Title: **Common methodological pitfalls and new developments in systematic review meta-analyses**

Word count: 1452

Authors: J Leonardi-Bee[1], C Flohr[2], EJ van Zuuren[3], L Le Cleach[4], LM Hollestein[5,6]

Affiliations: [1]Centre for Evidence Based Healthcare, Division of Epidemiology and Public Health, University of Nottingham; [2]Unit for Population-Based Dermatology Research, St John's institute of Dermatology, Kings College London; [3]Deparment of Dermatology, Leiden University Medical Centre; [4]Henri Mondor Hospital, University Paris Est Creteil; [5]Department of Dermatology, Erasmus MC Cancer Institute; [6]Dpartment of Research, Netherland Comprehensive Cancer Center

Corresponding author: J Leonardi-Bee, Centre for Evidence Based Healthcare, Division of Epidemiology and Public Health, Clinical Sciences building phase 2, University of Nottingham, Hucknall Road, Nottingham, NG5 1PB, UK; email: jo.leonardi-bee@nottingham.ac.uk

Funding sources: None declared

Conflicts of Interest: None declared

Introduction

Systematic reviews have become an increasingly popular tool to assess best evidence and have an established place at the top of the evidence pyramid, above individual randomised controlled trials. Most systematic reviews include a meta-analysis. Over recent years, the methodologies employed have become more complex, and we therefore review common methodological pitfalls and new developments in this editorial.

Meta-analyses are a 'statistical analysis that combines or integrates the results of several independent clinical trials considered by the analyst to the combinable'[1] and provides a statistical summary estimate of the effectiveness (called an effect size) of one intervention/treatment versus another, for a given population.[2] However, methods have been advanced to pool results from other study designs, for example, comparative observational studies for estimating an association between a risk factor and disease, or studies for estimating the prevalence of a disease. Meta-analyses are useful as they provide estimates of effect from more than one study and therefore will have increased statistical power (through larger sample size) and have improved estimates of precision (through smaller standard errors), thus yielding narrower confidence intervals for the effect estimate; therefore, the more certain we are that the effect estimate reflects the true effect size. Meta-analyses also help overcome controversies resulting from disparate results across studies and can generate new hypotheses to be tested in future studies. However, there are many decisions which need to be made before a meta-analysis is conducted as part of a systematic review. We describe some of the common pitfalls that authors may face below.

Common pitfall 1: Missing crucial steps in the conduct of a systematic review

There are five main steps in the conduct of a systematic review: i) framing the question; ii) identifying the literature; iii) assessing the methodological quality (or risk of bias) of the included studies using a validated tool appropriate to the design of the included studies; iv) synthesis of the findings and, where appropriate, rating of the certainty of evidence; and v) the interpretation of the findings. The most important step is framing the question, as the findings from a systematic review hinge on

whether the right question was asked. Research questions of healthcare interventions should aim to address feasibility, appropriateness, meaningfulness, and effectiveness.[2] Meta-analyses should only be performed, once all of the previous steps have been completed, using best method practices, such as comprehensive search strategies, which include searching electronic databases and grey literature, for example, conference proceedings abstracts, theses/dissertations, working/government reports, and contact with leading expert and/or relevant manufacturers of the intervention. It is also important to screen trial registries as positive studies are more likely to be published than negative ones, particularly in full text form. Whilst this can be very time consuming in practice, it is important that the searches are exhaustive and without language restrictions so that all of the relevant literature is identified. Furthermore, it is important that search strategies are reported in sufficient detail in the manuscript to allow replication. Two authors should be independently involved in the screening of papers for eligibility, data extraction, and the assessment of methodological quality (or risk of bias), to minimise errors.

Common pitfall 2: Use of surrogate outcome measures

Authors should be aware of apparent confirmatory systematic reviews, which are solely based on surrogate outcome measures or short-term outcome measures, rather than clinically important outcomes. For example, using transepidermal water loss as a surrogate outcome for assessing barrier function in eczema, solely reporting the reduction in actinic keratosis as a surrogate for assessing the prevention of cutaneous squamous cell carcinoma, and using sensitisation to peanut as a surrogate for clinically diagnosed peanut allergy. The primary outcome measures of the review should be chosen so that both clinically and patient-orientated outcomes are included, hence the primary outcomes may not be what is usually measured in a study. Whilst it can be very useful to consider surrogate outcome measures within a review, these should only be used as secondary outcome measures, since there is the potential for inconsistent results to be found between surrogate and clinical outcomes.

Common pitfall 3: Using heterogeneity measures to determine whether to conduct a meta-analysis

Often, meta-analyses are not performed based on the results of tests or measures of heterogeneity. However, as the effectiveness of a healthcare intervention will differ by participants or study characteristics, a degree of heterogeneity can be expected. Typical reasons for heterogeneity that can influence the estimate of effect are differences in the clinical setting (e.g. primary care versus secondary care), geographical setting (low/middle income versus high income countries), length of follow-up, baseline characteristics (e.g. severe psoriasis versus mild to moderate psoriasis), comedications (e.g certain comedications allowed versus no comedication allowed) or definitions in clinical outcomes (for example, *any* eczema flare versus *severe* eczema flare). Other important factors that can lead to variations in the estimates of effect are due to differences in study designs (RCTs versus quasi-RCTs) or other domains of methodological quality (e.g. allocation concealment, blinding, or attrition bias).

However, it must be noted that I^2, which is a commonly used measure for determining the quantity of inconsistency between studies in a meta-analysis[3], does not tell us how much variation there is in the effect estimates between studies. I^2 is the proportion of variation in the meta-analysis that is due to study results heterogeneity (ranges from 0-100%), and therefore should be used to describe how much overlap there is between the 95% confidence intervals of the studies included in the meta-analysis (for example, low I^2 values indicate that there is much overlap in 95% CI). I^2 values should not be compared between different meta-analyses since the value is dependent on the magnitude of sampling error within the included studies.

A valid rationale for not conducting a meta-analysis would be where it was decided *a priori* that it did not make clinical sense to estimate an average intervention effect. Therefore, if it makes clinical sense to perform the meta-analysis, then the analysis should be performed irrespective of the value of the I^2 statistic. However, it is important that a thorough investigation of heterogeneity is conducted to explore potential reasons for why differences in estimates of effect between the studies is seen.

As part of this assessment, a good first step is to re-check data entry as mistakes can happen, even when double data entry was used. Second, subgroup analysis can be used to explore whether the pooled magnitude of effect differ between specific subgroups measured at study level, for example, dose of treatment (high versus low), outcome definitions, and geographical setting. Subgroup analyses use a statistical test to assess whether differences exist in the pooled estimates. An extension to the subgroup analysis is meta-regression, where continuous study level factors, for example year of publication, can be included in the model to assess whether the effectiveness of the intervention varies as year of publication increases. Finally, sensitivity analyses can be used to explore whether the pooled results are robust when studies are restricted to those which meet specific criteria, for example, excluding studies with a high risk of bias for allocation concealment. Sensitivity analyses descriptively compare the pooled magnitude of effect between the overall meta-analysis including all studies and the meta-analysis restricted to studies meeting the criterion. The study level variables to be used in subgroup and sensitivity analyses should be pre-specified *a priori* in the protocol to minimise the chance of a false positive result. It is important to remember that subgroup and sensitivity analyses are observational associations, and not causal, as any observed effect could be due to confounding, bias or chance.

Common pitfall 4: Not adhering to conduct and reporting guidelines for systematic review meta-analyses

For systematic reviews to be considered for publication in the British Journal of Dermatology, authors are encouraged register the systematic review protocol either on the International Prospective Register Of Systematic Reviews (PROSPERO) database (https://www.crd.york.ac.uk/prospero/) or published in an appropriate peer-review journal. Prospective registration avoids duplication and reduce opportunity for reporting bias by enabling comparison of the completed review with was planned in the protocol.[4] Additionally, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines[5] should be adhered to when drafting the systematic review protocol (using PRISMA-P) and the final systematic review manuscript, to minimise the risk of not including pertinent information. Any changes in the methods between those reported in the protocol and the final systematic review need to be highlighted and fully justified. Finally, extensions to the PRISMA guidelines should be used for manuscripts with network meta-analyses,[6] individual patient data meta-analyses,[7] or diagnostic test accuracy meta-analyses;[8] and the MOOSE guidelines should be used for meta-analyses of observational studies.[9,10] We look forward to receiving your high quality systematic reviews in the BJD Evidence-Based Dermatology Section!

References

1. Huque MF. Experiences with meta-analysis in NDA submissions. *Proceedings of the Biopharmaceutical Section of the American Statistical Association*; 1988. pp. 28–33.
2. Aromataris E, Munn Z. Joanna Briggs Institute Reviewer's Manual. The Joanna Briggs Institute, 2017. Available at: https://reviewersmanual.joannabriggs.org/ (last accessed 9 April 2019).
3. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003: **327**:557-60.
4. Abuabaraa K, van Zuuren EJ, Flohr C. Why does the BJD require registration of systematic reviews and meta-analyses? *Br J Dermatol* 2019; **180**:249-50.
5. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *BMJ* 2009; **339**:b2535.

6. Hutton B, Salanti G, Caldwell DM, *et al*. The PRISMA Extension Statement for Reporting of Systematic Reviews Incorporating Network Meta-analyses of Health Care Interventions: Checklist and Explanations. *Ann Intern Med* 2015; **162**:777-84.
7. Stewart LA, Clarke M, Rovers M, *et al*; PRISMA-IPD Development Group. Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD Statement. *JAMA* 2015; **313**:1657-65.
8. McInnes MDF, Moher D, Thombs BD *et al*. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. *JAMA* 2018; **319**:388-96.
9. Stroup DF, Berlin JA, Morton SC, *et al*. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000; **283**:2008-12.
10. van Zuuren EJ, Fedorowicz Z. Moose on the loose: checklists for meta-analyses of observational studies. *Br J Dermatol* 2016; **175**(5):853-4.