The Role of Systematic Reviews and Meta-analysis in Dermatology

Katrina Abuabara[1,4], Esther E. Freeman[2,4], and Robert Dellavalle[3]

Journal of Investigative Dermatology (2012) **132**, e2. doi:10.1038/jid.2012.392

Systematic reviews of the medical literature are key tools in comprehensively summarizing data and helping clinicians and policy makers to make informed, evidence-based decisions regarding patient care and health policy.

Systematic reviews often contain a meta-analysis, a statistical method that synthesizes the available data from independent studies in order to answer a specific research question. According to the Cochrane Collaboration—an international organization dedicated to promoting and managing systematic reviews and evidence-based medicine—a meta-analysis (i) provides more precise estimates of the effects of an intervention than those from individual studies alone and (ii) allows for investigation of consistencies and differences across studies (Higgins and Green, 2011).

We review the basic methodology behind high-quality systematic reviews and meta-analyses, explain the statistical methods and analyses involved in meta-analyses, and emphasize how this methodological tool can be used in the field of dermatology.

ADVANTAGES OF SYSTEMATIC REVIEWS AND META-ANALYSES

- A systematic review is a comprehensive summary of available data pertaining to a specific question, organized through a rigorous design. Such reviews often contain a meta-analysis, which is a statistical method for synthesizing data from multiple studies.
- These techniques are used to answer specific research questions and help to minimize bias, improve precision of intervention estimates, and increase the statistical power of identifying a real effect. They may also help to settle controversy when individual studies show conflicting results and can be used to identify research gaps.

LIMITATIONS

- Limitations include the risk of misleading results if individual studies are biased or their reporting is not standardized. Although they are often useful for summarizing an intervention effect from randomized controlled trials, meta-analyses are less effective for capturing adverse effects or summarizing observational studies.

STUDY DESIGN METHODOLOGY: FORMULATING A CLINICAL QUESTION, ELIGIBILITY CRITERIA, AND SEARCH STRATEGY

Guidelines on how to report systematic reviews and meta-analyses have helped to establish criteria for developing, carrying out, and evaluating these studies. The first set of guidelines was published in 1999 under the title QUOROM (QUality Of Reporting Of Meta-analysis) Statement (Moher *et al.*, 1999). In 2009 these guidelines further evolved into PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses), which established a 27-item checklist to aid researchers in planning and reporting studies (http://www.prisma-statement.org) (Liberati *et al.*, 2009). The PRISMA statement is also a useful tool for readers to reference when evaluating a systematic review or meta-analysis.

Formulating an appropriate and specific research question is key to any study and underpins systematic reviews and meta-analyses. A research question should be feasible, interesting, novel, ethical, and relevant, known as the FINER criteria (Thabane *et al.*, 2009). A well-established format for structuring research questions is known by the acronym PICOT: Population, Intervention, Comparator, Outcome, Time-frame (Liberati *et al.*, 2009; Moher *et al.*, 2009). For example, in the systematic review and meta-analysis published by Hadley *et al.* (2006) in the *Journal of Investigative Dermatology* on the use of imiquimod for actinic keratosis, the population was patients with actinic keratosis (AK), the intervention was use of topical imiquimod, the comparator was placebo vehicle cream,

[1]*Department of Dermatology, University of Pennsylvania, Philadelphia, Pennsylvania, USA;* [2]*Department of Dermatology, Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts, USA and* [3]*Dermatology Service, Denver VA Medical Center, Denver, Colorado, USA*

[4]*These authors are co-first authors and made equal contributions.*

Correspondence: Katrina Abuabara, Department of Dermatology, University of Pennsylvania, 3400 Spruce Street, 2 Maloney, Philadelphia, Pennsylvania 19104, USA. E-mail: Katrina_abuabara@uphs.upenn.edu

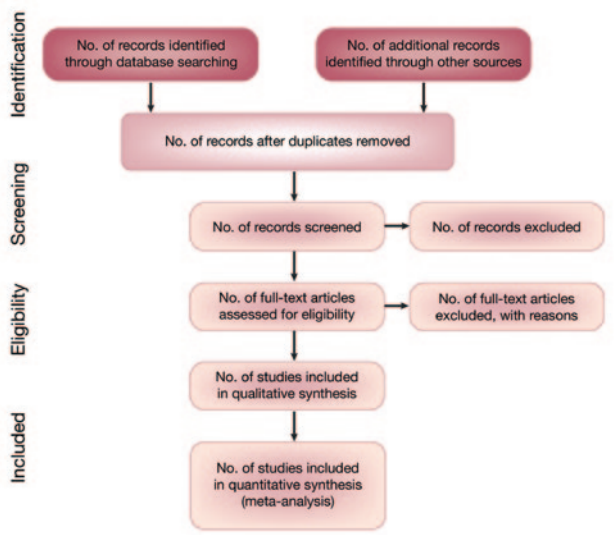

Figure 1. Flow diagram illustrating the search strategy and inclusion/exclusion criteria used in a systematic review and/or meta-analysis. Originally published with the PRISMA Statement (Liberati *et al.*, 2009).

the outcome measures were multiple and included complete clearance of lesions in treatment area, partial clearance, and complete clearance of a particular lesion, and the time-frame was number of weeks of treatment needed. Structuring a research question by addressing each of these elements is independently associated with better reporting quality in randomized controlled trials (RCTs) (Rios *et al.*, 2010) and should also be used in approaching systematic reviews and meta-analyses (Higgins and Green, 2011).

A hallmark of the systematic review is prespecified inclusion and exclusion criteria for studies, known as eligibility criteria. These criteria stem from the PICOT question above, defining the types of participants that will be included and the types of interventions to be studied and potentially also limiting the type of study design (Higgins and Green, 2011; Moher *et al.*, 2009). To use the same study example as above, inclusion criteria for the systematic review by Hadley *et al.* were randomized, double-blind trials investigating imiquimod for AK with efficacy or safety data. Exclusion criteria were reviews with clinical information published elsewhere, biochemical or immunological studies, abstracts, and studies of conditions other than AK.

Once the research question and eligibility criteria have been established, a search strategy should be laid out. This strategy will specify the electronic databases to be searched, as well as the search terms to be used. The Cochrane Collaboration recommends searching the electronic databases CENTRAL and MEDLINE as a minimum, together with EMBASE, if available. Searches should also include national and regional databases, gray literature (such as technical reports or working papers from research groups or committees), relevant journals, conference abstracts, other reviews and guidelines, and ongoing studies in trial registries, over a specified period of time (Higgins and Green, 2011). As per PRISMA, authors are encouraged to contact study authors for clarification and identification of further studies, and they should explicitly announce such contact, as well as specify the date last searched (Moher *et al.*, 2009). The search results should be presented in a flow diagram that clearly illustrates the number of studies included and the reasons for exclusion of studies that met initial search criteria (Figure 1).

After key studies have been identified for inclusion in the systematic review, two independent reviewers extract data from the studies. These results should be extracted on piloted standardized forms.

ASSESSING RISK OF BIAS IN INDIVIDUAL STUDIES

Having identified appropriate studies and extracted data in duplicate, the author should assess individual studies for bias. Both PRISMA and Cochrane caution against the use of numerical scales to assess the quality of studies, recommending a more descriptive approach of assessed methodological components. Studies should be assessed for selection bias (appropriate generation of random allocation sequence, concealment of this sequence, and intention-to-treat analyses), performance bias (blinding of participants and providers), detection bias (blinding of outcome assessment), attrition bias (loss to follow-up), and reporting bias (differences between reported and unreported findings) (Liberati *et al.*, 2009; Moher *et al.*, 2009). An excellent example of a bias assessment and discussion can be seen in the recent systematic review by Nankervis *et al.* (2012), "Prospective Registration and Outcome-Reporting Bias in Randomized Controlled Trials of Eczema Treatments: A Systematic Review."

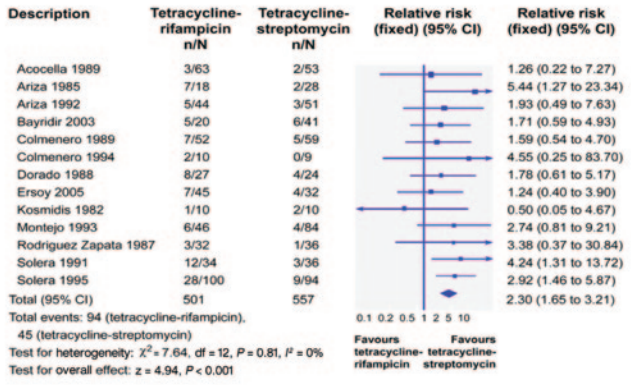

Figure 2. Forest plot. Used for presenting the results of a meta-analysis comparing two different antibiotic combinations for human brucellosis. The summary statistic for each study is shown as a square with a horizontal line indicating the confidence interval. At the bottom of the graphic, the overall intervention effect estimate is represented by a diamond, with the center showing the point estimate and the horizontal tips illustrating the confidence interval. The significance of each study and the overall estimate are highlighted by whether they cross a vertical line of no effect. Originally published with the PRISMA Statement (Liberati *et al.*, 2009).

STATISTICAL META-ANALYSIS

Meta-analysis allows data from multiple studies to be statistically combined. First, a summary measure is decided upon and a summary statistic is estimated for each study. Next, an overall intervention effect is calculated. Finally, to ensure that the results are valid and robust, researchers typically test for heterogeneity and publication bias and perform sensitivity analyses. Multiple software programs are available for meta-analysis; these may vary in terms of usability, although most give consistent numerical results (Bax *et al.*, 2007).

The first step in data analysis is to decide on a summary measure and calculate a corresponding summary statistic for each study. The choice of a summary measure will depend on the clinical question and type of data. For example, for dichotomous or "yes-or-no" data (e.g., did the patient develop melanoma: yes or no?), the risk ratio, odds ratio, absolute risk reduction, and number needed to treat all commonly use summary measures. For survival or time-to-event data (e.g., the number of months until melanoma development), the hazard ratio is the best summary measure. Researchers should choose the measure that is most appropriate for their data and that will give a consistent estimate of the treatment effect for the clinical situation (Deeks, 2002) and then calculate the same measure for each study. Each of the common summary statistics listed above can be reexpressed in terms of each other; however, when the unit of analysis differs among studies, care must be taken to avoid bias. For example, summary statistics may vary significantly based on the level of randomization (individual or group), number of treatment attempts, number of observations per patient, and whether multiple observations or body areas were included per patient.

Once summary statistics have been derived for each study, the overall intervention effect can be calculated. This effect is the pooled or weighted average of the effects estimated from the individual studies. Two commonly used approaches for combining data include fixed-effects and random-effects models. Fixed-effects models (such as Mantel–Haenszel and inverse variance approaches) account for only within-study variability. Random-effects models (such as DerSimonian, Laird, and Bayesian approaches) account for both within- and between-study variability and are generally considered more appropriate in settings of greater between-study heterogeneity, as described below. A more detailed summary of each of these statistical approaches can be found in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins and Green, 2011).

Results are typically presented in a forest plot (Figure 2), which provides a visual representation of the amount of variation between studies and the pooled estimate (Lewis and Clarke, 2001).

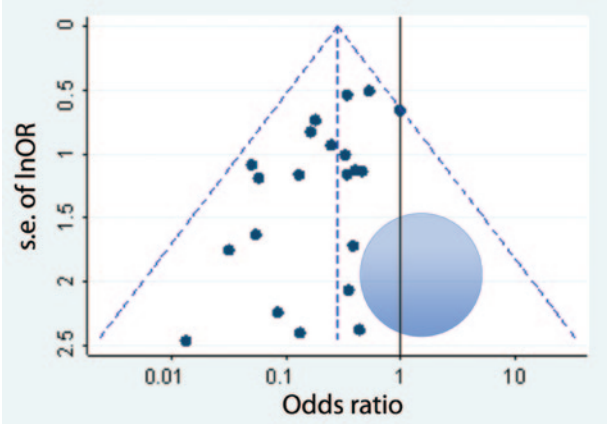

Figure 3. Funnel plot. The results of individual studies are plotted against a measure of the precision of the data. In this example, the odds ratio is plotted on the horizontal axis and the standard error of the natural log of the odds ratio—a measure of the precision of the data—is plotted on the vertical axis (this is a more precise estimate of statistical power than sample size alone). Because larger studies are more precise, most will fall within an area that forms a funnel or inverted V. In the presence of publication bias, fewer small studies that do not show a significant effect will be published, leading to a blank area in the bottom corner of the graph highlighted by the large blue circle. Reprinted and modified from Cochrane Handbook for Systematic Reviews of Interventions (Higgins and Green, 2011) with permission from Wiley. For additional discussion of funnel plots, see the TED video by Ben Goldcare on "Battling Bad Science" available at http://www.ted.com/talks/ben_goldacre_battling_bad_science.html.

Table 1. Rating the strength of clinical recommendations

GRADE
• Used by the World Health Organization in their guideline development process
• Classifies strength of recommendations into strong or weak. A strong recommendation means that based on the available evidence, clinicians are very certain that the benefits either do or do not outweigh the risks of an intervention
• Website: http://www.gradeworkinggroup.org/intro.htm
SORT
• Used by the American Academy of Dermatology
• Grades strength of recommendation into A, B, and C, with A-level evidence based on consistent, good-quality, patient-oriented evidence, B based on inconsistent or limited-quality, patient-oriented evidence, and C based on consensus, usual practice, or opinion
• Website: http://www.aad.org/education-and-quality-care/clinical-guidelines/guideline-development-process

Abbreviations: GRADE, grading of recommendations assessment, development and evaluation; SORT, strength of recommendation taxonomy.

The final steps in a meta-analysis are to assess for sources of bias and test the consistency of the results. Heterogeneity, or whether there is greater variation in the results between the studies than would be attributable to chance, may be due to differences between the participants, interventions, outcomes, study design and funding, or the statistical methods used. To assess heterogeneity, statistical methods such as the χ^2 test and I^2 test are commonly used. I^2 values <40% are generally considered low; however, the statistic may be imprecise when few studies are being compared. In addition to heterogeneity, publication bias may also affect the results. Smaller negative studies are less likely to be published, and the treatment effect may be overestimated if such studies are not included. Funnel plots are often used to assess for publication bias (Figure 3). Finally, sensitivity analyses may be used to gauge the robustness of results, which may involve removing certain studies that are considered heterogeneous or of lower quality and then repeating the analysis.

LIMITATIONS OF META-ANALYSIS

Meta-analyses may be performed for various reasons. Combining data from multiple studies may increase the power or chance of detecting a real effect, improve precision, and help to settle controversy where individual studies show conflicting results. Nonetheless, there is the risk that the results may be misleading if the individual studies are subject to bias, if there is significant heterogeneity among studies, or if there is publication or reporting bias that is not addressed appropriately. It is important for authors to reference any identified sources of heterogeneity, as well as whether there are subgroups for which evidence is stronger than for others, and then decide whether there is sufficient evidence to draw clear conclusions.

SUMMARY AND USE OF SYSTEMATIC REVIEWS AND META-ANALYSES

In a busy health-care setting, systematic reviews and meta-analyses allow a health-care provider or researcher to understand available evidence in a synthesized, coherent manner (Williams and Dellavalle, 2012). Whether you are performing a systematic review and meta-analysis or reading the literature in order to make a health-care or policy decision, it is important to understand the methodology behind these techniques. Understanding the methods will allow you to assess the quality of the systematic review or meta-analysis and the strength of the evidence behind it.

Finally, in using results from a systematic review and meta-analysis, either in a clinical practice or in a guideline development setting, conclusions are drawn, either implicitly or explicitly, regarding the strength of recommendations. There has been a recent movement to employ a systematic and explicit approach to making these judgments (Ebell *et al.*, 2004; Guyatt *et al.*, 2008). Table 1 describes two commonly used systems: GRADE and SORT. We refer to these scoring systems here because their use is increasing and they are of growing importance to informed dermatologists; Cochrane reviews now employ GRADE methodology in their summary of findings.

QUESTIONS

1. GRADE and SORT describe which of the following?
   - A. Techniques for assessing the risk of bias in individual studies
   - B. Guidelines on how to report systematic reviews and meta-analyses
   - C. Types of fixed-effect models used in statistical meta-analysis
   - D. Methods for rating the strength of clinical recommendations
2. Which of the following is true regarding random-effects models?
   - A. They are generally considered more appropriate if the I^2 value is >40%.
   - B. They are rarely used in meta-analysis.
   - C. They are often employed to assess the risk of bias in individual studies.
   - D. They are most appropriate when between-study heterogeneity is low.

Answers to the questions and an opportunity to comment on the article are available on our blog: http://blogs.nature.com/jid_jottings/.

CONFLICT OF INTEREST

The authors state no conflict of interest.

SUPPLEMENTARY MATERIAL

Answers and a PowerPoint slide presentation appropriate for journal club or other teaching exercises are available at http://dx.doi.org/10.1038/jid.2012.392.

REFERENCES

Bax L, Yu LM, Ikeda N, Moons KG (2007) A systematic comparison of software dedicated to meta-analysis of causal studies. *BMC Med Res Methodol* 7:40

Deeks JJ (2002) Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med* 21:1575–600

Ebell MH, Siwek J, Weiss BD, Woolf SH, Susman JL, Ewigman B *et al.* (2004) Simplifying the language of evidence to improve patient care: strength of recommendation taxonomy (SORT): a patient-centered approach to grading evidence in medical literature. *J Fam Pract* 53:111–20

Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schunemann HJ (2008) What is "quality of evidence" and why is it important to clinicians? *BMJ* 336:995–8

Hadley G, Derry S, Moore RA (2006) Imiquimod for actinic keratosis: systematic review and meta-analysis. *J Invest Dermatol* 126:1251–5

Higgins JPT, Green S (eds) (2011) *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 (updated March 2011). The Cochrane Collaboration. Available from http://www.cochrane-handbook.org

Lewis S, Clarke M (2001) Forest plots: trying to see the wood and the trees. *BMJ* 322:1479–80

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP *et al.* (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation

and elaboration. *PLoS Med* 6:e1000100

Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF (1999) Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet* 354:1896–900

Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 6:e1000097

Nankervis H, Baibergenova A, Williams HC, Thomas KS (2012) Prospective registration and outcome-reporting bias in randomized controlled trials of eczema treatments: a systematic review. *J Invest Dermatol,* e-pub ahead of print 26 July 2012

Rios LP, Ye C, Thabane L (2010) Association between framing of the research question using the PICOT format and reporting quality of randomized controlled trials. *BMC Med Res Methodol* 10:11

Thabane L, Thomas T, Ye C, Paul J (2009) Posing the research question: not so simple. *Can J Anaesth* 56:71–9

Williams HC, Dellavalle RP (2012) The growth of clinical trials and systematic reviews in informing dermatological patient care. *J Invest Dermatol* 132:1008–17