions, such as pharmacologic interventions, interventional procedures (including surgeries and procedures involving medical devices), nutritional therapy, physical therapy, and cognitive, behavior, and psychological therapy. Painful conditions, mental and neurodevelopmental disorders, conditions requiring surgery, and cardiovascular diseases are common groups of indications within which multiple-indication reviews are conducted. Pain, mortality, infection, and various other adverse events are among the most frequently examined outcomes.

Most of the included reviews focused specifically on effectiveness (37%) rather than on unintended effects (25%), whereas the remainder examined both types of end points (38%). Two-thirds of the reviews (69%) assessed the risk of bias or quality of studies they included, and of these, 40% used the results to inform their approach to the analysis. The majority of studies (65%) presented quantitative results in graphic format (eg, forest plots).

The multiple-indication reviews that we identified have adopted three broad approaches in terms of presentation and analysis of quantitative data:

1. Presenting results separately for each indication in a narrative approach.
2. Pooling of results across indications.
3. Examining the variation in effect sizes between different indications with or without pooling across indications.

The chosen approach appears to reflect the purpose of a given multiple-indication review. We shall expand on this issue in the case studies and discussion sections. Statistical pooling across indications was not attempted in 18 (35%) of the studies. An explicit reason for not pooling was given in only 4 of these 18 cases (perceived clinical heterogeneity

in three, while pooling was deferred for a separate article in the remaining case).

There are as yet no indexed terms in electronic databases for multiple-indication reviews. Different names such as “agenda-wide review” [11], “umbrella review” [2], and “panoramic meta-analysis” [6,12] have been used in the literature to describe multiple-indication reviews. We use “multiple-indication review” as it is the least ambiguous term and recommend its use in future studies for this reason. A result of lack of agreed terminology implies that our search strategy, despite several iterations of piloting, could not have uncovered all multiple-indication reviews. For example, many systematic reviews of adverse drug effects may have included data from trials conducted in different disease conditions without explicitly mentioning the multiple-indication nature of the review. Our intention is to identify a sufficient number of examples covering different clinical areas to inform a critical examination and discussion of the topic. Our search of recent literature suggests this approach to evidence synthesis is on the rise, and hence, a critique of the method is timely.

## 3. Case studies

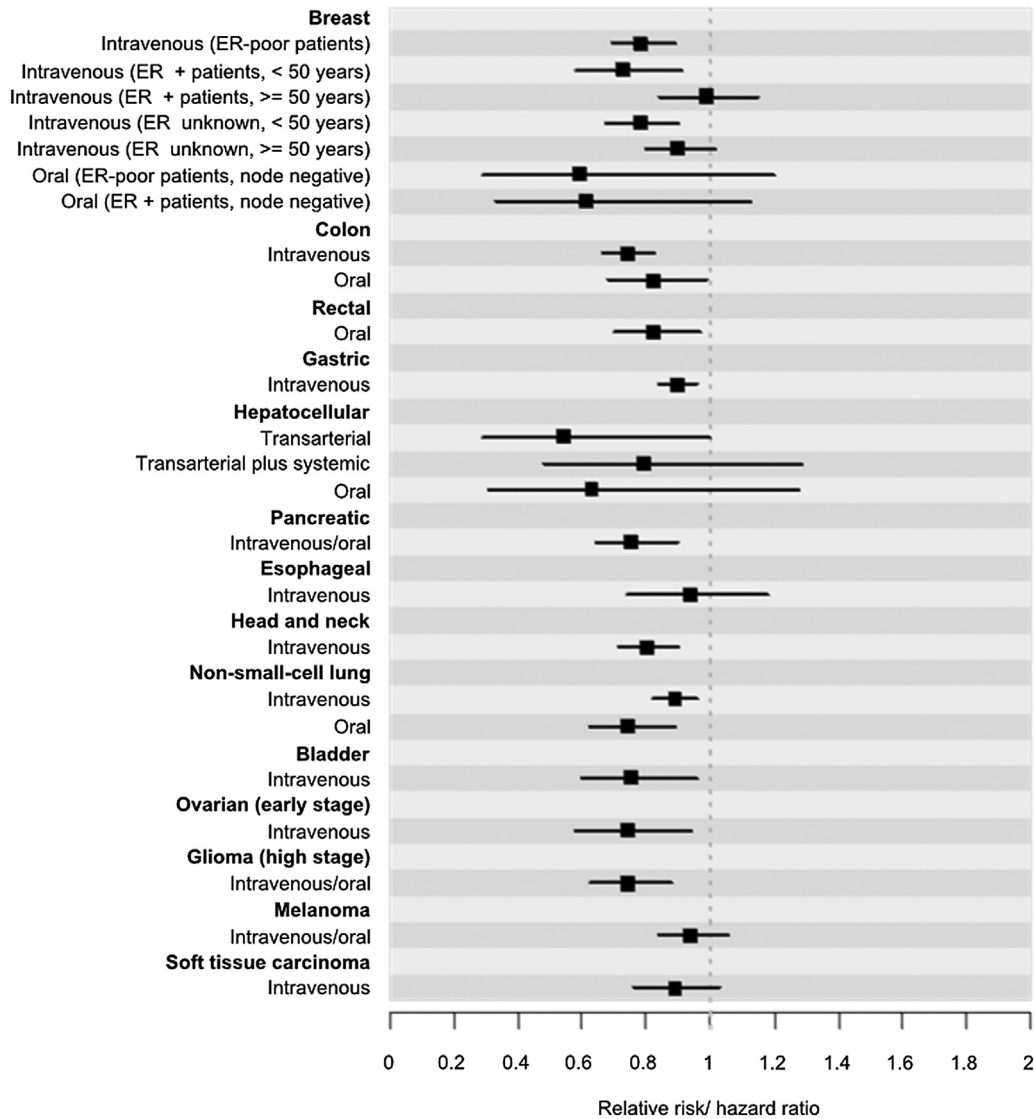
There are three broad nonexclusive uses of multiple-indication reviews (of RCTs or systematic reviews of RCTs)—to detect unintended effects, improve estimates of effectiveness, and examine heterogeneity of effect across disease groups. We have selected three case studies to demonstrate each of these uses of multiple-indication reviews.

### 3.1. Detecting unintended effects

It is often sensibly argued that although the effectiveness of a treatment is appropriately evaluated by RCTs, detection of unintended effects must usually rely on other methods [13]. First, the unintended effects will be expected to be the same regardless of the indication for the intervention [14]. Second, RCTs typically have low statistical power for the detection of unintended effects [14]. The Cochrane handbook states that “many adverse events are too uncommon or too long term to be observed within randomized trials” [15]. For these reasons, a typical systematic review of controlled trials focusing on a specific indication may not provide sufficient evidence on the adverse effects profile of an intervention. However, combining evidence from multiple indications will improve ability to detect unintended effects. Broadly, two types of unintended effects can be considered [16]:

- Rare unexpected effects, such as agranulocytosis in association with carbamazepine.
- Small, but important, increases in common symptoms or diseases, such as cardiovascular disease in association with COX-2 inhibitors.

The signal-to-noise ratio is higher with the former than the latter scenario. As a consequence, reporting systems



**Fig. 2.** Forest plot of meta-analyses concerning effectiveness of adjuvant chemotherapy after surgery vs. surgery alone in different cancer types. With kind permission from Springer Science + Business Media (Fig. 2 of Bowater et al.) [17].

can identify rare conditions, but a small increase in a common disease is much more problematic; these untoward effects cannot be detected by denominator-free reporting systems, yet they are likely to be missed by trials in single diseases [16]. It is in this second situation that comparisons across indications are particularly useful; by providing a substantial boost to sample size, they increase the probability of identifying a “signal in the noise.” The caveat here is that the necessary information must have been collected and reported for the trials that make up the studies included in the synthesis.

Our example here is a multiple-indication review examining the risk of cancer after treatment with tumor necrosis factor alpha inhibitors [8]. The authors boosted data from 74 RCTs in rheumatoid arthritis with 43 trials across a range of other conditions (providing a null result within

narrower confidence limits than would otherwise have been the case).

### 3.2. Improving estimates of effectiveness

Perhaps, a more controversial use of multiple-indication reviews is to improve estimates of effectiveness. Hemming et al. [6] suggested the idea of “borrowing strength” by comparing the estimates of a treatment’s effect size across a range of similar indications to better evaluate its effectiveness in an index indication, whereas Ioannidis and Karassa [11] argue that such an approach can reduce the risk of false-positive and false-negative study results; a null result in the index indication is less convincing if across many other similar indications, the treatment yields strongly positive results against the same comparator.

We illustrate this idea by a multiple-indication review that assessed the effectiveness of adjuvant chemotherapy after surgery vs. surgery alone, over many different cancer types [17]. It transpired that the evidence favored chemotherapy in most cancers; however, the confidence intervals included equivalence for some cancers. Fig. 2 immediately shows the nonsignificant results to be associated with wide confidence limits suggesting considerable uncertainty. This result appears more comparable with a common effect across all cancers than separate effects partitioned across individual cancers. By concentrating on each condition individually, guideline writers and opinion leaders may be too hasty in reaching a conclusion that the treatment should not be recommended—borrowing strength by extrapolating across cancer types may reduce the risk of false-negative study results, especially where the result is imprecise, and there are no compelling biological reasons for disease-specific effects. Conversely, cautions can be raised on adopting an apparently promising intervention with postulated benefits in many disease areas when examination of evidence does not suggest a consistent effectiveness across different indications [18].

### 3.3. Examining heterogeneity of effect across disease groups

Multiple-indication reviews can also be used to look for variations in treatment effects by classes of disease and in this way examine a scientific hypothesis. In the aforementioned example, no difference can be discerned in treatment effect (relative risk reduction) by histologic type—adenocarcinoma or squamous cell cancer or chemosensitivity of the tumor. Likewise, a recent overview of prophylactic perioperative antibiotics in both “clean” and “dirty” operations (Fig. 3) found no evidence of a difference in the odds ratio of postoperative infection according to the degree of contamination [19].

## 4. Discussion and critique

### 4.1. Deciding what indications to include

By definition, a multiple-indication review differs from a “typical” systematic review of a given intervention in that more than one indication is included. This raises a question as to which disease to include in the set. When looking for unexpected effects, it makes sense to include all the conditions for which the treatment has been used. This is because, with a few exceptions such as the interaction between infectious mononucleosis and ampicillin, adverse effects are not disease specific [14].

When examining the effectiveness of an intervention, diseases should be included on the ground that they are or may be linked by a theoretical construct—as in the examples of chemotherapy and antibiotics mentioned previously. Here, the purpose may be to “borrow-strength”—a method

that is commonly used in critical care research in which many conditions such as sepsis, pancreatitis, and massive trauma are believed to create organ damage through common pathologic pathways [20]. Conditions can also be included to test the hypothesis that effects differ according to theoretical construct, for example, clean vs. contaminated surgery in the aforementioned example. Likewise, the effects of antidepressants and cognitive behavior therapy have been compared across a range of conditions expressly to explore the theory that their effects are linked through a common pathway [21,22].

### 4.2. Sources of evidence

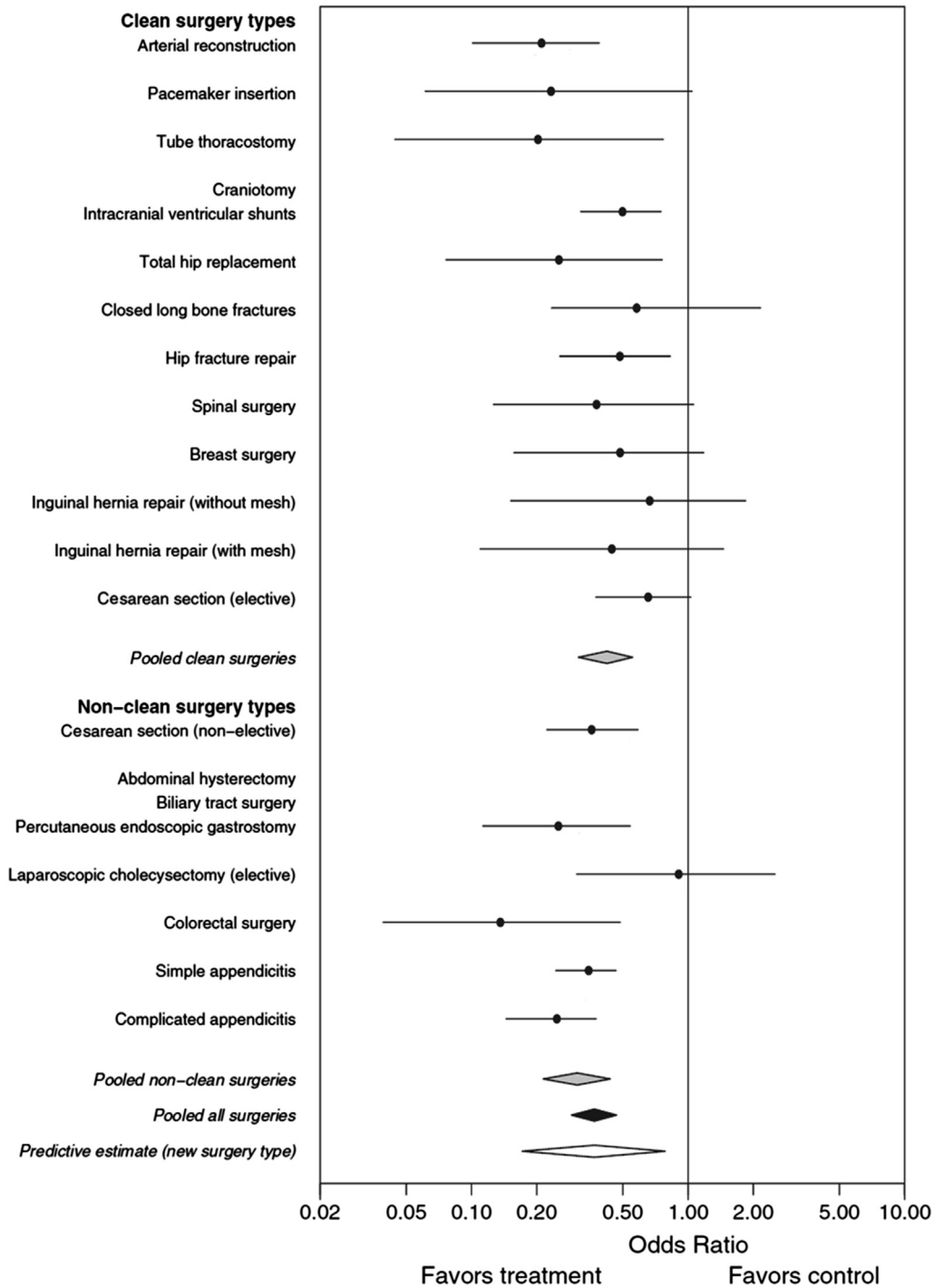
Multiple-indication reviews can be undertaken as systematic reviews of primary studies, in which intervention effects are pooled across indications and potential differences in effects are explored across subgroups of trials defined by indications. Alternatively, the growing number of published systematic reviews and meta-analyses offers an opportunity to conduct multiple-indication reviews through overviews of systematic reviews [23,24]. This approach provides a practical solution for dealing with large volume of evidence that would otherwise not be feasible to review within usual time and resource constraints. The development of methodology and potential issues associated with overviews of systematic reviews have received increasing attention, and comprehensive coverage of this topic is available elsewhere [25–29]. Multiple-indication reviews built up from systematic reviews encounter the problem of “overlapping” where the reviews include some, but not all, of the same studies. They share this problem with all studies where topic-specific systematic reviews are combined in an overview [12,25]. A strategy must be developed to deal with this issue when it arises, for instance, by taking into account quality, contemporaneity, and comprehensiveness of individual reviews. The problem of overlapping does not arise when the source data comprise individual studies, and this represents a clear advantage for such an approach where resources allow.

A multiple-indication review drawing data from existing systematic reviews will also need to deal with the extra level of complexity related to potential heterogeneity in the contributing systematic reviews (in addition to heterogeneity in the primary trial evidence) when comparison and pooling of data between indications is made (see next section). Reviews of reviews based on a program of systematic reviews undertaken using a similar, standardized approach (such as those of the Cochrane Collaboration) should mitigate these potential issues [29].

### 4.3. Recognizing and mitigating potential bias

The point of multiple-indication reviews is to examine for differences across diseases and, if this is present, to





**Fig. 3.** Forest plot of meta-analyses concerning effectiveness of antibiotic prophylaxis in different surgeries. Pooling over the surgery types, allowing for both between-study and between-surgery variation, results in a pooled odds ratio of 0.37 (95% credible interval: 0.29, 0.47), which is fairly convincing evidence that over all surgery types, prophylaxis is effective [6]. Similar results are obtained if the included studies are ordered by baseline infection risk (ie, control group infection risk) rather than classification system [6]. Adapted with kind permission from John Wiley & Sons Ltd (Fig. 1 in Hemming et al.) [6].

explore the cause of these differences, as was done in the cancer example. This type of heterogeneity is epistemic to a multiple-indication review and is to be distinguished from other sources of heterogeneity, which are a cause of bias in that they may obscure true heterogeneity across

diseases or create the appearance of such heterogeneity when none exists. Such bias may arise when trials of the same intervention adopt different treatment methods (such as dose and duration), use different control groups, and/or are conducted over different time epochs with varied

methodological rigor, across disease types. If the trials in one disease type were generally small and those in another were large, then this may lead to publication bias across indications. In addition, where a multiple-indication review is built up from data from existing systematic reviews of individual indications, the methodological quality of the contributing reviews, their completeness in evidence coverage, and how up-to-date they are may also add to heterogeneity. It is therefore important to examine and, where relevant, control for these potential confounders across indications when undertaking multiple-indication reviews. The impact of important potential confounders, such as methodological quality, can be investigated in the same way as in a conventional systematic review and raises the same issues—such as ecological bias. Techniques include subgroup analysis by the AMSTAR score to take account of the quality of individual systematic reviews [12] or a meta-regression approach to generate effect estimates with review-level covariates being adjusted for [6]. Multiple-indication reviews based on individual patient data allow more detailed examination of potential effect modification by these confounders and more comprehensive statistical adjustment than those built up from aggregated data [30].

From our perspective, the concerns about bias constitute grounds for caution in taking a wide-angled view, rather than an argument to eschew the multiple indication perspective in favor of considering each treatment effect in total isolation. We maintain that there is a middle ground between unquestioning extrapolation of results across diseases and a totally solipsistic focus on each disease, one at a time.

#### 4.4. Analytical approach

There are also methodological issues to consider in the presentation and synthesis of results. Approximately a third of the multiple-indication reviews that we found presented evidence individually for each indication without an overarching quantitative analysis. This might present a missed chance for improving the precision of effect estimates and reducing type 2 error [31]. Failure to synthesize results formally may be liable to the problems of “vote counting,” namely not sufficiently taking into account the weight of evidence and size of effect for individual indications [32]. Displaying results using a forest plot, in which each stem represents a meta-analysis for a different indication, gives a clear display of effect estimate, statistical uncertainty, and variation over indications. The analysis can be taken further by statistical pooling over indications. There are arguments for and against such a statistical approach. The argument against is that this may lend spurious accuracy, given the possibility that clinical factors and methodological quality may vary by indication, as discussed in the previous paragraph. The argument in favor of statistical pooling is that such heterogeneity can be explored, as in a conventional meta-analysis, before pooling. Pooling effect estimates

over indication should allow for both between-study and between-indication variability. In terms of the actual practicalities of implementation, the approaches have been illustrated in Hemming et al. [6]. A formal quantitative data synthesis can be undertaken using either a two-step frequentist approach or a full Bayesian approach. Both methods provide a single pooled estimate of the effect measure over all indications, along with estimates of degree of heterogeneity between indications. In the two-step approach, the data are first pooled (first step) within indications allowing for between-study heterogeneity and then pooled (second step) across indications allowing for between-indication heterogeneity. In the Bayesian approach, the data are modeled as a series of hierarchies similar to a generalized linear mixed model. These methods therefore allow for both between-study variability (if random-effects meta-analysis was used in the pooling of studies within indication) and between-indication variability (using random effects).

It is fully accepted that the average effect across many indications may hide important individual differences [33]. However, the possibility of hidden subgroup effects applies to any data set. Indeed, the current interest in stratified medicine represents a search for such subgroups using molecular mechanisms. At first thought, the “lumping” approach inherent in a review across indications seems to fly in the face of the ideal of stratified medicine. However, the two ideas may not be in opposition. Unremitting stratification is liable to yield diminishing returns as the number of subgroups increases. However, the risk of spurious positive and false null results can be reduced if the number of subgroups can be reduced in one dimension (say the organ of tumor origin), whereas increasing them in another dimension (say the molecular signature of the tumor). We argue that a review across indications is an investigative tool that has its place in science where theory is developed by synthesizing the results of particular studies to develop, modify, or refute theory [34].

Although this article has focused on RCTs and systematic reviews of RCTs of medical interventions, the idea could be applied more widely. For instance, studies of interventions could include observational designs where statistical techniques to model the effect of potential bias could be used [35].

## 5. Conclusion

The argument that multiple-indication reviews have an important role in the detection of adverse effects has attracted little peer criticism, and we believe the methods can be advocated without further ado. Hammad et al. [36] have developed a set of criteria, not yet included in the PRISMA statement, for the detection of unintended consequences of treatment in single-indication meta-analyses but do not mention multiple-indication reviews. However, we have encountered resistance to use of this method in the

assessment of effectiveness. The idea is counterintuitive to those who have been conditioned to “compare like with like”; the type of systematic review we are proposing sets out, quite deliberately, to compare (and possibly combine) things that are clearly not identical. The key, however, is to appreciate that things may be different in one respect (eg, the organ in which cancer has arisen, the specialty involved), while being similar in another (eg, sensitivity to chemotherapy). Whether they are indeed similar is, of course, precisely what the synthesis is designed to examine. In RCTs and “standard” systematic reviews, subgroup analysis based on patient characteristics (specified a priori on the basis of biological plausibility) is a widely accepted tool for exploring potential heterogeneity. The principle of multiple-indication reviews is no different—here, the individual indication plays the role of the patient characteristic (potential effect modifier) of interest. Detecting subgroup effects is one of the purposes of a “standard” systematic review, just as disease-specific effects can be examined in a multiple-indication review. We have illustrated the potential benefits of this wide-angled approach and highlighted issues that require caution when using this method. We maintain that there is a middle ground between examining each disease in isolation and unquestioning extrapolation across different diseases and that wider adoption of multiple-indication reviews in evidence synthesis is warranted.

### Acknowledgment

Authors' contributions: R.J.L. conceptualized the idea and worked with all the coauthors (Y.-F.C., K.H., P.J.C., K.K.G., and D.G.A.) to refine and finalize the article. K.K.G. conducted a preliminary search of the literature, and Y.-F.C. undertook the final search. Y.-F.C., K.H., P.J.C., and R.J.L. selected studies for inclusion. Y.-F.C. extracted data from included studies and carried out descriptive analysis. P.J.C. independently checked the extracted data. R.J.L. is the guarantor. The views expressed in this article do not necessarily match the views of the funders.

### Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.jclinepi.2014.08.007>.

### References

- [1] Patsopoulos NA, Analatos AA, Ioannidis JA. Relative citation impact of various study designs in the health sciences. *JAMA* 2005;293:2362–6.
- [2] Ioannidis JP. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. *Can Med Assoc J* 2009;181:488–93.
- [3] Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408–12.
- [4] Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
- [5] Shea B, Grimshaw J, Wells G, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10.
- [6] Hemming K, Bowater RJ, Lilford RJ. Pooling systematic reviews of systematic reviews: a Bayesian panoramic meta-analysis. *Stat Med* 2012;31:201–16.
- [7] Salanti G, Dias S, Welton NJ, Ades AE, Gollinopoulos V, Kyrgiou M, et al. Evaluating novel agent effects in multiple-treatments meta-regression. *Stat Med* 2010;29:2369–83.
- [8] Askling J, van Vollenhoven RF, Granath F, Raaschou P, Forred CM, Baecklund E, et al. Cancer risk in patients with rheumatoid arthritis treated with anti-tumor necrosis factor alpha therapies: does the risk change with the time since start of treatment? *Arthritis Rheum* 2009;60:3180–9.
- [9] Curtis SP, Ko AT, Bolognese JA, Cavanaugh PF, Reicin AS. Pooled analysis of thrombotic cardiovascular events in clinical trials of the COX-2 selective inhibitor etoricoxib. *Curr Med Res Opin* 2006;22:2365–74.
- [10] Moore A, Makinson G, Li C. Patient-level pooled analysis of adjudicated gastrointestinal outcomes in celecoxib clinical trials: meta-analysis of 51,000 patients enrolled in 52 randomized trials. *Arthritis Res Ther* 2013;15:R6.
- [11] Ioannidis JP, Karassa FB. The need to consider the wider agenda in systematic reviews and meta-analyses: breadth, timing, and depth of the evidence. *BMJ* 2010;341:c4875.
- [12] Hemming K, Pinkney T, Futaba K, Pennant M, Morton DG, Lilford RJ. A systematic review of systematic reviews and panoramic meta-analysis: staples versus sutures for surgical procedures. *PLoS One* 2013;8:e75132.
- [13] Loke YK, Price D, Herxheimer A. Systematic reviews of adverse effects: framework for a structured approach. *BMC Med Res Methodol* 2007;7:32.
- [14] Vandembroucke JP. When are observational studies as credible as randomised trials? *Lancet* 2004;363:1728–31.
- [15] Loke YK, Price D, Herxheimer A. Chapter 14: adverse effects. In: Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions* version 5.1.0 (updated March 2011). The Cochrane Collaboration; 2011. Available from: [www.cochrane-handbook.org](http://www.cochrane-handbook.org). Accessed 16 May 2014.
- [16] Jick H, Miettinen OS, Shapiro S, Lewis GP, Siskind V, Slone D. Comprehensive drug surveillance. *JAMA* 1970;213:1455–60.
- [17] Bowater RJ, Abdelmalik SM, Lilford RJ. Efficacy of adjuvant chemotherapy after surgery when considered over all cancer types: a synthesis of meta-analyses. *Ann Surg Oncol* 2012;19:3343–50.
- [18] Contopoulos-Ioannidis DG, Ioannidis JP. Claims for improved survival from systemic corticosteroids in diverse conditions: an umbrella review. *Eur J Clin Invest* 2012;42:233–44.
- [19] Bowater RJ, Stirling SA, Lilford RJ. Is antibiotic prophylaxis in surgery a generally effective intervention? Testing a generic hypothesis over a set of meta-analyses. *Ann Surg* 2009;249:551–6.
- [20] Lord JM, Midwinter MJ, Chen Y-F, Belli A, Brohi K, Kovacs EJ, et al. The systemic inflammatory response to acute trauma: overview of pathophysiology and treatment. *Lancet* 2014. In press.
- [21] Taylor D, Meader N, Bird V, Pilling S, Creed F, Goldberg D, et al. Pharmacological interventions for people with depression and chronic physical health problems: systematic review and meta-analyses of safety and efficacy. *Br J Psychiatry* 2011;198:179–88.
- [22] Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. *Clin Psychol Rev* 2006;26:17–31.
- [23] Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med* 2010;7:e1000326.

- [24] Silva V, Grande AJ, Martimbianco AL, Riera R, Carvalho AP. Overview of systematic reviews—a new type of study: part I: why and for whom? *Sao Paulo Med J* 2012;130:398–404.
- [25] Cooper H, Koenka AC. The overview of reviews: unique challenges and opportunities when research syntheses are the principal elements of new integrative scholarship. *Am Psychol* 2012;67:446–62.
- [26] Moore A, Jull G. The systematic review of systematic reviews has arrived!. *Man Ther* 2006;11:91–2.
- [27] Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. *BMC Med Res Methodol* 2011;11:15.
- [28] Pieper D, Buechter R, Jerinic P, Eikermann M. Overviews of reviews often have limited rigor: a systematic review. *J Clin Epidemiol* 2012; 65:1267–73.
- [29] Becker LA, Oxman AD. Chapter 22: overviews of reviews. In: Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions* version 5.1.0 (updated March 2011). Cochrane Collaboration; 2011. Available from: [www.cochrane-handbook.org](http://www.cochrane-handbook.org). Accessed 16 May 2014.
- [30] Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Eval Health Prof* 2002;25:76–97.
- [31] Ioannidis JP, Patsopoulos NA, Rothstein HR. Reasons or excuses for avoiding meta-analysis in forest plots. *BMJ* 2008;336: 1413–5.
- [32] Deeks JJ, Higgins JPT, Altman DG. Chapter 9: analysing data and undertaking meta-analyses. In: Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions* version 5.1.0 (updated March 2011). The Cochrane Collaboration; 2011. Available from: [www.cochrane-handbook.org](http://www.cochrane-handbook.org). Accessed 16 May 2014.
- [33] Popper KR. *The logic of scientific discovery*. 2nd ed. London, UK: Routledge; 2002.
- [34] Collins R, Peto R, Gray R, Parish S. Large-scale randomized evidence: trials and overviews. In: Weatherall DJ, Ledingham JGG, Warrell DA, editors. *Oxford textbook of medicine*. 3rd ed. Oxford, UK: Oxford University Press; 1996:21–32.
- [35] Turner RM, Spiegelhalter DJ, Smith GC, Thompson SG. Bias modelling in evidence synthesis. *J R Stat Soc Ser A Stat Soc* 2009;172: 21–47.
- [36] Hammad TA, Neyarapally GA, Pinheiro SP, Iyasu S, Rochester G, Dal Pan G. Reporting of meta-analyses of randomized controlled trials with a focus on drug safety: an empirical assessment. *Clin Trials* 2013;10:389–97.