

# Fixed or random effects meta-analysis? Common methodological issues in systematic reviews of effectiveness

Catalin Tufanaru MD, MPH, MClSci (EBHC), Zachary Munn PhD, Matthew Stephenson PhD and Edoardo Aromataris PhD

*The Joanna Briggs Institute, Faculty of Health Sciences, The University of Adelaide, Adelaide, 115 Grenfell Street, South Australia 5000, Australia*

## ABSTRACT

Systematic review aims to systematically identify, critically appraise, and summarize all relevant studies that match predefined criteria and answer predefined questions. The most common type of systematic review is that assessing the effectiveness of an intervention or therapy. In this article, we discuss some of the common methodological issues that arise when conducting systematic reviews and meta-analyses of effectiveness data, including issues related to study designs, meta-analysis, and the use and interpretation of effect sizes.

**Key words:** effect size, effectiveness, fixed effects, meta-analysis, random effects, systematic review

*Int J Evid Based Healthc* 2015; 13:196–207.

## Introduction

A systematic review aims to systematically identify, critically appraise, and summarize all relevant studies that match predefined criteria and answer predefined questions.<sup>1–5</sup> The most common type of systematic review is that assessing the effectiveness of an intervention or therapy.

Conducting a systematic review addressing a question about the effectiveness of an intervention or therapy is a complex research process. In this article, we discuss and provide guidance for some of the common methodological issues that arise when conducting systematic reviews and meta-analyses of effectiveness data.

## Inclusion of quasi-experimental and observational studies

Evidence about effects of interventions may come from three main categories of studies: experimental, quasi-experimental, and observational studies (the latter can

be further split into analytical and descriptive studies). Ideally, evidence about the effectiveness of interventions should come from good-quality randomized controlled trials (RCTs) which explore final clinical endpoints such as morbidity, mortality, and quality of life (rather than surrogate endpoints).<sup>6</sup> However, for many clinical interventions and conditions, RCTs are not available.

There are three approaches to considering which study designs to include that are equally reasonable when conducting an effectiveness systematic review. Option 1 – to consider just RCTs and quasi-experimental studies; this is the option that was favored in the past by the Cochrane Collaboration. Option 2 – if there are good-quality RCTs exploring the interventions, comparators, and the outcomes of interest for a systematic review of effectiveness evidence, reviewers may give priority to these RCTs over quasi-experimental or observational studies, and include in their review only the RCTs; however, if there are no (or limited) RCTs available following the search, reviewers may consider quasi-experimental studies for inclusion; if there are no (or limited) RCTs or quasi-experimental studies, reviewers may opt to include observational studies; this is the option that has historically been favored by the Joanna Briggs Institute. Option 3 – if reviewers want to include all study designs (RCTs and quasi-experimental and

*Correspondence:* Catalin Tufanaru, The Joanna Briggs Institute, Faculty of Health Sciences, The University of Adelaide, Adelaide, 115 Grenfell Street, South Australia 5000, Australia.  
E-mail: catalin.tufanaru@adelaide.edu.au

**Received** 1 April 2015 **Revised** 19 July 2015 **Accepted** 19 July 2015  
DOI: 10.1097/XEB.0000000000000065

observational studies) in their review, this inclusive approach is acceptable as it allows for examination of the totality of empirical evidence and may provide invaluable insights regarding the agreement or disagreement of the results from different study designs. In any case, the approach to be taken should be detailed in the a-priori systematic review protocol. Wherever feasible, we prefer and suggest reviewers consider option 3, the most inclusive approach.

It is important to note that the issues related to the agreement or disagreement of results from experimental and observational studies are complex. Empirical research has found that sometimes the results of RCTs contradict results from observational studies.<sup>6</sup> However, meta-analyses based on observational studies can produce estimates of effect that are similar to those from meta-analyses based on RCTs.<sup>7</sup>

### Inclusion or exclusion of studies with risk of bias

Evidence about the effects of interventions may come from studies with diverse risk of bias. There are two approaches which are equally reasonable when conducting an effectiveness systematic review: Option 1 – include only studies with low or moderate risk of bias, and exclude all studies considered at high risk of bias. In this case, reviewers have to provide in the review protocol clear and explicit justification with regard to how the risk of bias will be ascertained, what represents low, moderate, and high risk of bias, and whether any ‘cutoff’ scores will be used. Option 2 – include all studies regardless of their risk of bias and explicitly consider the risk of bias of different studies during data synthesis, presentation of the results, conclusions, and implications. Reviewers should be aware that there are statistical tools available for the incorporation of risk of bias in data synthesis and we recommend two such approaches, the quality effects model proposed by Doi *et al.*,<sup>8</sup> and the bias adjustment approach proposed by Thompson *et al.*<sup>9</sup> However, the issues of incorporating the risk of bias in the conduct of systematic reviews is complex.<sup>10</sup> In any case, the approach to be taken should be detailed *a priori* in the systematic review protocol.

### Issues related to cross-over trials, pre-post studies, and cluster randomized trials

Some systematic reviewers may not be aware that crossover trials, pre-post studies, and cluster randomized trials have specific characteristics aligned to both their design and statistical analyses that should be carefully considered whenever these study designs are included in systematic reviews and meta-analyses. We

recommend reviewers consult appropriate references regarding the design and analysis of these types of studies.<sup>11–15</sup> When including these types of studies in a systematic review, it is essential that reviewers consider the existing statistical guidance regarding meta-analysis of these types of studies.<sup>16–22</sup>

### Use and interpretation of effect sizes

Effect sizes refer to quantitative indicators of the direction and magnitude of the effect of the intervention on outcomes. Despite the vast amount of information regarding interpreting effect sizes, we have found that some systematic reviewers remain confused about the differences between risk (probability) and odds, and the differences between relative risk (RR) and odds ratio (OR), and use incorrect narrative descriptions for these effect sizes. These issues are discussed further and additional resources such as the *Users' Guides to the Medical Literature* and the *Tips for Learners of Evidence-Based Medicine* are also useful guides for systematic review authors.<sup>23,24</sup>

Common effect sizes reported in effectiveness systematic reviews are the OR, RR and risk difference (RD). Risk is defined as the probability that an event will occur; RR (also known as the risk ratio) is the ratio of risk in one group (e.g., intervention group) divided by the risk in another group (e.g., control group). Risk difference is defined as the difference between the risk in one group and the risk in the other group. Odds are the ratio of the probabilities of the two possible states of a binary variable, and an OR is the ratio of the odds for a binary variable in two groups of patients. If we consider the probabilities of the outcome being present and the probabilities of the outcome being absent (i.e., 1 minus the probability of the outcome present) in an intervention group and a control group then:

- (1) The odds of the outcome being present in the intervention group is the probability of the outcome being present in the intervention group divided by 1 minus the probability of the outcome being present in the intervention group.
- (2) The odds of the outcome being present in the control group is the probability of the outcome being present in the control group divided by 1 minus the probability of the outcome being present in the control group.
- (3) The OR of the outcome being present is the ratio of the odds of the outcome being present in the intervention group and the odds of the outcome being present in the control group.

### Absolute and relative effect sizes for meta-analysis of binary data

Relative effect sizes such as the RR are easily misleading in the absence of appropriate contextual information regarding the risk (probability) of the outcome in absence of any intervention, and the absolute difference of the risk (probability) of the outcome between those receiving an intervention and those receiving a different intervention. The RR only indicates the risk (probability) in the intervention group compared with the control group, for example, an RR of 0.5 indicates that the risk (probability) is reduced by half in the intervention group. The RD assists to put the RR into context. For example, an RR of 0.5 may mean the risk (probability) has reduced from 80% in the control group to 40% in the experimental group corresponding to a considerable RD of 40%. Alternatively, an RR of 0.5 could mean that the risk has reduced from 0.8% in the control group to 0.4% in the experimental group, an RD of only 0.4%. Therefore, if reviewers choose to use the RR, ideally they should also report the absolute RD effect size. Reviewers should provide correct and clear interpretation of the computed effect sizes (Tables 1 and 2). It is important to note that the RR is not symmetrical, resulting in potentially very different RRs for a positive outcome compared with its negative component.<sup>25</sup> We recommend that reviewers be aware of this issue as this can impact the presentation and interpretation of the results.

### Odds ratio: preferred effect size for the computation phase of meta-analysis of binary data

Often the terms 'odds' and 'risk' are used interchangeably; however, the OR and RR are calculated in different ways and it is important to understand this when interpreting the results of meta-analysis. Fleiss<sup>26</sup> discussed the statistical properties of the OR and concluded that the OR is the preferred effect size for the computation phase of the meta-analysis of binary data regardless of the study design of the studies. Also, Fleiss and Berlin<sup>27</sup> recommended OR as the preferred effect size for the computation phase of the meta-analysis of binary data, a view that is agreed to by others.<sup>28–30</sup> We agree with this position that OR should be used as the preferred effect size whenever possible for the computation phase of the meta-analysis of binary data. However, there is no universal agreement on this issue, and others prefer the RR over OR.<sup>31–33</sup>

Reviewers should be aware that OR is not easily interpretable, and they should be mindful of providing accurate and explicit interpretation of the ORs computed in meta-analysis.

In Tables 1 and 2, we provide examples of computation and interpretation of probability (risk), odds, RR, RD, and OR (examples for RR and OR >1 and examples for RR and OR <1).

### Reporting the results in natural (clinical) units for meta-analysis of continuous data

There are different 'difference' effect sizes for continuous data such as weighted mean difference (WMD), and the standardized mean difference (SMD).

The WMD is used in meta-analysis of continuous data if all studies included in the meta-analysis measured the same outcome on the same measurement instrument. For meta-analysis computation, the difference in means from each included study is used. The results are expressed in the natural (clinical) units used for the measurement instrument. For example, WMD may be used if all studies included in a meta-analysis measured blood pressure expressed in mmHg. Another example, WMD may be used if all studies measured intensity of pain on the same scale of measurement from zero to 100 units.

The SMD is used in the meta-analysis of continuous data if studies included in the analysis measured the same outcome but on different measurement instruments. The results are expressed in units of standard deviation. In order to facilitate the interpretation of the results reported in units of standard deviation, reviewers should convert the results into natural (clinical) units by multiplying the results expressed in units of standard deviation with the standard deviation of the scores on a known or most commonly used measurement instrument. Results will ideally be reported in both these formats: in units of standard deviation and also expressed in the natural (clinical) units for one measurement instrument. For example, there are diverse scales of measurement used to measure pain. Suppose that in one study, pain was measured on a scale from zero to 10; in another study pain was measured on a scale from 0 to 40; and, in yet another study, pain was measured on a scale from zero to 100. It is possible to use the SMD, to report results from all these studies in a standardized form, in units of standard deviation and later to convert the results back to clinical units by using one or each of these scales.

If WMD is used, reviewers should provide explanations regarding the interpretation of the results expressed in units used for the measurement instrument. The minimum score and the maximum score that are possible on the measurement instrument should be specified together with their interpretation. Suppose that intensity of pain was measured on a scale of

**Table 1. Example of computation and interpretation of probability (risk), odds, risk ratio, risk difference, and odds ratio (examples for risk ratio and odds ratio > 1)**

<p>Suppose that 100 patients are treated with a ‘new’ intervention and that 80 patients are cured of their disease and 20 patients are not cured of their disease. The probability (risk) of being cured is 80/100, that is, 80%. The probability (risk) of not being cured is 20/100, that is, 20%. The odds of the disease being cured are 80%/(100% – 80%) equivalent to 80%/20% = 8/2 = 4/1 or 4 to 1 in the intervention group. It is useful to remember that these odds of 4 to 1 for being cured in the intervention group means that there is 80% probability of being cured and 20% probability of not being cured.</p>
<p>Suppose that 100 patients are treated with a ‘control’ intervention and that 50 patients are cured of their disease and 50 patients are not cured of their disease. The probability (risk) of being cured is 50/100, that is, 50%. The probability of not being cured is 50/100, that is, 50%. The odds of the disease being cured are 50%/(100% – 50%) equivalent to 50%/50% = 5/5 or 1/1 or 1 to 1 in the control group. It is useful to remember that these odds of 1 to 1 for being cured in the control group means that there are 50% probability of being cured and 50% probability of not being cured.</p>
<p>The relative risk (RR) is computed as the probability (risk) of being cured in the intervention group divided by the probability (risk) of being cured in the control group, that is, 80%/50% equivalent to 8/5 or 1.6. This RR of 1.6 is interpreted as indicating that the probability of being cured in the intervention group is 1.6 times the probability of being cured in the control group.</p>
<p>The risk difference (RD) is computed as the probability (risk) of being cured in the intervention group minus the probability (risk) of being cured in the control group, that is, 80% – 50% equivalent to 30%. The RD of 30% is interpreted as indicating that the probability of being cured in the control group is 30% lower compared with the probability of being cured in the intervention group.</p>
<p>The odds ratio of the disease being cured are (80%/20%)/(50%/50%) or equivalently (80%/20%)/1 = 80%/20% = 8/2 = 4/1 = 4. This odds ratio of 4 is interpreted as indicating that the odds of the outcome being present in the intervention group are 4 times the odds of the outcome being present in the control group. In other words, for our example, the odds of being cured in the intervention group are 4 times the odds of being cured in the control group. It is important to remember that odds are not probabilities and that odds ratios are not ratios of probabilities but ratios of odds. In our example, in the intervention group, there is 80% probability of being cured and 20% probability of not being cured. In the control group, there is 50% probability of being cured and 50% probability of not being cured.</p>

**Table 2. Example of computation and interpretation of probability (risk), odds, risk ratio, risk difference, and odds ratio (examples for risk ratio and odds ratio < 1)**

<p>Suppose that 100 patients are treated with a new intervention and that 20 patients are cured of their disease and 80 patients are not cured of their disease. The probability (risk) of being cured is 20%. The odds of the disease being cured are 20%/(100% – 20%) = 20%/80% = 2/8 = 1/4 or 1 to 4 in the intervention group. It is useful to remember that these odds of 1 to 4 for being cured in the intervention group means that there is 20% probability of being cured and 80% probability of not being cured.</p>
<p>Suppose that 100 patients are treated with a control intervention and that 50 patients are cured of their disease and 50 patients are not cured of their disease. The probability (risk) of being cured is 50/100, that is, 50%. The odds of the disease being cured are 50%/(100% – 50%) = 50%/50% = 5/5 or 1/1 or 1 to 1 in the control group. It is useful to remember that these odds of 1 to 1 for being cured in the control group means that there is 50% probability of being cured and 50% probability of not being cured.</p>
<p>The risk ratio (RR) is computed as the probability (risk) of being cured in the intervention group divided by the probability (risk) of being cured in the control group, that is, 20%/50% equivalent to 2/5 or 0.4. This RR of 0.4 is interpreted as indicating that the probability of being cured in the intervention group is 0.4 times the probability of being cured in the control group.</p>
<p>The risk difference (RD) is computed as the probability (risk) of being cured in the intervention group minus the probability (risk) of being cured in the control, that is, 20% – 50% equivalent to –30%. This RD of –30% is interpreted as indicating that the probability of being cured in the intervention group is 30% lower compared with the probability of being cured in the control group.</p>
<p>The odds ratio of the disease being cured are (20%/80%)/(50%/50%) or equivalently 2/8 or 1/4 or 1 to 4 or 25% or 0.25. This odds ratio of 1 to 4 or 25% is interpreted as indicating that the odds of the outcome being present in the intervention group are 0.25 times the odds of the outcome being present in the control group. In other words, for our example, the odds of being cured in the intervention group are 0.25 times the odds of being cured in the control group. Or equivalently, the odds of being cured in the intervention group are 75% lower compared with the odds of being cured in the control group. It is important to remember that odds are not probabilities and that odds ratios are not ratios of probabilities but ratios of odds. In our example, in the intervention group, there is 20% probability of being cured and 80% probability of not being cured. In the control group, there is 50% probability of being cured and 50% probability of not being cured.</p>

measurement from zero to 100 units. In this example, the minimum score is zero, and the maximum score is 100. A score of zero may be interpreted as absence of pain; a score of 100 may be interpreted as unbearable pain.

In addition, reviewers should provide explanations regarding the interpretation of positive and negative scores. For example, sometimes, positive scores are used to express specific characteristics or degrees of these characteristics such as medication compliance, or existence of self-help skills, or positive patient behaviors, and negative scores are used for lack of medication compliance, absence of skills, or negative behaviors.

In Table 3, we provide a concise summary of the statistical properties of the effect sizes considered in this article.

**Meta-analysis: objectives of meta-analysis**

Essentially, in a systematic review of effectiveness, there are two synthesis options: meta-analysis and narrative summary or synthesis. Meta-analysis refers to the statistical synthesis of quantitative results from two or more studies. Many reviewers appear to adopt a narrow approach to meta-analysis, focusing exclusively on calculating estimates of effects. However, reviewers should be aware that there are different, legitimate objectives for a meta-analysis: to improve statistical power to detect a treatment effect, provide the closest estimate of an unknown real effect, identify subsets of studies (subgroups) associated with a beneficial effect, and explore if there are differences in the size or direction of the treatment effect associated with study-specific variables. Reviewers should explicitly consider and state the objective(s) of meta-analysis for their review.

**Clinical and methodological heterogeneity**

Meta-analysis is only appropriate when studies are considered similar enough from a clinical and methodological point of view (homogenous studies). If studies are heterogeneous from a clinical (i.e., population, intervention, comparator, and outcome) or methodological (i.e., study design and risk of bias) point of view, then it is uncertain if it is appropriate to synthesize the respective studies with meta-analysis. The judgment that studies are homogenous enough and that it is appropriate to combine the studies statistically should be based on the understanding of the review question, the characteristics of the studies, and the interpretability of the results. The decision should not be based just on statistical heterogeneity. Studies that are conceptually similar from a clinical point of view (but not necessarily identical) with regard to the participants, interventions, comparators, settings, outcomes, study design, and risk of bias may be

**Table 3. Common effect sizes used in meta-analysis of effectiveness evidence**

Effect size	For what type of variable	Type of effect size (relative or absolute)	Minimum theoretical possible value	Maximum theoretical possible value	Null value for the meta-analysis forest plot (line of no effect value)	Commentary
Relative risk or risk ratio (RR)	Dichotomous	Relative	Zero [log RR has as domain from minus infinity to plus infinity]	+ Infinity	1	In trials wherein the goal of the intervention is to prevent a negative outcome, for example to prevent death, an RR <1 indicates efficacy of the intervention; a RR of 1 indicates no difference between groups; in trials where the goal is to cause a positive outcome, such as disease cure or remission, an RR > 1 indicates treatment efficacy.

Risk difference (RD)	Absolute	-1	+1	Zero	In trials wherein the goal of the intervention is to prevent a negative outcome, an RD <0 indicates efficacy of the intervention; an RD of 0 indicates no difference between groups; in trials wherein the goal is to cause a positive outcome, an RD >0 indicates treatment efficacy.
Odds ratio (OR)	Relative	Zero (log OR has as domain from -infinity to +infinity)	+Infinity	1	In trials wherein the goal of the intervention is to prevent a negative outcome, for example, to prevent death, an OR <1 indicates efficacy of the intervention; an OR of 1 indicates no difference between groups; in trials wherein the goal is to cause a positive outcome, such as disease cure or remission, an OR >1 indicates treatment efficacy.
Difference of means (MD) (also called weighted mean difference WMD or mean difference)	Absolute	-Infinity	+Infinity	Zero	WMD is used if all studies included in a meta-analysis measured the measurement scale or instrument; in trials wherein the goal is to cause a positive outcome, a positive difference indicates that the intervention is effective; a WMD of 0 indicates no difference between groups.
Standardized mean difference (SMD)	Absolute	-Infinity	+Infinity	Zero	SMD is used if studies included in meta-analysis used different scales of measurement; in trials wherein the goal is to cause a positive outcome, a positive difference indicates that the intervention is effective; an SMD of 0 indicates no difference between groups.

\*A risk ratio is not a rate ratio.

**Table 4. Comparison between fixed-effects and random-effects model for meta-analysis**

Criterion	Fixed-effects model	Random-effects model
Goal of statistical inference (statistical generalizability of results).	Results apply only to studies included in meta-analysis.	Results apply beyond included studies.
Statistical assumption regarding the parameter.	There is one common, fixed parameter and all studies estimate the same common fixed parameter.	There is no common, fixed parameter and studies estimate different parameters.
Nonstatistical assumption regarding the comparability of studies from a clinical point of view (participants; interventions; comparators; outcomes).	It is reasonable to consider that studies are similar enough and that there is a common effect.	Studies are different and it is not reasonable to consider that there is a common effect.
Nonstatistical assumption regarding the comparability of studies from a methodological point of view (study design and risk of bias).	It is reasonable to consider that studies are similar enough and that there is a common effect.	Studies are different and it is not reasonable to consider that there is a common effect.
The nature of meta-analysis results.	The meta-analysis summary effect is an estimate of the effect that is common to all studies included in the analysis.	The meta-analysis summary effect is an estimate of the mean of a distribution of true effects; this estimate is not the shared common estimate that is common to all studies included in the analysis.

combined in meta-analysis. Where there are clinically similar but methodological dissimilar studies (such as in study design or risk of bias), subgroup analyses may be useful to determine whether these differences have an impact on the overall effect size.

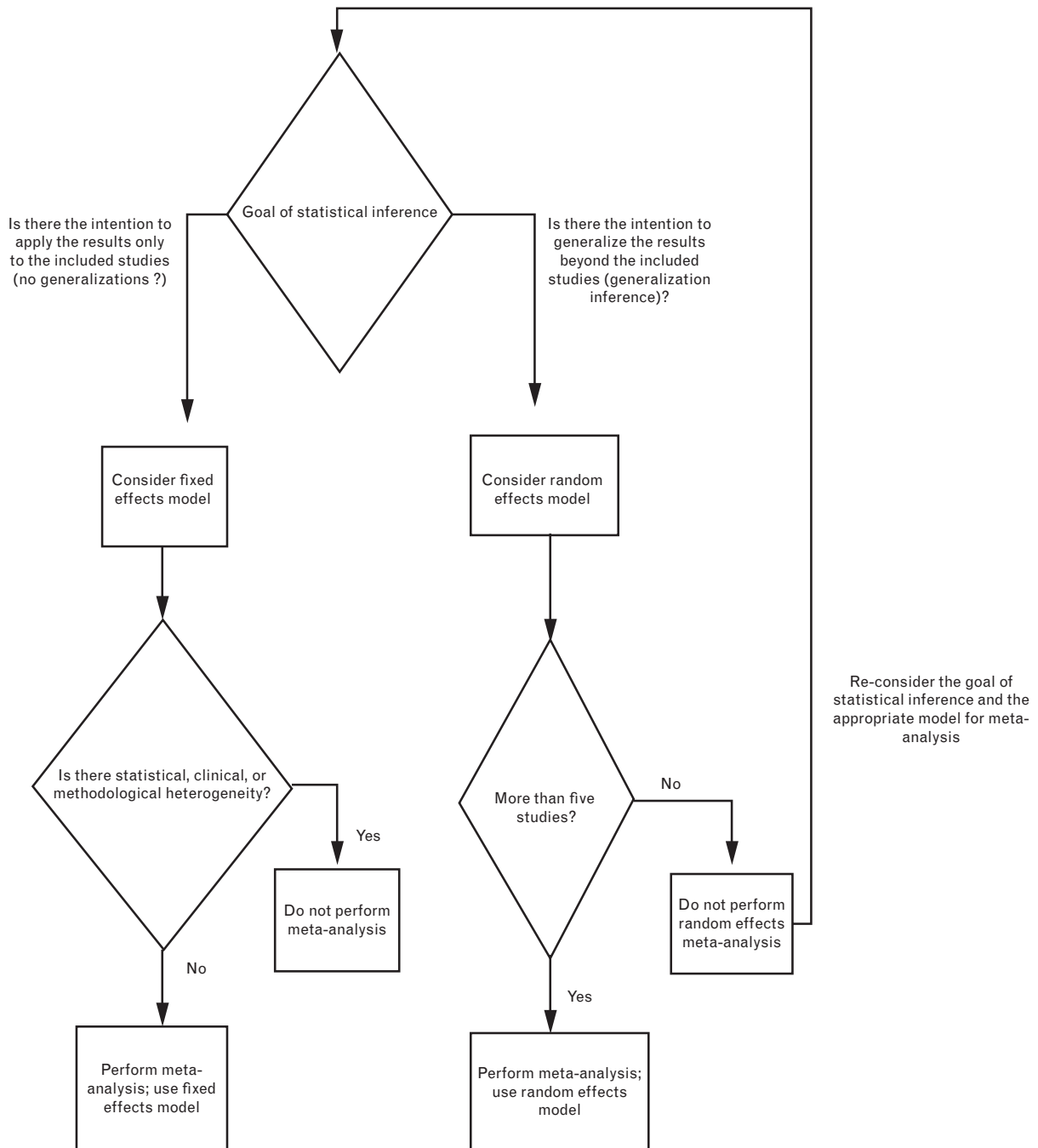
### Meta-analysis: statistical models (fixed-effects and random-effects models)

Fixed-effects and random-effects models are the most commonly employed statistical models for meta-analysis. In Table 4, we provide a concise summary of comparative characteristics of the fixed-effects and random-effects model. In Fig. 1, we provide a decision flow chart for the selection of the statistical model for meta-analysis. The decision to use one statistical model or another is complex and often subjective; however, there are criteria that can guide decisions about which model to use.

As the first criterion, reviewers should consider the goal of statistical inference: is there the intention to generalize the results beyond the included studies (generalization inference)? If the answer is 'yes', then the random-effects model is the appropriate statistical model for meta-analysis; is there the intention to apply the results only to the included studies (no generalizations)? If the answer is 'yes', then the fixed-effects model is the appropriate statistical model. As we assume that usually reviewers want to generalize the conclusions beyond the actual studies included in meta-analysis, we suggest that the default model for meta-analysis

in reviews should be the random-effects model. However, all decision criteria should be considered and the statistical model used should be appropriate from this multicriteria perspective. A second criterion to consider directly refers to the number of studies included in the meta-analysis. The fixed-effects model is the appropriate model when the number of studies is small. Random-effects models are appropriate when the number of studies is large enough, that is, enough studies to support generalization inferences beyond the included studies. It was suggested that the fixed-effects model should be used when the number of studies included in a meta-analysis is less than five.<sup>34</sup> A third criterion to consider refers to statistical heterogeneity. The fixed-effects model assumes that all studies included in a meta-analysis are estimating a single true underlying effect. If there is statistical heterogeneity among the effect sizes, then the fixed-effects model is not appropriate. The random-effects model should be considered when it cannot be assumed that true homogeneity exists.

Similarly, a fourth criterion refers to the likelihood of a common effect size. In fixed-effects models, we assume that there is one common effect. A random-effects model assumes each study estimates a different underlying true effect, and these effects have a distribution (usually a normal distribution). Fixed-effects model should be used only if it reasonable to assume that all studies shares the same, one common effect. If it is not reasonable to assume that there is one common effect



**Figure 1.** Decision flow chart for the selection of the statistical model for meta-analysis.

size, then the random-effects model should be used. If the studies are heterogeneous from a clinical and methodological point of view, it is unreasonable to assume that they share a common effect. Another criterion refers to the heterogeneity of sample sizes of included studies. The fixed-effects model is preferable when one study is

much larger (and usually it is presumed that it is more trustworthy) than one or more smaller studies.<sup>34</sup>

The use of the fixed-effects model and random-effects model presented here are based on a careful examination of the international literature.<sup>34–52</sup> It represents an accurate reflection of a classical or



**Table 5. Differences between statistical significance, practical significance, and clinical significance of the results**

	Statistical significance	Practical significance	Clinical significance	Comments
What it is about?	Results of statistical tests	Magnitude and direction of effect	The impact in a patient's life	Clinical significance is the most important; be aware of results that have statistical significance without practical and clinical significance.
Interpretation	A result of a statistical test was interpreted as indicating that the null statistical hypothesis is rejected	The magnitude and direction of effect size is judged as important by researchers and healthcare professionals	The magnitude and direction of effect size is judged by patients (or their families) as indicating real positive impact in a patient's life	For example, an odds ratio (OR) = 2 is considered significant from a practical point of view by some methodologists; for example, suppose that a trial reported that the intervention compared with the control intervention improved survival of patients with a specific type of cancer by 3 weeks and results are significant from a statistical point of view ( $P$ value $< 0.05$ for the statistical test that compared survival for the two groups); clinicians may say that results are not significant from a practical point of view if they adopt as standard for practical significance an improvement of minimum 3 months; however, patients may consider results as significant from their point of view.
How is it usually reported?	$P$ values; 95% confidence interval	Magnitude and direction of effect sizes; for example, an OR = 2.3 indicates a different magnitude and direction of effects compared with an OR = 0.23; a risk difference (RD) = +50 indicates a different magnitude and direction compared with a RD = -10	Commentaries with regard to the impact in a patient's life of the magnitude and direction of effect sizes; for example, patients may report that an improvement of a score on a specific scale (such as an improvement from 50 to 100 points on a scale from 0 to 100 points for pain; or an improvement of 50% of anxiety or depression score) is accompanied by a real change in their life	
Who evaluates/interprets it?	Researchers	Researchers and healthcare professionals	Patients (or their families)	

traditional view of the two meta-analysis models. It is worth acknowledging, however, that this traditional approach to meta-analysis has been critiqued by statisticians who suggested that this approach to meta-analysis is flawed and should be replaced.<sup>53–65</sup>

The traditional approach to meta-analysis as described in this article should be viewed as an acceptable simplification for novice reviewers without sophisticated statistical skills. The complexity of study designs, analysis approaches, and considerations related to risk of bias and the influence of moderator and mediators encountered in real statistical practice may require the use of more complex models for meta-analysis, some which include mixed-effects models, hierarchical models, and factorial models.<sup>37,42,44,45,47,49,66–70</sup> These approaches are complex and require sophisticated statistical skills. However, some of the newer approaches, including the inverse variance heterogeneity model and the quality-effects model proposed by Doi *et al.*<sup>55,56</sup> may be more accessible to novice reviewers, and should be used if possible.

### Statistical significance, practical significance, and clinical significance

Many review authors exclusively focus on statistical significance of the results. We recommend that the significance of the results should be considered from three different perspectives: statistical significance, practical significance, and clinical significance. Different authors use the terms practical significance and clinical significance with different meanings; our use of the terms is summarized in Table 5. Details with regard to these types of significance and a summary of international literature regarding guidance for interpretation of results (what is considered a significant OR, RR, RD, etc., from a practical point of view) are provided by Tufanaru *et al.*<sup>71</sup>

### Conclusion

Conducting a systematic review of effectiveness can be a difficult undertaking for the reviewer. The commentaries presented within this article are aimed to assist reviewers and it is hoped that it may improve the quality of the systematic review (meta-analysis) process, and presentation and interpretation of the results.

### Acknowledgements

The authors report no conflicts of interest.

### References

1. Aromataris E, Pearson A. The systematic review: an overview. *Am J Nurs* 2014; 114: 47–55.

2. Stern C, Jordan Z, McArthur A. Developing the review question and inclusion criteria. *Am J Nurs* 2014; 114: 53–6.
3. Aromataris E, Riitano D. Systematic reviews: constructing a search strategy and searching for evidence. *Am J Nurs* 2014; 114: 49–56.
4. Porritt K, Gomersall J, Lockwood C. JBI's systematic reviews: study selection and critical appraisal. *Am J Nurs* 2014; 114: 47–52.
5. Munn Z, Tufanaru C, Aromataris E. JBI's systematic reviews: data extraction and synthesis. *Am J Nurs* 2014; 114: 49–54.
6. Brignardello-Petersen R, Ioannidis JPA, Tomlinson G, Guyatt G. Surprising results of randomized trials. In: Guyatt G, Rennie D, Meade MO, Cook DJ, editors. *Users' guide to the medical literature. A manual for evidence-based clinical practice*, 3rd ed. New York: McGraw-Hill, 2015.
7. Shrier I, Boivin JF, Steele RJ, *et al.* Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. *Am J Epidemiol* 2007; 166: 1203–9.
8. Doi SA, Barendregt JJ, Khan S, *et al.* Advances in the meta-analysis of heterogeneous clinical trials II: the quality effects model. *Contemp Clin Trials* 2015. May 21. pii: S1551-7144(15)30008-2. doi: 10.1016/j.cct.2015.05.010. [Epub ahead of print].
9. Thompson S, Ekelund U, Jebb S, *et al.* A proposed method of bias adjustment for meta-analyses of published observational studies. *Int J Epidemiol* 2011; 40: 765–77.
10. Katikireddi SV, Egan M, Petticrew M. How do systematic reviews incorporate risk of bias assessments into the synthesis of evidence? A methodological study. *J Epidemiol Community Health* 2015; 69: 189–95.
11. Jones B, Kenward MG. *Design and analysis of cross-over trials*. 3rd ed. New York: CRC Press; 2015.
12. Senn SS. *Cross-over trials in clinical research*. 2nd ed. New York: Wiley; 2002.
13. Bonate PL. *Analysis of pretest-posttest designs*. New York: Chapman & Hall/CRC; 2000.
14. Hayes RJ, Moulton LH. *Cluster randomised trials*. Boca Raton: Chapman & Hall/CRC; 2009.
15. Eldridge S, Kerry S. *A practical guide to cluster randomised trials in health services research*. Chichester: Wiley; 2012.
16. Elbourne DR, Altman DG, Higgins JP, *et al.* Meta-analyses involving cross-over trials: methodological issues. *Int J Epidemiol* 2002; 31: 140–9.
17. Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychol Methods* 2002; 7: 105–25.
18. Curtin F, Altman DG, Elbourne D. Meta-analysis combining parallel and cross-over clinical trials. I: continuous outcomes. *Stat Med* 2002; 21: 2131–44.
19. Curtin F, Elbourne D, Altman DG. Meta-analysis combining parallel and cross-over clinical trials. II: binary outcomes. *Stat Med* 2002; 21: 2145–59.
20. Curtin F, Elbourne D, Altman DG. Meta-analysis combining parallel and cross-over clinical trials. III: the issue of carry-over. *Stat Med* 2002; 21: 2161–73.

21. Donner A, Klar N. Issues in the meta-analysis of cluster randomized trials. *Stat Med* 2002; 21: 2971–80.
22. Donner A, Piaggio G, Villar J. Statistical methods for the meta-analysis of cluster randomization trials. *Stat Methods Med Res* 2001; 10: 325–38.
23. Guyatt G, Rennie D, Meade MO, Cook DJ editors. Users' guide to the medical literature. A manual for evidence-based clinical practice. 3rd ed. New York: McGraw-Hill, 2015.
24. Barratt A, Wyer PC, Hatala R, et al. Evidence-Based Medicine Teaching Tips Working Group. Tips for learners of evidence-based medicine: 1. Relative risk reduction, absolute risk reduction and number needed to treat. *CMAJ* 2004; 171: 353–8.
25. Furuya-Kanamori L, Doi SA. The outcome with higher baseline risk should be selected for relative risk in clinical studies: a proposal for change to practice. *J Clin Epidemiol* 2014; 67: 364–7.
26. Fleiss JL. Measures of effect size for categorical data. In: Harris Cooper Hedges LV, editors. The handbook of research synthesis. New York: Russell Sage Foundation, 1994.
27. Fleiss JL, Berlin JA. Effect sizes for dichotomous data. In: Cooper H, Hedges LV, Valentine JC, editors. The handbook of research synthesis and meta-analysis. 2nd ed. New York: Russell Sage Foundation, 2009.
28. Senn S. Odds ratios revisited. *Evidence-Based Med* 1998; 3: 71.
29. Walter S. Odds ratios revisited. *Evidence-Based Med* 1998; 3: 71.
30. Olkin I. Odds ratios revisited. *Evidence-Based Med* 1998; 3: 71.
31. Altman D, Sackett D. Odds ratios revisited. *Evidence-Based Med* 1998; 3: 71–2.
32. Cummings P. The relative merits of risk ratios and odds ratios. *Arch Pediatr Adolesc Med* 2009; 163: 438–45.
33. Cummings P. Methods for estimating adjusted risk ratios. *Stata J* 2009; 9: 175–96.
34. Murad MH, Montori VM, Ioannidis JPA, et al. Fixed-effects and random-effects models. In: Guyatt G, Rennie D, Meade MO, Cook DJ, editors. Users' guide to the medical literature. A manual for evidence-based clinical practice. 3rd ed. New York: McGraw-Hill, 2015.
35. Cooper H, Hedges LV. Potentials and limitations of research synthesis. In: Cooper H, Hedges LV, editors. The handbook of research synthesis. New York: Russell Sage Foundation, 1994.
36. Higgins JPT, Green S editors. Cochrane handbook for systematic reviews of interventions. Chichester, UK: John Wiley & Sons, 2008.
37. Sutton AJ, Abrams KR, Jones DR, et al. Methods for meta-analysis in medical research. New York: Wiley; 2000.
38. Normand SL. Meta-analysis: formulating, evaluating, combining, and reporting. *Stat Med* 1999; 18: 321–59.
39. Hedges LV. Meta-analysis. *J Educ Behav Stat* 1992; 17: 279–96.
40. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods* 2010; 1: 97–111.
41. Shadish WR, Haddock CK. Combining estimates of effect size. In: Cooper H, Hedges LV. The handbook of research synthesis. New York: Russell Sage Foundation, 1994. pp. 261–281.
42. Cooper H, Hedges LV, Valentine JC. The handbook of research synthesis and meta-analysis. 2nd ed. New York: Russell Sage Foundation; 2009.
43. Petitti DB. Meta-analysis, decision analysis, and cost-effectiveness analysis: methods for quantitative synthesis in medicine. 2nd ed. New York: Oxford University Press; 2000.
44. Card NA. Applied meta-analysis for social sciences research. New York: The Guilford Press; 2012.
45. Cheung MWL. Meta-analysis. A structural equation modeling approach. Chichester: Wiley; 2015.
46. Schmidt FL, Hunter JE. Methods of meta-analysis. Correcting error and bias in research findings. 3rd ed. Los Angeles: Sage; 2015.
47. Mengersen K, Schmid CH, Jennions MD, Gurevitch J. Statistical models and approaches to inference. In: Koricheva J, Gurevitch J, Mengersen K, editors. Handbook of meta-analysis in ecology and evolution. Princeton: Princeton University Press, 2013. pp. 89–107.
48. Ringquist EJ. Meta-analysis for public management and policy. San Francisco: Jossey-Bass Wiley; 2013.
49. Lipsey MW, Wilson DB. Practical meta-analysis. Thousand Oaks: Sage; 2001.
50. Egger M, Davey SG, Altman DG editors. Systematic reviews in health care: meta-analysis in context. 2nd ed. London: BMJ Publishing Group, 2001.
51. Khan K, Kunz R, Kleijnen J, Antes G. Systematic reviews to support evidence-based medicine. How to review and apply findings of healthcare research. 2nd ed. London: Hodder Arnold; 2011.
52. Centre for Reviews and Dissemination (CRD). Systematic reviews: CRD's guidance for undertaking reviews in health care. York: Centre for Reviews and Dissemination, University of York; 2009.
53. Doi SA, Barendregt JJ, Rao C. An updated method for risk adjustment in outcomes research. *Value Health* 2014; 17: 629–33.
54. Stanley TD, Doucouliagos H. Neither fixed nor random: weighted least squares meta-analysis. *Stat Med* 2015; 34: 2116–27.
55. Doi SA, Barendregt JJ, Khan S, et al. Advances in the meta-analysis of heterogeneous clinical trials I: the inverse variance heterogeneity model. *Contemp Clin Trials* 2015. May 21. pii: S1551-7144(15)30007-0. doi: 10.1016/j.cct.2015.05.009. [Epub ahead of print].
56. Doi SA, Barendregt JJ, Khan S, et al. Advances in the meta-analysis of heterogeneous clinical trials II: the quality effects model. *Contemp Clin Trials* 2015. May 21. pii: S1551-7144(15)30008-2. doi: 10.1016/j.cct.2015.05.010. [Epub ahead of print].
57. Doi SA, Barendregt JJ, Onitilo AA. Methods for the bias adjustment of meta-analyses of published observational studies. *J Eval Clin Pract* 2013; 19: 653–7.

58. Doi SA, Barendregt JJ, Mozurkewich EL. Meta-analysis of heterogeneous clinical trials: an empirical example. *Contemp Clin Trials* 2011; 32: 288–98.
59. Al Khalaf MM, Thalib L, Doi SA. Combining heterogenous studies using the random-effects model is a mistake and leads to inconclusive meta-analyses. *J Clin Epidemiol* 2011; 64: 119–23.
60. Doi SA, Thalib L. A quality-effects model for meta-analysis. *Epidemiology* 2008; 19: 94–100.
61. Senn S. Trying to be precise about vagueness. *Stat Med* 2007; 26: 1417–30.
62. Cornell JE, Mulrow CD, Localio R, *et al.* Random-effects meta-analysis of inconsistent effects: a time for change. *Ann Intern Med* 2014; 160: 267–70.
63. Noma H. Confidence intervals for a random-effects meta-analysis based on Bartlett-type corrections. *Stat Med* 2011; 30: 3304–12.
64. Brockwell SE, Gordon IR. A simple method for inference on an overall effect in meta-analysis. *Stat Med* 2007; 26: 4531–43.
65. Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. *Stat Med* 2001; 20: 825–40.
66. Platt RW, Leroux BG, Breslow N. Generalized linear mixed models for meta-analysis. *Stat Med* 1999; 18: 643–54.
67. Aitkin M. Meta-analysis by random effect modelling in generalized linear models. *Stat Med* 1999; 18: 2343–51.
68. Turner RM, Omar RZ, Yang M, *et al.* A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2000; 19: 3417–32.
69. Sheu CF1, Suzuki S. Meta-analysis using linear mixed models. *Behav Res Methods Instrum Comput* 2001; 33: 102–7.
70. Stram DO. Meta-analysis of published data using a linear mixed-effects model. *Biometrics* 1996; 52: 536–44.
71. Tufanaru C, Huang WJ, Tsay S-F, Chou S-S. *Statistics for systematic review authors*. Philadelphia: Lippincott Williams & Wilkins; 2012.